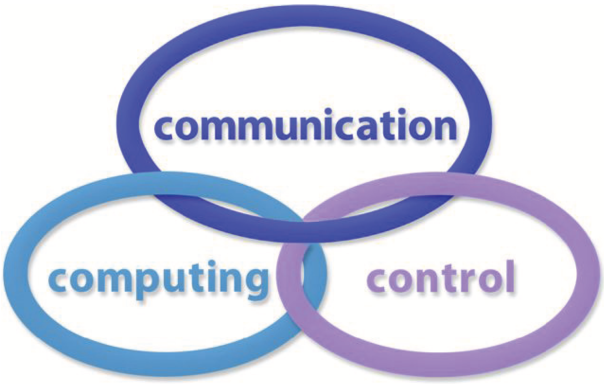
INTERNATIONAL JOURNAL

of

COMPUTERS COMMUNICATIONS & CONTROL

ISSN 1841-9836

ISSN-L 1841-9836



A Bimonthly Journal
With Emphasis on the Integration of Three Technologies

Year: 2012 Volume: 7 Issue: 5 (December)

This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).

C O P E COMMITTEE ON PUBLICATION ETHICS

Agora University Editing House

CCC Publications

URL: http://univagora.ro/jour/index.php/ijccc/

International Journal of Computers Communications & Control

EDITORIAL ADDRESS:
R&D Agora Ltd. / S.C. Cercetare Dezvoltare Agora S.R.L.
Piaţa Tineretului 8, Oradea, jud. Bihor, Romania, Zip Code 410526
Tel./ Fax: +40 359101032
E-mail: ijccc@univagora.ro, rd.agora@univagora.ro, ccc.journal@gmail.com
Journal website: http://univagora.ro/jour/index.php/ijccc/

International Journal of Computers Communications & Control

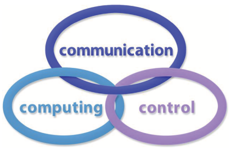
EDITORIAL BOARD

DATA FOR SUBSCRIBERS
Supplier: Cercetare Dezvoltare Agora Srl (Research & Development Agora Ltd.)
Fiscal code: 24747462
Headquarter: Oradea, Piata Tineretului Nr.8, Bihor, Romania, Zip code 410526
Bank: MILLENNIUM BANK, Bank address: Piata Unirii, str. Primariei, 2, Oradea, Romania
IBAN Account for EURO: RO73MILB0000000000932235
SWIFT CODE (eq.BIC): MILBROBU

International Journal of Computers, Communications & Control

Short Description of IJCCC

Title of journal: International Journal of Computers, Communications & Control
Acronym: IJCCC
Abbreviated Journal Title: INT J COMPUT COMMUN
International Standard Serial Number: ISSN 1841-9836, ISSN-L 1841-9836
Publisher: CCC Publications - Agora University
Starting year of IJCCC: 2006
Founders of IJCCC: Ioan Dzitac, Florin Gheorghe Filip and Mişu-Jan Manolescu
Logo:



Publication frequency: Bimonthly: Issue 1 (February); Issue 2 (April); Issue 3 (June); Issue 4 (August); Issue 5 (October); Issue 6 (December).
Coverage:

- Beginning with Vol. 1 (2006), Supplementary issue: S, IJCCC is covered by Thomson Reuters - SCI Expanded and is indexed in ISI Web of Science.
- Journal Citation Reports(JCR)/Science Edition:
  - Impact factor (IF): JCR2009, IF = 0.373; JCR2010, IF = 0.650; JCR2011, IF = 0.438.
- Beginning with Vol. 2 (2007), No.1, IJCCC is covered in EBSCO.
- Beginning with Vol. 3 (2008), No.1, IJCCC, is covered in Scopus.

Scope: International Journal of Computers Communications & Control is directed to the international communities of scientific researchers in computer and control from the universities, research units and industry.

To differentiate from other similar journals, the editorial policy of IJCCC encourages the submission of scientific papers that focus on the integration of the 3 "C" (Computing, Communication, Control).

In particular the following topics are expected to be addressed by authors:

- Integrated solutions in computer-based control and communications;
- Computational intelligence methods (with particular emphasis on fuzzy logic-based methods, ANN, evolutionary computing, collective/swarm
- intelligence);
  Advanced decision support systems (with particular emphasis on the usage of combined solvers and/or web technologies).

Contents

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):8-10, December, 2012.

Editorial Challenge: From a Quarterly to a Bimonthly Journal

F.G. Filip, I. Dzitac

Florin Gheorghe Filip
Editor in Chief of IJCCC,
Romanian Academy, Bucharest, Romania
ffilip@acad.ro

Ioan Dzitac
Associate Editor in Chief of IJCCC,
Aurel Vlaicu University of Arad, Romania &
Agora University, Oradea, Romania
rector@univagora.ro

Abstract:
Starting with issue 4 of volume 7(2012) *International Journal of Computers Communications & Control* (INT J COMPUT COMMUN, IJCCC) [4] is a member of, and subscribes to the principles of, the Committee on Publication Ethics (COPE) [2]. Beginning with issue 1 of volume 8(2013) IJCCC will be published as a bimonthly journal (6 issues/year) [5].

Keywords: ethics, bimonthly journal, impact factor.

1 Why a bimonthly journal?

In the period 2006-2012 the International Journal of Computers, Communications & Control, ISSN 1841-9836, has been published as a quarterly journal (4issues/year), with a supplementary issue every even year (5 issues/year) and has been covered in SCI Expanded [3] starting with the supplementary issue S of volume 1(2006) and has been indexed in Journal Citation Reports/Science Edition (JCR), with 3-years Journal Impact Factor 0.373 (in JCR2009), 0.650 (in JCR2010) and 0.438 (in JCR2011). This decreasing of the last impact factor can be explained by increasing of the number of published papers in issue 5 (December, 2010). Consequently many papers can be cited only in 2012, but no more in 2011, the year considered for Journal Impact Factor calculation in JCR2011. The correction of this anomalous fluctuation of the impact factor might be solved by a change of publishing policy of IJCCC, from a irregular frequency of publication (4-5 issues/year) to a bimonthly publication (6 issues/year) with a regular frequency. This new policy is sustainable because in the last two years we have received over 1,200 manuscripts.

Publication frequency starting with issue 1 of volume 8(2013): IJCCC will be published as a bimonthly journal (6 issues/ year), with a regular schedule such as: Issue 1 (February); Issue 2 (April); Issue 3 (June); Issue 4 (August); Issue 5 (October); Issue 6 (December).

2 Focus and Scope. Topics

IJCCC is directed to the international communities of scientific researchers in computer and control from the universities, research units and industry. To differentiate from other similar journals, the editorial policy of IJCCC encourages the submission of scientific papers that focus on the integration of the 3 "C" (Computing, Communication, Control). In particular the following topics are expected to be addressed by authors:
1. Integrated solutions in computer-based control and communications;
2. Computational intelligence methods (with particular emphasis on fuzzy logic-based methods, ANN, evolutionary computing, collective/swarm intelligence);

3. Advanced decision support systems (with particular emphasis on the usage of combined solvers and/or web technologies).

3 Author Guidelines

Length of a manuscript

The maximum number of pages of one article is 16. The publishing of a 6 page article is free of charge. For each supplementary page there is a fee of 50 EUR/page that must be paid after receiving the acceptance for publication.

Technical Instructions for Authors

The papers must be written in English. The first page of the paper must contain title of the paper, name of author(s), an abstract of about 300 words and 3-5 keywords. The name, affiliation (institution and department), regular mailing address and email of the author(s) should be filled as in [1].

Manuscripts must be accompanied by a signed copyright transfer form. The copyright transfer form is available at website.

Initial submission (manuscript for review: The manuscript can be write in LaTex by the IJCCC template or in MS Word format with the following specifications: paper A4, font TNR 12p, single column. The manuscript must be uploaded online in journal management and publishing system of IJCCC (open source software: Open Journal Systems developed by the Public Knowledge Project) [5].

Final submission (accepted paper for publication: Checklist of documents which must be send by e-mail: Completed copyright transfer form; Source files (One LaTeX file for the text, EPS files for figures - they must reside in a separate folder); Final PDF file (for reference).

Peer Review Process

The submissions will be revised independently by minimum two reviewers and will be accepted for publication only after end of the editorial workflow.

Evaluation period and rejection rate information:

- Evaluation period after paper submission: up to 6 months;
- Publication time is pending on the number of papers received. We aim at a time not longer than one year;
- Mean acceptance rate: of approximately 20%.

Bibliography

[1] Andonie R., Dzitac I., How to Write a Good Paper in Computer Science and How Will It Be Measured by ISI Web of Knowledge, *INT J COMPUT COMMUN*, ISSN 1841-9836, 5(4):432-446.

[2] Committee on Publication Ethics (COPE) (http://publicationethics.org).

[3] http://ip-science.thomsonreuters.com/cgi-bin/jrnlst/jlresults.cgi?PC=D&ISSN=1841-9836.

[4] http://www.journal.univagora.ro (IJCCC archive for 2006-2012).

[5] http://univagora.ro/jour/index.php/ijccc/ (New IJCCC website).

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):11-19, December, 2012.

Multi-period Customer Service Level Maximization under Limited Production Capacity

S. Babarogić, D. Makajić-Nikolić, D. Lečić-Cvetković, N. Atanasov

Sladjan Babarogić, Dragana Makajić-Nikolić
Danica Lečić-Cvetković, Nikola Atanasov
University of Belgrade, Faculty of Organizational Sciences
Serbia, 11000 Belgrade, Jove Ilića 154
E-mail: sladjan@fon.bg.ac.rs, gis@fon.bg.ac.rs
danica@fon.bg.ac.rs, nikola.atanasov@fon.bg.ac.rs

Abstract:
This paper will focus on a make-to-stock multi-period order fulfilment system with random orders from different classes of customers under limited production circumstances. For this purpose a heuristic algorithm has been developed aimed at maximizing the customer service level in any cycle and in the entire multi-period. In this paper, in order to validate the results obtained with this algorithm, a mixed integer programming model was developed that is based on the same assumptions as the algorithm. The model takes into account the priorities of customer groups and the balanced customer service level within the same group. The presented approaches are applied to a real example of Fast Moving Consumer Goods. Their comparison was carried out in several scenarios.

Keywords: limited production capacity, customer service level, heuristic algorithm, mixed integer programming.

1 Introduction

The distribution of available finished products among customer orders requires an efficient distribution system aimed at improving the effectiveness of the entire business. The effectiveness of business is directly related to products quantities and the profits through sales. In addition to profit-oriented decisions on the selection of orders to be met, it is necessary also to take into account the customer service level. According to [6], key performance indicators in manufacturing companies are identified in measuring the customer service level and customer satisfaction. Customers whose purchases represent a large share of the company's sales require special attention and the company should make sure that they achieve the highest possible fulfilment of each order. There are also customers that continuously increase their orders and based on that also expect a corresponding service level. Due to their large number, small customers influence the overall sales of manufacturing companies. Some of them also represent a potential for future sales growth and increase in revenues of the manufacturing company. These facts underline the importance of making the right decisions when selecting orders to be met. Therefore, a heuristic algorithm has been developed [5] that is used for decision-making concerning the customer service level in each cycle by taking into account the priorities of the customers. Traditional approaches to fulfil orders based on the make-to-stock (MTS) production system are described in [1] by taking into account the available supplies of finished products to satisfy customer orders following the principle of First Come - First Served (FCFS) without assigning priorities to customers and orders. The basic idea of the approach described in [8] is the segmentation of customers in order to increase the total revenue of the manufacturing company by accepting and delivering orders which provide maximum profit.

In this paper the customers are clustered into priority groups based on the size of their orders. Due to their potential growth, there is tendency to provide protected quantities of products to the customers with the lowest priority. If a manufacturing company has a limited production capacity, it is clear that the company will decide to reject some of the orders received which will have a direct impact on the profitability of the company. The decision to reject orders is made based on a comparison of orders where less profitable orders are rejected. According to [2] this issue has also been defined as dynamic models for managing orders under limited production capacity based on profitability analysis.

The problem discussed in this paper refers to meeting the demand in a multi-period, where the unmet demand in one cycle is not compensated in the following cycles, i.e. there are no backorders. Demand is a weekly phenomenon which requires dynamic decision-making. The heuristics shown in [9] refer to the problem of replenishment of multiple products in order to meet the demand when the storage capacity is limited. Authors in [13] present a mixed integer programming (MIP) model that applies to small and medium-sized enterprises with limited Available to Promise (ATP) quantities and which has to decide which customers they will accept and in each cycle which part of the demand of accepted customers they are going to meet.

A large number of MIP models include unlimited production capacity. The uncapacitated requirement planning model, with demand fulfilment flexibility, is shown in [7]. In each cycle separately a decision is taken regarding the launch of production and the part of demand of each customer that will be met in the respective cycle. A similar MIP model is presented in [12]. This model implies that the manufacturer may, in each cycle, decide whether to start the production, which order to fulfil and to what extent. The above mentioned papers consider the maximization of profit as the main criterion. The approach presented in this paper, however, aims at maximizing the customer service level. In [11] the orders of customers are clustered into two groups: small-size orders and large-size (divisible) orders In the presented basic MIP model, it is in each cycle determined which of the small-size orders will be fulfilled and what fraction of the large-size order will be met in order to maximize the customer service level. The multi-objective MI nonlinear mathematical model, in which the maximization of average customer service levels is one of the four objectives, is presented in [3].

The remainder of this paper is organized as follows: the second section defines the problem of allocating scarce resources in a manufacturing company. The allocation problem is related to the distribution of limited production capacity to customer orders in order to maximize the customer service level. The third section deals with the computational results and provides an analysis of these two approaches in a real example of Fast Moving Consumer Goods (FMCG). In the Conclusions section the authors lists the principal advantages of the proposed algorithm for the allocation of limited production capacity, as well as possible directions for the further development.

2 Problem definition and model formulation

The algorithm has been developed for solving the problem of FMCG industry products. By introducing minor modifications it can also be applied to the product allocation problem in other industries. The basic assumptions of the problem discussed in this paper are the following:

- It studies a multi-period and a set of customers that place order in all or almost all of the cycles. Demand is uneven and is known only for one cycle in advance;
- The production capacity is limited and constant in the entire period;
- If the incoming customer orders in a single cycle do not exceed the available stock of finished goods, the allocation is complete and all customer orders are fulfilled, while any surplus products are stored for the next cycle. Inventory holding costs are neglected;

- When the total of all orders exceeds the available stock of products, it is necessary to define the distribution of products i.e. rules based on which the allocation of products will be done according to the received customer orders. The allocation has to maximize the customer service level.

- Considering the type of products that are being studied, orders that have not been fully met in the reporting cycle shall not be compensated in the subsequent cycles;

- Order of customer priority is known. These are clustered into priority groups in which they have the same order of priority as the other group members;

- The product unit price is the same for all customers.

Based on the assumptions of problem, in previous research effort we have developed heuristic algorithm that introduces the concepts of partitions and tokens. The algorithm aims to maximize the cumulative customer service level with a balanced customer service level within the same group. The detailed explanation of the proposed algorithm is given in [5]. The important features of the algorithm are:

- Classification in groups, provides the order of allocation with primarily focus to satisfy customers that are important for the company.

- Application of mechanism of Partitions, ensures certain groups of lower priority within protected partitions to be involved in allocation so that the low-ranked customers would be at least partially satisfied.

- Application of mechanism of Group Memory Token, allows all customers within a marked group to receive the unsatisfied demand from the previous cycle, with the extended delivery lead time, with which they improve the overall customer service.

2.1 MIP customer service level maximization model

A mixed integer customer service level maximization model (CSL model) model was developed in order to validate the results obtained with this algorithm. The model is based on the same assumption as the algorithm. However, here the total demand of all customers in all cycles is known in advance. Given the purpose of the model, these are the assumptions on which the model is based:

- An r number of cycles is observed and the demand of any of n customers is known in each of those cycles, $OC_{il},\ i = 1, \dots, n,\ l = 1, \dots, r$. The demand might not be fulfilled and the demand that was not fulfilled is not compensated for in the next cycle;

- Production is limited and constant in all cycles and it equals $PT_l,\ l = 1, \dots, r$. If the production exceeds the total demand of the cycles, the surplus products are stored for the following cycle, so as that the total demand ST_l in any cycles equals $PT_l + max\{0, PT_{l-1} - OT_{l-1}\}$. Inventory holding costs are neglected;

- Customers are grouped according to priority. Lowest-priority customers are a protected group to which the amount of protected products $AP_{il},\ i = 1, \dots, n,\ l = 1, \dots, r$ is allocated in each cycle.

Model variables:

- z_{il} - customer service level, defined as the fraction of customer order demands OC_{il} delivered on time [10],
- AC_{il} - allocated quantity of the customer i in the cycle l.

CSL model:

$$\max \sum_{l=1}^{r} \sum_{i=1}^{n} w_i z_{il} \tag{1}$$

s.t.

$$AC_{il} - z_{il} \cdot OC_{il} = 0,\ i \in \{1, \ldots, n\},\ l \in \{1, \ldots, r\} \tag{2}$$

$$\sum_{i=1}^{n} AC_{il} \leq ST_l,\ l \in \{1, \ldots, r\} \tag{3}$$

$$AC_{il} \geq AP_{il},\ i \in \{1, \ldots, n\},\ l \in \{1, \ldots, r\} \tag{4}$$

$$z_{il} \leq 1,\ i \in \{1, \ldots, n\},\ l \in \{1, \ldots, r\} \tag{5}$$

$$AC_{il} \in \mathbb{Z}^+,\ i \in \{1, \ldots, n\},\ l \in \{1, \ldots, r\} \tag{6}$$

The objective function (1) represents the maximization of the total customer service level, where customers are differentiated by using weights. The w_i parameter, which is the weight coefficient, is used to cluster the customers into priority groups. The optimum solution of the CSL model is extremely sensitive to the given values of this parameter, in particular when production is significantly less than the total demand. The first constraint (2) refers to the percentage and absolute satisfaction of demand. The second constraint (3) models the total demand and the total fulfilment of demand in each of the cycles. The total demand in a cycle is made up of the production in the given cycle and the eventual surplus from the previous cycle. The third constraint (4) allows the creation of a protected partition. The AP_{il} parameter, which is the minimum quantity of products to be delivered to the customer i in the cycle l, represents the reserved quantity for the customer i which is in the protected partition. The value of this parameter for customers outside the protected partition is 0. The value of this parameter has a direct impact on the service level of customers from the protected partition and indirect impact on the service level of customers outside this partition. The last constraint (5) represents the upper bound of the fractional customer service level for the customer i in the cycle l.

3 Computational results and discussion

In order to analyze the impact of the production capacity to the customer service level, the algorithm and the CLS model were tested in three scenarios. All the parameters are the same, except for the production level which equals 1000, 1300 and 1400 FMCG units respectively. For the scenario when the production level reaches 1000 units, the supply is significantly lower than the total demand, for the 1300 units scenario the supply almost meets the minimum total demand. In case of the 1400 units scenario, there are unallocated products only in a few cycles.

Table 1 shows the demand of nine customers over a period of nine weeks. The customers are clustered into three groups. Customers A1 and A2 belong to the first group, while customers B1-B4 make part of the second one. The C-customers are in the third group. The first and the second group of customers are in the first partition, while the third one is in the second, protected partition.

Table 1: Input parameters

Week (cycle)	A1	A2	B1	B2	B3	B4	C1	C2	C3	Demand
W1	330	575	280	110	40	121	100	52	25	1633
W2	360	393	110	170	135	157	74	40	0	1439
W3	220	700	60	100	160	130	100	65	40	1575
W4	230	480	120	140	80	146	74	94	20	1384
W5	270	650	390	110	100	241	140	83	0	1984
W6	381	751	89	140	260	95	110	48	30	1904
W7	320	615	20	120	90	100	46	27	20	1358
W8	390	1.055	120	120	190	130	110	75	0	2190
W9	305	780	290	90	60	110	30	92	11	1768
TOTAL	2806	5999	1479	1100	1115	1230	784	576	146	

The parameter value KP (protective percentage quota) for all three scenarios is 0.95 for the first partition and 0.05 for the second one. These values are based on the decision of the company management which is founded on the realistic assumption that it is necessary to satisfy even the small customers in order to keep them in the system and take advantage of their potential growth. In this way the dependence on major customers would also be reduced. The given KP parameter values are used to determine the AP (amount of allocated products) parameter value . For each customer, the value of AP parameter is set to 5

For the purpose of benchmarking the presented algorithm it was necessary to set the weights in the CSL model that will provide the best balance of the customer service level for the given data. The customers from the protected partition are not included into this sensitivity analysis, as they belong to the lowest-priority group and in any weight distribution their weights are 1. The AP parameter values have a much greater influence on the fulfilment of their demands. The GNU Linear Programming Kit [4] was used to solve the model. The programming kit includes the branch-and-cut algorithm for solving the MIP problem. Figure 1 shows five value variants of weight coefficients. The first value in brackets represents the weight coefficient of the customers from the first group, while the second value is the weight coefficient of the second group. Given the fact that the first group is a higher priority group, the weight coefficients of the customers from this group have to be greater than the weight coefficients of the customers from the second group. The figure shows that the significant difference between weight coefficients has a negative effect on the balancing of the second group. However, small differences have a negative effect on the balancing of the first group. Using the Bisection method has shown that for the given data it is necessary to use the weight 65 for the first customer group and 10 for the second one.

Scenario 1

Table 2 presents the results obtained through the application of the algorithm, while Table 3 contains results of the CLS model optimization. The given production capacity in both cases was 1000 FMCG units. Both tables show the customer service level for every customer over a period of nine weeks. The last row provides the average customer service level for every customer.

By using the CLS model, the objective function value, which represents the weighted customer service level, is 1213.77. By weighting the results of the algorithm, the objective function becomes the value of 1092.77 (90.03% of the optimum value). This is an expected result because the CLS model maximizes the weighted customer service level. However, from the point of view of the company management it is more important that the customer service level for the customers in the same group is balanced.

Based on the results for customers in the first group (A1 and A2), it can be concluded that

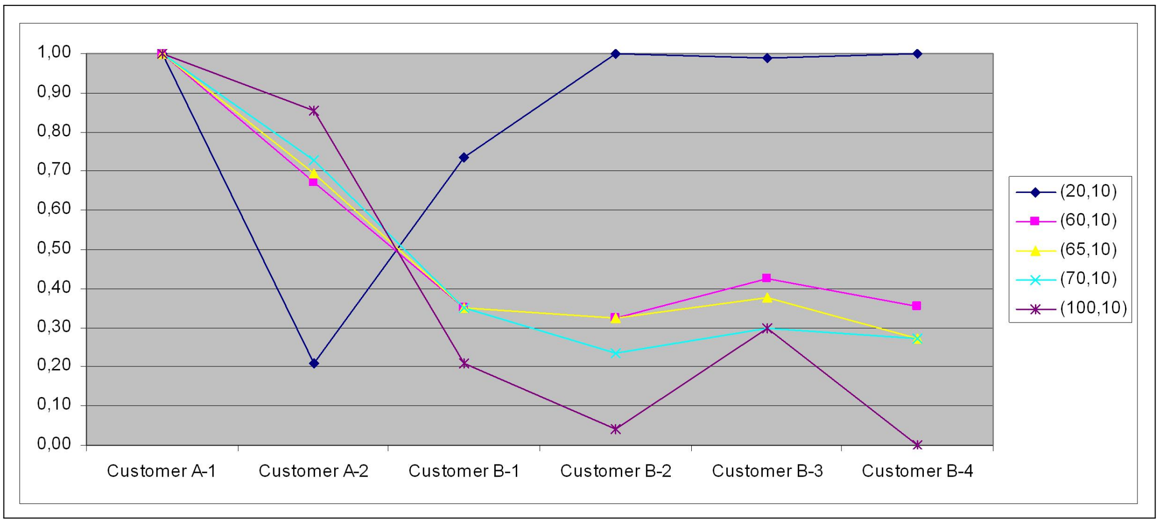


Figure 1: CLS sensitivity to weights

Table 2: Algorithm results for 1000 units

Week (cycle)	A1	A2	B1	B2	B3	B4	C1	C2	C3
W1	1.000	1.000	0.082	0.082	0.075	0.083	0.280	0.288	0.280
W2	0.786	0.784	1.000	0.594	0.274	0.707	0.432	0.450	-
W3	1.000	1.000	0.067	0.070	0.063	0.069	0.240	0.246	0.250
W4	0.843	0.846	0.467	0.664	1.000	0.829	0.270	0.266	0.250
W5	1.000	1.000	0.036	0.036	0.030	0.037	0.221	0.229	-
W6	0.496	0.498	1.000	0.757	0.373	1.000	0.264	0.271	0.267
W7	1.000	1.000	0.050	0.042	0.044	0.050	0.543	0.519	0.550
W8	0.438	0.440	0.158	0.958	0.453	0.731	0.273	0.267	-
W9	0.846	0.887	0.000	0.000	0.000	0.000	0.367	0.380	0.364
AVG	0.797	0.797	0.214	0.400	0.287	0.370	0.293	0.304	0.308

Table 3: Results of CLS model optimization for 1000 units

Week (cycle)	A1	A2	B1	B2	B3	B4	C1	C2	C3
W1	1.000	0.963	0.050	0.055	1.000	0.058	0.170	0.308	0.640
W2	1.000	1.000	1.000	0.053	0.644	0.051	0.230	0.400	-
W3	1.000	0.794	1.000	1.000	0.050	0.054	0.170	0.246	0.400
W4	1.000	1.000	1.000	0.236	1.000	0.055	0.230	0.170	0.800
W5	1.000	1.000	0.051	0.055	0.080	0.054	0.121	0.193	-
W6	1.000	0.487	1.000	0.050	0.050	1.000	0.155	0.333	0.533
W7	1.000	0.829	1.000	0.050	1.000	0.050	0.370	0.593	0.800
W8	1.000	0.187	1.000	1.000	0.053	1.000	0.155	0.213	-
W9	1.000	0.482	0.052	1.000	1.000	1.000	0.567	0.174	1.000
AVG	1.000	0.680	0.384	0.343	0.355	0.311	0.195	0.250	0.623

the use of algorithm keeps the customer service level completely balanced. By using the CLS models, customer A1, whose demand is much lower than the one of the customer A2, has a much higher customer service level. This is because the costs of high customer satisfaction within the same group are minimal when demand is the lowest, because the lowest demand causes the highest increase of the objective function. In the second group, when the CLS model was used, the customer service level for all nine weeks was more balanced than with the application of the algorithm. However, analyzing customer service level by week, it becomes clear that every week the algorithm assigns a certain amount of products to every customer from the second group and that the balance is very good every odd week. This is primarily due to use of tokens. On the other hand, based on the results of the CLS model, it is evident that every week at least one customer from the second group receives protected amount of products. In other words, looking at it by week, the balancing is inadequate. The third group belongs to a protected partition and always gets the guaranteed amount of products. The algorithm provides perfect balancing. However, based on the results of the CLS model the last customer in the third group has the highest customer service level due to its low demand which is often less than the guarantied amount of products. This happens due to the maximization of the satisfaction fractions and not the absolute amount of assigned products.

Scenario 2

The main feature of a scenario with 1300 FMCG units is that the production of each week still does not meet the total demand, but the lack of finished products is lesser than in the first scenario. Table 4 provides the average customer service level for each customer over a period of all nine weeks based on the results of the algorithm and the optimization of the CLS model.

Table 4: Results of the algorithm and the CLS model for 1300 units

	A1	A2	B1	B2	B3	B4	C1	C2	C3
Algorithm	0.937	0.929	0.539	0.706	0.556	0.675	0.434	0.452	0.429
CLS model	1.000	0.870	0.412	0.792	0.497	0.891	0.305	0.328	0.808

The weighted overall customer service level based on the results of the CLS model equals 1416.17. The results of the algorithm make this value reach 1361.42 (96.13 % of the optimum value). The balancing obtained by applying the algorithm is better than the balancing obtained through the model and it is even more prominent than in scenario 1.

Scenario 3

When the production reaches 1400 FMCG units, in two weeks (W4 and W7) the supply exceeds the demand and the surplus products are stored for the next week. The average customer service levels are shown in 5. Based on the results, the weighted overall customer service level for the CLS model equals 1463.77 and 1419.76 for the algorithm (96.99% of the optimum value).

Table 5: Results of the algorithm and the CLS model for 1400 units

	A1	A2	B1	B2	B3	B4	C1	C2	C3
Algorithm	0.971	0.968	0.627	0.733	0.575	0.765	0.458	0.542	0.548
CLS model	1.000	0.924	0.477	0.879	0.581	0.982	0.383	0.521	0.836

Looking at the data from the three scenarios above, it can be concluded that by increasing the production the total value of customer service level obtained through the algorithm, is nearing the optimum value. If observed from the balancing point of view, the advantage of the algorithm is increasingly more prominent compared to the CLS model.

4 Conclusions and Future Works

This paper presents benchmarking of a heuristic algorithm for the dynamic solving of the problem of allocating limited supplies of products to received customer orders with the aim of maximizing the customer service level. In order to validate the algorithm an MIP model was developed. The model is used for the distribution of products based on demands that are known in advance for the entire period. Computational results have indicated that the proposed algorithm, with the increase in the production capacity, ensures a value of the total customer service level that is closer to the optimum values obtained using the CLS model. In addition, in all three scenarios the balancing results of service level for customers from the same group achieved with the algorithm were better than the balancing obtained using the CLS model.

Further analysis of customer demand by week and the obtained customer service levels have shown that the further research might have to be directed towards the analysis of the correlation between the customer service level and the fluctuations in demand. Besides, the algorithm could be modified by introducing the assumption that the selling price depends on the customer affiliation to a certain group or on the quantity of products ordered. The growth rate of a customer's demand is one of the essential elements used by the management in manufacturing companies in planning the sales. The inclusion of this parameter into the problem would require the modification of the algorithm making it a more useful tool in the decision-making process.

Acknowledgement

This research was partially supported by the Ministry of Education and Science, Republic of Serbia, Project number: TR35045.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):11-19, December, 2012.

Bibliography

[1] Cederborg O., Rudberg M., Customer Segmentation and Capable-to-Promise in a Capacity Constrained Manufacturing Environment, *16th Int. Annual EurOMA Conference*, Goteborg, Sweden, 2009, http://www.iei.liu.se/prodek/forskning/iscaps/filarkiv /1.120209/CederborgandRUdbergEurOMA2009.pdf, accessed 12 January 2010.

[2] Chan F.T., Chung S.H., A Modified Multi-Criterion Genetic Algorithm for Order Fulfillment in Manufacturing Network, *Proceedings of the 9th Asia Pacific Industrial Engineering & Management System Conference*, APIEMS, Indonesia, 2221-2226, 2008.

[3] Chen C., Lee W., Multi-objective optimization of multiechelon supply chain networks with uncertain product demands and prices, *COMPUT. CHEM. ENG.*, ISSN 0098-1354, No 28: 1131-1144, 2004.

[4] GLPK - GNU Linear Programming Kit. http://www.gnu.org/software/glpk, accessed 25 December 2011.

[5] Lecic-Cvetkovic D., Atanasov N., Babarogic S., An Algorithm for Customer Order Fulfillment in a Make-to-Stock Manufacturing System, *INT J COMPUT COMMUN*, ISSN 1841-9836, 5(5): 983-791, 2010.

[6] Lin J., Chen J.H., Enhance Order Promising with ATP Allocation Planning Considering Material and Capacity Constraints, *JCIIE*, ISSN 2151-7606, 22(4): 282-292, 2005.

[7] Merzifonluoglu Y., Geunes J., Uncapacitated production and location planning models with demand fulfilment flexibility, *INT J PROD ECON*, ISSN 0925-5273, 102: 199-216, 2006.

[8] Meyr H., Customer Segmentation, Allocation Planning and Order Promising in Make-to-Stock Production, *OR SPECTRUM*, ISSN 01716468, 31(1): 229-256, 2009.

[9] Minner S., A comparison of simple heuristics for multi-product dynamic demand lot-sizing with limited warehouse capacity, *INT J PROD ECON*, ISSN 0925-5273, 118: 305-310, 2009.

[10] Pochet Y.,Wolsey L.A., *Production Planning by Mixed Integer Programming*, Springer, 2010.

[11] Sawik T., Integer programming approach to reactive scheduling in make-to-order manufacturing, *MATH COMPUT MODEL*, ISSN 0895-7177, 46(11-12): 1373-1387, 2007.

[12] Xiao Y., Taaffe K., Satisfying market demands with delivery obligations or delivery charges, *COMPUT OPER RES*, ISSN 0305-0548, 37(2): 396-405, 2010.

[13] Xiong M.H. et al, A DSS approach to managing customer enquiries for SMEs at the customer enquiry stage, *INT J PROD ECON*, ISSN 0925-5273, 103(1): 332-346, 2006.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):20-29, December, 2012.

Memetic Engineering for Permanent Education in Line with Sustainable Growth

C.I. Brumar, R.D. Fabian, M.-J. Manolescu, V. Chiş

Cristina I. Brumar, Ralf D. Fabian
Lucian Blaga University of Sibiu
Romania, 550024 Sibiu, 10 Bd.Victoriei
E-mail: crista.brumar@gmail.com, ralf.fabian@gmail.com

Mişu-Jan Manolescu
Agora University, Romania,
Piata Tineretului, 8, 410526 Oradea, Romania,
E-mail: mmj@univagora.ro

Violeta Chiş
Aurel Vlaicu University of Arad
Faculty of Exact Sciences
Department of Mathematics-Informatics
Romania, 310330 Arad, 2 Elena Dragoi
E-mail: viochis@yahoo.com

Abstract:
Given the recent point of view of the European Commission regarding the implementation of a new strategy for sustainable growth and jobs, this paper emphasises the opportunity and urgency of supporting the EU 2020 strategy, providing an appropriate educational tool for the knowledge society. The main objectives are: a. adapting memetic engineering expressed in terms of General System Theory to the teaching component of permanent education based on bounded rationality and "Just in Time" as key tools for fighting cognitive chaoplexity in the post-industrial era; b. facilitating the use of memetic engineering based on its double-faceted nature: as both positive and negative feedback; c. extending the applicability of memetic engineering to ecology as source of memes; d. exemplifying the above in primitive metamodels applying memetic engineering in ecology and highlighting the relevant design-space dimensions. Among the conclusions: a. to be sustainable in the long run permanent education must be modelled in line with learner bounded rationality, since bounded rationality is a psychological lasting feature; b. sustainable development depends on affordable permanent education; c. as a result, e-teaching should be systematically revisited through intense transdisciplinary research.

Keywords: sustainable growth, permanent education, bounded rationality, chaoplexity, memetic engineering.

1 Introduction

Given the recent point of views of the European Commission (March 2, 2012, Brussels [13]) regarding the implementation of a new strategy that "pursues both continued fiscal consolidation and determined action to boost growth and jobs; sustainable growth and jobs cannot be built on deficits and excessive debt levels" [13], this paper follows the requirements and urgency of implementing the strategy [13], namely strengthening and development of actions to achieve continuous growth of services and economy. At the European level, after renewing the Lisbon strategy through the EU 2020 strategy, lifelong learning is regarded as key pillar of sustainable growth. Permanent education based on skills acquired through dynamic knowledge, recognizes the failure of static knowledge-based lifelong learning through Learning Objects Repositories [16].

Since education is seen as a complex service in a post-industrial era, this paper aims to extend to ecology the general educational tool for knowledge society as proposed in [20], [19], [16], [6].

Dominant features of the post-industrial era are summarized and explained in [10] as: "a) Growing speed of change (due to the intense positive feedback entailed by Moore's law outcomes: Internet, broad-band technology, semantic Web, Google, etc.). b) Growing complexity (architectural, cognitive, structural). c) Globalization (expressed in IT context mainly through the modern enterprise paradigms). [...] Thus, modern IT environments, except for simple applications, move towards open and heterogeneous (resources are unalike and their availability is not warranted), dynamic (the pace of exogenous and endogenous changes is high) and uncertain (both information and its processing rules are revisable, fuzzy, and uncertain). Most situations to be controlled are complex and uncertain, and involve parallel processes." [10].

Thus, after updating related work in Section 2, the paper tracks the first objective in Section 3 investigating the main concepts related to memetic stability. On this groundwork, as bounded rationality and "Just-in-Time" were tools for e-teaching [16], Section 4 is build as paper core: it launches the concept of memetic engineering regarded as tool for research in ecology - since it matches with e-teaching environmental science. Section 5 highlights the design-space dimensions of a (primitive) metamodel for the "integral ecology" [12]. Conclusions and future work (Section 6) close the paper.

2 Rationale and Related Work

Considering the far range target of sustainable growth, the milestones are set up by the Europe 2020 strategy: "and if Europe is to emerge stronger from the crisis, we need more than ever to stimulate smart and green economic growth, underpinned by knowledge and innovation as its key drivers. [...] we need to focus on policies that give a chance to everyone to develop one's skills and live in dignity. [...] Our common aim is to provide a sustainable response to the many challenges facing us to transform the European Union into a knowledge-based, resource efficient and low-carbon economy" [2]. That means that the students have to learn how to be skilled (chaoplexity teaching for permanent education based on bounded rationality), how to reduce the ecological damage made by economic growth (ecological systems based on memetic engineering), and we need to prepare them for what comes next (students have to live based on very different principles).

As regards the concept of meme, a short definition might be: a meme is a unit of information residing in the brain and is the mutating replicator in human cultural evolution (Richard Dawkins, [8]). Slogans, riddles, songs, signs, inventions and fashions are typical memes. An idea or information pattern is not a meme until it causes someone to spread, to transmit it to another. "Why another new "....tics"? Memetics is a key Zeitgeist-component and is transdisciplinary par excellence; The "Self" memotype shows a cognitive complexity similar to the structural one of the genotype; The "Self-*" memeplex invaded modern IT ("star" = awareness) [...]; The thick-time meme is vigorous and old (preceded AI itself)" [1].

"According to the theory Memetic Engineering is, simply put, the analysis of an individual or individual's behaviour, the selection of specific memes and the distribution or propagation of those memes with the intent of altering the behaviour of others". [http://en.wikipedia.org/wiki/Memetic_engineering]. Including memetics in this paper (as shown in [16]) is reasonable because most paradigms in modern artificial intelligence have a memetic character and this phenomenon is ever more important in an increasing chaoplexic world [16]. Memes are transdisciplinary par excellence and they originate from myths that represent valuable sources of metaphors - an obvious source of stability.

Hence, the research about the role of permanent education for sustainable growth as well as the role of memetic engineering in permanent education [19], [16], [20], [6] has been continued in line with the following objectives: a) adapting memetic engineering expressed in terms of General System Theory to the teaching component of permanent education based on Bounded Rationality and "Just-in-Time" as key tools for fighting cognitive chaoplexity in post-industrial era [16]; b) facilitating the use of memetic engineering based on his double faceted nature: as negative feedback (through memetic stability) and positive feedback (through intense spreading of memes); c) extending the applicability of memetic engineering from myths and metaphors, including (paleo)linguistics [16], to ecology as source of memes; d) exemplifying the above in primitive metamodels applying memetic engineering in ecology and highlighting the relevant design-space dimensions. Sustainable growth depends on permanent education; this involves memetic engineering and bounded rationality, and requires changes in the three referred fields: cognitive psychology, IT, and Higher Education.

For the sake of self-containment the paper summarises some results of [16], [6], [20] underlining the transdisciplinary aspects related to: psychology, semantics, vital service, skill oriented and self-organizing (lifelong) learning, corpus of knowledge, intensely dynamic and uncertain environments.

3 Main Concepts related to Memetic Stability

Norbert N. Seel in "Encyclopedia of the Sciences of Learning" [22] proposed an interdisciplinary overview about learning and the connections between different fields of sciences: "Over the past century, educational psychologists and researchers have posited many theories to explain how individuals learn, how they acquire, organize and deploy knowledge and skills. [...] As the learning sciences became more specialized and complex, the various fields of interest were widely spread and separated from each other; as a consequence, even presently, there is no comprehensive overview of the sciences of learning or the central theoretical concepts and vocabulary on which researchers rely. Learning theories are not limited to psychology and related fields of interest but rather we can find the topic of learning in various disciplines, such as philosophy and epistemology, education, information science, biology, and - as a result of the emergence of computer technologies - especially also in the field of computer sciences and artificial intelligence." [22].

Three memetic features should be focused on to ease applying them in permanent education [16]: a) *lastingness* (they are a leitmotif in cultural history); b) *ubiquitousness* (they permeate all cultures); c) *effectiveness* (they are active now in education) [16], [6]. Even without prior knowledge of psychology it is well-known that behaviours are extremely lasting. Hence, they can illustrate the role of Bounded Rationality as "educational mechanism" much more convincingly than widespread topics - albeit very famous. The three steps are similar to those above: a) choosing the pervasive habit of scoring (instead of counting) - scoring was - and still is - certainly easier than counting; b) investigating the related memeplex; c) proposing a boundedly rational way to exploit simplicity in e-teaching.

Memetic engineering for permanent education tries to integrate transdisciplinary sciences. In recent papers [19], [20] bounded rationality and "Just-in-Time" were used as tool for e-teaching; this paper aims at proving that memetic engineering for permanent education is a tool for e-teaching environmental science. Ecologists use memetic stability because they want to preserve the environment.

As regards the concept of *chaoplexity*, it reveals its value in military operational research: "Chaoplexic warfare draws on the study of nonlinear phenomena of self-organization to propose a radical decentralization of armed forces through the adoption of the network form. [...] Information remains the central concept, and in this sense chaoplexity is an outgrowth of cybernetics; but the focus on change, evolution and positive feedback breaks with the cybernetic pioneers' concern for stability" [4]. To add a flavour to the syntagm "educational chaoplexity" (EDCHY): "In association with the respective technologies of the clock, engine and computer, the scientific theories of mechanism, thermodynamics, and cybernetics have all in turn been recruited to shape distinct approaches to the challenges of imposing order on the chaos of the battlefield. Today, it is on the basis of the new sciences of chaos and complexity that the latest regime of the scientific way of warfare is being erected" [4], [19]. In the abbreviation EDCHY, the focus is on the letters d and h because education has to take into account that we live in a dynamic and heterogonous environment. Concerning ecology, even if the time span is larger implied in EDCHY regarding learning, climatic changes take place during thousands of years. Hence "Just-in-Time" measures should be taken to ensure environment preservation - an obvious chaoplexic issues.

In short: permanent education relies on bounded rationality and "Just-in-Time". Bounded rationality can be used as a tool to educational chaoplexity and as common denominator for the two facets of permanent education, namely *e-teaching* and *e-learning*. "Just-in-Time", as connotation of "real time", is related with the role of *response* time (vital in any teacher-learner interaction) [5]. Permanent education involves "Just-in-Time" synchronization between e-teaching and e-learning and is basic for the shift from product-based to a service-based society [16], [6]. The links to memetic stability will become evident in the next section.

4 Double Faceted Nature

In line with common linguistic usage, positive and negative has connotations different from the scientific ones: stability (as negative feedback) has positive effect while creativity (as positive feedback) could become extremely risky - if not restricted.

4.1 Negative Feedback through Memetic Stability

"Negative feedback helps to maintain stability in a system in spite of external changes. It is related to homeostasis. For example, in a population of foxes (predators) and rabbits (prey), an increase in the number of foxes will cause a reduction in the number of rabbits; the smaller rabbit population will sustain fewer foxes, and the fox population will fall back." [http://en.wikipedia.org/wiki/Feedback]. On the same sight appear other correct sentences, but using the contrasting connotation: "In an electronic amplifier feeding back a negative copy of the output to the input will tend to cancel distortion, making the output a more accurate replica of the input signal" or even "Positive feedback amplifies possibilities of divergences (evolution, change of goals); it is the condition to change, evolution, growth; it gives the system the ability to access new points of equilibrium" [Wikipedia]. [...] Some assertions need badly to be improved (e.g., "Negative feedback, which tends to reduce the input signal that caused it, is also known as a self-correcting or balancing loop. [...] The terms negative and positive feedback can be used loosely or colloquially to describe or imply criticism and praise, respectively. This may lead to confusion with the more technically accurate terms positive and negative reinforcement, which refer to something that changes the likelihood of a future behaviour." [Wikipedia]).

Almost always in nature (e.g., *homeostasis* in living beings) and very often in technology (e.g., reducing noise and message *distortion* in communication systems), the target looked for is *preservation* (e.g., in living systems to prevent decay), or stability (e.g., in artificial systems to prevent deterioration). This basic kind of feedback is called - for obvious historical and physical reasons - negative feedback [16].

For permanent education, negative feedback is *sine qua non*, above all for the teaching process because it is *corrective*, promotes *stationariness, stability,* and *reversibility*; (e.g. ecologists are based on traditions and, some time, on people's inertia).

In short: IT and memetic engineering provides to ecologists tools to handle environmental issues and to decide what is sustainable or not sustainable for them. Memetics helps indirectly in the sense of tradition, because ecology implies environment stability.

4.2 Positive Feedback through Spreading of Memes

Positive feedback means creativity; using positive feedback, the parameter values are increased, the whole process is evolutive, innovative, generating chain reactions that increase instability and involve irreversibility. Concerning ecology, positive feedback refers to the countermeasures to be taken to oppose/face environment degradation. (E.g. Since the environment is preserved by fighting/opposing nuclear energy, specialists have to be creative to find methods for supplying energy without destroying the environment).

With no doubt language and culture are close related and have been subject of sociological, anthropological and memetic research [3], [8]. Language is determined by culture (ancient civilisations did not have words for computer, internet, phone, television, radio, etc.) but culture is determined by language too, if memes are regarded as fundamental replicators for their evolution.

Richard Dawkins emphasised in "The Selfish Gene" [8] the importance of thinking about evolution in terms of information that can be transferred from one person to another via imitation [8]. On this groundwork Susan Blackmore in "The Meme Machine" asserts: "Just as the design of our bodies can be understood only in terms of natural selection, so the design of our minds can be understood only in terms of memetic selection." [3] According to her, language developed as result of memetic evolution and is the principal medium used for spreading memes, i.e. replicating themselves in as many minds as possible.

Traditionally, spreading of memes was through word-of-mouth followed by written word and printed word. Today memetic propagation via language takes advantages of globalization and of a wide media spectrum (i.e., radio, television, internet, etc.) and of increasing propagation speed due to intense positive feedback (id est., Moore's Law and the new technologies based on it) [10].

Talking about internet, it is impossible to ignore the meme spreading power of search engines. Without them information would be hard to reach. Moreover, by email, newsgroups, message boards and social networks people can circulate their memes allowing adherents to unpopular/popular memes to come together ideologically. Companies already promote memes this way to advertise their products by *viral marketing* strategies.

Internet and IT&C in general "is becoming a tool for social interaction bridging the strands between online and offline activities, respectively, digital and social behaviour. [...] Information and Communication Technologies (ICT) experienced in people's everyday life sets a milestone for an active participation in the Knowledge Society" [17]. At European level, it is still a basic pillar of the strategy for growth: "High-speed Internet underpins all sectors of the economy and will be the backbone of the Digital Single Market. For every 10% increase in the broadband penetration the economy grows by 1 to 1.5%." [14]. Of course, this stimulates the speed of meme propagation.

Memes are spread by copying and suffer modifications in their evolution - similar to genes. Thus, the newness, the sparkle of creativity is within *mutation*, since it exceeds the possibilities of simple copying from *crossover* (when Archimedes came out of bath exclaiming "Eureka" he was not copying anyone). Thinking in terms of Genetic Algorithms, crossover is perceived as negative feedback, preserving stability by selecting and copying fittest the properties from parents to their offspring, whereas positive feedback manifests itself creatively in mutations, giving rise to new unpredictable, hence non-deterministic outcome. Corollary: memetic evolution (and also genetic algorithms) involves intrinsically non-determinism. That is why non-deterministic software is used in IT [16] - to be able to exploit positive feedback too.

In short, technology in first place, only eases rapid spreading and gives no clue about the effect of a meme. To decide if a meme is toxic or not, is still up to the end-user.

The contradiction between negative feedback - fundamental for stability required by ecologists - and positive feedback - fundamental for growth - can be solved only by approaching the problem through General System Theory (GST). Any sectorial, partial, local or parochial solutions for solving the problems are inappropriate, because these solutions are not only for a limited biocenosis but have to be for a global ecosystem. Consequently, holistic analyses of situations are *sine qua non* and the tools for handling these situations must also be holistic [16]. Examples and explanations inspired by *integral ecology* [12] are abridged below.

Based on a large public acceptance of renewable energy, the hype around wind energy developed significantly in the last years, promoting wind turbines and wind farms as one ecological sustainable solution for green energy [18]. A recent report praises the wind energy as "recession-busting industry" for its impact as "Green growth" on jobs and economy [15]. Beside benefits of green energy, e.g. no air pollution, moderate capital cost, low ongoing costs, energy independence, several side-effects came to fore. Wind farms need to be placed in wide areas for great efficiency. Reports of bird and bat mortality argue that wind turbines are artificial structures that interfere and destroy ecosystems, especially in regions where there are important flight paths for migratory bird [21], [9]. Furthermore, aesthetic impact, sound emission or even increase of night time temperatures [7] are part of the price to pay when it comes to green energy. To which extend the ecological impact is significant or not is still in question but it clearly leads to an "Environmental Paradox" [11] that ecologists have to deal with.

As regards environment, preservation is a legitimate endeavour whereas wild altering proved to be harmful. Memetic stability is consistent with the requirement for stability of environment preservation whereas creativity, in unconventional technological approaches, is shaping hope for finding proper solutions.

Climate change is linked subjective and objective to "Just-in-Time". That is, ecology has as main target conservation whereas climate change is perceived memetically as slowly. In this context "Just-in-Time" could be understood as a secondary restriction since there is enough time to fight climate change and find solution. Such arguments are perhaps convenient for emerging industries, pushing limits, conventions and boundaries as far as possible in the future (e.g. Kyoto Protocol). On the other hand, relying on the meme that climatic changes are slowly and considering that the last significant climate change of human history, the glacial period, dates of approximate 20 millenniums ago, means ignoring heavy side effects, like Hurricane Katrina or El Niño phenomena, with evolutions visible even in periods shorter than a human lifespan. Hence, solutions are addressable only through bounded rationality, "Just-in-Time" and chaoplexity together.

That is why holistic approaches are needed; any reductionist view leads to solutions that could hardly satisfy ecosystem stability and sustainable growth.

5 Metamodel

Nowadays, the world's natural resources are under pressure. Actions as soil erosion, acid rain, the extinction of species, have all contributed to the environmental system damage. "Economic production influences the environment in many ways, through the consumption of energy and natural, often non-renewable resources, and the production of pollution, toxic wastes, etc. They further stressed (not without opposition) that present environmental problems require a new type of development process which harnessed the benefits of economic growth without the damaging consequences which growth can have on the environment. Up to now, technology has been developed for the sole purpose of increasing economic and social standards, with little or no regard for its potential negative impact on the environment (e.g. exhaustion of non-renewable resources, extinction of species, eutrophication, acidification, ozone-depletion, etc.)" [23].

Understanding human diversity through cultural value and thinking, enables an approach to education that explains different outcomes of choices people make. The challenge is for explaining the thinking that led to certain behaviour and not for behaviour itself. Since memes being decision-systems transcending culture, society and evolution, the struggle is with the memes in humans that are at rumour. To name some memes in ecologic field: nuclear power plant Fukushima - economic efficiently but ecological disaster; whaling - coastal communities have long tradition and industrial level emerged as technology increased, but whaling is immoral, unsustainable, and should be banned.

As regards Memetic Engineering as antidote to vicious memes, it has been shown that "folklore can be harmful when myths are not dismantled before being disseminated as memes. The need to counteract them is obvious and memetic engineering is just the newest arm in the well assorted panoply of persuasive means, from ancient rhetoric to modern aggressive advertising." [16].

With the growing economy, we use a lot of resources which will regenerate in years, and therefore, if there are not taken certain measures from an ecological point of view, we exhausted all resources. Today's youth must be trained to meet this challenge, and for that we need to "e-teach" environmental science. The field of ecology is manifold and should be treated from a transdisciplinary point of view. Hence, it is appropriate to create a primitive metamodel for ecology (its relevant design-space metadimensions are shown in Figure 1). Sean Esbjörn-Hargens in his article [12] wrote about the existence of four correlated dimensions to approach ecological issues. "Integral Ecology inquires into all for quadrants, or four Terrains: Behavioral Terrain (behaviors at all levels of organization), Experience Terrain (experiences at all levels of perception), Systems Terrain (systems at all levels of ecological and social intersection), and Cultural Terrain (cultures at all levels of mutual resonance and understanding)" [12].

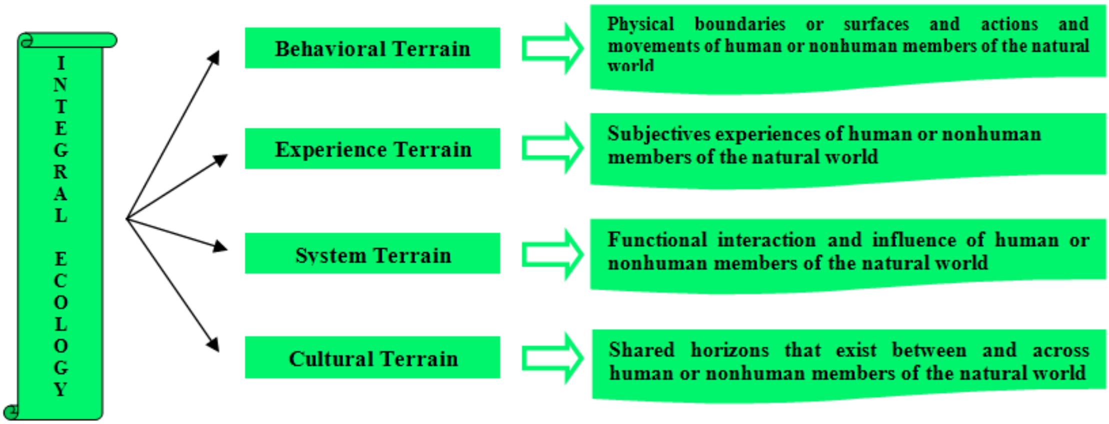

INTEGRAL ECOLOGY
Behavioral Terrain
Physical boundaries or surfaces and actions and movements of human or nonhuman members of the natural world
Experience Terrain
Subjectives experiences of human or nonhuman members of the natural world
System Terrain
Functional interaction and influence of human or nonhuman members of the natural world
Cultural Terrain
Shared horizons that exist between and across human or nonhuman members of the natural world

Fig. 1 Design-space (meta)dimensions.

In short: memetic engineering based on bounded rationality and "Just-in-Time" fits in all four terrains, but of course is the right and the duty of the end-user (i.e. researchers in ecology) to choose and to prioritise those terrains or maybe some blend of them.

6 Conclusions and future work

Starting of the outcomes of [16], the conclusions are: a) to be sustainable in the long run permanent education must be modelled in line with learner bounded rationality, since bounded rationality is a psychological lasting feature; b) for sustainable development e-teaching should be systematically revisited through intense transdisciplinary research; c) in line with the EU 2020 strategy, education for sustainability involves also to develop metamodels in the field of integral ecology using memetic engineering to alleviate the apparent opposition between economic growth and environment preservation; d) time has to be dealt with carefully since teaching and learning take place in different temporal frameworks. Any educational process that is changing and adapting every day has to be in accordance with the new strategy; memetic stability helps creating a new king of e-teaching. Some corollaries are: a) IT provides ecologists, in their attempt to preserve environment, a useful tool: memetic engineering is applied to put to work useful memes and to avoid vicious memes; b) to be sustainable on the long run lifelong learning needs e-teaching based on innovation and "out of the box" thinking; c) IT tools avoiding nondeterministic software are inefficient because of the evolutionary unpredictable processes, both memes and genes are based on.

As regards future work, to be reliable, the research should be separated regarding *research object* (ecology) and *method* (memetic engineering). In this respect, the intention is to develop methods based on memetic engineering where memes are more appropriate applied, namely in Decision Making adapting the way decisions will be described in a future work dedicated to strategic decisions for environment preservation.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):20-29, December, 2012.

Bibliography

[1] Bărbat, B.E., NEWTON, HUSSERL, WIENER: A Temporal Golden Braid (Invited paper at Int. Conf. on Comput. Commun. and Control, ICCCC 2010, Oradea), in *Abstracts of ICCCC Papers*, ISSN 1844-4334, pp. 12. 2010.

[2] Barroso J. M. D., Urban areas, drivers of growth and jobs, *5th European Summit of Regions and Cities*, Copenhagen, 22 march 2012.

[3] Blackmore S., *The Meme Machine*, Oxford University Press, 2000.

[4] Bousquet A., Chaoplexic warfare or the future of military organization, *International Affairs*, 84(5):915-929, Wiley Online Library, 2008.

[5] Brumar C. I., R. D. Fabian, Bărbat B.E., CSITAO Carnap-like Glossary: http://bcu.ulbsibiu.ro/digitale/doctorate/glossary_csitao.pdf.

[6] Brumar C. I., Sustainable Development in spite of Educational Chaoplexity. State of the Art, *First Technical Report for the PhD Thesis titled "Nondeterministic e-Teaching in Uncertain, Dynamic Environments. Experimental Model based on Memetic Engineering"*, LBUS, 2011: http://bcu.ulbsibiu.ro/digitale/doctorate/Cristina_Brumar_Ref1_Presentation.pdf.

[7] Carrington D., Wind farms can increase night time temperatures, research reveals, *The Guardian*, April 29, 2012, http://www.guardian.co.uk/environment/2012/apr/29/wind-farms-night-temperatures-study.

[8] Dawkins, R., *The Selfish Gene (30th Anniversary edition)*, Oxford University Press, 2006.

[9] Diac M., Proiectele eoliene sunt tot mai controversate din punctul de vedere al protectiei mediului, *Green Report*, November 25, 2011, http://www.green-report.ro/stiri/proiectele-eoliene-sunt-tot-mai-controversate-din-punctul-de-vedere-al-protectiei-mediului(in Romanian).

[10] Dzitac I., Bărbat B.E., Artificial Intelligence + Distributed Systems =Agents, *INT J COMPUT COMMUN*, ISSN 1841-9836, 4(1):17-26, 2009.

[11] Eilperin J., Mufson S., Renewable Energy's Environmental Paradox, *Washington Post*, April 16, 2009.

[12] Esbjorn-Hargens S., Integral Ecology, A post-metaphysical approach to environmental phenomena, *AQAL - Journal of Integral Theory and Practice*, Spring 2006, 1(1):305-378: http://jfk-integral-life.up.seesaa.net/image/Vol1_No1_Final_02_11_07_opt.pdf.

[13] European Council, Conclusions 1/2 March 2012, Brusells, http://www.consilium.europa.eu/uedocs/cms_Data/ docs/pressdata/en/ec/128520.pdf.

[14] European Union, *Digital Agenda: Commission opens public consultation on how to reduce the cost of rolling out high speed internet*, Press Release 27.04.2012: http://europa.eu/rapid/ pressReleasesAction.do?reference=IP/12/434&format=HTML&language=EN.

[15] European Wind Energy Association - report, *Green growth - the impact of wind energy on jobs and the economy*: http://www.ewea.org/fileadmin/ewea_documents/documents/publications/ reports/-Green_Growth.pdf, (retrieved April, 2012).

[16] Fabian R. D., *Bounded Rationality in Agent Orientation - "Just-in-Time"Visual Pattern Recognition*, PhD Thesis in Computer Science and Information Technology, Sibiu, 2011, Copyright: LBUS, Ralf D. Fabian, http://bcu.ulbsibiu.ro/digitale/doctorate/Ralf_Fabian_Phd_Thesis.pdf.

[17] Fabian R.D., M.J. Manolescu, L. Galea, G. Bologa, Bounded Rationality through the Filter of the Lisbon Objectives, *INT J COMPUT COMMUN*, ISSN 1841-9836, 5(5):710-718, 2010.

[18] Global Wind Energy Council, *Global Wind Report, Annual Market update 2011*, http://www.gwec.net/fileadmin/documents/NewsDocuments/Annual_report_2011_lowres.pdf.

[19] Oprean C., Brumar C. I., Canter M., Bărbat B. E., Sustainable Development: E-teaching (now) for Lifelong e-Learning, *Procedia - Social and Behavioral Sciences*, ISSN: 1877-0428, ELSEVIER, pp. 988-992, 2011, http://www.sciencedirect.com/science/article/pii/S1877042811020179

[20] Oprean C., Fabian R. D., Brumar C. I., Bărbat B. E., Bounded Rationality for "Just in Time" Education, *Procedia - Social and Behavioral Sciences*, ISSN: 1877-0428, ELSEVIER,pp. 983-987, 2011: http://www.sciencedirect.com/science/article/pii/S1877042811020167.

[21] Pearce-Higgins J. W. et al., Greater impacts of wind farms on bird populations during construction than subsequent operation: results of a multi-site and multi-species analysis, *Journal of Applied Ecology*, 49(2):386-394, April, 2012.

[22] Seel N. M., *Encyclopedia of the Sciences of Learning*, Springer, 2011.

[23] Ulhoi J. P., Henning M., Sustainable Development and Sustainable Growth: Conceptual Plain or Points on a Conceptual Plain? http://www.systemdynamics.org/conferences/1999/PAPERS/PARA197.PDF.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):30-38, December, 2012.

Disaster Prevention Integrated into Commonly Used Web Rendered Systems with GIS Capabilities

M. Cioca, S.C. Buraga, C. Cioranu

Marius Cioca
Lucian Blaga University of Sibiu
E-mail: marius.cioca@ulbsibiu.ro

Sabin-Corneliu Buraga
A. I Cuza University of Iasi
E-mail: busaco@infoiasi.ro

Cosmin Cioranu
UEFISCDI
E-mail: cosmin.cioranu@gmail.com

Abstract: The end of the 20th century brought a remarkable increase in the field of positioning techniques and communications, making them visible and available to the public, which led to an unprecedented interconnectivity. At the same time, disasters are part of our life. Regardless of their nature, measures can be taken, in order to prevent and mitigate their effects, by anticipative preparation or by avoiding the calamity area (if possible). To this end, this paper presents an integrated system, composed of a software component, a hardware component, and a decision-making human element, all having the declared role of diminishing or eliminating human and material losses.

Keywords: : DSS, Disaster, Open Layers, Apache.

1 Introduction

The end of the 20th century brought a remarkable increase in the field of positioning techniques and communications, making them visible and available to the public. Among them, one of the most spectacular - which revolutionized the way in which we relate to our planet - was in fact the introduction of GIS (Geographical Information System) positioning technologies on the consumer level, which were actively supported by communities of developers and which transposed this knowledge into a format that the ordinary user can access, starting with the mobile ones and ending with the old and well-known PC. (e.g. Google Earth) [16]. Fire, floods, earthquakes, storms, disasters caused by human negligence can occur and can give rise to irremediable damages, but in some cases, material and human losses can be avoided or, in the worst case, diminished through a set of measures that can be assisted or implemented by means of an information system with spatial capabilities, that collects data from different sources in order to develop a prevention system for such events, that finally has an dissemination role, that can act through deployment of resources, so that the event or the events concerned affect the normal course of life as little as possible [18].

2 Decision Support Technologies and Tools

The complexity of sustainable development requires rational decision-making, and decision-making becomes increasingly difficult especially regarding environmental issues. The progress in decision-making theory and decision support systems have determined the emergence of methods and tools that can assist the decision-maker in making the most adequate decisions. However, in helping decision-making on complex issues, these tools are not easily developed and built [1], [7], [8], [10], [11] and [12] . The use of general accepted techniques and methods in DSSs may improve the identification and prevention of complex risk situations. These imply making strategic decisions supported by expert boards and groups. Their activity is often burdened with physical, time and cognitive barriers. Moreover, these methods and techniques used by DSSs represent only general approaching instructions on specific decision-making situations. In practice, in order to be used efficiently, they need reinterpretations, refinements, adjustments and completions.

3 Technical Approach

To this end, this paper presents an integrated system (fig.1), informationally and spatially scalable from a community to a certain area, that stratifies both decisional and informational passive and active elements, in order to prevent human losses and/or danger situations, generated by natural or human calamities (floods, radiations, earthquakes etc.). For harmonization, these elements have been grouped into passive elements, meaning those elements that provide information related to environment, active elements, which in this case, are human decision-makers, that have as primary attributes the position in latitude and longitude system, as well as their type. Among the active elements of the system, we mention a list of spatially-localized decisional personnel, attached to each regional entity (village, commune, county), that can be alerted when the system detects a case of disaster, through specific algorithms. One of the main elements, of a great importance, are the specific approved sensors, both fixed and mobile, placed on the entire area of focus, that measure events like humidity level, landslides, earthquakes, radiations etc. These are equipped with point-to-point communication equipment like GSM, radio or other methods that ensure information transmission to the decisional cluster. Another passive element that brings a better accuracy to the system are the meteorologic maps, but which imply a high data volume, meaning that these should be served to the system in a semi-processed manner. The major problem of this system was the link between these elements, their harmonization being accomplished by integrating these data into a decision-making neural network, which concentrates the spatial information by a number of minimization operations, so that the result can be transposed into a GIS capable system and that generates a well-determined spatial answer. In such an approach, methodologies for this kind of wireless sensor-systems have to be used [4], security protocols for wireless sensor networks [2] have to be considered, as well as efficient algorithms, specific to this kind of mobile sensor networks [3].

4 Computer System Implementation

From the informational standpoint, the system is based on a web-type, client-server architecture, one of the most frequently employed in application development on one hand and sensor and meta-data set on the other hand. This approach ensures the separation of the function logic model of the application into smaller functional units, the usage being split into two major components: client and server. On the client-side, the users, both on administration level and also as common users, integrate browsing technologies, being compatible with 99,3% [17] of all users, using technologies such as Web2.0, HTML, CSS, JavaScript, OpenLayers. At server level, Apache, PHP, Python, Inkscape, ImageMagick, GDAL, MySQL are among the technologies employed [14]. Additionally, besides these technologies, the data server receives inputs on other communication channels such as GSM or radio. The architecture used and shown in figure 2 is

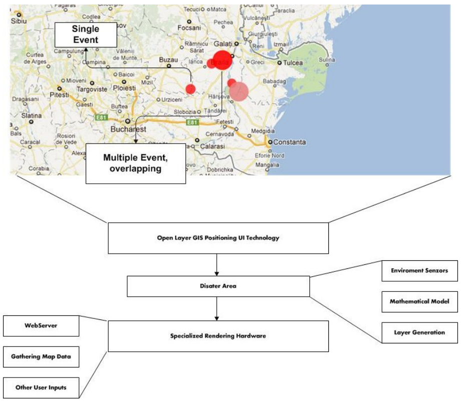


Figure 1: Integrated and scalable spatial information system

easy to approach. Likewise, the data central system can send alerts to some decision-making entities, firefighters, the police, responsible with disaster prevention and management etc. This architecture is one of the most versatile and can be accessed from: smartphones, desktop PC, laptops, tablets etc.

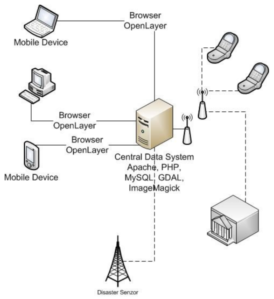


Figure 2: General system architecture

The presented system has a temporal role, giving the possibility to be loaded with statistical data. Once mature, it can play a crucial role in further cases.

From a functional standpoint, the system is divided into several distinct components:

a. the transfer component, communication; b. the alert component; c. collection hardware components passive elements, i.e. sensors, that collect states of the environment, processing systems etc.; d. Software components, i.e. the implementation of a mathematical model and a logic for the analysis of the risk of the various inputs from the external environment, all marked and positioned, using a GIS system; e. Finally, human components active, decision-making elements, which will be covered in this study only as constitutive element.

From the communication standpoint, two ways are used, namely:

a. GSM serial communication, usually for timely communications, status data or alert information, characterized by small data volume; b. or other TCP/IP wireless channels, for viewing and creating a graphical information map.

The advantage of GSM communications consists in a broader coverage area, but with a lower bandwidth, usually serial communications. They usually meet the need of speed and coverage for stationary sensors, that send the useful data in burst mode. The second is the wireless channel, which provides a wider bandwidth, but which is limited to a shorter distance. In terms of alert, this system integrates information capabilities of the decision-making elements, by using the following channels:

a. GSM channels, by SMS [5]; b. Ordinary channels, mail, newsletters, form-boards etc.

Industrial GSM modems ensure communication with the GSM network, but for a higher number of SMS messages, one can choose the SMS Bulk service of a local mobile network provider, which is a service provided by local mobile network operators, by which they ensure a way of interconnection of the local system with the mobile network operator. The disadvantage of such a system is the fact that this connection can be interrupted due to the way of implementation (cables), but can provide a greater number of alert messages even greater than a GSM Modem (which is, a mobile phone with advanced features of interconnectivity and interoperability with a computer) [6]. In terms of hardware component, the architecture is composed of sensors and/or other data sources, either automatic or human, which are afterwards integrated into the software architectural level. The internal structure of the software component is divided into the following (figure 3), in a top-dow approach:

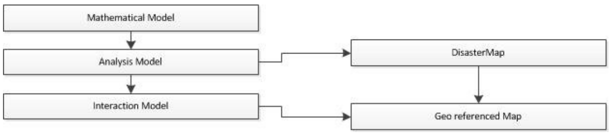


Figure 3: The internal structure of the software component

a. the mathematical model; b. the logic analysis model; c. the interaction model.

From the functional standpoint, this component uses accumulated data and by specific algorithms, it generates visually usable data, which can also be used in the analysis process.

The mathematical model has the following formal structure:

$$S = RxR \tag{1}$$

where S - geographical space, georeferenced WGS84 (as a rule) [16], but it can be transformed by using other utilities (e.g. GDAL) in any format compatible with OpenLayers [14].

$$P_i(\alpha, \beta) \in \mathrm{S} \tag{2}$$

The sensor coordinates, where α, β, geographic coordinates

$$V_{P_i(\alpha,\beta)} \in [\mathrm{V_{pmin}}, \mathrm{V_{pmax}}] \tag{3}$$

where V_{pmin}, V_{pmax} are the normal values. Above these values, maximum or minimum, a danger or disaster impact is to be generated.

Therefore, attached to each point (sensor) defined as such, the notion of impact is a specific function, defined as follows:

$$I_{P_i} = 0 \tag{4}$$

if $V_{P_i(\alpha,\beta)} \in [\mathrm{V_{pmin}}, \mathrm{V_{pmax}}]$.

$$I_{P_i} = \Psi(V_{P_i(\alpha,\beta)}) \tag{5}$$

if $V_{P_i(\alpha,\beta)} \notin [\mathrm{V_{pmin}}, \mathrm{V_{pmax}}]$.

and

$$\Psi \in RxR \tag{6}$$

In order to establish the implementation of thei Ψ, function, the following criteria will be taken into account:

- Risk level: Disaster, High risk, Medium risk, Low risk and Minor risk.
- Event type with human implications: Earthquake, Floods and Fire.

Taking into account the above-mentioned relations, we can define the disaster area, passing to an iteration through all the points where information can be collected, generating a layer shown in Figure 4, which overlapping with the already existing layers, generates figure 1.

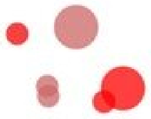


Figure 4: The layer generated on the basis of the information received from the collection points

In its turn, the logic model integrates analysis and decision-making components and prepares the needed elements for decision. At this level, one must specify the mathematical model that underlies the entire software component and has as a result a georeferenced layer, which includes the information of each important element that provides information about the environment.

The human element plays two roles, one is to be informed, namely the decision-making element, and the second is to inform oneself, in order to avoid a calamity area.

5 Technologies

Client [19], [14] - client requires a minimum number of software applications, which as we can see, are relatively inexpensive: a. Web Browser (Firefox, Internet Explorer >5.5, Nescape) to access the system residing on the server; b. Access to the server that hosts the system (Intranet, Internet) to ensure the interconnectivity for this station. Any communication media can usually be employed, from wire to wireless technologies; c. The required hardware is restricted to the above-mentioned software requirements; d. For interface software, Open Layer is used (in order to render the data).
Server [14] - the server software requirements are the following: a. Operation system: Linux Based -provides very good stability in terms of security and performance; b. Programming language: PHP 5.x; c. SGBD: MySQL 5.x; d. Web Server WEB : Apache 2.x; e. GDAL and ImageMagick for data display.

6 The Experimental Section. Results

The methodology for impact assessment is:

a. Establishing the area center or sensor coordinates; b. Establishing the limits affected on the 4 major directions N, S, E, W; c. Establishing the negative impact mitigation rates T_N, T_S, T_E, T_V

In order to calculate distance D between two coordinate points P1(lat_1, $long_1$) and P2(lat_2, $long_2$)) the haversine formula was used: [15], [20]

$$D = R * 2 * arcsin\big(\sqrt{sin^2(\frac{\Delta lat}{2}) + cos(lat_1) * cos(lat_2) * sin^2(\frac{\Delta long}{2})}\ \big) \tag{7}$$

where R is the Earth radius, which varies from the Equator to the Poles between 6378.14 and 6356.78, where the accuracy of the above-mentioned formula derives from, an error of about 0.5%.

Stage 1

Collecting the field data following the format below (in order to calculate distance D1).

- Location : P1 (45.017,27.525) => haversine [13] => $P_1(45^0\ 1'\ 1'', 27^0\ 31'30'')$

- Affected area

o Maximum coordinates: (N (45.097, 27.557), E (45.056, 27.613), S(45.012, 27.552), V(45.053, 27.482);

o Meters , from the point of origin [16]: N 5088m, with an impact lasting for up to 8 days, E 4661m, 6 days, S 4449m, 5 days, V 5792m, 8 days

- Impact parameters are indicated at each point.

Collecting the field data following the format below (in order to calculate the distance D2).

- Location: P2(45.391,27.76) => haversine [13] => $P_2(45^0\ 23'\ 28'', 27^0\ 45'\ 36'')$

- Affected area

o Coordinates: N(45.415, 27.757), E(45.395,27.783), S(45.375,27.757), V(45.395,27.783)

o Meters, from the point of origin: N 2178m, with an impact up to 6 days, E1897m, 6 days, S 1992m, 8 days, V 2136m, 9 days

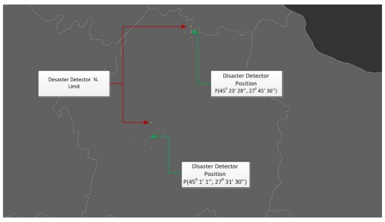

Desaster Detector N. Limit
Disaster Detector Position P(45⁰ 23' 28'', 27⁰ 45' 36'')
Disaster Detector Position P(45⁰ 1' 1'', 27⁰ 31' 30'')

Figure 5: Disaster positioning on the map

Stage 2

Stage 2 involves the development of the graphic management functions.

The progression calculation function for the affected area

$$D(t) = \frac{t_{max} - t}{t_{max}} d_{max} \tag{8}$$

where $t \in [0, t_{max}]$, $d \in [0, d_{max}]$

$$D(t) \in [0, d_{max}] \tag{9}$$

The impact function, calculates the impact level using the maximum distance, being applied on the 4 directions N, E, S, W

$$l(t, d) = \frac{t_{max} - t}{t_{max}} \frac{d}{d_{max}} \tag{10}$$

where $t \in [0, t_{max}]$, $d \in [0, d_{max}]$

$$l(t, d) \in [0, 1] \tag{11}$$

where 0 means there is no impact and 1, maximum impact.

Figure 6 results using the relations [2], [4]

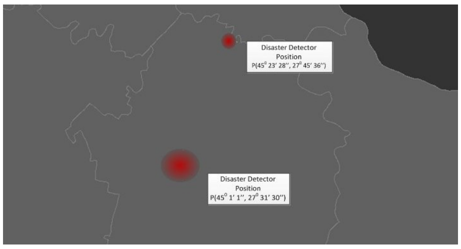


Figure 6: Disaster size at moment T_0 =0

The following figure represented in a logarithmic scale disaster cases, using the N direction.

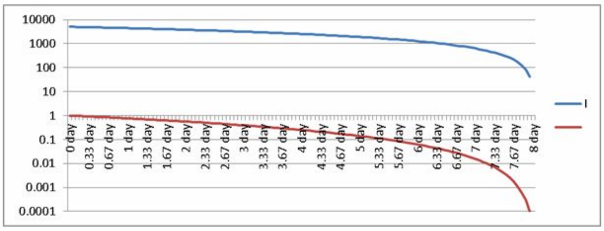


Figure 7: Represented in a logarithmic scale where: blue is space variation/day affected on direction N; red is intensity variation of the space affected /day on direction N

On zero-day of disaster the area was about 5000m and the eighth day this impact falls to 0.

On zero-day of the impact the disaster was labeled 1 then as time passed by, on the eighth day, it fell to 0.

The same calculation is used for the impact area P1 for the eastern element.

7 Conclusions and Future Research

In conclusion, we can say that, in order to develop high-quality and complex DSSs, able to cover as much as possible in case of disasters, on the one hand, a multidisciplinary approach is required, an approach which should bring together specialists in various fields, such as: environment, GIS, geographers, mathematicians, computer scientists, organizations in charge with such situations etc.; and on the other hand, a global approach which should reunite institutions and people from different countries, as such phenomena are not restricted to the borders of one specific country and they may affect large areas, encompassing people and goods from different states and at the same time, the transfer of know-how between partners/researchers from different countries brings benefits to such catastrophic situations, that might end up with material damages or worse, with the loss of human lives.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):30-38, December, 2012.

Bibliography

[1] Anica-Popa I., Cucui G., A Framework for Enhancing Competitive Intelligence Capabilities using Decision Support System based on Web Mining Techniques, *INT J COMPUT COMMUN*, ISSN 1841-9836, 4(4):326-334, 2009.

[2] Aseri T.C., Singla N., Enhanced Security Protocol in Wireless Sensor Networks, *INT J COMPUT COMMUN*, ISSN 1841-9836, 6(2):214-221, 2011.

[3] Ban D. , Yang W., Jiang J., Wen J., Dou W., Energy-Efficient Algorithms for k-Barrier Coverage In Mobile Sensor Networks, *INT J COMPUT COMMUN*, ISSN 1841-9836, 5(5):616-624, 2010.

[4] Chen J.I.-Z. , Chung Y.-N., A Data Fusion Methodology for Wireless Sensor Systems, *INT J COMPUT COMMUN*, ISSN 1841-9836, 7(1):39-52, 2012.

[5] Cioca, M., Cioca, L.I. Decision Support Systems used in Disaster Management, in *Decision Support Systems*, Chiang S. Jao (Ed.), ISBN: 978-953-7619-64-0, 2010.

[6] Cioca, M., Cioca, L.I., Buraga, S.C., Spatial (Elements) Decision Support System Used in Disaster Management, *Digital EcoSystems and Technologies Conference*, IEEE-DEST, ISBN: 1-4244-0467-3, IEEE Catalog Number: 07EX1418C, pp. 607-612, 2007.

[7] Filip, F.G. Decision Support and Control for Large-Scale Complex Systems, *Annual Reviews in Control* (Elsevier),ISSN: 1367-5788, 32(1): 61-70, 2008.

[8] Filip, F.G. *Sisteme suport pentru decizii*, ISBN 978-973-31-2308-8, Editura Tehnica, Bucuresti, 2007.

[9] Peng, Y., Gang Kou, Shi, Y., and Chen, Z., A Descriptive Framework for the Field of Data Mining and Knowledge Discover, *Int. J. of Information Technology & Decision Making*, 7(4):639-682, 2008.

[10] Peng, Y., Gang Kou, Wang, G., Wu, W., and Shi, Y., Ensemble of software defect predictors: an AHP-based evaluation method, DOI: 10.1142/S0219622011004282, *Int. J. of Information Technology & Decision Making*, 10(1):187-206, 2011.

[11] Power, D. J. *Decision support systems: Concepts and Resources for Managers*, Quorum Books, Westport, Connecticut, 2002.

[12] Suduc-Ana Maria, Bizoi, M., Filip, F.G., User Awareness about Information Systems Usability, *STUDIES IN INFORMATICS AND CONTROL*, 19(2):145-152, 2010.

[13] http://andrew.hedges.name/experiments/convert_lat_long/

[14] http://asrc.ro/imeteosat_beta (accessed on 07.04.2012)

[15] http://en.wikipedia.org/wiki/Haversine_formula

[16] http://support.google.com/earth/bin/answer.py?hl=ro&answer=148110 (accesed on 2012)

[17] http://www.articlesnatch.com/Article/Certification-Authorities-With-Browser-Ubiquity-Of-99-3-Are-Best-In-Industry-/253347

[18] http://www.cimec.ro/Resurse/Patrimoniu/Dezastre.htm (accessed on 5.04.2012)

[19] http://www.developer.com/java/web/print.php/10935_3528381_2 (accessed on 3.04.2012)

[20] http://www.movable-type.co.uk/scripts/gis-faq-5.1.html

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):39-46, December, 2012.

Data Dimensionality Reduction for Data Mining: A Combined Filter-Wrapper Framework

M. Danubianu, S.G. Pentiuc, D.M. Danubianu

Mirela Danubianu, Stefan Gheorghe Pentiuc
Dragos Mircea Danubianu
"Stefan cel Mare" University of Suceava
Romania, 720229 Suceava, 1 Universitatii
E-mail: mdanub@eed.usv.ro, pentiuc@eed.usv.ro
dragosdanubianu@yahoo.com

Abstract:
Knowledge Discovery in Databases aims to extract new, interesting and potential useful patterns from large amounts of data. It is a complex process whose central point is data mining, which effectively builds models from data. Data type, quality and dimensionality are some factors which affect performance of data mining task. Since the high dimensionality of data can cause some troubles, as data overload, a possible solution could be its reduction. Sampling and filtering reduce the number of cases in a dataset, whereas features reduction can be achieved by feature selection. This paper aims to present a combined method for feature selection, where a filter based on correlation is applied on whole features set to find the relevant ones, and then, on these features a wrapper is applied in order to find the best features subset for a specified predictor. It is also presented a case study for a data set provided by TERAPERS a personalized speech therapy system.
Keywords: data mining, feature selection, filters, wrappers.

1 Introduction

As an efficient way to find new and useful knowledge in data, knowledge discovery in databases (KDD) process, and implicitly data mining as its main step, have been the subject of extensive research. The main issue relates to model building process performances, and these performances are affected by factors such as type, quality and dimensionality of available data. As most data mining techniques may not be effective for high-dimensionality data, the solution consists in its reduction. In order to reduce the number of cases, one can use sampling or filtering, whereas feature reduction may be achieved by feature selection or feature composition. Feature selection aims to identify and to remove as many irrelevant and redundant features as possible with respect to the task to be executed, and it can be made using two approaches: filters and wrappers.

Filters do not consider the effect of selected features on the performance of the whole process of knowledge discovery, since the used feature selection criterion does not require a predictor evaluation for reduced data sets. Wrappers take into account the feed-back related to the performance of the selected set of features in the KDD process. So, it is used as criteria for feature selection the predictor performance. Wrappers often give better results than filters, because feature selection is optimized for the specific learning algorithm used, but if the computational complexity and execution time are considered, they are too expensive for large dimensional datasets since each selected feature set must be evaluated with the predictive algorithm used.

In these circumstances we aim to study if an apriori filtering of features based on their relevance related to the class may improve a closed loop feature selection process. Section 2 presents some theoretically aspects related to data mining and the influence of data dimensionality on its performances. There are also enumerated some data dimensionality reduction methods. Section 3 refers some aspects regarding feature selection whereas Section 4 provides a comparison between filters and wrappers. Section 5 presents a framework, which proposes a combination filter-wrapper and Section 6 offers some experimental results obtained by applying the proposed method over a dataset collected by TERAPERS - a computer-based speech therapy developed within the Center for Computer Research in the Stefan cel Mare University of Suceava, and used by the therapists from Regional Speech Therapy Center of Suceava from March 2008.

2 Data Mining and Data Dimensionality

Defined as the process of exploring and analyzing large volumes of data, in order to find new relationships within data or new patterns, data mining is a step in knowledge discovery in database (KDD). Its task is to analyze large volumes of data in order to extract previously unknown, interesting and potential useful patterns, and its performances are affected by factors such as: type, quality and dimensionality of data. Hypothetically, having more data, results are more precise, but practical experience with data mining algorithms has shown that this is not always true. On the one hand, the high dimensionality of data can cause data overload, and on the other hand if there are a lot of features, it is possible that the number of cases in data set to be insufficient for data mining operations. [1] This make some data mining algorithms non applicable. The solution for these problems is the reduction of data dimensions.

The size of a data set is determined both by the number of cases and by the number of features considered for each case. In order to reduce number of cases one can use sampling or filtering. Feature reduction may be achieved either by feature selection or by feature composition. These methods should produce fewer features, so the algorithms can learn faster. Sometimes, even the accuracy of built models could be improved. [2] Methods used for feature selection, can be classified as: filters or open loop methods, and wrappers or closed loop methods.

3 Feature Selection

Feature selection aims to identify and to remove as much irrelevant and redundant features as possible with respect to the task to be executed. It has the potential to be a fully automatic process, and brings some benefits for data mining, such as: an improved predictive accuracy, more compact and easily understood learned knowledge and reduced execution time for algorithms.

Feature selection methods are divided in two broad categories, filters and wrappers, and within these categories algorithms can be further individualized by the nature of their evaluation function and by the means the space of feature subsets is explored. Typically, feature selection algorithms perform a search through the space of feature subsets, and must solve four problems which affect such search:

- to select a point in the feature subsets space from which to start the search. A first choice, called forward selection, supposes to begin with no features and successively add attributes, whereas a second one, backward selection, begins with all features and successively remove them;

- even heuristic search strategies not guarantee finding the optimal subset, such strategies can give good results, and are more feasible than exhaustive search strategies which are prohibitive just for a small initial number of features;

- the most important factor which makes difference among feature selection algorithms is evaluation strategy. There are feature selection methods which operate independent of any learning algorithm, and irrelevant features are filtered before learning begins, based on general characteristics of the data to evaluate. Other methods use an induction algorithm combined with a statistical re-sample technique to estimate the final accuracy of feature subsets;

- each feature selection process must solve the problem regarding stop searching through the space of feature subsets. One might stop adding or removing features when none of the alternative improves upon the gain of current feature subset, or one might continue to alter the feature subset as long as the gain does not degrade.

4 Filters vs. Wrappers

The earliest and simplest approaches to feature selection were filters, called also open loop feature selection methods. Based on selecting features through class separability criteria, filters do not consider the effect of selected features on the performance of the whole process of knowledge discovery, as is presented in Figure 1(a). They provide usually a ranked list of features that are ordered according a specific evaluation criterion such as: accuracy and consistency of data, information content or statistical dependencies between features. [3] They give also information about the relevance of a feature compared with the relevance of other features, and do not tell to the analyst what is the desirable minimum set of the features. [4]

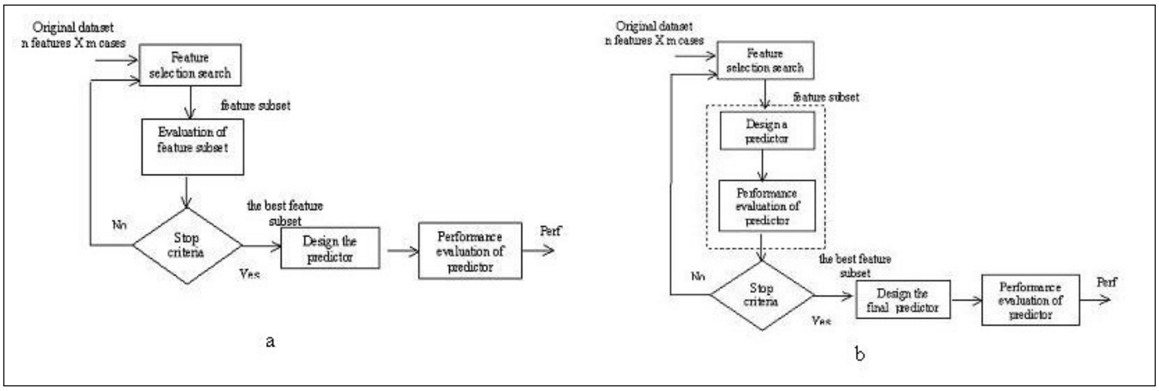


Figure 1: Open loop feature selection method (a), and closed loop feature selection method (b)

Wrappers, known also as closed loop feature selection methods take into account the feedback related to the performance of the selected set of features for the complete KDD process. They use the prediction performance as selection criteria, and evaluate the quality of selected features by comparing the performances for prediction algorithms applied on the reduced set of features and on the original one. Figure 1(b) presents a closed loop feature selection method. [2]

Regarding the final predictive accuracy of a learning algorithm, wrappers often give better results than filters, because feature selection is optimized for the specific learning algorithm used. But if the computational complexity and execution time are considered, wrappers are too expensive for large dimensional datasets, since each selected feature set must be evaluated with the predictive algorithm used. Additional, since the feature selection is closely coupled with a learning algorithm, wrappers are less general than filters and they must be run every time when one switch from one learning algorithm to another.

5 A Combined Approach Filter-Wrapper for Feature Selection

Studying the advantages and limitations for the two general feature selection methods we can conclude that, if improved performance for a specific learning algorithm is required, a filter can provide a reduced initial feature subset for a wrapper which contains only relevant features, as shown in Figure 2. This approach could produce shorter and faster search for the wrapper.

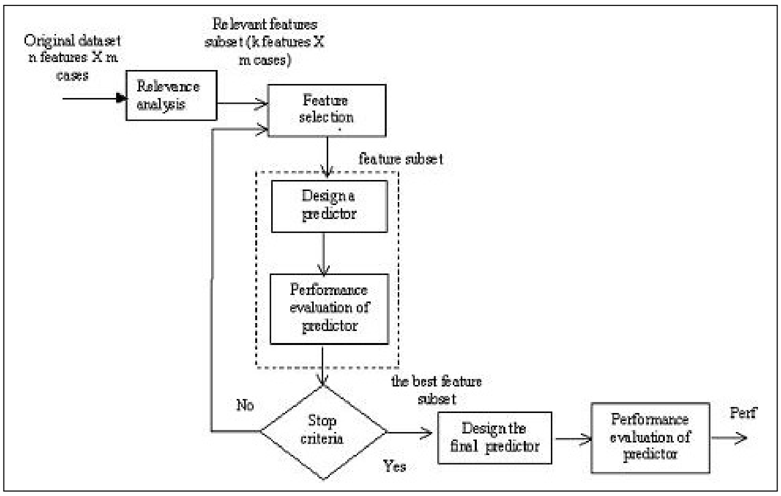

Original dataset n features X m cases
Relevance analysis
Relevant features subset (k features X m cases)
Feature selection
feature subset
Design a predictor
Performance evaluation of predictor
No
Stop criteria
the best feature subset
Yes
Design the final predictor
Performance evaluation of predictor
Perf

Figure 2: A framework which combines feature relevance analysis with closed loop feature selection

Since practice has demonstrated that irrelevant input features lead to great computational cost for data mining process and may cause overfitting, more feature selection researches have focused on extraction of relevant features from the whole data set in order to apply data mining algorithms upon these data. [5] [6] But how can we establish if a feature is relevant or not? In [7] is stated that features are relevant if their values vary systematically with category membership. That means that a feature is relevant if it is correlated with the class. Formally this was defined in [2] as follows:

Definition 1. A feature F_i is relevant iff there exists f_i and c for which $p(F_i = f_i) > 0$ such that

$$p(C = c|F_i = f_i) \neq p(C = c) \tag{1}$$

Relevance is usually defined in terms of correlation or mutual information. In order to define mutual information for two features we start from the concept of entropy, as a measure of uncertainty of a random variable. For a variable X the entropy is defined as:

$$E(X) = -\Sigma p(x_i) log_2(p(x_i)) \tag{2}$$

The entropy of a variable X after observing values of another variable Y is defined as:

$$E(X|Y) = -\Sigma p(y_i)\Sigma p(x_i, y_i) log_2(p(x_i|y_i)) \tag{3}$$

where $p(x_i)$ is the prior probability for all values of X, and $p(x_i|y_i)$ is the posterior probabilities of X given the value of Y. The value by which the entropy of X decreases, estimates additional information about X provided by Y. It is called information gain [8] and is calculated using the following expression:

$$I(X, Y) = E(X) - E(X|Y) \tag{4}$$

We take into account that for discrete random variable, the joint probability mass function is:

$$p(x_i|y_j) = p(x_i, y_j)/p(y_j) \tag{5}$$

and the marginal probability function, p(x) is:

$$p(x_i) = \Sigma p(x_i, y_j) = \Sigma p(x_i|y_j)p(y_i) \tag{6}$$

where p(x,y) is joint probability distribution function of X and Y, and $p(x_i)$ and $p(y_j)$ are the marginal probability distribution functions of X and Y respectively. Finally, for two discrete random variables X and Y , information gain is formally defined as:

$$I(X,Y) = \sum_j \sum_i p(x_i, y_j) log \frac{p(x_i, y_j)}{((p(x_i)p(y_j))} \tag{7}$$

According to this expression, one says that a feature Y is more correlated to feature X than feature Z if:

$$I(X,Y) > I(Z,Y) \tag{8}$$

It can be observed that information gain favors features with more values, so it should be normalized. In order to compensate its bias and to restrict its values to range [0,1] it is preferable to be used symmetrical uncertainty, defined as:

$$SU(X,Y) = 2\frac{I(X,Y)}{E(X) + E(Y)} \tag{9}$$

A value of 1 for symmetrical uncertainty means that knowing the values of either feature completely predicts the value of the other whereas a value of 0 implies that X and Y are independent. Starting from these considerations, in the proposed framework, first a relevance analysis is made using the symmetrical uncertainty $SU(F_i, C)$ between each feature F_i and the class C. [9] Based on this analysis one removes the irrelevant features, and one obtains a features subset containing only the relevant features. Then, on this dataset one applies closed loop feature selection methods, using as search strategy both forward and backward selection, and using a decision tree as predictor, both for feature selection and for the performance evaluation.

6 Experimental Results

We have applied the framework described above for a real dataset collected by TERAPERS system. This is a system which aims to assist the personalized therapy of dyslalia (an articulation speech disorder) and to track how the patients respond to various personalized therapy programs. Implemented in March 2008, the system is currently used by the therapists from Regional Speech Therapy Center of Suceava.

An important aspect of assisted therapy refers its adaptation according to individual patients' characteristics and evolution, for which therapist must perform complex examination of children, materialized in recording of relevant data relating to personal and family anamnesis. These collected data may provide information relative to various causes that may negatively influence the normal development of the language. Further, one provide to the personalized therapy programs data such as number of sessions/week, exercises for each phase of therapy and the changes of the original program according to the patient evolution. The tracking of child progress materializes data which indicate the moment of assessing the child and his status at that time. All these data are stored in a relational database, composed of 60 tables.

Data stored in the TERAPERS database is the set of raw data that can be the subject of data mining process. It might be useful, because as it was shown in [10] one can use classification in order to places the people with different speech impairments in predefined classes (if attribute diagnosis contain the class label one can predict a diagnosis based on information contained in various predictor variables), one can use clustering to group people with speech disorders on the basis of similarity of different features and to help therapists to understand who are they patients, or one can use association rules to determine why a specific therapy program has been successful on a segment of patients with speech disorders and on the other was ineffective. For our experiments one considers a data set consisting of 72 features with numeric and descriptive values and 312 cases. These are anamnesis data or data derived from complex examination on which one intend to build a classification model to predict, in order to suggest to therapist the diagnosis for future cases. On this data set one have applied the feature selection method described above. Shown in Figure 3, such experiment is designed and implemented in WEKA. [11]

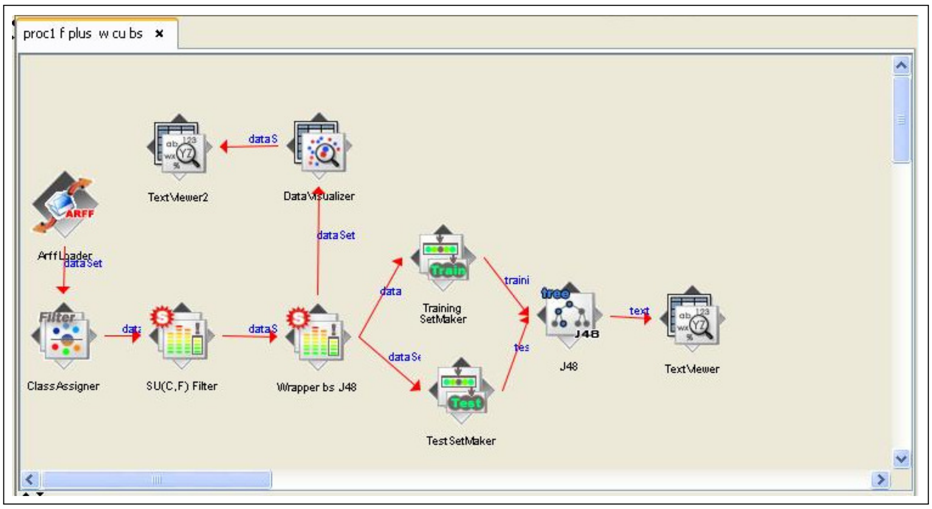
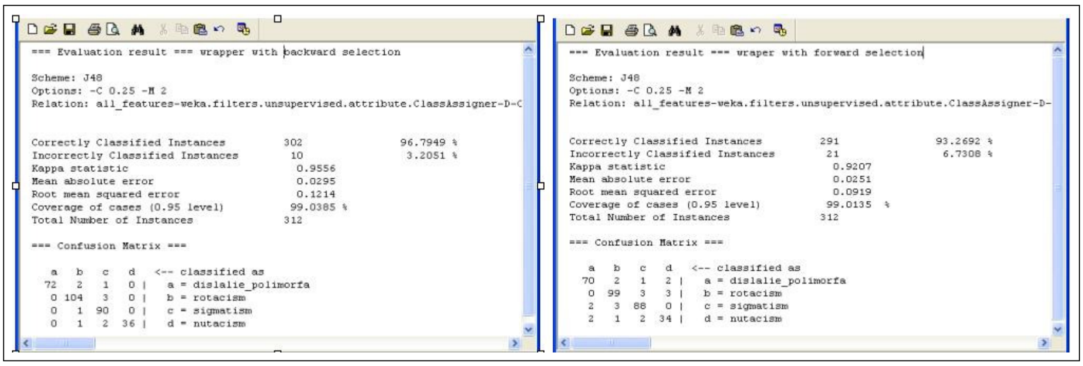


Figure 3: WEKA knowledge flow for the proposed framework

One has considered the attribute *diagnosis* as class label, and we have built a feature selection process in two steps.In the first stage one have applied a unsupervised filter against the whole set of features. Based on values for symmetrical uncertainty $SU(F_i, C)$ there were retained 52 relevant features. In the second stage over this feature subset one have applied a wrapper which uses as estimation predictor a decision tree classifier (J48). For this wrapper one applies alternatively the two search strategies forward and backward search.

To analyze the influence of data reduction we need to know what we gain or what we lose so, we must compare computing times and the accuracy for the model built for reduced data sets obtained using the described approaches. Figure 4 presents the performances of the classifiers built on feature subset produced by wrapper using a backward selection strategy and on features obtained from the same wrapper which use, this time, a forward selection strategy.

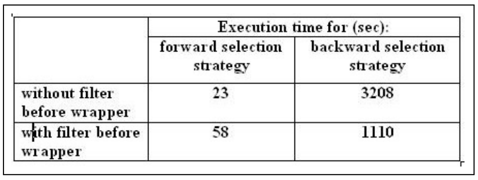
As one can see in Figure 4 for the subset obtained by backward selection, the predictions' performance, measured in correctly classified instances, are better than those for the feature subset obtained by forward selection. In Figure 5 a comparison of the execution time for forward and backward selection in both cases, with filter before wrapper, and without filter before wrapper is made. One see that for backward selection, which provides better performance for prediction, the process which considers as input for wrapper the data subset provided by filter is three times faster than the other process.



```
=== Evaluation result === wrapper with backward selection

Scheme: J48
Options: -C 0.25 -M 2
Relation: all_features-weka.filters.unsupervised.attribute.ClassAssigner-D-C


Correctly Classified Instances         302               96.7949 %
Incorrectly Classified Instances        10                3.2051 %
Kappa statistic                          0.9556
Mean absolute error                      0.0295
Root mean squared error                  0.1214
Coverage of cases (0.95 level)          99.0385 %
Total Number of Instances              312

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
  72   2   1   0 |   a = dislalie_polimorfa
   0 104   3   0 |   b = rotacism
   0   1  90   0 |   c = sigmatism
   0   1   2  36 |   d = nutacism
```

```
=== Evaluation result === wraper with forward selection

Scheme: J48
Options: -C 0.25 -M 2
Relation: all_features-weka.filters.unsupervised.attribute.ClassAssigner-D-


Correctly Classified Instances         291               93.2692 %
Incorrectly Classified Instances        21                6.7308 %
Kappa statistic                          0.9207
Mean absolute error                      0.0251
Root mean squared error                  0.0919
Coverage of cases (0.95 level)          99.0135  %
Total Number of Instances              312

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
  70   2   1   2 |   a = dislalie_polimorfa
   0  99   3   3 |   b = rotacism
   2   3  88   0 |   c = sigmatism
   2   1   2  34 |   d = nutacism
```

Figure 4: Performances of classifier built on feature subsets obtained by proposed method



	Execution time for (sec):	
	forward selection strategy	backward selection strategy
without filter before wrapper	23	3208
with filter before wrapper	58	1110

Figure 5: Performances of classifier built on feature subsets obtained by proposed method

7 Conclusions and Future Works

As a possibility to reduce the number of feature considered by data mining algorithms, in order to make them more efficient, this paper presents a method which uses a combination filter-wrapper. We have used a correlation based filter on the whole set of features, then on relevant subset of features we have applied a wrapper which uses a decision tree classifier for prediction. As a case study we have applied this method on data collected by TERAPERS a system which aims to assist speech therapists on personalized therapy of dyslalia. The process was designed and implemented in WEKA. We have compared the performances obtained both for feature selection by the described method, and for feature selection using only the same wrapper as in first case. We have achieved clearly superior performances for execution time, when we have used for feature selection the combined approach and backward selection as search strategy for wrapper. The positive results obtained for the considered data encourage us to continue our work. We will try to improve these execution times by parallelization of feature selection operations.

Acknowledgments

This paper was supported by the project "Progress and development through post-doctoral research and innovation in engineering and applied sciences PRiDE - Contract no. POSDRU/89/1.5/S/57083", project co-funded from European Social Fund through Sectorial Operational Program Human Resources 2007-2013.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):39-46, December, 2012.

Bibliography

[1] Danubianu M., Pentiuc S.G., Tobolcea I., Schipor O.A., Advanced Information Technology - Support of Improved Personalized Therapy of Speech Disorders, *INT J COMPUT COMMUN*, ISSN 1841-9836, 5(5): 684-692, 2010.

[2] Kohavi R., John G., Wrappers for feature subset selection, *Artificial Intelligence*, Special issue on relevance, 97(1-2):273-324, 1997.

[3] Hall, M., Correlation-based feature selection for discrete and numeric class machine learning, *Proc. of International Conference on Machine Learning*, 359-365, Morgan Kaufmann, 2000.

[4] Douik A., Abdellaoui M., Cereal Grain Classification by Optimal Features and Intelligent Classifiers, *INT J COMPUT COMMUN*, ISSN 1841-9836, 5(4):506-516, 2010.

[5] Peng H. Long F., Ding C., Feature Selection based on mutual Information: Criteria of Max-Dependency, Max-Relevance and Min-Redundancy, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 27(8):1226 - 1238, 2005.

[6] John G.H., Kohavi R., Pfleger P., Irrelevant features and the subset selection problem, *Machine Learning: Proceedings of the Eleventh International Conference*, 121-129, Morgan Kaufman, 1994.

[7] Gennari J.H., Langley P., Fisher D., Models of incremental concept formation, *Artificial Intelligence*, (40):11-16, 1989.

[8] Quinlan J.R., *C4.5: Programs for Machine Learning*, Morgan Kaufman, 1993.

[9] Yu, L., Liu, H., Efficient Feature Selection via Analysis of Relevance and Redundancy, *Journal of Machine Learning Research*, 5:1205-1224, 2005 .

[10] Danubianu M., Pentiuc St. Gh., Socaciu T., Towards the Optimized Personalized Therapy of Speech Disorders by Data Mining Techniques, *The Fourth International Multi Conference on Computing in the Global Information Technology ICCGI 2009*, Vol: CD, 23-29 August, Cannes - La Bocca, France, 2009.

[11] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I., The WEKA Data Mining Software: An Update, *SIGKDD Explorations*, 11(1):10-18, 2009.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):47-55, December, 2012.

Public Discourse Semantics. A Method of Anticipating Economic Crisis

D. Gîfu, D. Cristea

Daniela Gîfu
Alexandru Ioan Cuza University of Iaşi, Faculty of Computer Science
16, General Berthelot St., 700483 Iaşi, Romania
E-mail: daniela.gifu@info.uaic.ro

Dan Cristea
1. Romanian Academy - the Iasi branch, Institute for Theoretical Computer Science
2, T. Codrescu St., 700481 Iaşi, Romania
and
2. Alexandru Ioan Cuza University of Iaşi, Faculty of Computer Science
16, General Berthelot St., 700483 Iaşi, Romania
E-mail: dcristea@info.uaic.ro

Abstract:
This paper provides a proof that anticipation of an economic crisis by analysing public discourses (in particular, speeches on economic issues) is feasible. It proposes a method of text classification and semantic interpretation based on natural language processing techniques that could be used to trace, over a period of time, the print press discourses, with the aim to valuate the perspective of occurrence of crises. Classification is the task of assigning tags (words, expressions) to the texts that make up a corpus. In our case, we were interested to identify among the texts under scrutiny those belonging to classes like *financial*, *economic*, *nationalism*, etc. This approach is sustained by the fact that public discourses can be characterized from a rhetorical perspective, depending on the specific strategies their authors have chosen: orientation to change opinions or to determine action, ratio between rational (*logos*) and emotional (*pathos*), etc. We are sugesting an automatic analysis of the content of the public language, by using quantitative measures. Our purpose was to develop a computational tool able to offer to researchers in the economic, social or political sciences, but, not less, to the public at large, the possibility to measure the acuity of different accents of a written public discourse (*financial*, *emotional*, etc.), as mean to anticipate the threat of financial waves. Such a tool could help the processes of decision making in the analysis of crisis. Although our analysis used as data the journalistic and economic environments of Romania, it could easily be extrapolated to other languages/countries.

Keywords: public language, text categorization, semantic analysis, economic crisis.

1 Introduction

In the atempt to divulge ante-factum crises in public discourse, primarily the voices of those entities must be listen to which are most influencial on the financial and economic domains. These entities, clearly, are: The Romanian National Bank (in the internal context) and the World Bank (in the international context)[1]. The voices of these entities are best listen to in the public speeches of governors and, in many cases, of journalists specialised on economic-financial issues.

A public discourse arguing on some extremely important moment-related issue is, most of time, an amalgam of arguments, rational forms, descriptions, stylistic procedures, which are

[1] "In times of internal or international crisis (...), we talk about managing various symbolic aspects of the role of: guardian of institutions, guarantor of national unity, moderator." [3]

intended to inform or to prepare a receptor in front of a problematic reality. But, as close to the subject a discourse would be, often it hides, in subtle ways, the true nature of the subjective thinking of the emitter. For instance, an exaggerated trust in the fresh energy of the society, in the benefits the loans on mortgage could bring to ordinary people, on the exceptional rise of the rate of interests, or on the incredible high bonuses certain banks are offering to their highly ranked employees could simultaneously bring the negative news, that something wrong is in the air, that a crisis is insinuating. Decoding of this hidden message, which is most of the time transmitted unintentionally, could be done only by someone extremely sensible to all facets of the financial and economic life. Signals for economic crises are issued by the central banks (e.g. Federal Reserve System, Central Bank of U.S., European Central Bank, etc.). During the period 2001-2008, when the banking system issued large but artificially cheap credits, there have been many public rhetoric appearances favourable to this behaviour which tried to set up an economic development investment with questionable prudence. Slogans like "a home for every American" (U.S.) or "credit with only an ID" (Romania), addressing a wide range of borrowers but having extremely low interest rates, could have been taken as signals of an economic potential crisis.

In this study we address the question: *Can an economic crisis be anticipated by evaluating public discourses from a lexical-semantic perspectives?*

We are interested to pursue a content analysis of the public language, using for that investigation tools that belong to the domain of natural language processing (NLP) and addressing: vocabulary (key words, frequent words), semantics (classes of concepts arranged in a hierarchy) and rhetorical-pragmatic discursive strategies (presence of the person I, preference for vague statements, generalities, etc.).

In U.S., the tradition of quantitative analysis is very strong, its roots being defined by Lasswell [5]. In Europe the interest grew more towards theoretical investigation of the semiotics of discourse ([1], [10], [1]). Modern content analysis is not only an illustration of a theory of text, but, should be rooted on empirical data. On the other hand, the American analysis is often neutral, technical, comparative, while the European analysis (especially the Critical Discourse Analysis model[2]) has a critical component and a strong enough ethicist.

In the perspective of our study, we are interested on public discourses (speeches), in written form, given by specialists on economy or by journalists, on economic issues. It is known that economy crises succeeds either a period of economic thrive or, as happened recently, a previous crisis. In our investigation we have used texts produced by most pertinent spokesmen which appeared in press materials issued by the Romanian National Bank (BNR), the most legitimate voice on economy issues in Romania. The other major filter in selecting the texts that should populate our corpus was the economic context (e.g. economic stability vs. economic crisis). A text categorization application filtered a stream of news that was considered of interest for our research. Some of the topics of interest have been: "credit ID only", "real estate boom", "mortgages", and "transactions with land or housing".

On another hand, at the base of our quantitative investigation was laid a lexical-semantic database. In order to assure generality, in acquiring it we had to use rather neuter sources, not necessarily tight to our specific corpus of texts. As such, the lexicon and the semantic classes have been collected from different sources usually dealing with economy themes: the BNR publications, already mentioned, but also a collection of dailies, *Ziarul financiar*, *Curierul Naţional*, *Bursa*, that have been monitored for a long period of time.

Current empirical approaches in analysing the public language put at work NLP techniques,

[2] "Critical theories, thus also CDA, are afforded special standing as guides for human action. They are aimed at producing enlightenment and emmancipation. Such theories seek not only to describe and explain, but also to root out a particular kind of delusion. Even with differing concepts of ideology, critical theory seeks to create awareness in agents of their own needs and interests." [11]

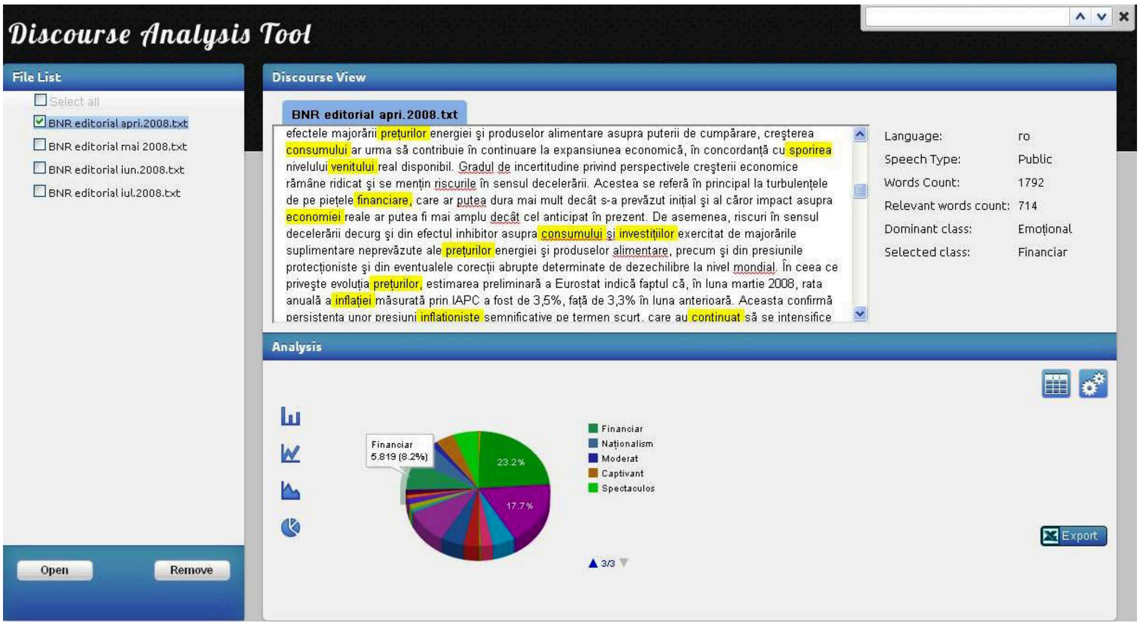
by which a multitude of features of the discourse were extracted and interpreted. The domain of NLP includes a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications. In this paper we describe a platform (*Discourse Analysis Tool* - DAT) specialised in the interpretation of the public discourse, which integrates a range of language processing tools with the intent to build complex characterisations of the public discourse. The idea behind it is that the vocabulary betrays discursive tonalities, this way allowing interpretations over the speakers orientation.

The paper is structured as follows. Section 2 shortly describes the previous work. Section 3 presents the DAT software and section 4 discusses an example of comparative analysis of economic discourses, elaborated during one year (2007-2008). Finally, Section 5 highlights interpretations anchored in our analysis and presents conclusions.

2 Previous Work

The aim of an interdisciplinary approach such as analysing the language of public speeches is to define and explain different discursive contexts, in our case, reflected in the print media. The studies in this direction have mainly concentrated on three tasks. The first had to do with a cognitive side and, often, with an emotional side, of how humans acquire, produce, and understand language. The second aimed at understanding the relationship between the linguistic utterance and the world, and the third - at understanding the linguistic structure of the language as a communication device. Linguistics has usually treated language as an abstract object which can be accounted for without reference to social or political concerns of any kind [9].

As we will see, one aspect of the platform that we present touches a lexical-semantic functionality, which has some similarities with the approach used in *Linguistic Inquiry and Word Count* (LIWC), an American software used to analyse the elections in United States in 2008. There are, however, important differences between the two platforms. LIWC-2007 basically counts words and increments counters associated with their declared semantic classes. DAT performs part-of-speech (POS) tagging and lemmatization of words. The lexicon contains a collection of lemmas (over 8800) for the POS categories of verb, noun, adjective and adverb, each being associated with one or more semantic classes. In the context of the lexical semantic analysis, the pronouns, numerals, prepositions and conjunctions, considered to be semantically empty, have been left out. Then, a special section of the lexicon includes expressions. An expression is defined as a sequence of stems of words. DAT includes now 33 semantic classes, chosen to fit optimally with the necessities of interpreting the public discourse, five of them having been added recently (`failures`, `nationalism`, `moderation`, `firmness`, `spectacular`). Then, another range of differences between the two platforms regards the user interface. In DAT, the user is served by a friendly interface, offering a range of services: opening and displaying one or more files, editing and saving the text, functions of undo/redo, functions of editing the lexicon, visualization of the mentioning of occurrences of certain semantic classes in the text, etc. The menus offers a whole range of output visualization functions, from the tabular form to graphical representations and to printing services. And finally, and most importantly, to help the user to interpret different authors simultaneously, she/he can chose among a collection of formulas that facilitate comparative studies.



Discourse Analysis Tool

File List

□ Select all
☑ BNR editorial apri.2008.txt
□ BNR editorial mai 2008.txt
□ BNR editorial iun.2008.txt
□ BNR editorial iul.2008.txt

Discourse View

BNR editorial apri.2008.txt

efectele majorării prețurilor energiei şi produselor alimentare asupra puterii de cumpărare, creşterea consumului ar urma să contribuie în continuare la expansiunea economică, în concordanță cu sporirea nivelului venitului real disponibil. Gradul de incertitudine privind perspectivele creşterii economice rămâne ridicat şi se mențin riscurile în sensul decelerării. Acestea se referă în principal la turbulențele de pe piețele financiare, care ar putea dura mai mult decât s-a prevăzut inițial şi al căror impact asupra economiei reale ar putea fi mai amplu decât cel anticipat în prezent. De asemenea, riscuri în sensul decelerării decurg şi din efectul inhibitor asupra consumului şi investițiilor exercitat de majorările suplimentare neprevăzute ale prețurilor energiei şi produselor alimentare, precum şi din presiunile protecționiste şi din eventualele corecții abrupte determinate de dezechilibre la nivel mondial. În ceea ce priveşte evoluția prețurilor, estimarea preliminară a Eurostat indică faptul că, în luna martie 2008, rata anuală a inflației măsurată prin IAPC a fost de 3,5%, față de 3,3% în luna anterioară. Aceasta confirmă persistenta unor presiuni inflationiste semnificative pe termen scurt, care au continuat să se intensifice

Language:	ro
Speech Type:	Public
Words Count:	1792
Relevant words count:	714
Dominant class:	Emoțional
Selected class:	Financiar

Figure 1: The DAT interface: in the left window appear the selected files, in the middle window - the text from the selected file, and in the right window, information about the text (language, word count, dominant class, etc.). Bellow, a plot chosen from a range of graphical styles is displayed. By selecting a specific class in the middle window, all words assigned to that class are highlighted in the text.

3 The DAT Platform

The Discourse Analysis Tool (DAT, currently at version 3) considers the public discourse from two perspectives: lexical and semantic. We describe shortly our platform which integrates a range of language processing tools, with the intent to build complex characterisations of the public discourse. The concept behind this method is that the vocabulary used by a speaker betrays the authors sensibility, her/his level of culture, her/his cognitive world, and, by this, to the semantic spectrum of her/his speeches, while the syntax may reveal the level of culture, intentional persuasive attitudes towards the public, etc. Some of these means of expression are intentional, aimed to deliver a certain image to the public, while others are unintentional.

Figure 1 shows a snapshot of the interface showing a semantic analysis, during a working session. To display the results of the lexical-semantic analysis, the platform incorporates two alternative views: graphical (pie, function, columns and areas) and tabular (Microsoft Excel compatible).

The vocabulary of the platform covers 33 semantic classes (**swear**, **social**, **family**, **friends**, **people**, **emotional**, **positive**, **negative**, **anxiety**, **anger**, **sadness**, **rational**, **intuition**, **determine**, **uncertain**, **certain**, **inhibition**, **perceptive**, **see**, **hear**, **feel**, **sexual**, **work**, **achievements**, **failures**, **leisure**, **home**, **financial**, **religion**, **nationalism**, **moderation**, **firmness**, **spectacular**), considered to fulfil optimally the necessity of interpreting the public discourse in different contexts. Some of these categories are placed in a hierarchical relation.

Linguistic processing begins by tokenization, part of speech tagging and lemmatization. Only the words belonging to the lexicon are considered relevant and therefore count in establishing the weights of different semantic classes. In response to the text being sent by the user, the system returns a compendium of data which includes: the language of the document, the number of words, and the type of discourse detected, a unique identifier (usually the file name), and a

Table 1: Examples of phrases on economy issues, on BNR editorials

Classes		Original in Romanian	English equivalent
financial	positive	creşterea PIB, expansiunea economiei mondiale, investiţii, scăderea ratei şomajului, expansiunea economică	PIB growth, global economic growth, investments, unemployment has declined, economic growth
	negative	moderarea ritmului de creştere a salariilor, gradul de incertitudine, turbulenţe pe pieţele financiare, efectul inhibitor asupra consumului şi investiţiilor	moderate the wage growth, uncertainty, financial markets turmoil, dampening impact on consumption and investment

report of the lexical-semantic analysis.

Our interest went mainly in determining those discursive attitudes able to betray an approaching recession. But the system can be parameterised to fit also other conjunctures: the user can define at will her/his semantic classes, which, as indicated, are partially placed in a hierarchy. Thus, for example, for the lemma *economist*, the following classes are assigned: 2 = `social` and 5 = `people`. The class `people`, is a subclass of the class `social`. These classes and their hierarchy are defined in a XML-like manner:

```
<class name="social" id="2">
<class name="people" id="5" parent="2">
```

Whenever an occurrence belonging to a lower level class is detected in the input file, all counters in the hierarchy, from that class to the root, are incremented. In other words, the lexicon assigned to superior classes includes all words/lemmas of its subclasses.

4 A Comparative Study

4.1 The corpus

The corpus used for our investigation was configured to allow a comparative study over the discursive characteristics of economic-financial themes, by including economy texts published on the BNR site in three different periods:

1. April-June 2007, when Romania crossed a period of economic stability.
2. April-June 2008, when Romania was near the economic crisis.
3. July 2008, when the Romanian president declared the economic recession.

Table 1 presents examples of phrases in the economy domain that exhibit two different discourse moods: positive emotional and negative emotional.

The analyzed texts were essentially dealing with the topics social and financial. After processing the texts with the DAT software, the following classes proved to have preponderant occurrences: `financial`, `social`, `work`, `emotional` (`positive` and `negative`), `rational` (`intuition`, `determine`, `uncertain`, `certain` and `inhibition`) and `nationalism`. To stress the distinguishing features, only these classes were finally left on the graphics.

4.2 The lexical-semantic analysis

We show in this section the results outputted by DAT when analysing the streams of textual data belonging to the three sections of the corpus (presented in section 4.1). For that, we have

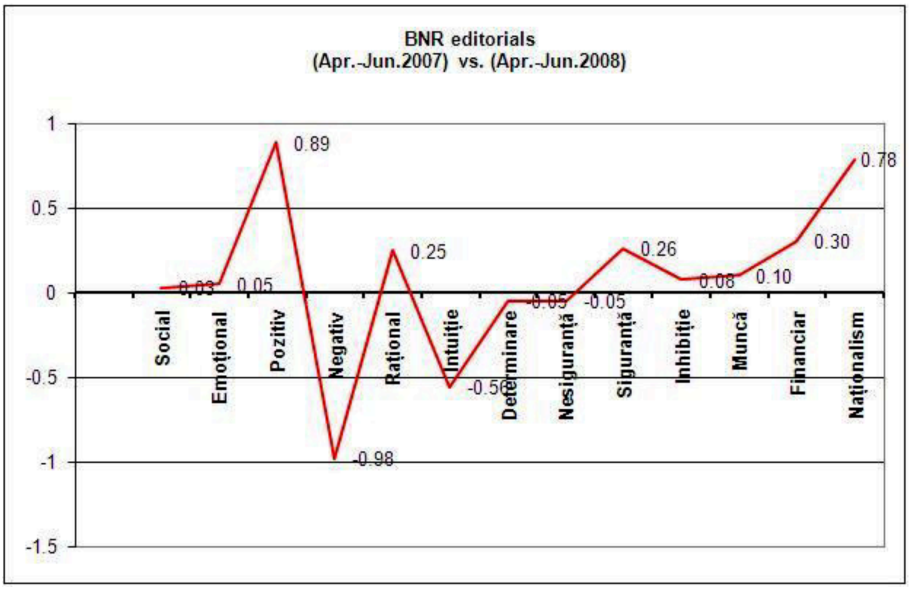


Figure 2: Difference between the occurrence of semantic classes in BNR editorials: one year before the economic recession versus three months before.

used the DAT feature of performing comparative studies. The values are supposed to reflect correctly the indicated classes, because they were computed by averaging on the whole collections of texts, not just a single text. The graphics considered for the interpretation computed one-to-one differences, as given by Formula 1, included in the DAT Mathematical Functions Library:

$$Diff_{x,y}^{1-1} = average(x) - average(y) \tag{1}$$

where x and y are two streams; $average(x)$ and $average(y)$ are the average frequencies of x and y over the whole stream, and the difference is computed for each selected class. Since a difference can lead to both positive and negative values, these particular graphs should read as follows: values above the horizontal axis are those prevailing at the first element more than at the second element, and those below the horizontal axis show the reverse prominence. A zero value indicates equality. Our experience showed that values below the threshold of 0.5% should be considered as irrelevant and, therefore, were ignored in the interpretation.

So, the graphical representation in Figure 2, in which the editorials (Apr.-Jun. 2007) are compared against the editorials (Apr.-Jun. 2008) should be interpreted as follows: in 2007 the BNR discourse was extremely optimistic (high difference values of the class `positive`) and they were giving high importance to Romanian specific aspects (class `nationalism`), while in 2008 (nearly recession time) the BNR discourse had become rather pessimistic (class `negative`) and speculative (class `intuition`) with respect to the Romanian economic future.

In the following we will compare the same 2007 discourse against their discourse immediately after the recession.

The graphical representation in Figure 3, in which the editorials (Apr.-Jun. 2007) are compared against the editorials (July 2008) should be interpreted as follows: the difference in optimism between the BNR discourse one year before the recession and that of the moment the crisis was officially declared (class `positive`) is more pregnant (1.25% here versus 0.89% in Figure 2). However, although the pessimistic tone (class `negative`) is more pronounced in July 2008 than in the period of stability, it has weakened in intensity. We could say that BNR is caution to push too much on the distress pedal, because its voice could influence the fixing and, by that, worsen

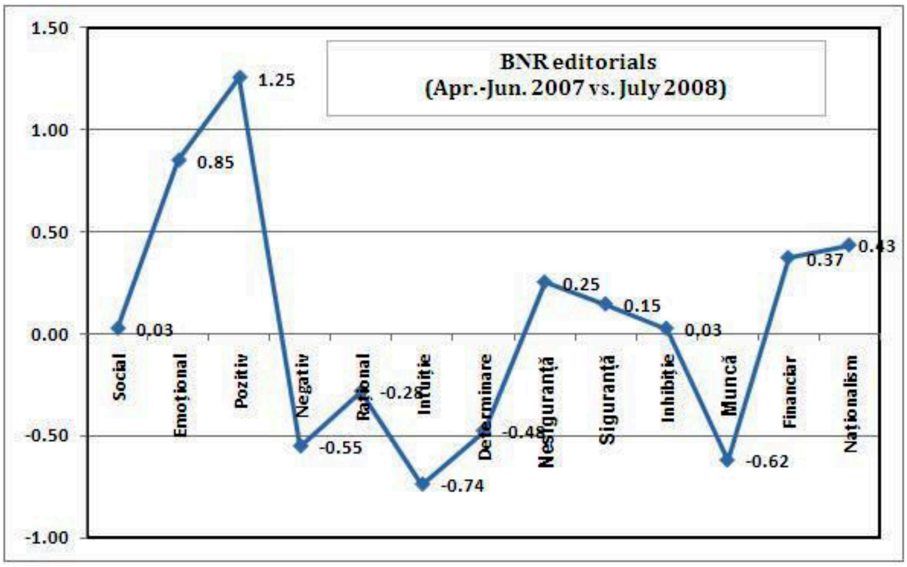


Figure 3: Difference between the occurrence of semantic classes in BNR editorials: one year before the economic recession versus one month after the economic recession.

the financial market even more. Moreover, BNR is offering a possible immediate solution, by accenting on the job sphere (class `work`).

5 Conclusions and Future Work

In this paper we presented a quantitative method and an application that strengthen the idea that crises can be anticipated by monitoring public speeches produced by representative entities.

We are aware that some of the differences which we have evidenced in our comparative study should partially be attributed to idiosyncratic rhetorical styles. However, when the traits inventoried acquire the regularities of patterns, then they could be used as measure apparatuses and, properly used, could emit useful signals to a receptive society.

There are a number of ways in which we think our research could be continued. First, we want to add new features to the platform, with a special emphasis on the syntactic and rhetorical levels of analysis. The new release of DAT should help the user to identify and count patterns of use at the syntactic and rhetorical level. Another line to be continued regards the evaluation metrics, which have not received enough attention till now. We are currently studying other statistical metrics able to give a more comprehensive image on different facets of the public discourse. A weakness of the present system is the fact that the unequal sizes of the lexicons characteristic to semantic classes can influence the decisions: the more entries in the lexicon a certain class contains, the higher its influence could be foreseen. To this problem, the solution is not to balance the classes in their number of entries, because the language makes them intrinsically unequal, but to find calibration techniques that bring their values on equivalent ranges, irrespective of the dimensions of the lexicons. Let's note that in the present study we have counterbalanced somehow this skew by using the difference-based formulas (and thus avoiding absolute values).

Surely, the problem of characterising public speeches receives no final solution with our approach. We believe, however, that our method sheds an interesting light on possibilities of automatically interpreting discourses and, equally, it opens new perspectives.

Acknowledgements

In pursuing this research the authors had partial support from the projects POSDRU-63663, ICT-PSP 250467-ATLAS and ICT-PSP 270893-Metanet4U.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):47-55, December, 2012.

Bibliography

[1] Bürger, C.,*Textanalyse als Ideologiekritik, Zur Rezeption zeitgenössischer Unterhaltungsliteratur*, Frankfurt am Main, Athenäum, 1973.

[2] Cristea, D., Răschip, M., Forăscu, C., Haja, G., Florescu, C., Aldea, B., Dănilă, E., The Digital Form of the Thesaurus Dictionary of the Romanian Language, in *Proceedings of SpeD 2007 Speech Technology and Human-Computer Dialogue*, Iaşi, May 10-12, 2007.

[10] Dijk van, T. A., *Sémantique générale et théorie des textes*, Linguistics, 62, 66-95, 1970.

[3] Gerstlé, J., *Comunicarea politică*, trad. Gabriela Cămară Ionesi, Institutul European, Iaşi, 94, 2002.

[4] Gîfu, D., Cristea, D., Computational Techniques in Political Language Processing: AnaDiP-2011, in *J.J. Park, L.T. Yang, and C. Lee (Eds.), FutureTech 2011*, Part II, CCIS 185, 188-195, 2011.

[5] Lasswell, H. D., *Politics: Who Gets What, When, How*, McGraw-Hill, New York, 1936.

[6] Lazarsfeld, P. F., Berelson, B., Hazel, G., *The Peoples Choice: How the Voter Makes up His Mind in a Presidential Campaign*, 3d ed., New York, Columbia University Press, 1944.

[7] Perelman, C., Olbrechts-Tyteca, L., *Traité de l'argumentation*, Éd. de l'Institut de Sociologie de l'Université Libre de Bruxelles, 72, 1972.

[1] Plett, H. F., *Ştiinta textului şi analiza de text*, trad. Speranţa Stănescu, Ed. Univers, Bucureşti: 72, 1983.

[9] Romaine, S., *Language in society. An Introduction to Sociolinguistics*, Oxford University Press Inc., New York, 1994.

[11] Wodak, R., *Critical Linguistics and Critical Discourse Analysis*, Handbook of Pragmatics, Benjamins, 2006.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):56-64, December, 2012.

Visual Depth Perception of 3D CAD Models in Desktop and Immersive Virtual Environments

F. Gîrbacia, A. Beraru, D. Talabă, G. Mogan

Florin Gîrbacia, Andreea Beraru
Doru Talabă, Gheorghe Mogan
Transilvania University of Brasov
Romania, 500036 Brasov, Eroilor, 29
E-mail: garbacia@unitbv.ro, aberaru@unitbv.ro
talaba@unitbv.ro, mogan@unitbv.ro

Abstract:
In this paper is presented an experimental study that aims to compare the depth perception of virtual prototypes in immersive virtual environments with the depth perception of CAD models using 2D LCD display. First, a multipurpose solution of a large-scale interactive multi wall projected virtual environment named Holo-CAVE is described and then the conducted experiments are presented. The experiments carried out highlight that perceived depth values estimated for virtual prototypes are significantly influenced by the 3D stereoscopic visualization. Another interesting result of the study is that the estimated depth accuracy increases with the depth size that has to be perceived. The results of experimental study illustrate that the use of immersive stereoscopic visualization is useful during Computer Aided Design related activities.

Keywords: virtual reality, computer aided design, immersive 3D systems.

1 Introduction

Current Computer Aided Design (CAD) systems offer extremely rich modeling features and functions for the development of 3D virtual prototypes, which increase the productivity of the new products design. While the geometrical database is 3D since long ago, the user interaction within this software has not significantly changed. At present time CAD tools use standard WIMP (Window, Icon, Menu, Pointer) desktop-based Graphical User Interfaces (GUI), and the interaction is made through keyboard, mouse and CRT/LCD display which are solely 2D devices.

In the last years, Virtual Reality (VR) technology became a dynamic field of research that has began to be used to a certain extent in industrial applications. An important goal of the current worldwide research efforts is to facilitate the implementation of VR in industrial product development processes and to asses the impact and its feasibility into the workplace and everyday life contexts in terms of cost-effectiveness, human-machine interaction and side-effects on the users, as well as their impact on the actual working environment, at both individual and organizational level. VR provides new perspectives for user interaction with CAD tools. It enhances the immersion feeling and the depth perception of 3D objects, providing information with less perceptive ambiguities. This opportunity is important for a CAD application where users must have a direct and thus better appreciation of object shapes and dimensions. Many research activities are currently focused to integrate CAD architecture inside VR-systems in order to enhance the immersion feeling and the user interaction interface [1], [7], [10].

In many applications, VR technologies are used only for visualization and analysis of previously created CAD models [7], [10]. Another emerging category of VR design applications are the VR-CAD integrated systems, which allow the creation, modification and manipulation of 3D models directly in the VR environment [1]. Despite of the intensive research activities, none of them produced a significant impact for the development of the next generation of CAD systems.

Therefore, it is necessary to develop various experimental researches in order to evaluate the impact of Virtual Reality technologies in the design process, and analyze their advantages and shortcomings. The last generation of commercial CAD systems still uses 2D CRT/LCD displays for visualization in most of the cases. The disadvantage of these devices is the lack of depth perception of 3D models.

Generally, the commercial visualization systems offer better visualization conditions comparing to the reduced possibilities of 2D displays. On the other hand, the performance of various technical solutions is different, each of them being appropriate to be used in special defined situations. Therefore, regarding CAD related activities, the development of an evaluation study of the visual perception is necessary, that will highlight the impact of VR technologies on visual perception of 3D CAD models during the design process of products.

There are previous extensive researches on depth perception of 3D models using immersive or volumetric 3D environments [3], [4], [5], [6], [8], [9], but the novelty of this experimental study is the comparison of the depth perception of CAD models using two distinct display modalities: monoscopic desktop system and 3D stereoscopic immersive environment.

The conducted experimental study analyzes the depth perception appreciation (value dimension along Z axis) of 3D CAD models using immersive CAVE-like stereoscopic visualization systems compared to the usage of 2D traditional display. It is known that the real perception of dimensions in Computer Aided Design related activities plays an important role in the decision-making process of a design solution. This experimental study is helpful in evaluating the benefits of immersive visualization system compared with monoscopic visualization using traditional 2D LCD/CRT equipment.

2 Description of Immersive Holo-Cave System

The conventional CAD systems use for the visualization of the generated CAD model a traditional CRT/LCD 2D display. The disadvantage of this type of display for CAD systems is the lack of depth cues. Immersive Virtual Reality CAVE-like systems [2] are 3D stereoscopic displays that significantly improve the way users can visualize, navigate and interact in virtual environments. Compared to other devices like Head Mounted Display (HMD) or volumetric displays, the CAVE-like systems offer several advantages: improvement of the immersion awareness; obtaining high-definition stereoscopic images; large filed of view; collective visualization; collaboration between several users.

A multipurpose architecture was developed at VR lab of Transilvania University of Brasov, that is able to provide both possibilities for the 3D visualization: four walls CAVE-like and Holobench [12] functionality. Therefore, the system is called "Holo-CAVE" (figure 1). This solution allows making experiments related to the study of product engineering tasks that are performed by a human operator in the posture "seated" in the case of Holobench functionality or, alternatively, in a "standing" posture when the system is configured as CAVE. The developed system enables the visualization of large scale, high-resolution 3D stereoscopic images with a large field-of-view. Another advantage is the improvement of immersion awareness and the possibility to visualize the CAD models in their natural dimensions.

The physical structure of the Holo-CAVE has the dimensions of 2.8 x 2.8 x 3 meters. The hardware architecture of the VR Holo-CAVE is presented in figure 2. The used screens were Screen-Tech type, rigid back-projection with the dimension of 2.7 x 2 meters. The mechanical frame of the system was constructed from wood material because it was the simplest and easiest construction free of magnetic field that could be built. The screens were attached to the frames using tubes and glue. Considering the high price and the variety of manufacturers, it was decided to use eight Hitachi CPX1350 high-end projectors (two for each screen for displaying the passive

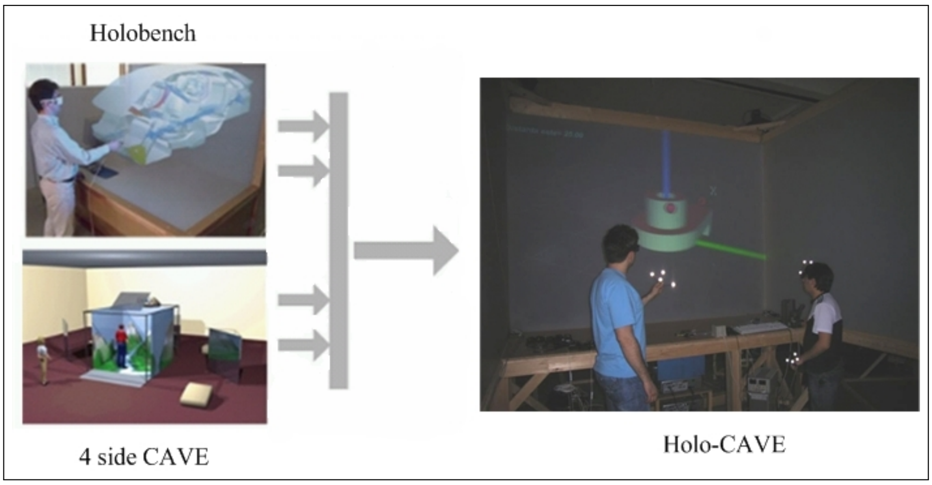


Figure 1: The Holo-CAVE immersive system

stereoscopic images) capable of displaying images with a resolution of 1400 x 1050. The physical projection distance between the projector and the screen is of 5.2 of meters. Mirrors have been used to cut down the required distance. In order to calculate the exact locations and dimensions of the mirrors and the projectors a CAD model was used. A PC Cluster was used as a computer system, composed of one server with a dual 2.4 GHz CPU and eight PC with 3 GHz CPU and dedicated video cards.

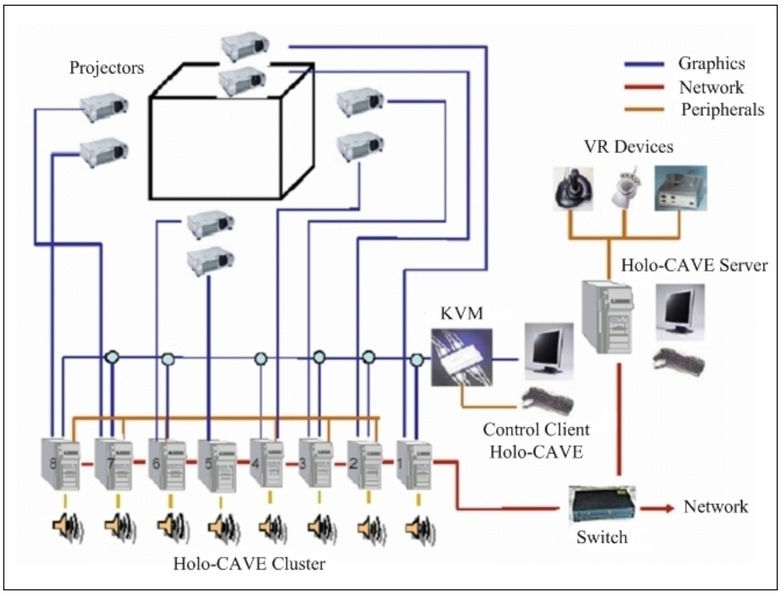


Figure 2: Architecture of the Holo-CAVE immersive system

The Holo-CAVE software is capable to load and display in a synchronized way a 3D scene on the multi-wall display environment, to display passive stereo 3D images and plug in different VR devices. It also provides methods by which the user can manipulate, add or remove objects in virtual environment. The Holo-CAVE software architecture is designed as a distributed highly modular network based on the strict separation of its VR system management into two layers: a Multi User Server that performs the administration of the 3D model, the users' interaction and a Virtual Environment Server that coordinates local projections and navigation devices. The 3D representation is full VRML2.0 (Virtual Reality Modeling Language [11]) thus compliant with all VRML sensors, events and sounds that can be used. Because VRML is the data format, VRML events are used for communication. Based on this approach, all VRML sensors, thus environmental sensors (time, proximity, visibility, and collision node), pointing device sensors (plane, cylinder, sphere, anchor, touch) and embedded JavaScript/ECMA Script can be used.

3 Experiment Description

This study tries to answer the following research questions:

1. Is immersive 3D visualization useful for the design engineers?

2. What is the performance of 2D display devices compared to 3D immersive visualization for perception of dimensions of 3D CAD models?

3. Which is the most intuitive and natural interface for the visualization of 3D CAD models?

We have devised and conducted two experiments to measure and record the estimated depth value of several CAD models using two types of displays. The former is the traditional desktop workspace with 2D input (keyboard and mouse) and 2D output (computer screen) peripherals. The latter consists of a multimodal immersive interface of an integrated VR-CAD system that uses the immersive 3D Holo-CAVE system. The results of these experiments will allow answering the three research questions.

4 Experiment Procedure

In order to evaluate the perception of depth in a CAD model an experiment was conducted, involving eight subjects (three women and five men) with the average age of 28 years and with a healthy sense of vision. None of the subjects used VR immersive stereoscopic 3D visualization for the perception of 3D CAD models before. Instead, they had extensive experience in using CAD software and good computer skills. In the conducted experiment six 3D CAD models were used, each of them composed from a parallelepiped part with variable dimensions. In order to give the subjects the opportunity of appreciating depth, the models were placed on a virtual table with the size of 300 x 200 x 150 cm. The solid models were visualized using two types of devices: a universal 2D LCD display with the diagonal of 15.4" for desktop interface (figure 3a) and a Holo-CAVE system for immersive perception (figure 3b). To display the 3D environment in the first case a SolidWorks CAD system was used and in the second case dedicated software was used: BSContact Stereo VRML visualization player integrated in the Holo-CAVE. In the beginning, each subject was informed about the purpose of the experiment and specific instructions were given regarding the method of depth estimation. The subjects were asked to assess the depth of six objects using centimeter as measurement unit. In order to estimate the depth of CAD models, the subjects were informed about the size of the virtual table where virtual objects were positioned. In the case of Holo-CAVE immersive system, the distance between the viewpoint of the user and the projection screen was kept constant, 2 m (figure 3b). For each subject were displayed in a random order the CAD models. Each subject that participated to this experiment filled a questionnaire in which they were asked to provide information about age, experience of using VR equipment, experience in using CAD systems and computer skills.

Half of the users estimated first the depth of virtual objects using the traditional desktop CAD system, then, after a break of 20 minutes, they were asked to estimate the depth of virtual CAD objects using the Holo-CAVE stereoscopic visualization system. Simultaneously, the other half of subjects estimated first the depth of 3D CAD objects using stereoscopic system and then using monoscopic desktop system. The value of the estimated depth was recorded in a text file that was used afterward for the assessment of the results.

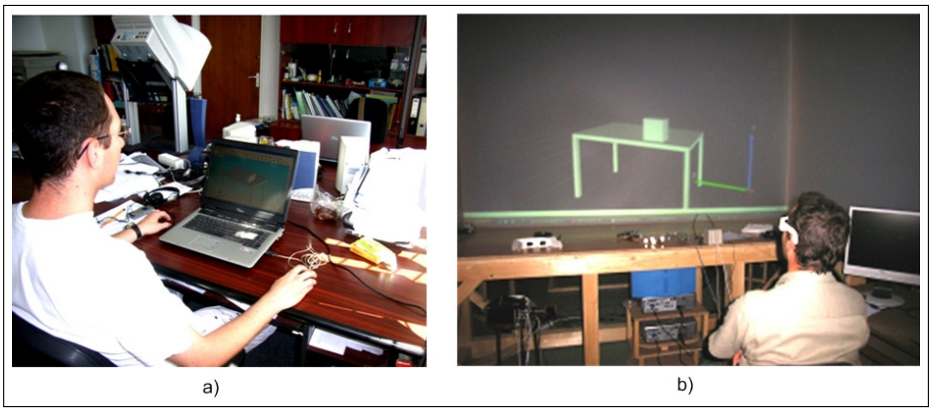


Figure 3: The subject estimating depth using 2D display(a) The subject standing inside the CAVE-like visualization system(b)

5 Results Evaluation

Figure 4a presents the difference between the estimated values of the depth and the real dimensions of objects. After analyzing the data, the drawn conclusion is that for small values of depth (less than 35 cm) the subjects overestimated the depth of CAD 3D models. Another significant result is that for higher values of depth, the average of estimated depth was more accurate when using stereoscopic 3D immersive visualization. An interesting result obtained by using the immersive 3D visualization, was that for all models the subjects overestimated the depth value.

Figure 4b shows the accuracy of depth estimation that was obtained by using the value of relative error. The relative error was calculated using the following formula:

$$E_r = (D_p - D_r)/D_r \tag{1}$$

in which Er - is the value of the calculated relative error, Dp - the value of the estimated depth, Dr - the real value of virtual objects depth.

Figure 4: Difference between the estimated values of the depth and real object dimensions(a); Accuracy of depth estimation(b)

If the value of relative error is positive, the subject overestimated the depth of the CAD models, and if the value of relative error is negative then the subject underestimated the depth of the CAD models. The highest value of the relative error was obtained for the depth value of 35 mm and was due to the overestimation of depth. The conclusion drawn from this experiment is that the precision of depth estimation for stereoscopic viewing is lower for CAD models with small depth values, but increases significantly when CAD models depth value is higher.

After conducting the experiments, each of the subjects was asked what viewing equipment he/she prefers. Most subjects would use the immersive stereoscopic system because of the superior intuitive way of visual perception. However, few subjects considered as a shortcoming the need to wear glasses for passive stereoscopic visualization. We can conclude that the subjects estimated the depth of 3D CAD models with greater accuracy using the Holo-CAVE stereoscopic immersive visualization compared to monoscopic traditional desktop display. Another interesting result drawn from the experiment was the increasing of estimated depth with the dimensions of the 3D CAD models.

In order to emphasize the results of the experiments described above, there was a new series of experiments conducted on the same experimental set-up already presented, namely the Holo-CAVE system. The experiment was dedicated to assert the variation of stereopsis depth perception. The observers viewed the image by wearing polarizing glasses. The position of the observer was tracked by using a magnetic Ascension Flock of Birds tracking system with 6 DOF. The observer was standing inside the CAVE-like visualization system facing the screen (Fig. 5).

The viewing distance was set to three predefined values (1.5, 2.0, and 2.5 meters). The stereoscopic image consisted of two cubes, a red and a blue one having the sides of sizes 50 and 35 cm respectively. The arrangement of the cubes was such that the smaller one was set to a distance of 2 m (behind) with respect to the bigger one. The cube displayed to the left eye had a range of disparities added to it by shifting its horizontal position. The values of the disparities were 1, 6, and 11 cm. When the observer fuses left and right image he always perceives the cubes being in front of the screen. Each observer was tested individually having the task to estimate the depth of the scene, namely which is the distance he perceives to the red cube and to the blue cube. The dimensions of the cubes and the relative position of one to the other were not made known to the participants. Free eyes movement and as much time as required to estimate the depth of the scene were allowed. Each observer for all values of the viewing distance and disparities repeated the task.

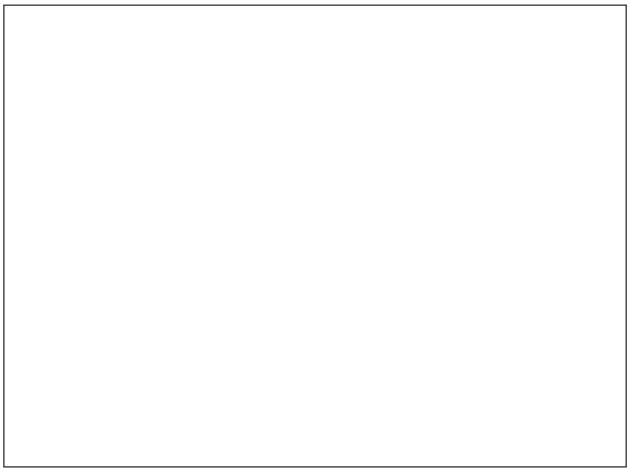


Figure 5: The observer standing inside the CAVE-like visualization system

Depth of a scene can be determined by using a simple arrangement as in figure 6, where e is the interpupillary distance, D is the viewing distance, d is the disparity distance and L is the depth distance from the observer. The following equation expresses the depth L as function of variable d:

$$L = eD/(e + d) \tag{2}$$

For D and e constants, the depth is affected only by the variation of disparity d. Therefore, when disparity becomes smaller the object tends to be farther away from the user and vice versa. This is also illustrated in figure 7 where the predicted depth is represented as a function of the disparity distance for all the three cases of the viewing distance. For the calculation, an average value of the interpupillary distance of 6 cm has been considered.

Figure 6: Determination of depth distance

Figure 7: Calculated depth

The results of these experiments are summarized in figures 8a and 8b. The graph in figure 10 shows a good correspondence between the average values of the perceived depth as a function of disparity with the calculated values of the depth for the same value of the viewing distance D = 2m. The open circles represent the average values of the perceived depth whereas the full circles are the calculated values of the depth. For D = 1.5m, one can found the same good correspondence between the theoretically estimated values of depth and the perceived values. For

Figure 8: Depth as a function of disparity(a);Depth as a function of the viewing distance(b)

D = 2.5m the agreement between theoretical and experimental values is not so good anymore, in this case the observers reporting difficulties in estimation of the depth.

In figure 8b it is displayed the dependence of the perceived depth on the viewing distance for a constant value of the disparity d = 1 cm. For the sake of the comparison, the calculated depth is displayed too. The same trend it is observed for all the other values of the disparity. Concerning the precision of the depth perception of the viewer, it is observed that the users presented more accurate stereopsis when the value of the disparity is small while increasing the disparity value leads to more imprecise stereopsis.

6 Conclusions and Future Works

Realistic perception of the models depth in Computer Aided Design plays an important role in decision making of design engineers. In this paper was presented an experiment aimed to estimate the depth of 3D CAD models. From the performed experiment, we can emphasize that the perception of CAD model depth is significantly influenced by the stereoscopic visualization. The subjects estimated depth of 3D models with greater accuracy using the immersive stereoscopic Holo-Cave system compared to traditional desktop display. The accuracy of depth perception is not considerably improved when the depth of CAD models is small, but it increases significantly corresponding to a higher depth. As a general conclusion, we can declare that the alternative of replacing the 2D desktop systems with 3D VR visualization systems can be considered a viable alternative.

7 Acknowledgments

The research activities where supported by the Sectoral Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number POSDRU/89/1.5/S/59323.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):56-64, December, 2012.

Bibliography

[1] Bourdot, P.; Convard, T.; Picon, F.; Ammi, M.; Touraine, D.; Vezien, J.-M.(2010); VR-CAD integration: Multimodal immersive interaction and advanced haptic paradigms for implicit edition of CAD models, *Comput. Aided Des*, 42(5): 445-461.

[2] Cruz-Neira, C.(1995); Virtual Reality Based on Multiple Projection Screens: The Cave and its Applications to Computational Science and Engineering, Ph.D. Dissertation, University of Illinois at Chicago, Chicago, IL, USA. UMI Order No. GAX95-32383.

[3] Foley, L. M.(1991); Stereoscopic distance perception, *Pictorial communication in virtual and real environments*, Stephen R. Ellis (Ed.). Taylor & Francis, Inc., Bristol, PA, USA, 558-566.

[4] Grossman, T.; Balakrishnan R.(2006); An evaluation of depth perception on volumetric displays, *Proceedings of the working conference on Advanced visual interfaces (AVI '06)*, ACM, New York, NY, USA, 193-200.

[5] Hoskinson, R.; Akai C.; Fisher, B.; Dill, J; Po B.(2004); Causes of depth perception errors in stereo displays, *Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, ACM, New York, NY, USA, 164-164.

[6] Lang, M.; Hornung, A.; Wang, O.; Poulakos, S.; Smolic, A.; Gross, M.(2010); Nonlinear disparity mapping for stereoscopic 3D, *ACM Transactions on Graphics*, 29(4): 1-10.

[7] Raposo, A.; Soares, L; Wagner, G.; Corseuil, E.; Gattass, M.; Santos I.;(2009); Environ: integrating VR and CAD in engineering projects, *IEEE Comput. Graph. Appl.*, 29(6): 91-95.

[8] Reichelt, S.; Haussler, R.; Futterer, G.; Leister, N.(2010); Depth cues in human visual perception and their realization in 3D displays, *Three Dimensional Imaging, Visualization and Display*, Bahram Javidi and Jung-Young Son(Ed.), Proc SPIE 7690, 134-144.

[9] Svarverud, E.; Gilson, S.J.; Glennerster, A. (2010); Cue combination for 3D location judgements, *Journal of Vision*, 10(1): 1-13.

[10] Weidlich, D.; Cser, L.; Polzin, T.; Cristiano, D.; Zickner, H. (2009); Virtual reality approaches for immersive design, *Int. J. on Interactive Design and Manufacturing*, 3: 103-108.

[11] www.web3d.org/x3d/specifications/vrml/ISO-IEC-14772-VRML97/

[12] http://www.barco.com/fr/virtualreality/product/961

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):65-73, December, 2012.

Strategic Decision Models Cross-Validation by Use of Decision Reports Information Extraction

L. Hancu

Lucian Hancu
1. "Babes-Bolyai" University of Cluj-Napoca
Romania, Cluj-Napoca, 1 M. Kogalniceanu, and
2. SoftProEuro Ltd. Cluj-Napoca
Romania, 400614 Cluj-Napoca, 1 Lacul Rosu
E-mail: lhancu@softproeuro.ro

Abstract:
From all the events in the life of a business entity, the Mergers and Acquisitions transactions are one of the most challenging ones, as they drastically affect the life of the involved entities, but also their business stakeholders (like clients or suppliers). The Merger transaction can be seen as a growth crisis in the life of the buyer entity and a strive for survival in the life of the acquired company. Studying such transactions are being a constant preoccupation for both academia and practitioners, modeling mergers in order to predict them - one of the most ambitious task. In this paper, we present our technique of cross-validating the results of our model and use several boosting methods for improving the computed decisions scores.

Keywords: Mergers and Acquisitions, Quantitative Models, Cross-Verification, Boosting Algorithm, Growth Crisis, Business Survival

1 Introduction

The strategic decisions of the type Mergers and Acquisitions are of crucial importance for the life of both the entities involved in such a process and their stakeholders ones (clients, suppliers, or even competitors). Predicting such transactions are, thus, of great importance for the participants to the economic activities, as the changes in the market conditions can drastically affect the entities, especially the small competitors.

Bearing this in mind, we have previously built a model of predicting future mergers and acquisitions based on the financial statements of the entities involved in such a strategic process and on the correlations between the two entities activity's codes (the so-called Business Dependencies Map [4]). In addition, data regarding previously completed acquisitions are available on the Web and can be easily downloaded and analyzed. Information Extraction from such data can be of great help in cross-verifying the quantitative models for Mergers and Acquisitions.

In this paper, we apply a cross-validation mechanism in order to correlate data manually extracted from the Competitors Council merger decisions reports from 2003 up to 2008 of the type Buyer (the entity who bought another entity) - Target (the entity who was bought) - Seller (the entity who sold its ownership of the Target entity to the Buyer entity) with the results of the MAVOC (Mergers, Acquisitions, Virtualizations or Conservations) quantitative model. The cross-validation occurs only when the Buyer and Target are both Romanian entities and their financial statements are present in the previously-computed database of the Top of the Romanian Entities, so it is possible to compute the MAVOC quantitative score.

Prior to performing such a step, an automatic cleansing is performed on the data extracted from the Competition's Council Decisions Reports, that assures that the entities are found on the databases collecting the financial statement. The cleansing step is crucial as many entities change names after the completion of an acquisition transaction, which makes difficult (or even impossible) the finding of the financial information regarding the specified entity.

In addition to the cross-validation task, a boosting algorithm is used in order to improve the results of the Mergers and Acquisitions MAVOC model. The boosting algorithm is based on the risk profile of both buyer and the seller and it takes into consideration the risk associated with the two entities activity codes. The boosting algorithm has the scope of improving the acquisition score and downgrading the other scores (especially the inverse acquisition **A-** and conservation **C**), so as the MAVOC model would output **Acquisition** for the two companies extracted from the Decision Reports. This boosting technique is required when the acquiring company's financial strength is similar to (or weaker than) the acquired company's financial strength, which would conduct to a false inverse Acquisition (**A-**) recommendation if used alone without boosting.

The paper is organized as follows. In the following section we provide a brief introduction to the mergers and acquisitions research. This research usually is the result of consultant companies and it takes several years of investigation, when analyzing transactions from several decades. In the subsequent section, we briefly explain our methodology of modeling mergers and the technique of extracting data from decisions reports from the Competition Council, that are later used in cross-validating the model. The fourth section describes our boosting techniques aimed at improving the results of the decision model, whereas the paper ends with a brief summarization of the discussed topics and depicts directions for future work.

2 Mergers and Acquisitions Transactions

A thorough analysis of more than 20 years of Italian mergers and acquisitions transactions is done in [7]. The research takes into consideration the interval between 1998 and 2010, in which the authors analyze transactions from various sectors including banking and public services like electricity and gas. It is analyzed the context of such transactions - the Italian economy in the analyzed period, in which the small and medium entities occupy a large percentage of the total amount of business entities. The context is, to some extent, similar to the one of the Romanian market: some transactions were done as privatizations, between the State agencies and private entities (most of them being foreign entities), others being transactions between foreign entities (one which previously acquired the company and which is now willing to exit from the investment).

A special role in these transactions are occupied by investment funds, business entities that acquire several percentages of companies, develop a new business, then sell the company to a third-party investor. While transactions between two companies (competitors, clients or suppliers) are easier to be analyzed - as one could extract features like financial indicators, position on the market, coverage of the market, the case of private funds (also known as investment funds or equity funds) remain a distinct subject of research and it shall be left behind during our research.

The main motives of mergers transactions are depicted in [2]: to affect more rapid growth, gain economies of scale, increase market percentage, expand in the territory, increase stock market value, expand or improve the mix of products, spread risk through diversification, enhance the power and influence of the entity, invest the entity's idle capital, acquire technical knowledge and expertise, counter cyclical of seasonal revenues, obtain managerial talent, gain from tax advantages, obtain more control over the supply sources and/or the retail outlets or to defend against a possible takeover.

Some of these motives are summarized also in [1]. The main aim of the Merger transaction is to realize value, by managing risk and exercising power. The mergers transactions from the market power perspective are also analyzed in [3] and [8]. The former research states that markets are passing through several stages in their way to consolidation, when having almost 90% of the market power concentrated into the industry giants. The latter research, instead, focuses on exploting niche markets for mergers transactions.

The research literature of Mergers and Acquisitions coming from both practitioners but especially from academic researchers is crucial in deriving the criteria for future modeling of mergers and acquisitions transactions. By analyzing the research literature, we have figured out that several mergers motives can be modeled by quantitative variables, while others (denoting especially human resource-related questions) focus on more qualitative results. In this article, we shall concentrate on modeling mergers through quantitative models and improve them by cross-validations with data previously extracted from the decisions reports of the National Competition Council, that are published through their public Web Information Systems.

3 Gathering Decision Reports and Cross-Verification of Merger Model

Modeling mergers decisions, based on the financial indicators of the analyzed entities and the business dependencies of the various business activities (which constitutes the *Business Dependency Map* [4], has been first explained in [5]. The **M***ergers* **A***cquisitions* **V***irtualizations or* **C***onservations* **MAVOC** model consists in extracting the financial indicators of the involved entities from the available public Web sources (like the Ministry of Finance or Registry of Commerce).

The model makes use of the two-business entities activity codes, number of employees, financial resources (turnover, tangible assets, intangible assets), market share - whether the entities are part of the local (or national) top of the entities corresponding to the county of each business entity. According to the relations between the financial indicators of the two business entities, a score for each alternative (MAVOC) is computed, the score which is higher ranked is returned as the suggested alternative for the two business entities for the specified year.

One method of verifying the quantitative mergers model would be to extract data from the decision reports published by the Competition Council and cross-verifying the results of the model with the data extracted from these decision reports. In the following paragraph, we briefly explain the methodology we use in extracting data from these reports.

3.1 Extracting Information from Decision Reports

Strategic decisions like mergers or takeovers are analyzed by the country's Competition Council (in the case that the two entities are from the same country) or by each country's specific Competition Council (in the case that the two business entities are coming from two different countries). Upon each analysis, a decision report is issued (and usually available using Web Information Systems) which describes the details of the *transaction* and the Council's acceptance or rejection.

The technique of extracting information from these reports has been thoroughly explained in [6]. We extract information from these reports then cross verify the extracted data with our data sources, in order to obtain relationships of the type **Buyer** (the entity which bought another entity) - **Target** (the entity who was bought) - **Seller** (the entity which previously owned the **Target**). In some cases, the relationship is restricted to only Buyer - Target, as there is no entity which owns the Target company.

3.2 Cross-Verification of the model with previously-extracted data

The extraction of the data resulted in 740 relationships of the form Buyer - Target - Seller, from which $20,27\%$ of them were business entities coming from Romania. In order to further perform cross-validation tasks, we had to check that these entities are business entities (and not public agencies owned by the state), so that we can obtain the financial indicators from the external information sources.

The results of the cross-validation baseline experiment are depicted in the first row of the Table 1 and in the Figure 1. The table synthesizes the percents of the final results of the model for the input coming from the decisions reports: $2,67\%$ were classified as Mergers, 14% were classified as *inverse acquisitions* (if A and B are the entities, we denote *direct acquisition* or **A+** the acquisition of B made by A and *inverse acquisition* or **A-** the acquisition of A made by B).

In addition, $26,67\%$ were classified as Virtualizations (creation of short-term virtual entities for profiting from a business momentum), 30% as direct acquisitions and $26,67\%$ as errors. The errors are due to the fact that several values are missing in our Information Systems which contain financial data on business entities. Instead of adding an auto-updater tool to our Information Systems for correcting these errors, we have performed several boosting strategies in order to correct or limit these errors, strategies that are explained in the next section.

The last two columns of the Table 1 represent the relative number of correct answers with respect to all the output of the model including the errors, and the relative number of those answers with respect to the output excluding the errors. The former is computed by dividing the M, A+ and V answers to the M, A+, A-, V and Error answers, while the latter is computed by dividing the M, A+ and V answers to the M, A+, A- and V answers. As previously noted in our experiments, with the introduction of the 250 top of the entities (instead of the previous top of 10 entities in the original model [5]) the number of Conservations (C) results is zero. We consider here the M, A+ (direct acquisitions) and V as correct answers for the cross-validations with the decision reports data extraction. During our boosting experiments we shall concentrate in minimizing the number of errors and the number of inverse acquisitions A-.

In the Figure 1 we depict the analytical scores of the model - the maximum between the M, A+ and V scores in comparison with the A- and C scores. The image serves at an overview of the baseline experiment. We shall improve the score and present an updated version of the score, during the boosting section.

4 Boosting Techniques

Several approaches were required for improving the results of the decisions scores. In this section, we briefly summarize the boosting approaches that we are using.

Experiment 2 Downward model - in order to improve the number of correct answers, we apply a downward strategy to our model: we take the initial year of the transaction and the fiscal IDs of the two entities from the extracted decision reports (see previous paragraph for the methodology). If the model outputs an error, we apply the same model to the previous years from the range 2003..2008 in order to obtain better results (non-error result of the model). The reason for applying this boosting technique is: several transactions completed at the end of a fiscal year are analyzed only in the next year, when one of the entity might not be a valid entity any more, in the mean time being absorbed by the other entity, and, thus, making impossible the gathering of its financial indicators for the current year. The results are depicted in the second row of Table 1 and reveal a slight improvement of the results of the experiments.

Experiments 3 Upward model - a similar mechanism of upward strategy is also used, when the model outputs the same error (indicating that one or both business entities are missing from the Information Systems business data), with the sole difference that we analyze the years after the transaction and output the model decision score, when available.

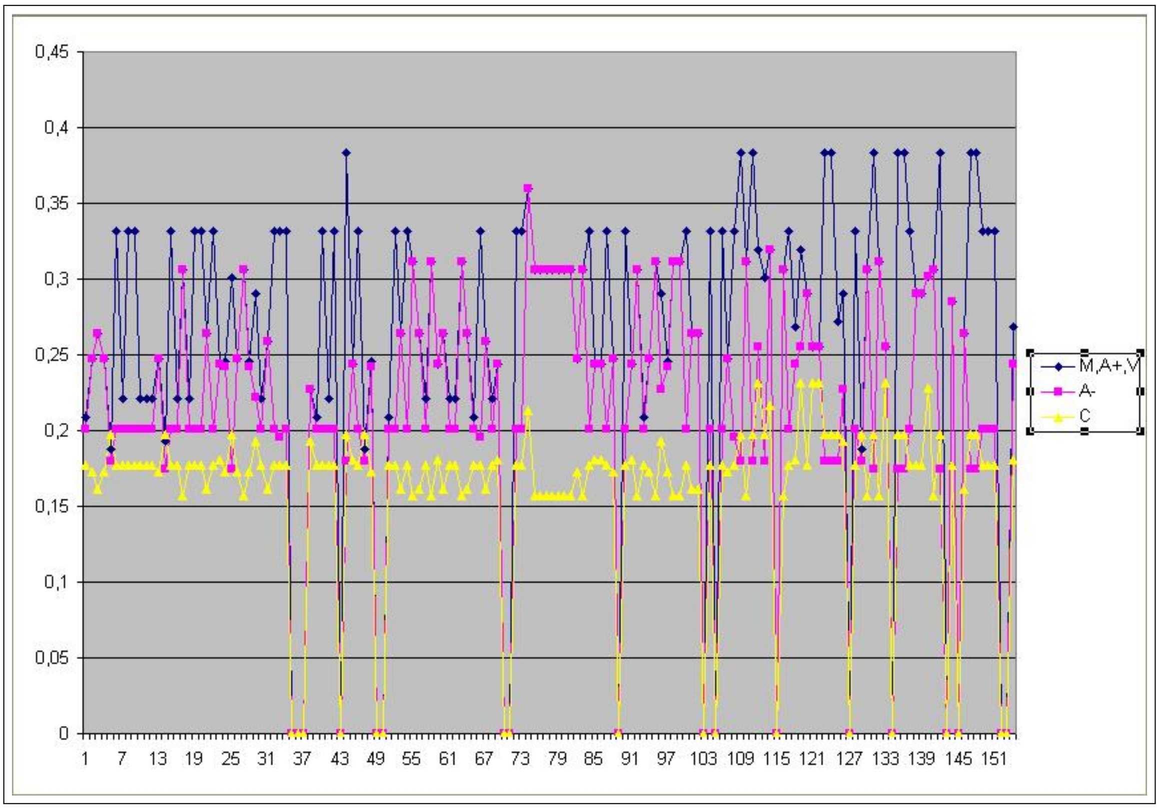


Figure 1: Model score **before** boosting

Experiments 4 and 5 Using **Total incomes** instead of **Turnover** into the model. By manually inspecting the false negative results from our experiments, we have discovered that a number of entities had much higher total incomes with respect to the turnover, which suggested us to use the other indicator into the model in order to improve the results. The difference between the two experiments is a small change in the importance of the Business Classes of the two entities, that affected the results of our experiments; we shall use this strategy in the boosting techniques, discussed below.

Risk-based Boosting: Experiments 6-12 The main idea of this paper was to apply risk-based boosting techniques in order to improve the decision scores of our model. The motivation for this approach relies on the fact that businesses coming from certain business sectors are more willing to accept higher risks than businesses coming from the more conservative business sectors. We have previously developed various techniques for computing the dependency-based risks from the Virtualized Supply Chains; in this article we shall limit at applying boosting techniques to the high risk-tolerant business classes, without insisting in the classification low-medium-high risk Business Class. Thus, higher risks from mergers and acquisitions transactions are supposed to be accepted for the high risk-tolerant business classes.

The Risk-Based Boosting techniques are summarized below:

Model's parameters When the Business Class of the A company is one of the high risk-tolerant business classes, we use different thresholds for the parameters K_1, K_2, K_3, K_4 (in the relations 1, 2, 3, 4, as follows: for the first relation we use the lower-bound parameter for higher risk business classes, allowing that weaker business entities acquire stronger ones

	M	**A-**	**V**	**A+**	**Err**	$Proc_1$	$Proc_2$
1	2,67	14	26,67	30	26,67	59,34	80,91
2	2,67	16	30,67	36	14,67	69,34	81,25
3	2,6	14,94	31,17	39,61	11,69	73,38	83,08
4	2,6	16,23	30,52	38,96	11,69	72,08	81,62
5	2,6	12,34	31,82	41,56	11,69	75,98	86,03
6	1,3	12,99	31,17	42,86	11,69	75,33	85,29
7	5,23	13,07	26,8	43,14	11,76	75,17	85,19
8	4,55	12,34	26,62	44,81	11,69	75,98	86,03
9	1,3	10,39	37,01	39,61	11,69	77,92	88,23
10	12,34	7,14	29,22	39,61	11,69	81,17	91,91
11	7,14	7,14	29,22	44,81	11,69	81,17	91,91
12	2,6	5,19	29,87	50,65	11,69	**83,12**	**94,12**

Table 1: Boosting experiments

(but within the specified threshold between the intangible assets AC_A and the sum between tangible and intangible assets AIM_B and AC_B. The relation 2 specifies that the lowest parameter should be use in less risk-tolerant business classes, whereas the highest should be used in the higher risk-tolerant ones. We point out that the more risk-tolerant business class refers to the business class of A and not to the risk class of B. Similar relations are 3 and e4 when we use the number of employees. More risk-tolerant business classes are those that are willing to assume a higher risk during the acquisition transaction. This could be contracting a credit for the acquisition of the company, or having a weaker company willing to acquire a stronger one for gaining market share.

$$AC_A > K_1 * (AIM_B + AC_B), K_1 \in \{0.5, 1.0\} \tag{1}$$

$$AC_B > K_2 * (AIM_A + AC_A), K_2 \in \{1.0, 2.0\} \tag{2}$$

$$EM_A > K_3 * EM_B, K_3 \in \{0.5, 2.0\} \tag{3}$$

$$EM_B > K_4 * EM_A, K_4 \in \{2.0, 4.0\} \tag{4}$$

Importance of criteria The criteria used in our MAVOC model can be divided into two categories: risk-dependent ones and non-risk-dependent ones. For the risk-dependent ones like the financial criteria, human resources criteria and business classes criterion we use different importance weights than in the standard risk classes. This assures that the risk-dependent variables are more weighted in constructing the final decision score.

Weights of alternatives When using the risk-dependent criteria we provide various weights for the 5 alternatives, that are tuned during the boosting **Experiments 6-12**. Higher weights are given to the alternatives that express a direct acquisition, whereas gradually lower weights are given to the ones expressing inverse acquisition **A-**.

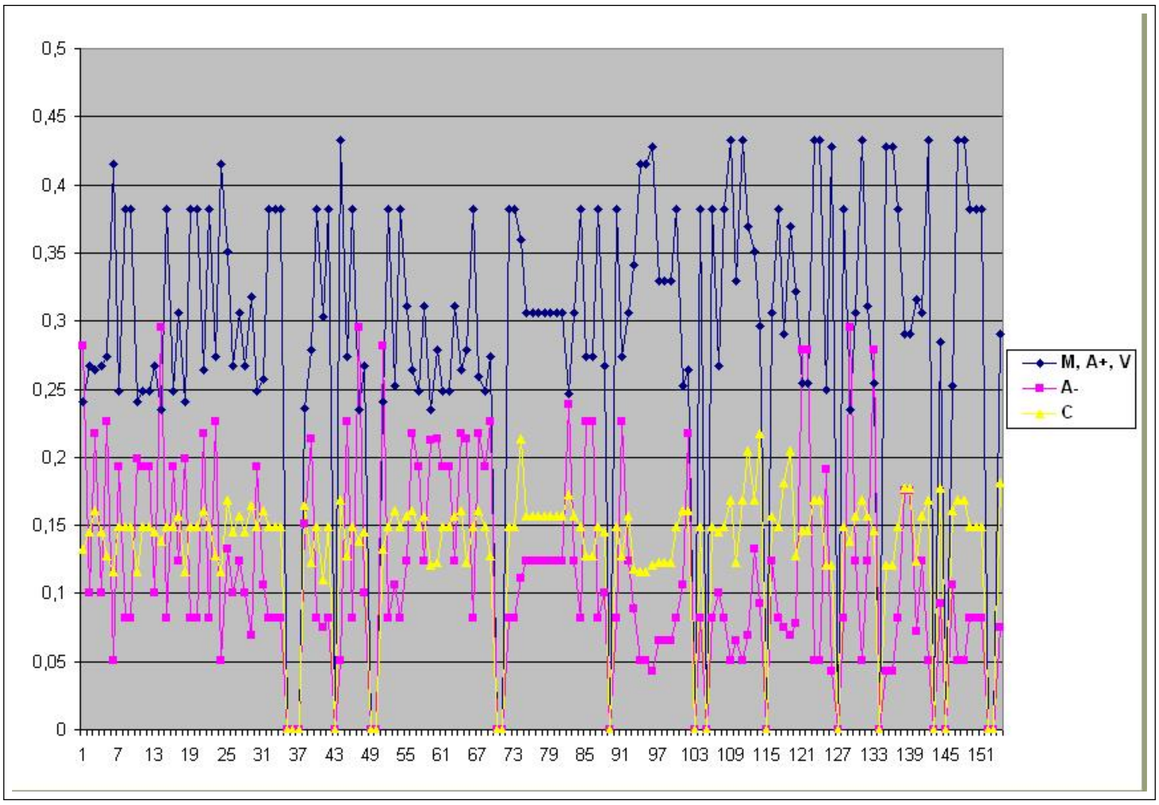


Figure 2: Model score **after** boosting

The results from our experiments are summarized in Table 1 and Figure 2. The figure shows that the cross-validated results are better than the initial ones. The reason for summing the results of the scores **M, A+, V** is that we can consider the Merger and Virtualization score as similar to the acquisition score. As stated in the experiments from [5], the Merger score is given to almost equal in financial and human strength of business entities with similar activity codes, whereas the Virtualization score is given to the same kind of business entities, but having different and correlated activity codes.

5 Conclusions and Future Works

In this paper we have presented a technique for boosting the results of the merger scores obtained by cross-validating the MAVOC model decision scores when the business classes are ones from more risk-tolerant business classes.

Several factors limit the results of our experiments: the relatively high number of transactions in which one or two of the businesses from the Buyer-Target-Seller relationship are foreign entities (or local businesses hidden behind foreign offshore companies). In these cases, it was not possible to apply the decision score and boost its results. In our future works, we plan to also investigate the possibility to include foreign companies into the model, by enhancing the Information Systems database which contains the financial statements of the analyzed business entities.

Acknowledgment

This work was partially supported by the **CNCS-UEFISCDI** under the grant **PN II-RU-PD-397/2010**. We would like to thank the *Combinations of companies* M.Sc. students and their academic mentors from the Faculty of Economic Sciences and Business Administration for their contribution in manually extracting the Mergers and Acquisitions informations from the Decisions Reports.

Bibliography

[1] D. Angwin, *Mergers and Acquisitions*, Blackwell Publishing, 2007, pg. 21.

[2] H. K. Baker, T. O. Miller, B. J. Ramsperger, *A Typology of Mergers Motives*, in Akron Business and Law Review, 12(4), 1981, 24-29, reprinted in J. A. Krug, Mergers and Acquisitions, SAGE, 2008, pp. 67-76.

[3] G.K. Deans, F. Kroeger, S. Zeisel, *Winning the Merger Endgame, A playbook for Profiting from Industry Consolidation*, A.T. Kearney, 2003, pp. 22-95.

[4] L. Hancu, *Data-Mining Techniques for Supporting Merging Decisions*, in International Journal of Computers, Communications and Control, Suppl. Issue, 2008, pp. 322-326.

[5] L. Hancu, *Pruning Decision Trees for Easing Complex Strategic Decisions*, in Annals of the Tiberiu Popoviciu Seminar, Volume 6, 2008, pp. 194-203.

[6] L. Hancu, *Mining Strategic Decisions Information Systems for Predicting Future Market Concentrations*, in Proceedings of the International IADIS Information Systems Conference, March 2012, Berlin, Germany.

[7] KPMG, *20 anni di M&A - Fusioni e acquisizioni in Italia dal 1988 al 2010* (20 years of M&As - Mergers and Acquisitions in Italy from 1988 to 2010), EGEA, 2010.

[8] F. Kroeger, A. Vizjak, M. Moriarty, *Beating the Global Consolidation Endgame*, A. T. Kearney, 2008.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):74-82, December, 2012.

Using Opinion Mining Techniques for Early Crisis Detection

A. Iftene, A.L. Ginsca

Adrian Iftene, Alexandru-Lucian Ginsca
"Alexandru Ioan Cuza" University of Iasi,
Faculty of Computer Science
E-mail: adiftene@infoiasi.ro,
lucian.ginsca@infoiasi.ro

Abstract: The goal of our research is to investigate the use of internet monitoring in crisis management using linguistic processing and text mining techniques. We present a system that detects and classifies events on topics and, using an altered opinion mining workflow, detects geographical entities related to these events and the sentiments expressed towards them. The results are displayed in customized GoogleMaps views, indicating areas with a potential risk, such as natural disasters, unfavorable weather or threatening protests. All the processing is done in real time and, depending on the monitored sources, our work could be of used as a population warning system, but it could also be useful for regional or local authorities in managing intervention time and resources by prioritizing the situations for which they have to act.

Keywords: Opinion mining, Event detection, Crisis management.

1 Introduction

In recent years, an increasing trend regarding Internet monitoring for multiple types of crises (political, weather, terrorist or health related) can be observed. An example of the use of web mining in conflict detection that fits in such a trend is described in [5], in which the authors focus on the 2011 African protests. The main differences between their approach and ours are that we propose a system that can be easily adapted to different types of crises, identifies threats at a more localized level (districts, streets) and that we use a purely automatic approach, whereas they combine web mining with human reports.

The main components of our system allow us to monitor a collection of newspapers and to save on our computers all the news. After they are processed locally, we detect the main topics and we find the most relevant topic(s) for a particular crisis scenario. In next step, named entities and users opinions are identified, and based on them the risks are identified. Accordingly to the values attached to locations, a Google Map is created and a set of "islands", some with potential risks and some without risks are generated on the map. In the last step a user receives an alternative path for a pair of (start location, end location), which avoids as much as possible the islands with negative scores (those drawn in red, with potential risks) and that approaches the "islands" with positive scores (those drawn in green, without potential risks).

In the following chapters, we will present the main components of our system. Given the fact that the main purpose of the research described in this paper is to incorporate opinion mining elements in crisis detection systems, we will insist more on those components that are used for this task such as event detection and a proper graphical representation of the results and less on those concerning strictly opinion mining. Also, we present the new resources especially created for this task, such as those used for the identification of streets and the detection of opinions related to crisis management.

2 System Description

Below, we present the most important components of our system. We will also show how these are used for monitoring protests that took place in Romania between 13 and 26 January, 2012.

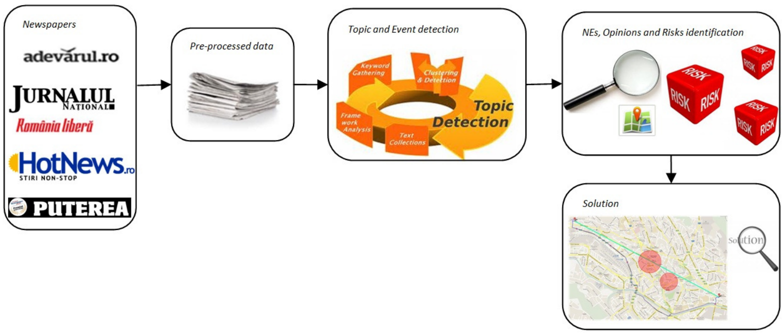


Figure 1: *System architecture*

2.1 Newspaper Monitoring

A number of newspapers are monitored using RSS feeds and the articles are gathered using a crawling component. For our case study, we monitored five newspapers Adevarul, Hotnews, Jurnalul, Puterea and Romania Libera (see Figure 1). This data was stored locally and it was preprocessed (from html pages were removed links, photos, menus, special characters).

2.2 Identification of potential risk events

After the initial processing step, we identify news articles containing mentions of potential risk events. For this task, we propose a novel approach that uses topic models for event classification and semantic similarities for event recognition. On a collection of news articles, after a couple a preprocessing steps, such as lemmatization and stop word elimination, a topic model is applied in order to identify a predefined number of topics present in that collection. In general, topic models describe the topics by a list of relevant words for each topic [1]. This raises a problem, because we want to be able to identify a particular event with minimum human intervention. In order to solve this issue, we use semantic similarities between the words that describe the topic, given by the topic model, and a small manually built vocabulary for an event. Such a similarity measure will be able to find the topic most related to the followed event. In the next sections, we will detail each of these components.

2.3 Topic Detection

To identify groups of topics we have used the Latent Dirichlet Allocation (LDA) topic model. LDA represents documents as mixtures of topics that generate words with certain probabilities [2]. We have applied LDA on our protests corpus. Although the news articles were taken so that they correspond to this scenario, we wanted to evaluate the results of LDA over this corpus. For our experiment, we assumed that we track 3 topics.

In the word clouds from Figure 2, we have put the first 10 words for each topic in the descending order of their relevance. The words have been translated into English and their size is directly proportional to the LDA weight. As it can be observed from Figure 2, although 3 topic clusters were formed, all of them have words related to the street protests. This result indicates that LDA performs well even if the number of topics in not known in advance.

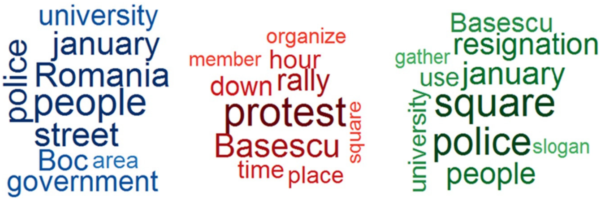


Figure 2: *Topics terms word clouds*

Semantic similarity

LDA offers a set of terms for each detected topic in the descending order of their relevance for the topic. For our event detection task, we want to establish which of the detected topics is the most related to our scenario. We address this problem by using semantic similarity measures between the first n LDA words for a topic, and a small vocabulary describing the scenario. This vocabulary can be entirely manually built or it can be automatically extended, although, as it can be seen from our experiments, a cardinality of 5 for the vocabulary is sufficient. For computing the semantic similarity between two terms, we have tested three WordNet semantic similarity algorithms, Wu, Resnik and Lin. Next, we give more details about these measures.

Wu and Palmer measure. The Wu & Palmer measure calculates semantic similarity by considering the depths of the two synsets in the WordNet taxonomies, along with the depth of the least common subsumer [10]. The formula is as follows:

$wuPalmerScore(t_1, t_2) = \frac{2 \times depth(lcs(s_1, s_2))}{depth(s_1) + depth(s_2)}$, where:

s_1: the synset of the first term, s_2:the synset of the second term, $lcs(s_1, s_2)$: the synset of the least common subsumer. This means that $0 < wuPalmerScore \leqslant 1$. The score can never be zero because the depth of the least common subsumer is never zero. The depth of the root of a taxonomy is one. The score is one if the two input synsets are the same.

Resnik measure. This measure also relies on the idea of a least common subsumer (LCS), the most specific concept that is a shared ancestor of the two concepts. The Resnik [9] measure simply uses the Information Content of the LCS as the similarity value:

$resScore(t_1, t_2) = IC(lcs(t_1, t_2))$, where $lcs(t_1, t_2)$: the least common subsumer.

$IC(t) = -log\left(\frac{freq(t)}{maxFreq}\right)$, where:

$freq(t)$: the freaquecy of term t in a corpus, $maxFreq$: the maximum frequency of a term from the same corpus. The Resnik measure is considered somewhat coarse, since many different pairs of concepts may share the same LCS. However, it is less likely to suffer from zero counts (and resulting undefined values) since in general the LCS of two concepts will not be a very specific concept.

Lin measure. The Lin measure augments the information content of the LCS with the sum of the information content of concepts A and B themselves [7]. The Lin measure scales the information content of the LCS by this sum.

$linScore(t_1, t_2) = \frac{2 \times resScore(t_1, t_2)}{IC(t_1) + IC(t_2)}$

Topic set similarity. For computing the semantic similarity between two sets of words using one of the three measures described above, we use the lemma of each term. We propose two different methods for computing the global similarity. In the first case, the final similarity score is obtained using a weighted average over the maximum score obtained by applying a semantic similarity measure on each combination of a term from the first set and one from the second set. In second one, we simply add all the similarity values between each combination of terms. This is suitable for situations where there are a consistent number of similar words with scores less, but close to the maximum and that would have been ignored by the first formula:

$$globalMaxSim(T_1, T_2) = \frac{\sum_{t_1 \in T_1} max(sim(t_1,t_2) \in T_2))}{|T_1|}$$

$globalAddSim(T_1, T_2) = \frac{\sum_{t_1 \in T_1} \sum_{t_2 \in T_2} sim(t_1,t_2)}{|T_1|}$, where T_1: first set, T_2: second set, $sim(t_1, t_2)$: one of the Wu and Palmer, Resnik or Lin similarity measures.

Event detection evaluation

For evaluation, we have used 10 of the 20 topics from the "The 20 Newsgroups", a widely used corpus for text classification [6]. We have included the following topics: "rec.sport.baseball", "rec.motorcycles", "talk.politics.guns", "alt.atheism", "comp.graphics", "sci.electronics", "sci.med", "sci.space", "soc.religion.christian", "talk.politics.mideast".

In order to evaluate the similarity measures, we have observed which measure captures the similarity between 2 sets of words describing the same topic, while lowering the similarity between sets describing different topics. For the experiments, we have chosen the "baseball", "guns" and "motorcycle" topics. We compare the first $n (5 \leqslant n \leqslant 50)$ relevant words for each topic as identified by the LDA topic model and a set containing the following words: "bike", "tire", "motorcycle", "helmet", "drive". In a first series of experiments in which we compared the Resnik, WuPalmer and Lin similarity measures, Resnik was the single one that found a higher similarity between the sample vocabulary and the "motorcycle" topic words for every instance of n and disregarding the global similarity measure that was used.

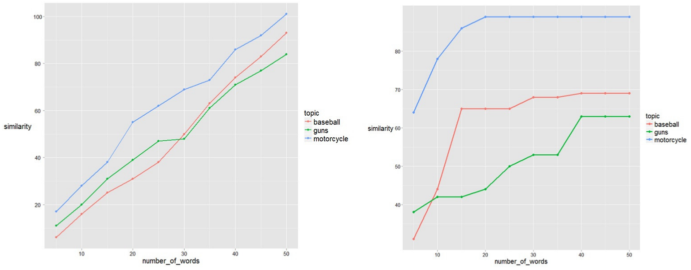


Figure 3: *(a) Progress of globalAddSim (b) Progress of globalMaxSim*

In the next series of tests, we wanted to establish which of the two global similarity measures provides the best results when using Resnik as a similarity between two words. In Figure 3 (a), we present the evolution of the $globalAddSim$ similarity between the sample vabulary and the LDA words for each of the 3 topics. In Figure 3 (b), we track the evolution of the $globalMaxSim$ similarity measure. As it can be seen from the two figures, the $globalMaxSim$ provides the best separation between the correct similarity (represented in blue) and the others. Based on the previous experiments, we have chosen the globalMaxSim measure with the Resnik similarity and we keep the first 20 LDA relevant words.

For the experiments, we have used 80% of the data to train the topic model and 20% to evaluate it. Due to the fact that the results of LDA depend on the random initialization of the initial topic distributions, in Figure 4 we have tracked the average accuracy over 3 runs when using a number of topics varying from 2 to 10. The sudden drop in accuracy when using 9 and 10 topics appears due to the inclusion of the "christian" topic, which shares a high number of terms with the "atheism" topic.

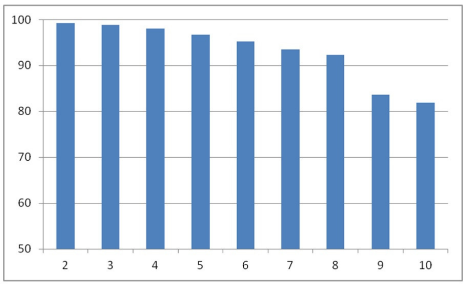


Figure 4: *LDA Accuracy*

2.4 Data Processing

Identifying locations, regions, keywords: Named entity identification is a crucial component of our application. The correct identification of "islands" with potential risks on a map depends on the accuracy of this component. For this, we use the Romanian language specific resources [4] that contain cities (Iasi, Bucuresti, Ploiesti, etc.), regions (Bucovina, Moldova, Transilvania, etc.). Additionally, we have added a new type of named entity, "street", for which we have created specific resources (containing the major streets of big cities "Iasi, Bulevardul Independentei", "Bucuresti, Calea Victoriei", etc.) and specific rules to identify streets (Street + *entity*, Boulevard + *entity*, etc.). To refine the localization to smaller inner city regions, we have added a new category, "area" that captures locations such as Pacurari district, center of Iasi, Arch of Triumph Square, etc. Using rules designed for this specific type of entities, our system is able to capture location related expressions, such as "the area between street A and street B" or "the area of the building A".

The quality of the module responsible with NE identification and with NE classification remains the same, after the adding of a new type of named entity "street". Thus our evaluation on 538 files with 2,806 entities of "street" type shows that the quality of NE identification component is around 92% and the quality of NE classification component is around 67%. Problems in NE identification: incorrect spelling (Pieta Universitatii), anaphora resolution (only Piata or only Universitate are not identified), other problems (Cotroceni, Primaria, Prefectura, sediul PDL, in fata simbolului Iasiului, Palatul Culturii, Piata Romana). The most frequent problems in NE classification are related to ambiguity situations when from the context we cannot conclude that the NE is a person name or a street name.

2.5 Identification of Opinions

While the majority of the opinion mining systems have in common the use of a sentiment lexicon, a distinction can be made between rule based and statistical approaches [8]. Due to the fact that we propose a general architecture that needs to be easily adapted to different crisis situations, we use the first type of approach. In this case, switching from a crisis scenario to another will require only the changing of the lexicon, whereas in the statistical approach, a significant training corpus would be required for each scenario. We use manually built resources to identify opinion keywords that signal the (good, bad, etc.), amplifiers (most, more, etc.), diminishers (less, etc.), Negation (not, never, etc.) [3]. Additionally, for our "street protests" test case we have added 21 specific words for conflict monitoring, such as "protest", "conflict", "fight".

The application described in [3] allows us to calculate the valences for groups of feelings and pairing named entities with scores based on the distance, punctuation and context. Based on these values we are able to classify named entities previously identified based on the opinion expressed towards them. Although obtaining a general opinion, as defined by the opposites positive/negative still can provide valuable clues concerning a potential threat, by adapting the context to a specific issue (protests, weather etc.) and introducing a relevant seed vocabulary, we can shift the semantics of the opinion towards the problem in hand. For example, we will be able to present the results in terms of degrees of danger.

2.6 Building a Customized GoogleMaps Map for Events

The purpose of this component is to create a map based on GoogleMaps, in which the locations and critical values calculated for them will be placed. Depending on these "islands", we will inform concerned people of the potential risks that appear and we find a solution which can be adopted. In order to build the GoogleMap we use JavaScript and accordingly with sentiment values associated to locations we create "red islands" (when the values are negative) with potential risks and "green islands" (when the values are positive) without potential risks.

3 Results

For the street protests scenario, our application has identified 698 news articles in this topic from the 13 to 26 January 2012 time span a total of 21,156 named entities, out of which 1,166 locations. In Figure 5 (a) we can see how cumulated sentiment values were greater with negative values, between 15 January and 19 January, and similar the numbers of mentions per days and per entities are higher in the same period (Figure 5 (b)). After analyzing the newspapers we see how between 15 January and 19 January were fights and confrontations between guardians and football teams supporters and a part of protesters.

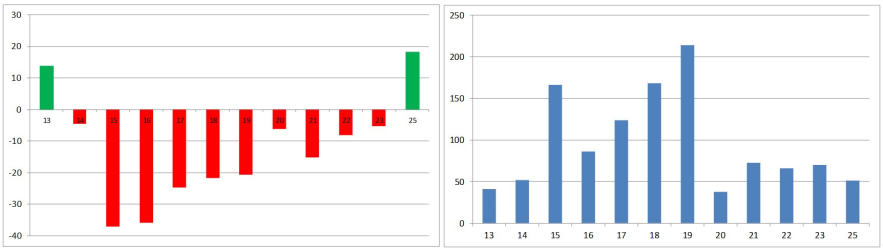


Figure 5: *(a) Cumulated sentiment values by days (b) Location type entities mentions by day*

After aggregation of 1,166 values on location entities we obtained 198 unique entities. From 198 location entities, 61 represent countries, 5 represent cities outside Romania, in 12 cases entities are marked as cities, but because we didn't perform anaphora resolution we didn't know at which cites they are referring to, and in 12 cases we classify wrong identified these entities in ambiguous cases. In the end we have 101 cities, with 51 with negative values associated, 23 with value 0 associated, and with 27 with positive values. The cities with lowest values are Bucuresti (-38.8), Cluj-Napoca (-19.41), Tulcea (-10.13), Sibiu (-9.09), Slobozia (-8.57). For these cities we can see in the next Figure the associated red circles. The cities with highest values are Mangalia (2.61), Medgidia (2.61), Bacau (2.48), Vaslui (2.05), Brasov (1.45) and we can see in next Figure 6 the associated green circles. Although a part of these entities are without diacritics, GoogleMaps API is able to identify correct the entity and to put it on the map.

Optimal path between cities

In the above scenario after we put red and green circles, we want to find an optimum path

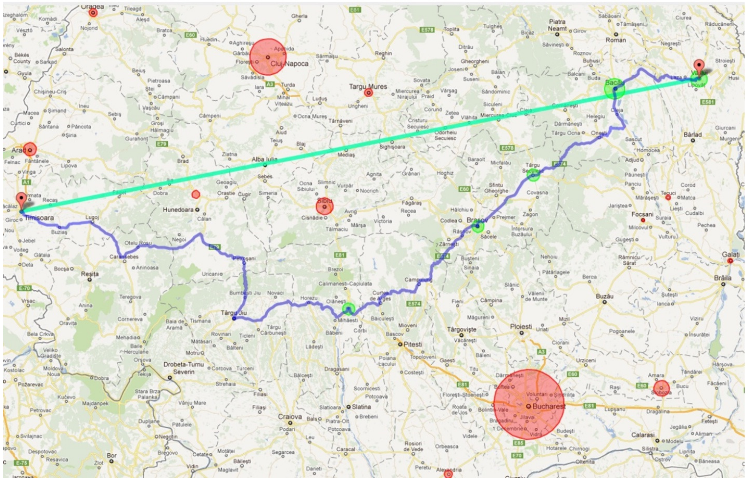


Figure 6: *The "optimum" path which passes near red islands and pass through green islands between cities*

from Vaslui to Timisoara. In this scenario, the shortest way uses a part of the cities which have associated red circles. Our algorithm is able to find another path (longer) which passes near the "red islands" and prefers the ways near the "green islands". The algorithm uses a graph structure corresponding to the streets network, in which the nodes correspond to cities and edges to streets. In this structure we associate sentiment values to nodes and length values to edges. When all nodes have associated values equals with zero, the solution identified by our algorithm corresponds to the shortest path. The graph structure was built by us starting from the main streets from Romania. On this structure we apply our algorithm which built step-by-step partial solutions starting from start node, until the partial solution reach the end node and the partial solution become the final solution. At every step is possible to insert penalties when the partial solution crosses red islands (with potential risks) and add bonuses when the partial solution crosses green islands (without potential risk).

In the above case we can see how our solution doesn't prefer red islands (Sibiu, Deva, Cluj-Napoca, Tirgu-Mures) which are on the shortest way and prefer green islands (Vaslui, Bacau, Tirgu Secuiesc, Brasov, Rimnicu Vilcea) from a longer way. A first type of problems with our algorithm is related to the following situation: even if some cities have negative values close to zero our algorithm prefers to pass near them. For these situations we must identify a threshold value, and under this value we will ignore the negative value. Other problems are with some cities which have false positive values and in a wrong manner attract routes, and other cities have false negative values and influence the final solution.

Optimal path between streets locations in the same city

When we want to find a solution for a path between two locations which are in the same city the things is different. At street level we identify 2,806 entities (172 streets, 150 boulevards, 2,299 squares, 185 areas). If in the cases of streets and boulevards the GoogleMaps API is able to put these entities on the map, for specific squares and areas it is not able to do this. In such cases we built an additional resource which specifies the GIS coordinates for them. In this way we can generate red islands at city level.

4 Conclusions

We have described in this paper a system that, starting from an opinion mining architecture, can be used to detect and localize different types of threats and which offers an expressive visualization for a rapid and targeted intervention.

We have proposed a global system architecture that can be easily adapted for the detection of multiple crisis situations with minimum human intervention. An important aspect of our approach is the detection of crisis events. Due to the fact that the opinions expressed towards geographical entities are strongly related with that context, by correctly identifying the context, even without using a dedicated vocabulary for a particular situation, and we can use any opinion mining configuration with the results being relevant for the detected context. We have proposed a novel use of topic models with semantic similarities to indentify and classify the main topics from a news collection. Our results, both for English and for Romanian, have shown that by using Latent Dirichlet Allocation we can obtain an accurate language independent topic distribution and by using a WordNet based semantic similarity, we can successfully correlate a discovered topic with any given topic. More than that, we identify various inner city location and we offer a clear visualization suggesting alternative routes to bypass potential dangerous areas.

Bibliography

[1] D. Blei, J. Lafferty, Topic models, *Text Mining:Theory and Applications*, Taylor and Francis, London, UK, 2009.

[2] D. Blei, A. Ng, M. Jordan, Latent Dirichlet allocation, *Journal of Machine Learning Research*, 3:993-1022, 2003.

[3] A. L. Ginsca, et al., Sentimatrix - Multilingual Sentiment Analysis Service, *In Proceedings of the 2nd Workshop ACL-WASSA*, 2011.

[4] A. Iftene, D. Trandabat, M. Toader, M. Corici, Named Entity Recognition for Romanian. *In Knowledge Engineering, Principles and Techniques. Selected Papers*, 49-60, 2011.

[5] F. Johansson, et al., Detecting Emergent Conflicts through Web Mining and Visualization, *In Proceedings of the European Intelligence and Security Informatics Conference*, 2011.

[6] K. Lang, Newsweeder: Learning to filter netnews, Proceedings of the Twelfth International Conference on Machine Learning, pages 331-339, 1995

[7] D. Lin, An information-theoretic definition of similarity, *In Proceedings of the International Conference on Machine Learning*, Madison, August, 1998.

[8] B. Pang, L. Lee, Opinion mining and sentiment analysis, *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135, 2008.

[9] P. Resnik, Using information content to evaluate semantic similarity, *In Proceedings of the 14th International Joint Conference*, 1995.

[10] Z. Wu, M. Palmer, Verb semantics and lexical selection, *In 32nd Annual Meeting of the Association for Computational Linguistics*, pages 133-138, Las Cruces, New Mexico, 1994.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):83-97, December, 2012.

Multi-criteria Receiver Self-Election Scheme for Optimal Packet Forwarding in Vehicular Ad hoc Networks

R.H. Khokhar, M.A. Ngadi, M.S. Latiff, K.Z. Ghafoor, S. Ali

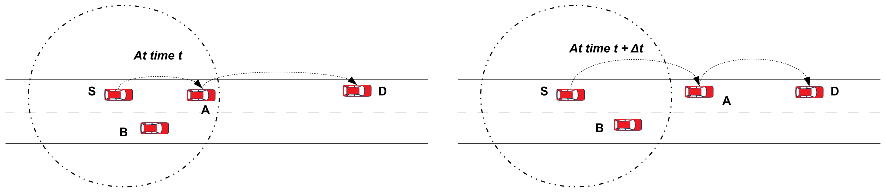
Rashid Hafeez Khokhar, Md Asri Ngadi
Mohammad Shafie Latiff, Kayhan Zrar Ghafoor, Saqib Ali
Faculty of Computer Science and Information Systems,
Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
hkrashid2@live.utm.my, dr.asri@utm.my, shafie@utm.my, zgkayhan2@live.utm.my, asaqib2@live.utm.my

Abstract: In most of the existing geographical forwarding methods of Vehicular Ad hoc NETwork (VANET), a node periodically sends "hello" messages to determine the positional information of its direct neighbors. Each node stores and maintains more or less accurate information of its direct neighbors in a table. However, due to high mobility vehicles and traffic congestion the stored neighbors information is quickly outdated, failure notification increases significantly, and leading sub-optimal path. Furthermore, the transmission of periodic "hello" messages and table maintenance consume resources, which is not suitable for sensitive VANET. In this paper, we propose a geographical forwarding mechanism based on Multi-criteria Receiver Self-Election (MRSE) scheme to find best next hop without sending the periodic "hello" messages and maintaining neighbors information in the table. The selection of best next hop is based on the multi-criteria waiting function. In this function, the four key parameters including link life time, optimal distance from sender to receiver, optimal transmission range, and received power are determined to enable the next candidate node to make packet forwarding decisions. The simulation results show that the MRSE scheme performs up to 22% better in terms of packet delivery ratio as compared to some existing schemes. In terms of average delay, MRSE scheme performs best, with as much as 81% decrease compared to some existing schemes.
Keywords: Vehicular Ad hoc Networks, VANET Routing, Geographical Forwarding, Multi-Criteria Waiting Function.

1 Introduction

Vehicular networks are emerging as a new promising field of wireless technology, which aims to deploy vehicle-to-vehicle and vehicle-to-infrastructure communications for different applications such as roadway safety, dynamic route planning, mobile sensing, and in-car entertainment. Vehicular Ad hoc NETworks (VANETs) provide true ubiquitous communication networks with great features as these networks are self configurable, infrastructureless, and rapidly deployable. These promising applications and features of VANETs require an efficient routing protocol for vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications. Because of the unique characteristics of VANETs i.e., highly dynamic topology, frequently disconnected network, and various communications environments, the traditional mobile ad hoc network (MANET) routing protocols such as AODV [1], DSR [2], OLSR [3] are not suitable in VANET. The main problem in VANET is route instability, leading to frequent route breaking due to high vehicle speeds and traffic congestion in urban environments. The geographical routing protocols such as GPSR [4], GFG [5], GOAFR [6], GPCR [7], GpsrJ+ [8], GeoCross [9], FAST [10]and offer a suitable solution to handle these problems. However, the geographical forwarding in these protocols do not perform well if they cannot find next hop due to high mobility.

Proposed recovery strategies [4,7] of geographical routing protocols in the literature are not as effective in high mobility where the network topology frequently changing. Figure 1 shows that route breaks between nodes S and A are due to either high mobility or staleness of neighborhood information. In this case, node B should be used instead of A to forward data packets to destination node D. Initially, the route $S \rightarrow A \rightarrow D$ established at time t, and the route breaks when node A moves out of the transmission range of node S after time $t + \lambda t$. Despite the better path stability of geographical forwarding methods, these methods still do not perform well in city environments ([11, 12]).



(a) Sending message at time t (b) Sending message at time $t + \lambda t$

Figure 1: Frequent route breaking in high mobility using traditional node centric routing protocols

Existing geographical forwarding methods [13–16] use one, two, or three criteria based forwarding scheme to select best next hop. These methods do not consider error-prone wireless channels, low connection time between vehicles, and optimal transmission range. For example, the hop-count based greedy geographic forwarding approaches [4, 17, 18] has received a great deal of attention in the vehicular ad hoc networking research community. These forwarding approaches have shortcomings due to sub-optimality of packet forwarding, a transmitter tends to select node with poor link quality. As a result, many data packets are dropped and the overhead increases significantly due to route failure and repair notification. For this reason, there has been a growing acceptance that the traditional purely greedy forwarding approaches are not optimal in most practical settings where the unit disk assumption or a perfect reception-within-range does not hold true. Some link-aware routing schemes have been recently reported [19–21]. However, the trade-offs between greediness and link quality has not been thoroughly studied. Furthermore, high mobility shortens the link duration between vehicles in the vicinity and might lead to performance degradation of the network. Therefore, in packet forwarding link life time should be considered to give higher priority to a candidate node which has higher link duration with the packet carrier node.

Moreover, the nature of traffic distribution in vehicular environments is heterogeneous (sparse and dense). In dense environments, routing protocols suffer from high over head due to proactive hello message broadcasting. The transmission of periodic "hello" messages under increased congested network consume resources, which can significantly affect the performance of VANET routing protocols on road segments especially during peek working hours in urban environment. Figure 2 illustrates this problem with the help of simple city scenario. The source node S broadcasts "hello" messages to direct neighbors to find routes for destination node D. Flooding may be required to get updated routing information. Each neighbor node transmits its own information including location and IP address to neighboring nodes. If there are only a few nodes on road segments, the messages can easily be forwarded to next hops within a short time. However, in case of traffic congestion, the average delay significantly increases because each node is transmitting information at the same time.

The greedy forwarding mode of geographical routing protocols such as GPSR [4], GDBF [16], GPSR+AGF [22], GRANT [23] handle traffic congestion in such a way that the neighbor node which is closest to destination node is selected to forwarded messages. For example, as shown in the hatched area in Figure 2, node N_1 forwards a message to N_2 as N_2 is the shortest distance from destination node D. In the first transmission, the total number of nodes transmitting "hello" messages with each other is nine and each node also maintains it's routing table before forwarding this message to the next available node. N_2 forwards the message to destination node via N_3 and N_4. Finally, the source node S establishes the route for destination node D through nodes $N_1 \rightarrow N_2 \rightarrow N_3 \rightarrow N_4$. The routing table of each node is updated every time they receive new messages in a congested network. As a result, the performance of packet delivery ratio and average delay are significantly affected.

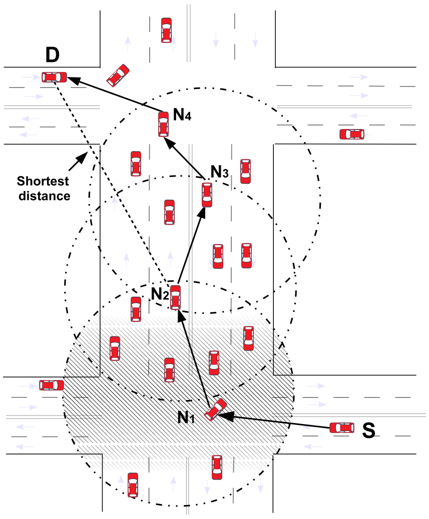


Figure 2: Traffic congestion problem in city scenario

In this paper, we have proposed a Multi-criteria Receiver Self-Election (MRSE) scheme to tackle the issues of high mobility and traffic congestion by suppressing the "hello" message and giving packet forwarding decision to the candidate receivers. In self-election process, a multi-criteria waiting function uses four key parameters such as link life time, optimal distance from sender to receiver, optimal transmission range, and received power to determine the best next hop from all neighbors nodes. We assign the different weight values dynamically to each parameter and the next hops will use the same values of these parameters. It works as, the greater weight value has more impact than the parameter has in the self-election scheme. In previous schemes [14,15,24], the static values are used for these factors. However, we have determined and adjusted the weight values according to the local traffic density information. This information is determined by calculating the number of nodes within the communication range of transmitter. The MRSE allows for the transfer of data packets quickly between intersections (streets) which significantly improve routing performance.

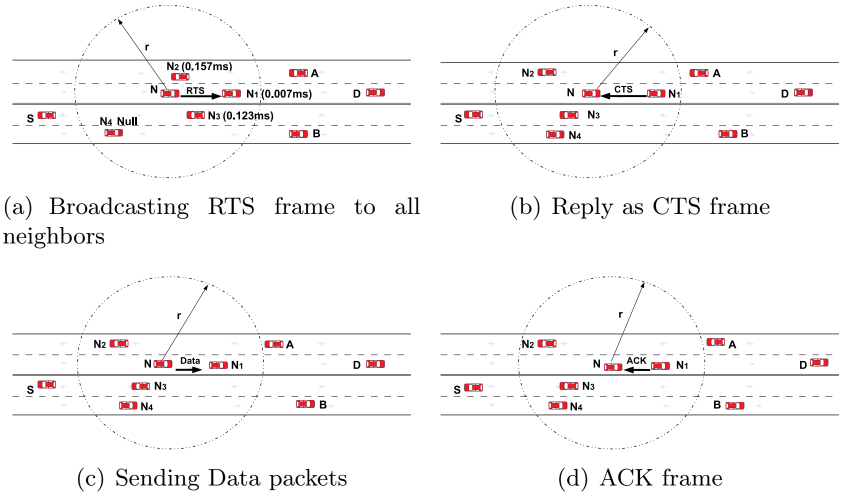
The rest of this paper is organized as follows, Section 2 presents the proposed multi-criteria receiver self-election scheme, its design, an example of an urban VANET scenario, and optimization. In Section 3, after describing an evaluation methodology, we present the performance analysis of the proposed scheme with two related receiver self-election and source selection schemes. The paper is concluded in Section 4.

2 Proposed Multi-criteria Receiver Self-Election Scheme

In this section, we present Multi-criteria Receiver Self-Election (MRSE) Scheme to determine the best next hop from all potential candidates. MRSE is a distributed process where the next relaying node is selected using four criteria including link life time, optimal distance, optimal transmission range and received power for non-uniform radio propagation in vehicular networks. We used IEEE 802.11 DCF (distributed coordination function) RTS/CTS (request-to-send/clear-to-send) frames [25] to select a best next hop with less overhead.

2.1 Receiver Self-Election using RTS/CTS

It has already been discussed in literature that the route breaks due to high mobility and the network becomes congested through the frequent sending of "hello" messages. Our proposed MRSE scheme is based on a receiver side relay election approaches [13–16, 24] that selects alternative nodes to handle route breaks caused by high mobility and implicitly eliminates the overhead by frequently sending "hello" messages. First, sender node broadcasts the RTS frame including the positions of the sender and destination nodes to all neighbors. Each receiving node calculates a waiting time, and this waiting time is sent back as a reply the CTS frame to the sending node. A waiting time assigned to each node basically determines how close to perfect this node is as best next node. The assignment of waiting time depends on the multi-criteria parameters that we will explain in Section 2.2. The node with the shorter waiting time will be considered as best node and will answer first by replying CTS to sender node. In the next step, the sender starts forwarding the data packets and receiver node acknowledge the data by sending ACK frame. Figures 3(a)-3(d) illustrate the whole procedure with the help of city scenario.
In Figure 3(a), node N receives a message from source node S for destination node D, looking to forward a message to the best next hop. This node broadcasts an RTS frame including its current position, the position of the destination node, and the transmission time of the RTS frame. The neighbor nodes calculate their waiting time to reply to node N by sending a CTS frame once they receive the RTS frame. The nodes that are farther from the destination node than the sender are not involved in this process, for example node N_4 in Figure 3(a). Node N_1 is closest to the destination and has a shorter waiting time (0.007ms), thus it replies with CTS first to node N. Nodes N_2 and N_3 will automatically cancel their timers when they overhear the CTS from N_1. Furthermore, the neighbor nodes of N_1, which are A and B, will not send any messages before the transmission is completed. The neighbor list updates accordingly, if any node moves out of the communication range of node N. Node N starts sending data packets after receiving the CTS from N_1. At the same time, the neighbor nodes (i.e., N_2, N_3) of N will not send any messages until N_1 finishes sending the ACK frame to N. In this example, we have



(a) Broadcasting RTS frame to all neighbors

(b) Reply as CTS frame

(c) Sending Data packets

(d) ACK frame

Figure 3: Multi-criteria receiver self-election example

illustrated how the best next hop can be effectively selected without sending “hello” messages in forwarding method.

The following is the proposed MRSE algorithm using multi-criteria parameters. In this algorithm, we have used three types of times such as (1) Δ_{RTS}, Δ_{CTS}, Δ_{DATA}, and Δ_{ACK} shows time to transmit RTS, CTS, DATA, and ACK frame; (2) τ_i shows waiting time of a node node n_i; and (3) Time to Live (TTL) of data packet. Initially, sender node n_c looking for the next hop broadcasts a very short RTS frame including its positions and position of destination node. Each neighbor node call waiting function and calculate its waiting time τ_i which is set using weight values according to local density information. If a node is still receiving data packet then postpone transmission for DATA and RTS frame. A node which has less waiting time will first send-back a CTS frame to node n_c and other node automatically discard their timers. After receiving CTS frame, the node n_c will send DATA to best next node. Finally, current node will send ACK frame to complete transmission. This step continues until the data packet will reach at destination node.

Notations:

$\Delta_{RTS}, \Delta_{CTS}, \Delta_{DATA}, \Delta_{ACK}$: time to transmit for$RTS, CTS, DATA, and$ ACK frame

τ_i = waiting time of node n_i

p_i = position of node n_i

p_d = position of the destination node

n_c = current node ID that finds next hop

T_FLAG = Transmision status, (True/False)

Proposed MRSE Algorithm:

```
Initialy $T\_FLAG$ set to True
if($T\_FLAG$ & Receiving $RTS(p_i, p_d, \Delta_{DATA})$ from node $n_c$ then
    Call waiting function and calculate $\tau_i$
    Set timer to $\tau_i$ using weight values according to local density information
    if(RTS frame transmission complete) then
        Set $T\_FLAG$ to True
    else
        Postpone transmissions for $\Delta_{DATA} + \Delta_{RTS}$
        Set $T\_FLAG$ to False
    end if
else
if($T\_FLAG$ & Receiving $CTS(n_j, n_c, \Delta_{DATA})$ from node $n_j$ before the
        timeout) then
    Cancel timer /*$n_j$ is the best next hop candidate */
    if(CTS frame transmission complete) then
        Set $T\_FLAG$ to True
    else
```

```
        Postpone transmissions for $\Delta_{DATA}$
        Set $T\_FLAG$ to False
    end if
else
if($T\_FLAG$ & Overhearing $DATA$ from node $n_c$) then
    if(DATA transmission complete) then
        Set $T\_FLAG$ to True
    else
        Postpone transmissions for $\Delta_{ACK}$
        Set $T\_FLAG$ to False
    end if
end if
Upon timout :
Broadcast $CTS(n_i, n_c, \Delta_{DATA})$ /* $n_i$ is the best next node */
```

2.2 Multi-Criteria Waiting Function

In this section, we illustrate how the waiting time is determined which is based on multi-criteria waiting function to select the best next hop. In multi-criteria waiting function, we try to achieve the following objectives:

1. The best next hop should reply first with shortest time.

2. To avoid time collisions of second best next hop, the waiting time difference should be reasonable.

3. The waiting time should not be too long that creates unnecessary delays.

Four key parameters, including link life time, optimal distance from sender to receiver, optimal transmission range, and received power are used to achieve these goals. Finally, different weights are assigned and dynamically adjusted based on the traffic and network conditions.

Link Life Time

The link life time estimates the link quality of the wireless channels between each vehicle. This duration represents the minimum time in which two direct neighbor vehicles can exchange information with guaranteed delivery. Since vehicles are travelling at high speed, this time interval can be very short. Therefore, we place a great deal of reliance on this time interval to give higher priority to those nodes whose connection time lasts longer. We consider the distance between two nodes as the distance between two vectors rather than the distance between two static points. The position vector of $\vec{p_i}$ of each node N_i is defined by the following equation:

$$\vec{p_i} = (x_i + \nu_x t, y_i + \nu_y t) \tag{1}$$

where t is the time since the initial position and speed vector of node N_i were $\vec{p_i} = (x_i, y_i)$ and $\vec{\nu_i} = (\nu_{ix}, \nu_{iy})$ respectively. The function of the distance square $\delta^2(t)$ is defined as:

$$\delta^2(t) = (\vec{p_{1x}} - \vec{p_{2x}})^2 + (\vec{p_{1y}} - \vec{p_{2y}})^2 \tag{2}$$

$$\delta^2(t) = [(x_1 - x_2) + t(\nu_{1x} - \nu_{2x})]^2 + [(y_1 - y_2) + t(\nu_{1y} - \nu_{2y})]^2 \tag{3}$$

Suppose $\lambda_x = (x_1 - x_2)$, $\lambda_y = (y_1 - y_2)$, $\lambda_{sx} = (\nu_{1x} - \nu_{2x})$, $\lambda_{sy} = (\nu_{1y} - \nu_{2y})$, by putting these values in Equation 3, we have the simplified form of distance function $\delta^2(t)$ as follows:

$$\delta^2(t) = (\lambda_x^2 + \lambda_y^2) + t^2(\lambda_{sx}^2 + \lambda_{sy}^2) + 2t(\lambda_x\lambda_{sx} + \lambda_y\lambda_{sy}) \tag{4}$$

The Equation 4 is second degree polynomial that only assumes non-negative values. Therefore, the smallest value of $\delta^2(t)$ occurs when its derivative $\delta^{2\prime}(t)$ equals to zero and the value of t is determined as:

$$\bar{t} = \frac{-(\lambda_x\lambda_{sx} + \lambda_y\lambda_{sy})}{(\lambda_{sx}^2 + \lambda_{sy}^2)} \tag{5}$$

The $\bar{t}$ gives the two types of connection times, if the value of $\bar{t}$ is positive then the nodes are getting closer to each other; if negative, the nodes are moving away from each other. We calculate the link life time until the node goes out of communication range.

Optimal Distance from Sender to Receiver

This parameter determines the optimal distance between a sender node S and intermediate node N_i for destination node D. A single criteria receiver election schemes [13, 14, 17] usually used this parameter. The optimal distance is defined as follows:

$$d_i = d_{SD} - d_{N_iD} \tag{6}$$

where d_i, d_{SD}, and d_{N_iD} are the distances between nodes S-N_i, S-D, and N_i-D respectively. These distances actually denote the progression towards the destination node, if node N_i is the next hop and closest to the destination.

Optimal Transmission Range

The optimal transmission range f_i of a node N_i describes the probability that the data packet is successfully received by a node. Wireless channels are error-prone and do not provide any guarantee that signals out-side of particular range will successfully transmitted. There are many factors that may obstruct the radio signals. For example, a node away from the nominal range may receive a RTS frame but may not receive data packets successfully. This problem happen in real wireless radios channels because it does not follow unit disk assumption [26].

To find the optimal transmission range, we use a translation function proposed by [24]. In this function, the distance from the sender node is used as an input and the optimal transmission range is the output. For example

$$f_{trans}(x) = \begin{cases} x + R_{td} & if\ \ x \leq\ \ R_{ot} \\ -x + R_{\max} & if\ \ x >\ \ R_{ot} \end{cases} \tag{7}$$

where R_{ot} shows the optimal transmission range of a sender node, R_{max} the estimated maximum transmission range with acceptable error rate, and R_{td} is the translation distance. For more dynamic results, these parameters can be set according to the network conditions in the area.

Received Power

The received power p_i of sender node provides signals with real channel quality. Many researchers are trying to calculate the optimal transmission range with received power. However, different obstacles such as big buildings, trees, advertisement boards, traffic lights, etc may block the radio signals in the real-life deployment of V2V and V2I communications. The received power at a particular node can differentiate the nodes at comparable distances. The reasoning is that if the vehicle is moving, the quality of the reported data is not affected by the received signal power; however the real reason is that the distance travelled by a vehicle is negligible when it receives the RTS frame.

Finally, the multi-variable function is used and customize it into a four variable polynomial of the selected parameters. The waiting time of any node t_i returned by this function within the time interval $[0, T_{max}]$ (where T_{max} is the maximum waiting time) is determined as:

$$f(\bar{t}_i, d_i, d_{SN_i}, p_i) = A\bar{t}_i^{w_1} d_i^{w_2} f_i^{w_3} p_i^{w_4} + T_{\max} \tag{8}$$

where $A = \frac{-T_{\max}}{\bar{t}_{\max}^{w_1} d_{\max}^{w_2} f_{\max}^{w_3} p_{\max}^{w_4}}$ and $w_i (i = 1, 2, 3, 4)$ are the weights of each parameter. The weight values of these parameters are adjusted using the mapping function proposed by [14]. In mapping function, the main objective is to compute a single ranking scale through the use of an aggregating function that weighs all criteria into a single unit. We assign the different weight values dynamically to each parameter and the next hops will use the same values of these parameters. It works as, the greater weight value has more impact than the parameter has in the self-election scheme. In previous schemes [14, 24], the static values are used for these factors. However, we have determined and adjusted the weight values according to the local traffic density information. Each node independently computes the weight values according to the number of vehicles within its radio coverage. This local traffic density information is estimated based upon each node's Contention Window (CW). According to the Ke, et al. [27] the value of CW is higher in heavy traffic density, which means that frequent retransmission occurs between contending nodes due to an increase of the probability of collision. On the contrary, lower CW implies that the light traffic density. Accordingly, when vehicles within the communication range of the source receives RTS frame, they simply check their CW in order to dynamically adjust the weight values. It is note worthy that this method of dynamic weight adjustment does not generate any network and computational overhead.

2.3 Impact of Multi-criteria Waiting Function

The main purpose of using multi-criteria waiting function is to determine a single value based on the above mentioned key parameters. An optimal decision is taken based on this final single value. The impact of multi-criteria waiting function is compared with forwarding progress only, as shown in Figures 4 and 5. The waiting times are determined when vehicles receive an RTS frame request between transmitter at position (0, 0) and destination at position (600, 200). As Figure 4 shows the forward progress without multi-criteria also determined the waiting time out of the transmission range that may cause the loss of many data packets. The waiting time value is around 0.9 seconds and the transmission range is more than 400 meters in this case.

The Figure 5 shows the nodes waiting time using proposed MRSE when an RTS frame is received from different positions near the transmitter. We adjusted weight values w_1, w_2, w_3, and w_4 according to the network condition in multi-criteria waiting functions, in this example the weight values are $w_1 = 0.45$, $w_2 = 0.3$, $w_3 = 1.4$, and $w_4 = 0.02$. We considered the optimal transmission rage to be 300m. However, we did not assume perfect reception due to interference generated by the urban environment. We used a Shadowing propagation model proposed by [28]

Figure 4: Waiting time using forwarding progress only

to calculate the received power. The Figure 5 illustrates the shortest waiting time and optimal transmission range using the proposed MRSE scheme with four key parameters. In case of four criteria waiting function, the waiting time is around 0.5 seconds within transmission of range 285 meters. The results show that the proposed MRSE prefers nodes with optimal transmission range and shortest waiting time as compared to forward progress method only.

Figure 5: Waiting time using four criteria (link life time, optimal distance from sender to receiver, optimal transmission range, and received power) waiting function

3 Performance Evaluation

This section presents a detailed analysis of the proposed MRSE scheme applied to our recently proposed Reactive Traffic-Aware Routing Strategy (ReTARS) [29] for real-time urban vehicular environments. The MRSE scheme is used to reduce traffic overhead in the streets, helping to transfer data packets quickly. ReTARS leverages prior global knowledge of real-time vehicular traffic to create paths between each vehicle. In ReTARS, the critical decisions are taken at the road intersections where the decision making node evaluates the best possible routes towards the destination based on the prior global knowledge of real-time vehicular traffic. ReTARS is fairly well-understood and can be used in this domain because it accommodates the frequent network disconnections and traffic congestion that are observed in many vehicular networks.
All experiments are conducted on a map of Chicago city with simulation dimension 3968m×1251m. The area contains 370 road segments with a total length of 3630394.0 meters extracted from TIGER/LINE line databases [30]. The main reason for the big area is to test the performance of MRSE in high mobility and increased traffic congestion. The integrated, configurable, and scalable Swans++ simulator [31] with IEEE.11b DCF standard is used to evaluate the performance of the proposed MRSE scheme. To generate and evaluate a large number of different scenarios, the numbers of vehicles and data packet rates are varied for with obstacles in urban environ-

Table 1: Parameter values used in simulation for proposed MRSE scheme

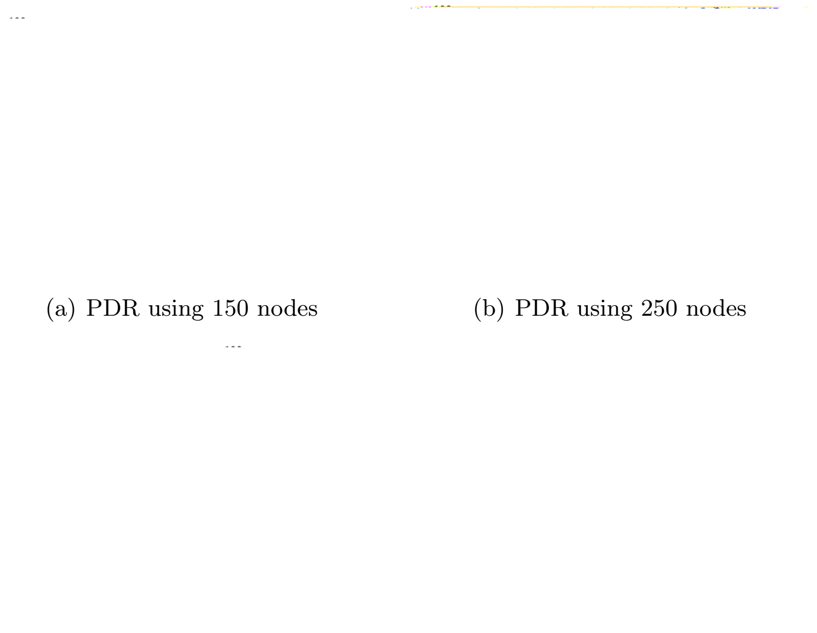
Parameter	**Value**
Simulation dimension	3968m × 1251m
Simulation area	3630394.0m
Number of vehicles	150, 250, 350
Warning packet size	512 bytes
Normal packet size	1024 bytes
Packet sending frequency	1 per second
Transmission range	400m
Simulation time	400s
Vehicle speed	20-80 m/h
Mobility model	STRAW
Routing protocol	ReTARS
MAC protocol	IEEE 802.11 DCF

ment. The total simulation time for a single flow was 400 seconds which is a reasonable time for this area of the map and the number of nodes. However, to obtain a more accurate result, the first 100 seconds of simulation are discarded. The STreet RAndom Waypoint (STRAW) mobility model [32] is used for node mobility. The nodes were placed on the map using the random placement model and experiment was repeated for 15 flows. In each experiment 10 source and destination nodes pairs with different CBR and UDP packets are selected randomly. In the proposed MRSE scheme, the packet carrier node needs to have both self and destination locations for packet forwarding purposes. To gain access of such location information, we utilized the implemented scalable and distributed location service in Swans++ packet level network simulator. Similarly, each node can compute its self direction and speed vectors by using the implemented street mobility module in Swans++ network simulator. In MRSE, each node in the radio coverage of the packet carrier node calls the waiting function when it receives a modified RTS frame. When a node receives a modified RTS frame, waiting function is called to compute waiting time for the selection of best next hop. This setting helps to get more accurate neighbor information. Table 1 describes the simulation parameters used in all experiments.

3.1 Simulation Results in Urban Environment (With Obstacles Scenario)

The performance of the proposed MRSE scheme is evaluated using PDR and average delay by varying the number of vehicles for with obstacles urban scenario. Figures 6(a)-6(c) show the simulation results that determine the PDR of proposed MRSE method by comparing these results with the two most related receiver self-election schemes [14,24] and source-selection scheme that periodically send "hello" packets to all neighbors. A slight modification of IEEE 802.11 (with DCF standard) RTC/CST frame is used to select the best next hop using receiver self-election scheme. As shown in Figures 6(a)-6(c), the PDRs of proposed MRSE scheme steadily increased from 65% to 81% when node density increased from 150 to 350 nodes.

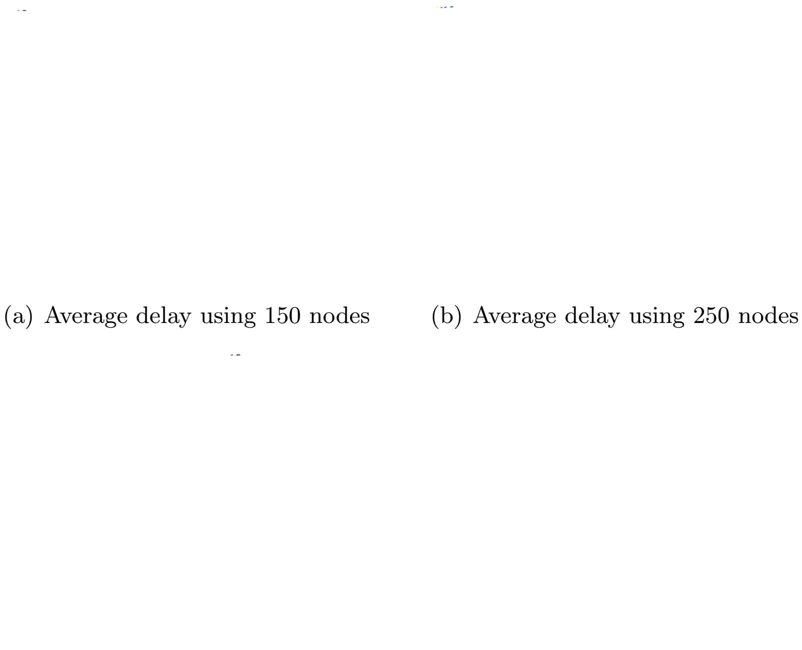
The PDR of proposed MRSE scheme is consistently higher than Nzouonta et al., Egoh & De, and source-selection schemes in all cases. This is because the weight values were carefully assigned to each parameter to compute waiting time. The PDR of Nzouonta et al., self-election scheme is about 8% to 11% lower than MRSE in all cases. The main reason for the low results of Nzouonta et al., scheme is the assignment of static weights values to determine a waiting time using multi-criteria waiting function. The PDR of Egoh & De self-election scheme is apparently lower than MRSE, which is about 15% when node density is 150, (Figure 6(a)) and 20% in case of



(a) PDR using 150 nodes (b) PDR using 250 nodes

(c) PDR using 350 nodes

Figure 6: Packet delivery ratio for proposed MRSE, Nzouonta et al., [24], Egoh & De [14], and Source-Selection schemes for with obstacles scenario



(a) Average delay using 150 nodes (b) Average delay using 250 nodes

(c) Average delay using 350 nodes

Figure 7: Average delay for proposed MRSE, Nzouonta et al., [24], Egoh & De [14], and Source-Selection schemes for with obstacles scenario

high density, as depicted in Figures 6(b) and 6(c). This is because two criteria based forwarding node selections, hop progress (greediness) and reachability (link quality) were used to determine waiting time to select best next hop. Similarly, the PDR of source selection method is lower than other protocols, as even the PDR starts falling below 50% when the CBR is just around 2 packets rate/sec. This is due to the frequent broadcast of "hello" messages that each node needs to create and to maintain a list of neighbors. The overall PDR for all schemes are not so high, as was expected. The implementation of more accurate RPM with obstacles that reduces the contention level is the main reason for this low percentage of PDR.

The average delay of proposed MRSE, Nzouonta et al., [24], Egoh & De [14], and source-selection schemes are shown in Figures 7(a)-7(c). The average delay of the proposed MRSE is consistently lower than other protocols in all cases. The average delay of Nzouonta et al., receiver self-election scheme is slightly higher than our proposed MRSE, as even node densities increased from 150 to 350 nodes. For comparison, Egoh & De receiver-self election scheme has around 7-12 seconds average delay when node densities increase from 150 to 350 nodes, which is about 5 seconds higher than MRSE. Source selection schemes have three times higher delay as compared to MRSE. The main reason for the better performance of MRSE is the careful adjustment of weight values to each selection parameter that selects the best next hop candidate within the shortest time to avoid unnecessary delays. As a result, better link utilization is received for data transfer. Also, it leads to improved delays as it needs less retransmission and backoffs.

4 Conclusions

In this paper, a multi-criteria receiver self-election (MRSE) scheme is proposed to suppress the sending of periodic "hello" messages that significantly degrade the network performance of end-to-end data transfer rates. The waiting function is defined using multi-criteria parameters to select best next hop. Four key parameters are used including link life time, optimal distance from sender to receiver, optimal transmission range, and received power. The simulation results using packet delivery ratio and average delay show that the proposed scheme offer better performance as compared to two receiver self-election using RTS/CTS-based schemes and source selection using "hello" packet scheme. The proposed MRSE scheme forwards data between intersections (streets) and it performs better in real time urban environments where large obstacles such as buildings may block radio signals. These results show that distributed applications that generate moderate or high traffic can be successfully implemented in VANETs.

Bibliography

[1] C.E. Perkins and E.M. Royer, Ad-hoc on-demand distance vector routing, *Proc.of the Second IEEE Workshop on Mobile Computer Systems and Applications*, pp.90-100, 1999

[2] David B. Johnson and David A. Maltz, *Dynamic Source Routing in Ad Hoc Wireless Networks, Kluwer Academic Publishers*, 1996.

[3] P. Jacquet, P. Muhlethaler, T. Clausen, A. Laouiti, A. Qayyum, and L. Viennot, Optimized link state routing protocol for ad hoc networks, *Technical report, HIPERCOM Projet, INRIA Rocquencourt*, 62-68, 2001.

[4] Brad Karp and H. T. Kung, Gpsr: greedy perimeter stateless routing for wireless networks. In *MobiCom '00: Proc. of the 6th annual int. conf. on Mobile computing and networking*, New York, NY, USA, 243-254, 2000.

[5] Prosenjit Bose, Pat Morin, Ivan Stojmenovic, and Jorge Urrutia, Routing with guaranteed delivery in ad hoc wireless networks, *Wireless Networks*, 7(6):609-616, 2001.

[6] Fabian Kuhn, Roger Wattenhofer, Yan Zhang, and Aaron Zollinger, Geometric ad-hoc routing: of theory and practice, In *PODC '03: Proc. of the twenty-second annual symposium on Principles of distributed computing*,New York, NY, USA, 63-72, 2003.

[7] Christian Lochert, Martin Mauve, Holger Fussler, and Hannes Hartenstein, Geographic routing in city scenarios, *SIGMOBILE Mob. Comput. Commun. Rev.*, 9:69-72, January 2005.

[8] K.C. Lee, J. Haerri, Uichin Lee, and M. Gerla, Enhanced perimeter routing for geographic forwarding protocols in urban vehicular scenarios. In *Globecom Workshops*, 2007 IEEE, 1-10, 2007.

[9] Kevin C. Lee, Pei-Chun Cheng, and Mario Gerla, Geocross: A geographic routing protocol in the presence of loops in urban scenarios, *Ad Hoc Networks*, 8(5):474-488, 2010.

[10] R. Khokhar, R. Md Noor, K. Ghafoor, C-H Ke, and N. Md Asri. Fuzzy-assisted social-based routing for urban vehicular environments, *EURASIP Journal on Wireless Communications and Networking*, 2011(1):178, 2011.

[11] C. Lochert, H. Hartenstein, J. Tian, H. Fussler, D. Hermann, and M. Mauve, A routing strategy for vehicular ad hoc networks in city environments, *Intelligent Vehicles Symposium, 2003. Proceedings. IEEE*, 156-161, 2003.

[12] Tonghong Li, S.K. Hazra, and W. Seah, A position-based routing protocol for metropolitan bus networks, *Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st*, 4:2315-2319, 2005.

[13] Holger Fubler, Jrg Widmer, Michael Ksemann, Martin Mauve, and Hannes Hartenstein, Contention-based forwarding for mobile ad hoc networks, *Ad Hoc Networks*, 1(4):351-369, 2003.

[14] K. Egoh and S. De. A multi-criteria receiver-side relay election approach in wireless ad hoc networks. In *Military Communications Conference*, MILCOM 2006, IEEE, 1-7, 2006.

[15] Komlan Egoh and Swades De, Priority-based receiver-side relay election in wireless ad hoc sensor networks. In *IWCMC '06: Proc. of the 2006 int. conf.on Wireless communications and mobile computing*, New York, NY, USA, 1177-1182, 2006

[16] Mohit Chawla, Nishith Goel, Kalai Kalaichelvan, Amiya Nayak, and Ivan Stojmenovic, Beaconless position based routing with guaranteed delivery for wireless ad-hoc and sensor networks. In *Ad-Hoc Networking, IFIP International Federation for Information Processing*, Springer Boston,212:61-70. 2006.

[17] M. Zorzi and R.R. Rao, Geographic random forwarding (geraf) for ad hoc and sensor networks: multihop performance, *Mobile Computing, IEEE Transactions on*, 2(4):337-348, oct. 2003.

[18] S. De, On hop count and euclidean distance in greedy forwarding in wireless ad hoc networks. *Communications Letters, IEEE*, 9(11):1000-1002, 2005.

[19] Karim Seada, Marco Zuniga, Ahmed Helmy, and Bhaskar Krishnamachari, Energy-efficient forwarding strategies for geographic routing in lossy wireless sensor networks. In *Proc. of the 2nd int. conf. on Embedded networked sensor systems, SenSys '04*, 108-121, 2004.

[20] Seungjoon Lee, Bobby Bhattacharjee, and Suman Banerjee, Efficient geographic routing in multihop wireless networks. In *Proc. of the 6th ACM int. symposium on Mobile ad hoc networking and computing, MobiHoc '05*, New York, NY, USA, 230-241, 2005.

[21] M.R. Souryal and N. Moayeri, Channel-adaptive relaying in mobile ad hoc networks with fading, In *Sensor and Ad Hoc Communications and Networks, 2005. IEEE SECON 2005. 2005 Second Annual IEEE Communications Society Conference on*, 142-152, 2005.

[22] Valery Naumov, Rainer Baumann, and Thomas Gross, An evaluation of inter-vehicle ad hoc networks based on realistic vehicular traces, In *MobiHoc '06: Proceedings of the 7th ACM international symposium on Mobile ad hoc networking and computing*, New York, NY, USA, 108-119, 2006.

[23] S. Schnaufer and W. Effelsberg, Position-based unicast routing for city scenarios. In *World of Wireless, Mobile and Multimedia Networks, 2008 International Symposium on a*, 1-8, 2008.

[24] J. Nzouonta, N. Rajgure, Guiling Wang, and C. Borcea, Vanet routing on city roads using real-time vehicular traffic information. *Vehicular Technology, IEEE Transactions on*, 58(7):3609-3626, 2009.

[25] The Institute of Electrical and Electronic Engineers (IEEE), Wireless lan medium access control (mac) and physical layer scpecifications.

[26] Gang Zhou, Tian He, Sudha Krishnamurthy, and John A. Stankovic, Impact of radio irregularity on wireless sensor networks, In *MobiSys '04: Proceedings of the 2nd international conference on Mobile systems, applications, and services*, 125-138, 2004.

[27] Chih-Heng Ke, Chih-Cheng Wei, Kawuu W. Lin, and Jen-Wen Ding, A smart exponentialthreshold- linear backoff mechanism for ieee 802.11 wlans, *Int. J. of Communication Systems*, 24(8):1033-1048, 2011.

[28] T. S. Rappaport, *Wireless Communications Principles and Practice*, Prentice Hall, 2nd edition edition, 2002.

[29] R. Khokhar, A. Ngadi, M. S. Latiff, and M. A. Amin, Reactive traffic-aware routing strategy for urban vehicular environments, *International Journal of Ad Hoc and Ubiquitous Computing*, 10(3):149-163, 2011.

[30] Tiger. tiger/line and tiger-related products. u.s. census bureau, 2011.

[31] Swans++. Swans++ - extensions to the scalable wireless ad-hoc network simulator, 2011.

[32] David R. Choffnes and Fabian E. Bustamante. An integrated mobility and traffic model for vehicular wireless networks. In *VANET '05: Proceedings of the 2nd ACM international workshop on Vehicular ad hoc networks*, New York, NY, USA, 69-78, 2005.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):98-111, December, 2012.

Client Side Internet Technologies in Critical Infrastructure Systems

I. Lendak, N. Ivancevic, S. Vukmirovic, E. Varga, K. Nenadic, A. Erdeljan

Imre Lendak, Srdjan Vukmirovic,
Ervin Varga, Aleksandar Erdeljan, Kosa Nenadic
Faculty of technical sciences
Serbia, 21000 Novi Sad, Trg Dositeja Obradovica, 6
lendak@uns.ac.rs, srdjanvu@uns.ac.rs, evarga@uns.ac.rs,
ftn_erdeljan@uns.ac.rs,

Nikola Ivancevic
EXPRO I.T. Consulting Llc
Serbia, 23300 Kikinda, Svetosavska 43, I/2
ivancevic.nikola@gmail.com

Abstract:
This paper assesses the applicability of client side Internet technologies in software solutions for critical infrastructure systems (CIS). It contains an in-depth analysis of four significant and well known development platforms, namely JavaScript with jQuery, the Google Web Tookit, Microsoft's Silverlight and Adobe's Flash/Flex. They were compared by using the ISO software quality characteristics as comparison criteria. Each of the technologies was applied in a real-life project and the results summarize the authors' experience. The ultimate goal of this research is to enable software engineers to more easily choose a client-side Internet technology when developing a new software solution for the CIS domain.
Keywords: critical infrastructure systems, Rich Internet Applications, comparative analysis.

1 Introduction

Electric power systems, water distribution and telecommunication networks are large-scale critical infrastructure systems (CIS), which can be categorized as safety critical [1, 2]. Their improper use can lead to a discontinuity of critical services (e.g. power outages), or in extreme cases, even to the loss of human lives (e.g. improper safety measures can harm the maintenance crew members). They provide essential services and usually are operated from computerized control centers utilizing various communication systems for data acquisition and control, different software solutions for asset, crew and outage management, and simulation software allowing engineers to plan system extensions, find week spots and optimize the operation of the CIS.
It is a complex task to build software solutions for the aforementioned control center operating a safety critical system, and usually those solutions are highly distributed by their nature. When developing a Web based Human Machine Interface (HMI) for such systems, the system architect can choose from various Internet technologies suitable for production use. Viable technologies exist on both sides of the spectrum: client and server. Relying on a proven technology does speed up and improve development, and it is also a superb tactic for delivering quality into the final product. It comes as no surprise that more and more critical software systems are built by leveraging Internet technologies. Finally, recent tendencies show that even those software intensive systems, which traditionally offered data exchanges solely via some proprietary mechanisms, like a Supervisory Control and Data Acquisition (SCADA) system, are opening up, thus allowing access to their various services through the Internet.
The main motivation behind this work is the authors' aspiration to share with the readers their

extensive experience in incorporating Internet based technologies into software intensive systems built for the electric power and financial domains. That endeavor is still under way, since the transition from a pre-Internet stage to a fully Internet based solution is usually carried out gradually. This transition process is positively influenced by advances in Internet technologies, as new possibilities for solving old problems emerge quite frequently. Thankfully also to the abilities of modern Internet browsers the performance of the contemporary Web based applications is comparable to their desktop counterparts [3, 4].

Modern Internet applications are quite different from the conventional notion of client-server systems, i.e. the client is a lot more than a simple node merely capable of displaying some fully processed static content from the server. Nowadays, the delineation between clients and servers is pretty blurred. Clients are now totally equal with other nodes inside some multi-tiered distributed Web application. This is the principal reason why the authors have chosen to analyze client side Internet technologies in the context of CIS. This paper elaborates some of the recognized client side Internet technologies as development platforms without the intent of showing a definite and complete list of each and every such technology. The analysis is limited to the following technologies:

- Google Web Toolkit (GWT) - an open-source source development kit (SDK) [5] for developing complex cross-platform JavaScript-based front-end applications. The GWT-based applications are written in Java and then compiled into JavaScript by the GWT compiler.

- Javascript/jQuery (JJ) - a scripting language supported by all the latest web browsers. Best suitable for those who require complete control over each line of the client side code. Development is made a lot easier with modern tools, like jQuery [6], a cross-browser scripting library.

- Flex - a user interface (UI) source development kit (SDK) [7] for rich cross-platform Internet applications based on the Adobe Flash platform [8].

- Silverlight (SL) - Microsoft's web technology for Rich Internet Application (RIA) development [9], [10] based on the Microsoft.NET Framework. It offers user interface functionality which is nearing the desktop level.

While in referemce [11] the focus was on a general comparison of client side Internet technologies, this paper performs a similar analysis with CIS in mind, i.e. reflecting on the aspects essential for developing large, reliable and sensitive software intensive systems. The set of characteristics taken into account in each of these technologies were selected from the standard quality model defined in [12]. The following characteristics were selected for this analysis:

- Functionality: interoperability, security

- Reliability: maturity

- Usability: learnability, attractiveness

- Efficiency: time behavior, resource utilization

- Maintainability: stability

- Portability: installability

These characteristics were identified as applicable in the direct comparison of client side Internet technologies when used in large scale critical infrastructure systems. The examination of the

remaining characteristics (e.g. operability, suitability, accuracy, fault tolerance, recoverability, changeability, testability, understandability, adaptability, analyzability) defined by the ISO quality standard was outside the scope of this paper, as they were deemed not entirely applicable for comparing software technologies, i.e. they are more applicable for comparing the software solutions which are developed by using those technologies.
Apart from this introduction and the vital conclusion at the end, this paper is comprised from a detailed discussion of real-life solutions developed with the identified client side Internet technologies, and a comparative analysis of these technologies from the standpoint of industrial systems.

2 Real Life Case Studies

The following three sections contain an in-depth report about each of the chosen Internet technologies when applied in real life project implementations. The first case study describes a modern control center solution for electric power systems [13] developed in both JavaScript and GWT. It is followed by the descriptions of a Meter Data Management (MDM) [14] solution developed in Silverlight, and finally a web based financial system developed in (Adobe) Flex. Each section contains a short description of the project/problem which had to be solved with the application of one of the above listed Internet technologies. These descriptions are followed by short discussions of the applied technologies, focusing on the ISO quality model characteristics listed in the previous section, and aim to analyze the technology itself (e.g. GWT, Flex), instead of the applications developed.

3 DMS Web User Interface

Control center solutions for the operation of truly Smart Grids [16] require varying level of web access. Some customers are content with a web based reporting system available to the management, while others require (nearly) complete SCADA functionality to be available on the Internet and accessible from everyday Internet browsers. While working on various control center projects, the authors learned that in the electric power system domain, web based user interfaces are often required to perform some or all the following operations:

- Visualize geographic diagrams of the grid (or its parts) [17]
- Visualize single-line diagrams of the grid (or its parts) [18]
- Visualize detailed substation single-line diagrams [19]
- Allow easy access to equipment details, e.g. the relevant characteristics of a breaker: nominal current, nominal voltage, maximum breaking current, phases connected, etc.
- Monitor measurement values (obviously not in real-time, but with a certain delay)
- Show simulation function results, e.g. the results of state estimation, load flow [20], [21]
- Allow insight into operations performed by other users, e.g. oversee switching sequences executed by dispatchers

Geographic and single-line views give a high level overview of the complete power system. They provide a means for navigation, zooming and viewing different parts of the network. Detailed substation diagrams allow the users to access information about substation equipment which

*[tbhp]

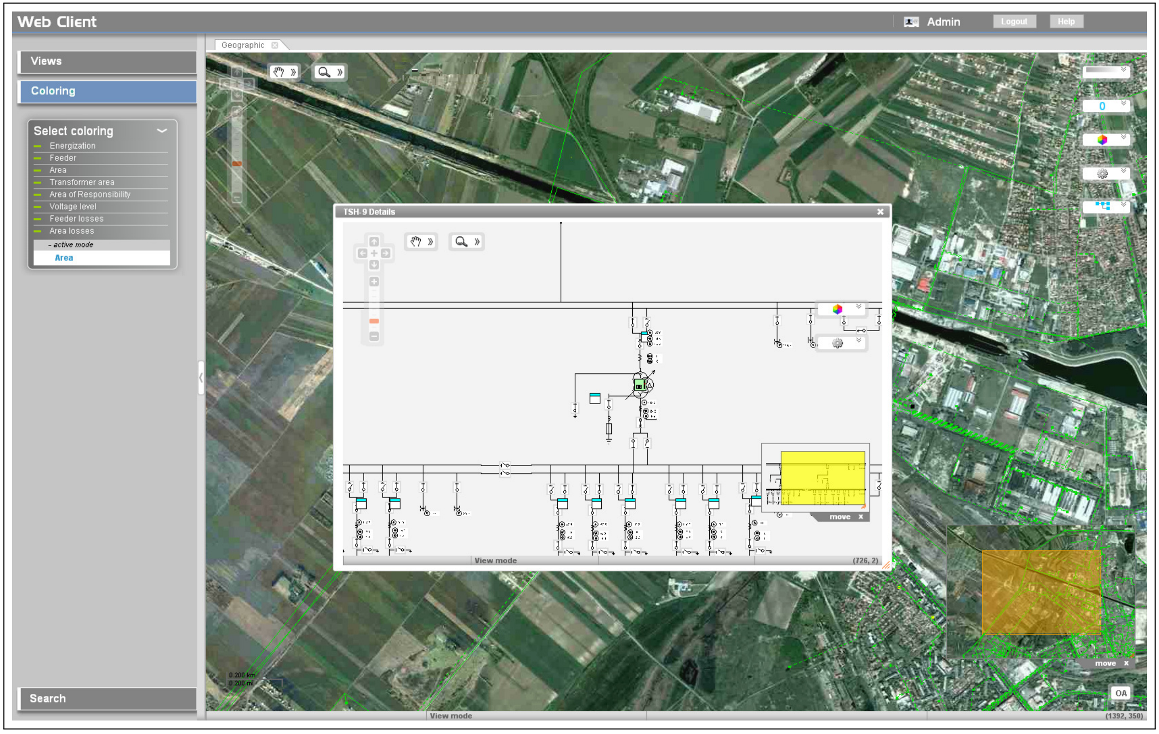


Figure 1: Web view of substation one-line diagram with geographic view in background

is essential in case of outages. Equipment details should be accessible from lists and from the graphical diagrams also. Analytic function results are necessary both online (finding weak spots or overloads not directly measured) and offline (network optimization, planning). Measurement value monitoring allows the user to see these values refresh in (near) real-time.

Fig. 1 shows a snapshot taken from the web user interface developed as part of the Telvent DMS software solution [22]. Apart from presenting the up-to-date state of the electric power system in both geographic (visible in the background in Fig. 1) and one-line diagram views, it also allows insight into the internal structure of substations (as shown in the foreground in Fig. 1) and details of each relevant piece of equipment. Fast navigation in large diagrams is available through the pilot view in the lower right corner. Users can easily navigate to particular network devices using the search tool which allows name and identifier based searches. Besides that, there is a set of handy tools placed on each diagram which improve user experience (e.g. zoom in/out, quick access buttons).

Two versions of this web based user interface were developed: the first technology of choice was the GWT for creating a functional prototype. The second technology of choice was JavaScript/jQuery (JJ). Although quite similar, these two solutions will be discussed separately for clarity. Both are truly thin client applications, fetching diagrams in a graphical format and additional information (e.g. equipment details, names, etc.) in the JavaScript Object Notation (JSON) format [23] from the server side.

3.1 Google Web Toolkit

The Google Web Toolkit (GWT) is an open-source source development kit (SDK) for developing complex cross-platform JavaScript-based front-end applications running in regular web browsers. The GWT-based applications are written in Java and then compiled into JavaScript

by the GWT compiler. GWT was used to build one of the early prototypes of the web based DMS solution, as a kind of a proof of the concept version of the system.
Interoperability was excellent and the GWT proved to be a truly cross platform development platform, allowing the client applications to work on a wide range on destination systems and browsers (Windows, Linux, Apple iOS, Symbian). *Security* was average, e.g. client side code security available through code obfuscation. *Maturity* was moderate in early 2010 and it has considerably improved since then.
Learnability is a matter of knowing or not knowing Java as a programming language and platform, i.e. GWT is straightforward for Java developers and not so accessible for the rest. GWT enables usage of any mature Java tooling system for source code editing, refactoring, testing and debugging. As far as Integrated Development Environment (IDE) support goes, GWT can utilize the well-known Java IDEs. Comparing to a pure JavaScript approach, GWT makes the development process highly efficient. *Attractiveness* is guaranteed by the numerous built-in widgets and libraries which make it possible to implement solutions which resemble the look and feel of desktop applications.
Time behavior on the client side was entirely dependent on JavaScript engine performance, the generated code itself did not have any issues. The GWT compiler optimizes the output code and resource utilization to an extent not necessarily possible when writing JavaScript code manually. The GWT profiler (Speed Tracer) allows developers to fine-tune their GWT applications for the best performance. The execution speed of the JavaScript engine in the latest versions of Internet browsers (Internet Explorer 8, Safari 5, Chrome 19, Firefox 13) allows programmers to develop very complex desktop-like user interfaces with significant computational demands.
Stability was excellent during the complete development cycle of the prototype. *Resource utilization* was considerable at times, slowing down the PC based clients and making it necessary to implement a standalone application for mobile phones instead of running the application in their built-in Internet browsers.
Installability of the solutions built with GWT was excellent and no plug-ins were necessary In order to satisfy one mandatory system requirement from the specification that a full control is necessary over the source code of the application the GWT was (temporarily) abandoned. Although GWT does offer the JavaScript Native Interface (JSNI), which allows developers to write parts of the application in (native) JavaScript, this was not sufficient to satisfy the stringent source code control related system requirement. Nevertheless, by judging the pace with which GWT is improved from version to version, the authors of this web based interface are considering to implement the next version of the system fully in GWT, which might reduce complexity and maintenance costs (compared to JavaScript - see its description below). The stability and quality of the latest version of the GWT obviates the need for such rigorous source code control aspects of the system.

3.2 JavaScript/JQuery

Due to the high level of complexity of the web based Distribution Management System (DMS) [24] graphical user interface (GUI), and because the developers required absolute control over the JavaScript code. Therefore the decision was made to migrate the solution to pure JavaScript/JQuery. This transition was completed in 2010 and the JavaScript version was successfully tested on mobile devices (e.g. mobile phones) just as well as on personal computers.
Interoperability was just as excellent as with the GWT and the JJ solution proved to be a truly cross platform development platform, allowing the client applications to work on a wide range on destination systems and browsers (Windows, Linux, Apple iOS, Symbian). Some limitations to cross-browser support exist and developers in certain cases have to write browser specific

JavaScript code. *Security* was above average, e.g. client side code security is available through code obfuscation.

Maturity of the JavaScript engine and tools was excellent, which came as no surprise as they were around for more than a decade.

Learnability is a major disadvantage of JJ, both because of language limitations and the lack of truly integrated and advanced development environments. *Attractiveness* of JJ as a technology is very limited - people might get put back because of the steep learning curve. Other than that, JJ gives full control and can utilize the rendering capabilities of web browsers to their full extent.

Time behavior was similar to the one achieved with GWT and entirely dependent on JavaScript engine performance - the full control given to JJ developers might mean a slight advantage over the GWT, as the performance of these applications is entirely under the developer's control. With well optimized Document Object Model (DOM) [25] structures one can implement responsive client applications. It is often better to use the JavaScript Object Notation's (JSON) [23] simple text format for data exchange with the server side as it incurs less parsing overhead and bandwidth use than the eXtensible Markup Language (XML). *Resource utilization* was similar to GWT, i.e. considerable at times, slowing down the PC based clients and making it necessary to implement a standalone application for mobile phones instead of running the application in their built-in Internet browsers. *Stability* was unquestionable during the complete endeavor.

Installability of the JJ based solutions was excellent as no plug-ins were necessary.

At the moment of writing, the latest, JavaScript version of the discussed web based GUI was in its production phase with successful installations worldwide as part of the Telvent DMS software platform [22].

4 Meter Data Management User Interface

Smart metering systems are information systems used in electric power systems which gather measurement values from distant measurement devices (usually 'smart' meters) and integrate them into a complex and unified system for acquisition and control in power distribution systems. These systems are essential components of most Smart Grid [16] initiatives. With the introduction of modern meters it is now possible to control consumption per individual consumers. Meter Data Management (MDM) systems within smart metering infrastructures gather meter data and perform data processing (validation, estimation, editing - VEE).

Large power utilities can have millions of smart meters which periodically report their current consumption rate and receive commands. For the application of algorithms which allow fine grained control of distribution management systems, it is often necessary that these meters send consumption rates as often as possible. This raises significant problems as the millions of meters sending a few measured and status values consume bandwidth, processing and storage resources. Data loss and even compression is not allowed as meter data is used for various mission critical purposes, e.g. billing, load-shedding, etc. The data gathered by smart meters has to be easily accessible to other services within a more complex smart grid (e.g. simulation functions taking into account latest consumption data) as well as to financial people. Therefore, very efficient data processing and storage solutions, a modern user interface and application interfaces built by following the latest industry standards are all necessary components of these systems.

Telvent DMS' Meter Data Management (MDM) system [14] was developed to address the following requirements:

- Search meters and access detailed meter data
- Group meters into groups and areas

- Run validation and estimation functions (where values are not available - linear or spline interpolation is used)
- Display meters on a geographical background
- Data exchange with other software components, e.g. simulation functions, outage management, customer information system, work order management
- Send commands to meters, e.g. on demand read, disconnect, reconnect, ping

The actual implementation of this MDM solution relies on a number of state-of-the-art technologies and tools:

- Meter data is stored in a PI database [26]
- Data access is provided through Microsoft Windows Communication Foundation (WCF) written in Microsoft's C# programming language
- The user interface components were developed in Microsoft Silverlight in combination with Microsoft RIA Services [27]
- Meter visualization on a geographical background is implemented with a specialized component developed by Miner & Miner for their ArcFM (an extension of ESRI's ArcGIS for power utilities) [28]

Fig. 2 shows a SL form from the MDM solution's user interface which displays meter readings with tree view selection and data displayed in both a grid view and a chart. The screenshot was taken during a test session run in the Internet Explorer web browser. The user interface is a thin client running in any web browser with a Silverlight plug-in. The client side consists of approximately fifty forms utilizing advanced components, e.g. grids, charts and the above mentioned geographical background. Drag and drop, direct data editing in grids and deep zoom (i.e. load only visible elements at a certain zoom level) are supported. The thin client was optimized for web based use. The WCF interface is used to exchange custom meter objects which are in line with industry standards [29]. Data security is ensured by data encryption (HTTPs) and user groups with varying access levels.

An MDM form can contain data of up to twenty meters thereby reducing load times. When more than twenty meters should be fetched, paging is used on the client side. The geographical view allows the visualization of up to 3000 meters only on deeper zoom levels.

Interoperability of SL applications is limited to the platforms for which Microsoft issues the latest Silverlight plug-ins. *Security* is inherited from the .NET development platform and the developer has excellent tools for directly influencing security behavior.

As far as *maturity* goes, at the moment of writing Silverlight was becoming a stable and well recognized RIA development platform. The authors' experience was positive, with only a few glitches and memory leaks.

Learnability is excellent due to the use of .NET programming languages which can be easily mastered, the excellence of development tools (i.e. Microsoft Visual Studio) and the high availability of both official learning material and free online resources. *Attractiveness* is excellent and the built applications have a truly desktop-like look and feel. Numerous user interface controls are immediately accessible after installation. .NET version 4.0 comes with an extended set of UI controls which cover most of the needs for business style web applications: grid, chart, tree view control, etc. There is a wide range of open source projects offering additional components, e.g. scheduler, extended charts and grids, etc.

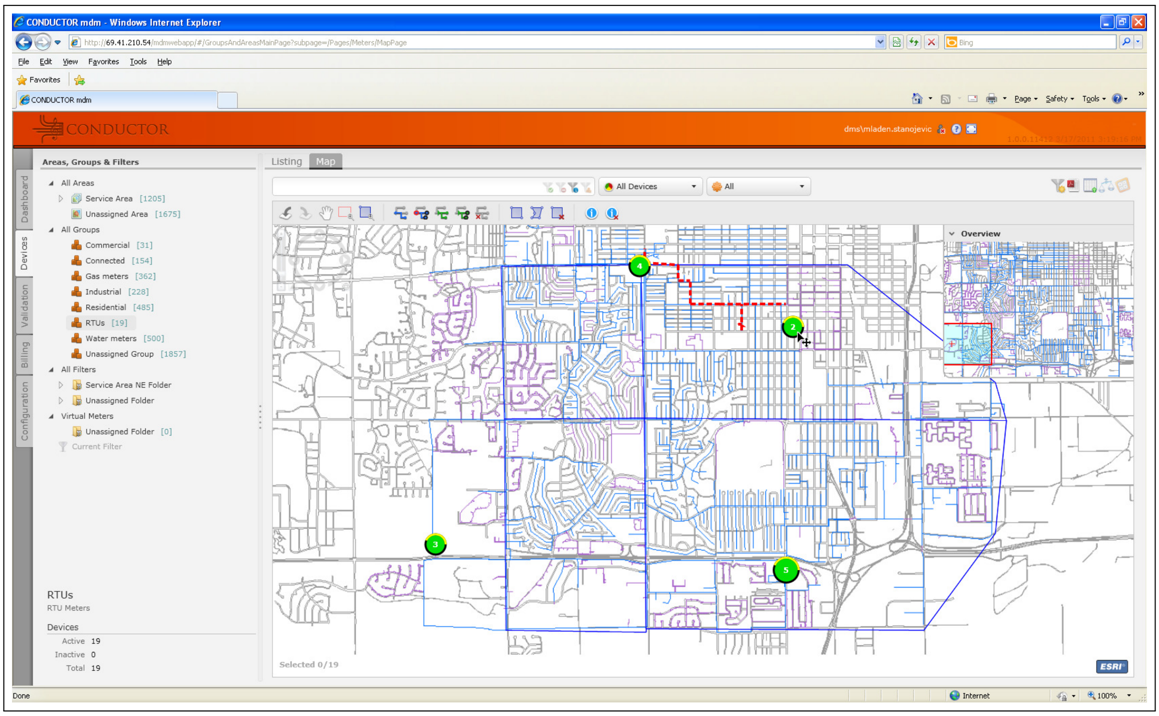


Figure 2: Meter reading data screen developed in Microsoft Silverlight

Time behavior on the client side was above average and quite acceptable even for large numbers of meters with a few optimizations (e.g. paging). *Resource utilization* proved to be moderate which came as no surprise knowing that the code itself is executed by a runtime environment. Network bandwidth utilization is configurable and entirely depends on the developer: the choice of available protocol and serialization formats is wide and can be tailored to specific needs.
Stability was very good on the client side during the complete development lifecycle. There were a few memory leaks.
Installability is somewhat impaired as a plug-in is necessary to run Silverlight based applications. Plug-ins also might not be available for all available hardware and operating system platforms. At the moment of writing the above discussed MDM project was in a release candidate status.

5 Flex in Live Betting

Live, online betting supports the modification of betting instruments (betting markets and prices) during the course of a sporting event. Live betting systems provide the following functionalities:

- management of betting entities (bookmakers, sports, competitions, teams, betting events)
- management of betting instruments (betting markets, prices)
- risk management
- betting events and markets publishing
- accepting players' bets
- betting history management

- processing financial transactions

One such live betting solution is Juliet [30] which the authors of this paper developed in Flash/Flex. It consists of the following major components:

- Core - the component responsible for maintaining the Juliet data model and processing all requests made by the other Juliet components
- BO Console (the back-office) - the back-end application for maintaining the live betting entities (betting events, bookmakers, sports, competitions, etc.)
- Live Console - the back-end application used by bookmakers for creating and maintaining betting events' status, markets and prices
- database (DB) - the Juliet system database
- channel server (CS) - the intermediate component between the UI instances and the rest of the Juliet components
- the user interface (UI) enables players to watch betting events' status, markets and prices and to place bets

The requirements for the Juliet UI were stringent:

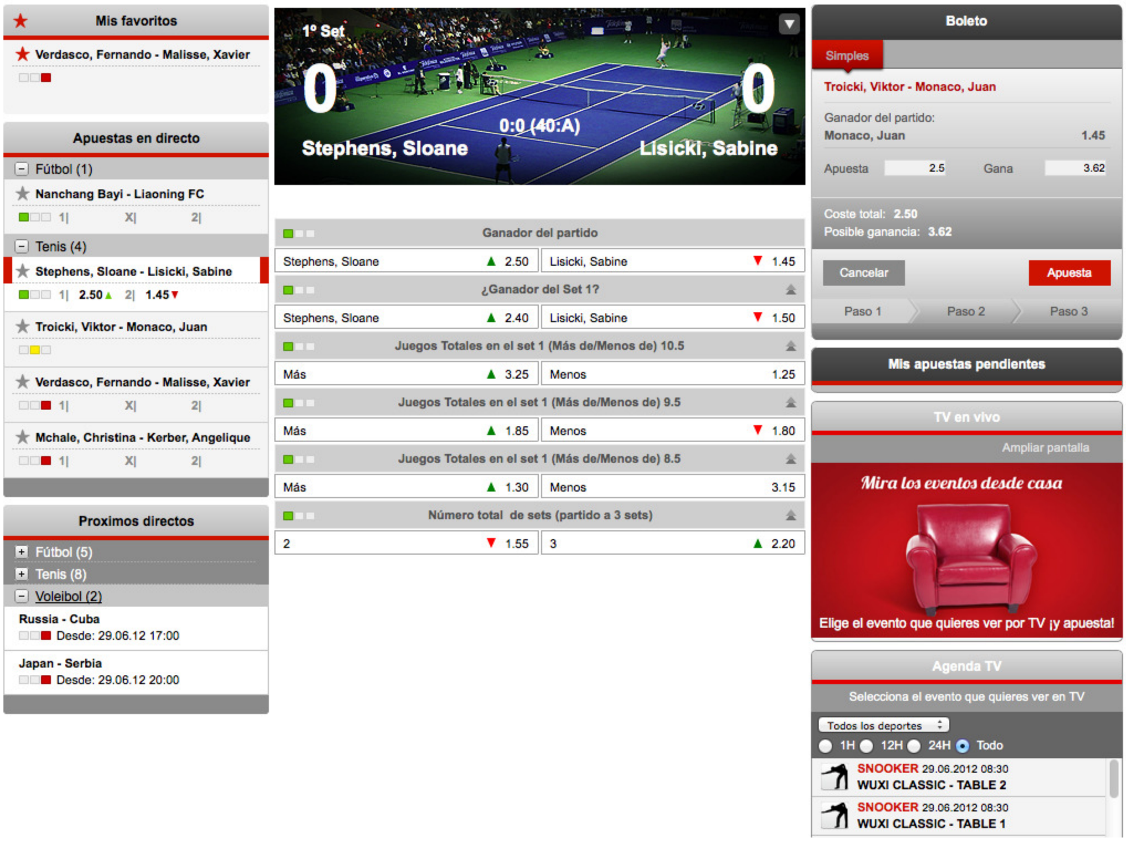
- cross-platform application that supports all major browsers
- display information about betting events and markets in a real-time fashion
- easy integration in an existing web site
- possibility to change the look and feel of the UI to match the design of the hosting web site

Fig. 3 shows a screenshot of the Juliet user interface. The left-most block displays the list of betting events (sporting events) available for betting, grouped by the match's sport. The list is automatically updated and contains the betting events which have already started or are about to start in a short time. The status of every betting event in the list (the start time or the current match time, the score, etc.) is shown and updated in real-time. These messages are pushed to the UI from the channel server.

The UI components are constantly connected to the CS and they are receiving notifications as soon as some change occurs. The user interaction is executed completely within the client side (within the hosting HTML page) and the server-side (the CS) is involved only when it is absolutely necessary (e.g. sending a place-bet request). The Juliet UI (see Fig. 3) is implemented as a set of components which can be placed in an HTML page as one or more Flash (SWF) files. This allows an easy integration of the UI components into virtually any website layout the hosting HTML page can have. Also, in order to fit the design of the hosting HTML page, the visual appearance (the look and feel) of the UI components is implemented using a new Flex skinning technique.

The communication layer (the event bus) for exchanging notification events between the UI components and to/from the CS is implemented in JavaScript using the jQuery library.

The Flex experience can be boiled down to the following: the highest level of *interoperability* is ensured by the existence of the Flash player plugin implementations for all major browsers (Internet Explorer, FireFox, Chrome, Safari, Opera) on different destination systems (Windows, Linux, iOS, Android). *Security* of the client side code is excellent as it is stored in compiled



★ Verdasco, Fernando - Malisse, Xavier

□ Fútbol (1)

★ Nanchang Bayi - Liaoning FC
1| X| 2|

□ Tenis (4)

★ Stephens, Sloane - Lisicki, Sabine
1| 2.50 ▲ 2| 1.45 ▼

★ Troicki, Viktor - Monaco, Juan

★ Verdasco, Fernando - Malisse, Xavier
1| X| 2|

★ Mchale, Christina - Kerber, Angelique
1| X| 2|

+ Fútbol (5)
+ Tenis (8)
□ Voleibol (2)

Russia - Cuba
Desde: 29.06.12 17:00

Japan - Serbia
Desde: 29.06.12 20:00

0:0 (40:A)
Stephens, Sloane — Lisicki, Sabine

Ganador del partido	
Stephens, Sloane ▲ 2.50	Lisicki, Sabine ▼ 1.45
¿Ganador del Set 1?	
Stephens, Sloane ▲ 2.40	Lisicki, Sabine ▼ 1.50
Juegos Totales en el set 1 (Más de/Menos de) 10.5	
Más ▲ 3.25	Menos 1.25
Juegos Totales en el set 1 (Más de/Menos de) 9.5	
Más ▲ 1.85	Menos ▼ 1.80
Juegos Totales en el set 1 (Más de/Menos de) 8.5	
Más ▲ 1.30	Menos 3.15
Número total de sets (partido a 3 sets)	
2 ▼ 1.55	3 ▲ 2.20

Troicki, Viktor - Monaco, Juan
Ganador del partido:
Monaco, Juan 1.45
Apuesta 2.5 Gana 3.62
Coste total: 2.50
Posible ganancia: 3.62

SNOOKER 29.06.2012 08:30
WUXI CLASSIC - TABLE 2

SNOOKER 29.06.2012 08:30
WUXI CLASSIC - TABLE 1

Figure 3: Juliet live betting user interface

meta-code (the byte code for the Flash player platform).

Maturity is excellent as the Flex SDK has been present and constantly improved since 2004.

Learnability is directly affected by mastering new languages that the Flex SDK supports - ActionScript and MXML. In spite of this, the learnability is good because of Adobe's visual tools and excellent knowledge base and documentation. *Attractiveness* is above average as creating visually attractive rich user interfaces quickly, a wide set of server-side integration components and the abundance of third-party widgets makes the Adobe Flex attractive choice.

Time behavior on the client side, once the framework is loaded and initialized, is excellent. The rendering performance of the user inteface is constant and smooth even in cases of great load. The Flex compiler compiles the Flex-based application into SWF byte code while performing possible optimizations. The Flash player itself is able to exploit hardware graphic acceleration to increase the execution speed of graphically intensive application tasks. *Resource utilization* is considerable at the initialization time and load, when the plugin has to download and initialize the Flex framework before the application is ready to start.

Stability is in direct relation to the stability of the Flash player plugin and it was very high even in case of complex demands.

Installability was a bit impaired due to the necessity to install a plug-in. The situation is further aggravated by the fact that some platforms do not have a Flash plug-in, i.e. support for Flash based applications.

The decision to develop this web based, highly sensitive solution in Flex has proven to be correct, as Juliet is online, quite functional with no major issues or downtime during its lifetime.

6 Comparative Analysis

The overview of the client side Internet technologies when utilized in critical systems is shown in Table 1. The table neither contains the definite list of tools for client side application development, nor is the list of applied criteria complete. Each of the analyzed characteristic is assigned a mark between one and five, one being poor and five being excellent. The marks are based on the authors' experience working with the latest versions of the discussed technologies during the complete development cycle of the critical systems discussed in the previous sections of this paper. These marks are a rough indicator of the usability of these tools for developing client side user interfaces in critical systems.

It is visible even after a brief look at the marks in Table 1 that both Silverlight and JavaScript are a 'roller coaster' experience: although they have some excellent characteristics which stand out, they also have serious limitations hampering their immediate use in certain situations, e.g. cross-platform support with Silverlight and learnability and changeability with JavaScript. While Flash has a smoother curve connecting the marks, the smoothest curve and the highest score in general goes to the Google Web Toolkit, which does not excel in all of the analyzed characteristics, but does not disappoint in any of them either.

The marks shown in Table 1 were given based on the experience gained during the development of multiple web based solutions for critical systems. JavaScript/JQuery gives complete control to the developer but needs more time to master then the rest of the tools. Full cross platform support, i.e. operating system and hardware platform independence can be achieved with GWT and JavaScript/JQuery. On the other hand, these two technologies usually incur a bit higher network load, as part of the provided functionality might have to be delivered through bitmap transfer.

Plug-ins are necessary for Silverlight and Flex. These tools themselves and their development environments also require some form of licensing. Application startup is slower as it takes some time while the complete application is loaded into the browser from the server. Silverlight might require a bit higher resource utilization due to the additional overhead introduced by the .NET Common Language Runtime (CLR) running the code.

Each of the above client side technologies has its pros and cons. For very specific and highly specialized user interfaces it might be necessary to choose JavaScript/JQuery or in certain cases GWT, which is only slightly less flexible. On the other end of the range are Flash and Silverlight which are quite modern and allow fast development cycles. Their most notable downsides are the lack of complete cross-platform support and developer freedom.

Table 1: Client side overview (* = poor, ***** = excellent)

Criterion/Technology	Flash	GWT	Silverlight	JavaScript
Interoperability	**	****	*	****
Security	***	***	*****	***
Maturity	***	**	**	****
Learnability	**	**	****	*
Attractiveness	****	***	*****	*
Time behavior	**	****	**	****
Resource utilization	*	***	*	***
Stability	***	***	**	***
Installability	**	*****	*	*****

Applications requiring large amounts of graphical data (see section 3 for an example) might be equally complicated to develop whichever of the shown technology is chosen. Experience gained in the development of such graphically intensive solutions shows that partial, on-demand loading of graphical data (e.g. loading and refreshing only the visible parts of the visualized system) might need custom development with all four presented technologies.

7 Conclusion

The results of this work help mitigate risks in software projects building Internet based critical systems by providing a pragmatic guidance in choosing the right client side Internet technology. The analysis is based on the authors' own practical experience in the domain of critical (infrastructure) systems. A set of ISO standardized quality attributes were used in the assessment of these technologies. They were described through real life case studies from the electric power systems and financial systems domain.

The conclusion is that although modern Internet based client side technologies still have their weaknesses, they are gaining a foothold in critical systems. Applications relying on these technologies can be made fast, secure, reliable and maintainable. The technologies analyzed in this paper will surely become even more widely adopted, as a vehicle in supplying software solutions for critical systems in the near future.

As a future work, the authors plan to extend the list of technologies to be researched, and to create a similar comparison of server side Internet technologies used in the CIS domain.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):98-111, December, 2012.

Bibliography

[1] Karsai, G.; Massacci, F.; Osterweil, L. J.; Schieferdecker, I. (2010); Evolving embedded systems, *IEEE Software*, 43: 34?40.

[2] Parks, R.C.; Rogers, E. (2008); Vulnerability Assessment for Critical Infrastructure Control Systems, *IEEE Secuirty & Privacy*, 6: 37-43.

[3] Fraternali, P.; Rossi, G.; Sánchez-Figueroa, F. (2010), Rich Internet Applications, *IEEE Internet Computing*, 14: 9-12.

[4] Melia, S.; Gómez, J.; Pérez, S.; Díaz, O. (2010); Architectural and Technological Variability in Rich Internet Applications, *IEEE Internet Computing*, 14: 24-32.

[5] Google Web Toolkit; http://code.google.com/webtoolkit; accessed 2011-03-29.

[6] jQuery library; http://jquery.com; 2011-03-29.

[7] Adobe Flex; http://www.adobe.com/products/flex; accessed 2011-03-29.

[8] Adobe Flash; `http://en.wikipedia.org/wiki/Adobe_Flash`; accessed 2011-03-29.

[9] Microsoft Silverlight; http://www.silverlight.net; accessed 2011-03-29.

[10] Pendleton, C. (2010); The World According to Bing, *IEEE Computer Graphics and Applications*, 30: 15-17.

[11] Lammarsch, T. et al (2008); A Comparison of Programming Platforms for Interactive Visualization in Web Browser Based Applications, *12th International Conference Information Visualization*, 194-199.

[12] International Standardization Organization (ISO); ISO/IEC 9126-1:2001 Software engineering - Product quality - Part 1: Quality model.

[13] Bose, A (2010); Smart Transmission Grid Applications and Their Supporting Infrastructure, *IEEE Transactions on Smart Grid*, 1: 11-19.

[14] Vukmirovic, S.; Erdeljan, A; Lendak, I.; Capko, D (2010); A novel software architecture for smart metering systems, *Journal of Scientific & Industrial Research*, 69: 937-941.

[15] Vukmirovic, S.; Erdeljan, A.; Kulic, F.; Lukovic, S. (2010); Software architecture for Smart Metering systems with Virtual Power Plant, *2010 15th IEEE Mediterranean Electrotechnical Conference*, 448 ? 451.

[16] Santacana, E.; Rackliffe, G.; Tang, L.; Feng, X. (2010); Getting Smart, *IEEE Power and Energy Magazine*, 8: 41-48.

[17] Ong, Y.S.; Gooi, H.B.; Chan, C.K. (2000); Algorithms for Automatic Generation of One-line Diagrams, *IEE Proceedings Generation, Transmission and Distribution*, 147: 292 ? 298.

[18] Lendak, I.; Erdeljan, A.; Capko, D.; Vukmirovic, S. (2010); Algorithms in electrical power system one-line diagram creation, *2010 IEEE International Conference on Systems, Man, and Cybernetics*, Istanbul, Turkey, 2867-2873.

[19] Yongli, Z.; Malik, O.P. (2003); Intelligent Automatic Generation of Graphical One-Line Substation Arrangement Diagrams, *IEEE Transactions on Power Delivery*, 18: 729-735.

[20] Cheng, S.; Shirmohammadi, D. (1995); A three phase power flow method for real time distribution system analysis, *IEEE Transactions on Power Systems*, 10: 671?679.

[21] Lendak, I.; Varga, E.; Erdeljan, A.; Gavric, M. (2010), RESTful Access to Power System State Variables, *2010 IEEE Region 8 International Conference on Computational Technologies in Electrical and Electronics Engineering (SIBIRCON)*, Irkutsk, Russia, 450-454.

[22] Telvent DMS Llc official website; http://www.telventdms.com; accessed 2011-03-28.

[23] JSON.org; Introducing JSON; http://www.json.org/; accessed 2012-06-22.

[24] Popovic, D.; Varga, E.; Perlic, Z. (2007); Extension of the Common Information Model with a Catalog of Topologies, *IEEE Transactions on Power Systems*, 22: 770 ? 777.

[25] World Wide Web Consortium (W3C); Document Object Model (DOM); http://www.w3.org/DOM/; accessed 2012-06-22.

[26] OSI Soft; What is PI; http://www.osisoft.com/software-support/what-is-pi/what_is_pi_.aspx; accessed 2011-03-29.

[27] Microsoft; WCF RIA Services; http://msdn.microsoft.com/en-us/library/ee707344(VS.91).aspx; accessed 2011-03-29.

[28] ESRI; ArcGIS: A complete integrated system; http://www.esri.com/software/arcgis/index.html; accessed 2011-03-29.

[29] IEC (2003); IEC 61968-1: Application integration at electric utilities - System interfaces for distribution management - Part 1: Interface architecture and general requirements.

[30] Juliet Live Betting system; http://www.parspro.com/fp/products/live-betting; accessed 2011-03-21.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):112-119, December, 2012.

Mobile Network QoE-QoS Decision Making Tool for Performance Optimization in Critical Web Service

C. Lozano-Garzon, C. Ariza-Porras, S. Rivera-Diaz, H. Riveros-Ardila, Y. Donoso

Carlos Lozano-Garzon, Christian Ariza-Porras
Sebastián Rivera-Díaz, Horacio Riveros-Ardila,
Yezid Donoso
Universidad de los Andes, Bogotá, Colombia
E-mail: ca.lozano968@uniandes.edu.co
cf.ariza975@uniandes.edu.co, s.rivera57@uniandes.edu.co
lh.riveros102@uniandes.edu.co, ydonoso@uniandes.edu.co

Abstract:
Regardless of the type of service that a company offers the customer satisfaction is a factor for success, if these services are in a highly competitive environment. This situation encourages companies to develop strategies to improve the Quality of the Experience (QoE) of their users. Strategies include improving their processes, or infrastructure for provisioning the services. Take these kind of decisions is very difficult because they ignore how the Key Performance Indicators (KPI) services are correlated with the information about user experience. This problem is approached from the perspective of mobile telecom operators, who have addressed this challenge through the Quality of Service (QoS) concept. Unfortunately, the QoS is only characterized by technical aspects, the user's criteria are not included. Into a highly competitive environment, the user's loyalty is a key component to be considered in the operator's development plan. Nowadays, the mobile telecom operators focus their efforts to ensure not only the QoS but also the QoE.
The aim of this paper was the develop a decision making tool that allows the mobile telco operators support their determinations about the maintenance of network infrastructure, as well as the expansion of the same, specifically for their critical web services; based in a correlated information between QoS and QoE. This tool was developed on the basis of the Pseudo Subjective Quality Assessment (PSQA) methodology.
Keywords: Decision making tool, Pseudo subjective quality assessment, Quality of experience, Quality of service, Web services.

1 Introduction

The rapid evolution that has made the telecommunications industry in the world, represented in technological development in networks and the emergence of IP as a fundamental part of both fixed and mobile networks, has led the sector to a converged environment that enables businesses to provide new services. This new environment base its performance in what is now known as Next Generation Mobile Networks (NGMN) which is intended to support the growing needs of both technical and quality demands for new services.This environment generates new challenges not only in the design and implementation of the network, also this should allow the provision of services independent of location, time or device from accessing this and besides challenges in the way that operators can implement mechanisms to ensure quality of Service (QoS) and quality of experience (QoE) under these different technology platforms.

The term QoS is widely used in the environment for communications networks, it was defined by the ITU-T [1] as the collective effect of service performance which determine the satisfaction of a service user. Associated with the conceptualization of QoS Harry in [2] defines three concepts of QoS: intrinsic QoS perceived QoS and evaluated QoS, the definite relationship between these three concepts are general QoS model proposed by the ITU-T.

Stankiewicz, Cholda, & Jajszczyk in [3] describes the intrinsic QoS, it is known as "network performance" by the ITU and ETSI in Recommendation E.800, it covers all the features of service determined by the efficiency of the network. The intrinsic QoS is key to the quality perceived and evaluated by the customer. The perceived QoS reflects the customer experience in using a particular service through the ITU Recommendation G.1000 has four important perspectives: the QoS required by the client, the QoS offered by the provider, the QoS achieved by the supplier and the QoS perceived by the customer. Finally the evaluated QoS starts when the customer decides whether to continue using a service or not, this decision depends on the perceived quality, the price of the service, and supplier responses (problems and complaints), the ITU defines the guidelines of Quality of Experience (QoE) in Recommendation P.10.

Even though QoS and QoE measurements are quite different, they have a high degree of correlation; nevertheless, some mobile operators have not yet implemented tools for incorporating the "feel" of a user based on the QoS parameters measured for a specific service.

The mobile phone companies have improved the deployment and delivery of their products influenced by the quality of service (QoS), regardless of user perception. However, the mobile telco operators know that user satisfaction is a key success factor for the loyalty and positioning of the company against its competitors. This fact has led these operators to develop strategies for adoption of the perception of its users in their decision processes to the tuning of their infrastructure for the provision of services. This work aims to provide to mobile telecom operators a QoE-QoS decision making tool that will allow them to support their determinations about the maintenance of network infrastructure, as well as the expansion of the same, specifically for critical web services.

The remainder of this paper is organized as follows. In Section 2 shown some works related with the measurement of quality of experience and the correlation of it with quality of service for some specific services. The PSQA methodology is presented in Section 3. The proposed methodology used to develop the QoS-QoE decision making tool is presented in Section 4 and the implementation of this methodology is described in Section 5. Finally, the experimental results of the proposed tool are presented in Section 6 and the Section 7 presented the conclusions and future work.

2 Related Works

Stankiewicz, Cholda, and Jajszczyk in [3] notes that QoE assessment methods could be classified into Qualitative (Subjective) and Quantitative (Objective) Methods. Qualitative methods are built with the participation of people, a representative sample of the population, whom used a particular service. In these methods the service is assessed in a controlled environment and people fill out a survey with numerical values qualification. Quantitative Methods provide a QoE assessment based on the measurement of several parameters related with service quality indicators in the signal at the output of the transmission channel. However, the Institut National de Recherche en Informatique et en Automatique (INRIA) in France proposed a Hybrid Method between subjective and objective assessments of QoE called PSQA [4].

Other papers examine the relation between QoS parameters and user perception. For example the research generated by Telenor in [5] about the conceptual difference between QoS, considered the quality associated with technical performance parameters; and QoE understood as a measure of user performance, based on objective and subjective measures of the use of Information and Communications Technology (ICT) product or service. In [5], [6], and [7] it is evident that between these two concepts clear relationship that allows us to express the quality of the experience in terms of quality of service, which is the starting point for further studies made in order to establish a relationship or a relational model between the two. Some of these studies have focused on the relationship can be identified for a specific service such as the IPTV [8] and [9], transmission media [10] or public Internet [11], other studies have sought to establish a generic model of relationship between them as shown in [12], [13] and [14].

As a result of these studies, progress has been made in studies seeking to establish an objective measure based on a subjective view by users. Within the work developed in this field we find the following: [15] focused on the development of the concept of the quality of experience focused on the measurement and communication requirements for industrial use, [16] measurement studies quality of experience based on ontologies and [17] which presents a look at the main techniques for measuring the quality of experience focused on the methodologies and tools available free. Finally in [18] the authors propose a new methodology called Pseudo-Subjective Quality Assesment (PSQA) based on Random Neural Networks, to quantify the quality of a video or audio transmission over the Internet. They discuss the results concerning PSQA-based dynamic quality control and conversational quality assessment.

A general system developed to evaluate QoE on IP networks was shown in [19]. Their system architecture is designed to be capable of emulating multi agent networks and dynamically changing conditions, in a Web Browsing QoE experiment. The experiment was conducted on the basis of ITU-T Recommendation G.1030 , and aimed to update the perceptual model provided in this Recommendation to today context.

As can be seen both of these studies: relationship between Quality of Service and Quality of Experience, as well as metrics for quality of experience; are an important part of an environment using All-IP network.

3 Pseudo-Subjective Quality Assessment (PSQA)

Knowing that mobile operators require a method that allows them: correlate QoS parameters and subjective perception of the user, it can be used without creating a new testbed, and generating a set of reports that support their decisions about network infrastructure. After making a comparison between the existing metrics the Hybrid Model (specifically PSQA metric) was selected because it takes the best of the subjective and objective models, the results are in terms of Mean Opinion Score (MOS), is a not intrusive method, obtains real-time data and its implementation phase is low at the time.

This model is divided into three big steps: Firstly the application of a subjective evaluation in a controlled environment where the samples are distorted in periods of time; in the second step the samples go through a statistical process where the elements out of range are detected and removed; and finally the results are used to train a statistical learning tool, a random neural network (RNN), that learns the correlation between configurations and MOS values defined, related among the parameters that cause distortion and the perceived quality.

4 Proposed methodology for development of QoE-QoS Decision Making Tool

Based on the PSQA methodology we propose an adaptation of it, in order to develop a making decision tool that allow us to combine the knowledge of end-user experience and technical parameters values for the decision making in mobile operators. It is hoped that this new model assigns to the samples (QoS Parameters) a QoE value very close to the value that an average human observer would give.(see Figure 1).

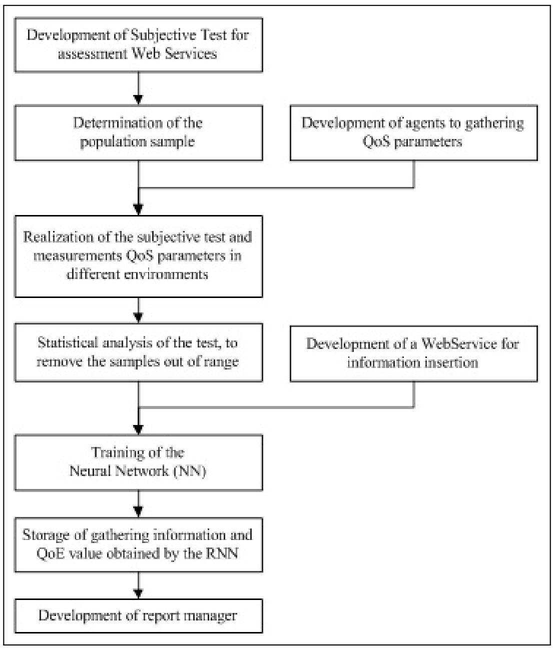

Development of Subjective Test for assessment Web Services
Determination of the population sample
Development of agents to gathering QoS parameters
Realization of the subjective test and measurements QoS parameters in different environments
Statistical analysis of the test, to remove the samples out of range
Development of a WebService for information insertion
Training of the Neural Network (NN)
Storage of gathering information and QoE value obtained by the RNN
Development of report manager

Figure 1: Flow Chart of QoE-QOS Decision Making Tool

4.1 Development of Subjective Test for assessment Web Services

When we started the development of this work, we didn't find any documented work related to the existence of a MOS test oriented to web services. For this reason, we decided the developing of a subjective test for critical web services based on some MOS existing tests. For the development of this test, we considered the following web contexts: a text-only page, a page with images and text, a video and download a file.

4.2 Determination of the Population Sample and Statistical analysis of the test

Based in [20] the number of samples must be calculated so as to ensure a confidence interval of at least 95% and an error no greater than 5%, taking into consideration the following formula:

$$n = \frac{(z_{1-\frac{\alpha}{2}})^2}{a^2} \cdot (\frac{s}{mean(x)})^2 \tag{1}$$

where n is the number of samples, $z_{1-\frac{\alpha}{2}}$ is the $1 - \frac{\alpha}{2}$ percentile of the standard normal distribution, s is the expected standard deviation, mean(x) is the expected mean value, and a is the relative accuracy.

After we collect the samples through the designed test application, the results are passed through a statistical process in order to detect and remove users who present data out of range.

4.3 Development of agents to gathering QoS parameters

The agents are the responsible of making the measurements of the QoS parameters defined for the Web browsing services. These parameters were selected according to [20]and are: bandwidth, latency, signal strength, trademark of device and the cell where the device is located.

Another important characteristic of these agents is the need to send the measurements collected to the server in order to use these in a first time like initial information to training the neural network and later as the principal source of information to calculate the QoE through the neural network.

4.4 Training of the Neural Network

Given that the neural network should behave as a classifier, we propose the use of a multilayer perceptron whose input will correspond to the five (5) QoS parameters defined above and the output will correspond to one (1) value of QoE (excellent, very good, good, fair, or poor). Once the neural network has been trained and validate, it is hoped that the results produced will be very similar to the results of people's subjective tests.

4.5 Storage of information

In order to maintain a historical record of both: the values of QoS parameters collected and the QoE values calculated; it is necessary to implement a repository of information. This repository will be the primary source of information for the generation of reports that will support decision making.

4.6 Development of report manager

The report manager should enable operators to generate the required documentation for making decisions regarding the maintenance of network infrastructure, as well as its expansion, specifically for critical web services. Some of the basic reports considered are the following: information of QoE by base station or by the device type and the correlation between QoS parameters and the estimated QoE. Also the operators need that the generator can produce new reports easily according to their requirements.

5 QoE-QoS Decision Making Tool implementation

As a first step towards the implementation of this tool we designed the software architecture to be used, it is built by six (6) components: the mobile agent, the agent listener, the persistence component, the classifier, the report generator and the presentation component. These components and their connections can be seen in Fig. 2. Additionally, in order to support connectivity between the agents and the server, we propose a client - server network architecture of three layers (See Fig. 3). The implementation of the tool was performed according to previously proposed architectures. In the server component is used, among other tools: Weka for the classifier, webservices developed to receive data from agents and classified using the trained models and BIRT as a development tool for reports and report viewer. While the mobile agent components were implemented in Java.

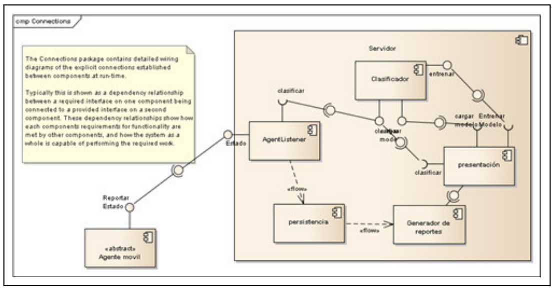

cmp Connections
The Connections package contains detailed wiring diagrams of the explicit connections established between components at run-time.
Typically this is shown as a dependency relationship between a required interface on one component being connected to a provided interface on a second component. These dependency relationships show how each components requirements for functionality are met by other components, and how the system as a whole is capable of performing the required work.
Servidor
Clasificador
entrenar
clasificar
cargar modelo
Entrenar Modelo
clasificar modelo
AgentListener
Estado
presentación
clasificar
«flow»
persistencia
«flow»
Generador de reportes
Reportar Estado
«abstract» Agente movil

Figure 2: Software Architecture Connection Diagram

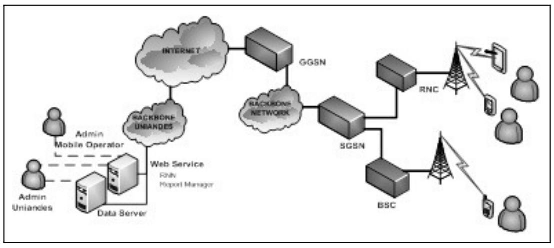


Figure 3: Network Infrastructure Proposed

6 Experimental Results

As a first step we performed the collection of about 120 subjective tests in 5 different cells of the operator. Subsequently we made a statistical analysis on data collected and all data that were outside three times the standard deviation were removed, because they were atypical behaviors that could influence the model training. With the remaining surveys, we conducted the relationship with data taken by the agents and estimated the value of QoE general survey taking an average of the evaluations of the test (text, text and images, video and downloads) and an average compared against the modified test.

Once the data were purged we proceeded to do the training of the neural network. The classification algorithm used was the multilayer perceptron with 50 nodes in the hidden layer and 5000 epochs (times) as a limit. From 50 additional samples, which are selected to valid the model, after executing the classifier (neural network) had a success rate in the classification of 90% aand the mean absolute error is close to 4.7%. In Fig 4 we show the relationship between the calculated QoE and QoE assessment of the user to test Video.

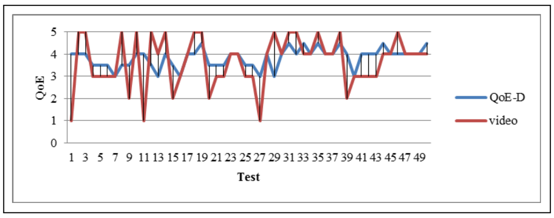


Figure 4: Relationship between the QoE assesment calculated and the test user to the video

7 Conclusions and Future Works

In conclusion, through the decision making tool developed the mobile telecom operators could estimate the users' subjective opinion based in network QoS parameters. This information allows the generation of specific reports with the aim of supporting decisions oriented to prevent product or service rejections by the users. Some of the most important decisions are related with the determinations about: the tuning of network infrastructure, the expansion of this infrastructure, the use of some specific equipment, among others. As future work we plan to work in two actions, the first is related to the exploration of other training algorithms for neural network that allows us to achieve better results, and the second seeks to expand this study to other data transmission services.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):112-119, December, 2012.

Bibliography

[1] ITU - T. , Terms and Definitions Related to Quality of Service and Network Performance including Dependability, *International Telecommunication Union., Recommendation E.800*, 1995.

[2] W. C. Hardy, *QoS: Measurement and Evaluation of Telecommunications Quality of Service:Baffins Lane*, Chichester, United Kingdom: John Wiley & Sons, Ltd., 2001.

[3] R. Stankiewicz, P. Cholda, and A. Jajszczyk, QoX: What is It Really?, *IEEE Communications Magazine*, 49(4):148 - 158, 2011.

[4] G. Rubino. The PSQA project. [Online]. http://www.irisa.fr/armor/lesmembres/Rubino/myPages/psqa.htm

[5] B. Hestnes, P. Brooks, and S. Heiestad, QoE (Quality of Experience) - measuring QoE for improving the usage of telecommunication services, Telenor, Research Report 2009.

[6] A. van Moorsel, *Metrics for the Internet Age: Quality of Experience and Quality of Business*, Hewlett - Packard Laboratories, Palo Alto, California, USA, HPL-2001-179, 2001.

[7] K. Bharrathsingh, Quality of experience as an integral part of network engineering, Focus in Convergence , vol. 1, February 2005.

[8] J. Kim, T.W. Um, Ryu W., and B. Sun Lee, Heterogeneous Networks and Terminal-Aware QoS/QoE- Guaranteed Mobile IPTV Service, *IEEE Communications Magazine*, 46(5):110 - 117, 2008.

[9] H.J. Kim and S.G. Choi, A Study on a QoS/QoE Correlation Model for QoE Evaluation on IPTV Service, in *The 12th International Conference on Advanced Communication Technology (ICACT 2010)*, Gangwon-Do, Korea, 2:11077 - 1382, 2010.

[10] M. Siller and J.C. Woods, QoS Arbitration for Improving the QoE in Multimedia Transmission, *Int. Conf. on Visual Information Engineering (VIE 2003)* , 238 - 241, 2003.

[11] S. Khirman and P. Henriksen, Relationship between Quality-of-Service and Quality-of-Experience for Public Internet Service, *Passive and Active Measurement Conference*, Palo Alto, California, USA, 1 - 6, 2002.

[12] M. Fiedler, T. Hossfeld, and Phuoc Tran-Gia, A Generic Quantitative Relationship between Quality of Experience and Quality of Service, *IEEE Network*, 24(2):36 -41, March - April 2010.

[13] H.J. Kim et al., The QoE Evaluation Method through the QoS-QoE Correlation Model, *Fourth Int. Conf. on Networked Computing and Advanced Information Management (NCM '08)*, 2:719-725, 2008.

[14] C. Guo, Y. Liu, and Y. Liu H. Du, Research on relationship between QoE and QoS based on BP Neural Network, *IEEE Int. Conf. on Network Infrastructure and Digital Content (IC-NIDC 2009)*, 312 - 315, 2009.

[15] P. Brooks and B. Hestnes, User Measures of Quality of Experience: Why Being Objective and Quantitative Is Important, *IEEE Networks*, 24(2): 8 - 13, March - April, 2010.

[16] A. Sánchez-Macián, D. López, J. E. López de Vergara, and E. Pastor, A Framework for the Automatic Calculation of Quality of Experience in Telematic Services, *Proc. of the 13th HP-OVUA Workshop, Côte d'Azur*, 1-6, 2006.

[17] R. Kooij, D. De Vleeschauwer, K. Brunnström, and F. Kuipers, Techniques for Measuring Quality of Experience, *WWIC 2010*, 216 - 217, 2010.

[18] G. Rubino, P Tirilly, and M..Varela, Evaluating Users' Satisfaction in Packet Networks Using Random Neural Networks, *Proceedings of ICANN'06*, Athens, Greece, 303-312, 2006.

[19] E. Ibarrola, F. Liberal, I. Taboada, and R. Ortega, Web QoE Evaluation in Multi-agent Networks: Validation of ITU-T G.1030, *Fifth Int. Conf. on Autonomic and Autonomous Systems (ICAS '09)*, 289 - 294, 2009.

[20] European Telecommunications Standards Institute, "Speech Processing, Transmission and Quality Aspects (STQ); User related QoS parameter definitions and measurements., European Telecommunications Standards Institute, Sophia Antipolis Cedex - FRANCE, Standard ETSI EG 202 057-2 V1.3.1, 2009.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):120-127, December, 2012.

Building a Cloud Governance Bus

V.I. Munteanu, T.-F. Fortiş, A. Copie

Victor Ion Munteanu, Teodor-Florin Fortiş, Adrian Copie
1. West University of Timişoara
Romania, Timişoara, bvd. V.Pârvan 4, and
2. Institute e-Austria, Timişoara
Romania, Timişoara, bvd. V.Pârvan 4
E-mail: vmunteanu@info.uvt.ro,
fortis@info.uvt.ro
adrian.copie@info.uvt.ro

Abstract: Thought still at its first steps, cloud governance lays the foundation upon which business innovations can be built. It fills in the gaps left by cloud providers and allows major players on the IT market to be challenged by small and medium-sized enterprises (SMEs) for their share.
At the core of cloud governance, its bus enables interaction and communication between various services and governance components. The cloud governance bus is a step forward for the enterprise service bus (ESB) into the cloud environment, addressing data integration and full implementation of enterprise integration patterns.
This paper covers current requirements for ESB migration to the cloud environment and proposes a cloud governance architecture that meets the given requirements.

Keywords: cloud governance, cloud management, cloud governance bus, enterprise integration patterns.

1 Introduction

Cloud migration is an ongoing process to which small and medium-sized enterprises (SMEs) must adhere such that they can benefit of the advantages given by its economic model. This adoption can enable them to challenge large enterprises by creating niche solutions or grouping themselves in order to provide complex applications that are tailored for their customers' needs.

According to [13], cloud computing can be summed up in five core characteristics: on-demand self-service, ubiquitous network access, location independent resource polling, rapid elasticity and pay-per-use. The last one, pay-per-use, is a clear incentive as to why cloud adoption is desired. Cloud adoption is also driven by technical characteristics like virtualization, service orientation, link with business models, strong fault tolerance, and loosely coupling, as identified in [10] .

The large amount of proprietary technologies used by cloud vendors and the lack of cloud standards has lead to the fragmentation of cloud environments making development hard. By having multiple deployment models (public, community, hybrid and private clouds), the gap is further enlarged because of the different type of policies that need to be implemented for each of them.

Several solutions that are built on-top of cloud infrastructures (IaaS) come in aid by offering flexible cloud-independent development environments and partially handling de facto things like resource provisioning, management and monitoring. Unfortunately, these platform-as-a-service (PaaS) solutions lack the functionality that is required to have a complete cloud management solution and require a set of complementary services, as exposed in [6–8].

Furthermore, current cloud applications run in isolation or in a small clusters [19] even though there is a demand for application integration at SaaS level [4, 5, 19]. This leads to the necessity of a central entity whose purpose is to enable both service and data integration and create a unitary ecosystem where applications can be easily created, managed, discovered and can easily interact one another, the necessity for cloud governance.

Cloud governance, a step forward for service oriented architecture (SOA) governance, is essential for full cloud adoption, and even the lack of a partial solution can lead to serious challenges [14]. While not part of SOA governance itself, an enterprise service bus (ESB) is a flexible connectivity infrastructure for integrating applications and services [1]. Similarly to SOA, cloud governance can benefit from the use of such a bus.

This paper focuses on defining the requirements for a cloud governance bus while providing partial solutions in the form of already available software. The remainder of the paper is organized as follows. Our motivation and related work is covered in Section 2. Section 3 introduces the mOSAIC project and its component, the Cloud Agency (CA). Our proposed cloud governance architecture is covered in section 4. The main results are presented in Section 5 and conclusions and future work are presented in Section 6.

2 Motivation and Related Work

2.1 Cloud management and governance

Cloud Management is covered in Distributed Management Task Force's (DMTF) white papers [7,8] which identify concerns and issues related to aspects of cloud service lifecycle, components in the architecture for managing clouds etc. The white papers describe management requirements in close relationship with governance ones.

The growing interest in cloud management solutions has lead to an abundance of PaaS solutions, like mOSAIC[1], OpenShift[2], Cloud Foundry[3], or Morfeo 4CaaSt[4]. However, little interest is payed o cloud governance related concerns like data and security management, logging and audit, event management and others.

A clear place for cloud governance in relation to a generic cloud management architecture is specified in [7]. Important information related to cloud governance covering Service Level Agreements (SLAs), security patterns and controls are covered in [6].

Several enterprises have taken interest in cloud governance and have integrated it as part of their PaaS solutions: enStratus[5], WSO2 Stratos[6] and Fiorano Cloud Platform[7].

2.2 Enterprise Service Bus in the cloud

Building an enterprise service bus is a challenge for any developer because of the complexity of integrating multiple services in one environment, the use of different technologies etc. There is a high variety available of ESBs ranging from commercial ones to open source ones like IBM WebSphere Message Broker[8], ORACLE ESB[9], Fuse ESB[10], Mule ESB[11], Petals ESB[12], JBoss ESB[13], OpenESB[14]. In his paper, GarcĂa-JimĂŠnez et. al. [9] compares some of the open source ESBs.

[1] http://www.mosaic-project.eu

[2] https://openshift.redhat.com/app/

[3] http://www.cloudfoundry.com/

[4] http://4caast.morfeoproject.org/

[5] http://www.enstratus.com/

[6] http://wso2.com/cloud/stratos/

[7] http://www.fiorano.com/products/ESB-enterprise-service-bus/Fiorano-Cloud-Platform.php

[8] http://www-01.ibm.com/software/integration/wbimessagebroker/

[9] http://www.oracle.com/technetwork/middleware/service-bus/overview/index.html

[10] http://fusesource.com/products/enterprise-servicemix/

[11] http://www.mulesoft.org/

[12] http://petals.ow2.org/

[13] http://www.jboss.org/jbossesb

[14] http://openesb-dev.org/

While not originally designed for the cloud, ESBs are slowly making their way into cloud environments. Some PaaS providers offer them alongside their products either built in or as a service. Some of the commercial ESB providing solutions are WSO2 Statos, Fiorano Cloud Platform, Netperspective Cloud ESB[15] and others, while open source ones are Mule ESB[16].

3 mOSAIC

mOSAIC[17] is an FP7-ICT project [12], which is developing a platform that promotes an open-source Cloud application programming interface (API) and a platform targeted for developing multi-Cloud oriented applications. Its goal is to provide enough freedom both at resource and programming level such that cloud-based services can be easily developed and deployed.

The architecture of the platform [16] is designed around the use of open and standard interfaces. Its main goal is to provide a unified Cloud programming interface which enables the flexibility needed to build inter-operable applications across different Cloud providers [15]. mOSAIC is comprised of the mOSAIC API and the Cloud Agency.

The Cloud Agency [2,17,18] is a multi-agent system that has been designed to handle resource provisioning and monitoring and also to handle reconfiguration of resources. The Cloud Agency is easily accessible to the mOSAIC platform through a REST interface. Built around a semantic engine, the Cloud Agency has capabilities that allow dynamic discovery and mapping of cloud providers. The Cloud Agency works at an IaaS level within the mOSAIC platform.

4 Cloud Governance Architecture

The proposed cloud governance architecture (Figure 1) is built in close relation with mOSAIC's Cloud Agency and is designed to offer a variety of services which complement it. This architecture closely follows DMTF's white paper [7] and is built as a multi-agent system. The Cloud Agency exposes itself within the ecosystem as services.

A clear representation of the system is depicted in Figure 1, and is composed of four subsystems: Service Management, Security Management, Audit Management and Governance Management. Each of the these subsystems is made of several agents, each agent being able to serve several of them.

The Service Management subsystem is in charge of service lifecycle management (publishing, brokering, instantiation/commissioning, etc.). Of the agents which compose it, the Service Management Agent is the most important one as it stores all service related information in the Service Datastore.

Security Management handles all security for our governance solution. The Service Management Agent is the core of this subsystem as it handles storing, retrieving, generating all the security information within the system.

The Audit Management subsystem covers all governance monitoring, ranging from cloud resource monitoring to service monitoring. It also uses a set of policies in order to notify the system or human administrator of possible errors/faults.

Governance Management manages the system based on setting and policies. It makes sure that all agents are running and that there is a sufficient number so that all systems work properly.

Several issues have been thought of when designing the system:

[15] http://www.netspective.com/netspective-cloud-esb-overview

[16] http://www.mulesoft.org/

[17] Open source API and Platform for multiple Clouds

- complete integration of cloud management (cloud resource management, scaling, monitoring, reconfiguration);
- complete service management and lifecycle related issues (including scaling, monitoring and reconfiguration);
- complete security and privacy management;
- compliance with business practices and standards.

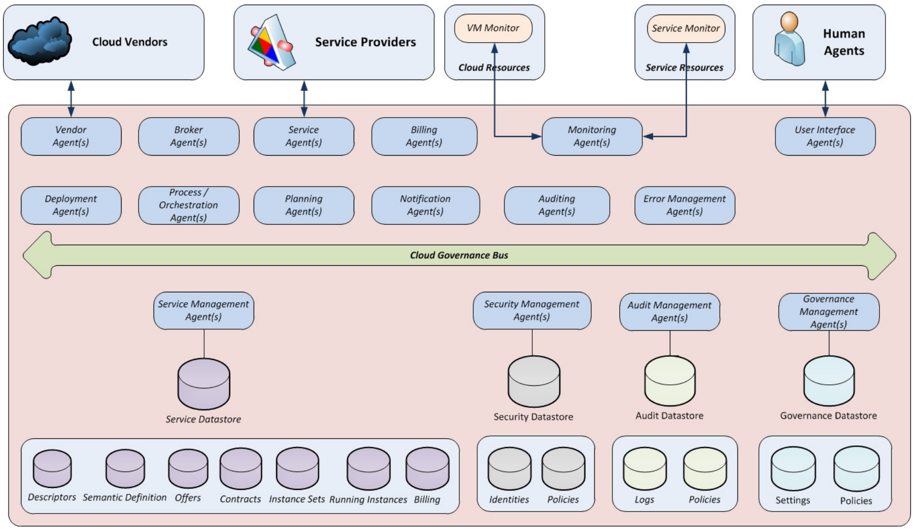


Figure 1: Cloud Governance

5 Cloud Governance Bus

The traditional role of an ESB in a SOA environment is to simplify access by hiding the complexity of the underlying system and providing a generic way for querying, accessing and interacting between services. This is achieved by handling the routing and monitoring of messages between services, handling service deployment and versioning etc.

Similarly to the ESB, a cloud governance bus (CGB) needs to be able to handle messages (queuing, sequencing), security, exceptions, protocol conversion and provide an adequate level of quality of services (QoS). Unlike traditional ESBs, our proposed CGB implements enterprise integration patterns (EIP) as well as data integration (extraction, transformation, loading, mapping) which enables easy access to datastores as well as other components like the integrated Scala implementation of ActiveMQ provided by Apache Apollo[18].

The CGB is first and foremost designed to handle the internal communication of our proposed cloud governance solution in a secure manner. Having ESB-like features is a secondary goal. As SOA allows for the development of both tightly coupled and loosely coupled services, having the opportunity to integrate them in our CGB is nice to have, but not a priority.

The following list summarizes what CGB features we would like/are a must having:

- Support both synchronous and asynchronous interaction between services
- Allow message operations like filtering, routing, translating

[18] http://activemq.apache.org/apollo/

- Allow various forms of message routing including, but not limited to, static routing, content-based routing, rules-based routing, policy-based routing
- Allow both statically and dynamically bound services
- Allow any type of data to be handled
- Handle semantic transformation if required
- Allow the possibility to define message channels
- Separate system messages from service messages
- Allow various ways in which endpoints can be defined

In his paper, Kiran Kanetkar discusses several functions that an ESB must handle [11]: routing, transformation, adaptation, messaging, orchestration, UDDI registry, security, consumer integration, service integration, metrics and management, and B2B. However, for building our cloud governance system, we only need to handle the most important functions as well as EIP.

In [3], Rob Barry identifies several problems an ESB has to face when being deployed in a cloud environment. Because of the various deployment environments (public, hybrid or private clouds) an ESB must adopt specific security policies (encryption) when dealing with the messages or authentication within the system. Another issue is the latency that can arise from sending messages between various clouds and the transport protocols the ESB needs to know.

5.1 Using Akka and Apache Camel

In order to address the issues related with building a cloud governance bus, two technologies that can cover them were identified, solutions that are event-driven and enable EIP and data integration. One of them, Akka[19], is an event-driven middleware in Scala[20]. While not a traditional, FIPA compliant, multi-agent system, Akka can be used successfully for building high performance and reliable distributed applications.

Akka's architecture allows easy mapping of agents to its Actor system. Its event driven system allows building reactive agents, facilitating them with mailboxes, another feature needed for an CGB. Akka's high-performance, self-healing, transparent-distributed system is complemented by features like support for various development libraries that enhance it like REST, Comet, Spring, Guice, Lift, Apache Came™, Persistence and AQMP libraries.

Akka's Apache Camel™ module allows easy integration with it. Apache Camel™ is a versatile open-source integration framework based on known Enterprise Integration Patterns.

It enhances our CGB by enabling the definition of routing and mediation rules in a variety of languages. It can be easily integrated into any kind of transport of messaging model that our CGB employs, and enhances or cloud governance architecture's ability to communicate with 3rd party applications and partners.

6 Conclusions and Future Work

Unlike large enterprises which have the resources (financial and otherwise) to build and maintain their own infrastructure, SMEs find themselves lacking and looking elsewhere for support. That support is found in the Cloud, where they can delegate infrastructure management to cloud

[19] http://akka.io/

[20] http://www.scala-lang.org/

providers benefiting from the given pay per use economic model. However they are somewhat limited and need a governance solution that enables them to group and provide complex, targeted services tailored for their customers' needs.

Cloud governance is complementary to cloud management through the services it provides. By having a cloud governance bus as the core of our cloud governance architecture, we enable a new approach to business integration and to building a highly complex and business oriented ecosystem.

This paper tried to cover requirements for a cloud governance bus from the perspective of our proposed governance architecture. Future work will cover patterns for building highly interdependent cloud services by using our bus to route and translate their messages.

Acknowledgments

This work was partially supported by the grant of the European Commission FP7-ICT-2009-5-256910 (mOSAIC), FP7-REGPOT-CT-2011-284595 (HOST), and Romanian national grant PN-II-ID-PCE-2011-3-0260 (AMICAS). The views expressed in this paper do not necessarily reflect those of the corresponding projects' consortium members.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):120-127, December, 2012.

Bibliography

[1] Wohl Associates. SOA governance. An IBM white paper. White paper. 2006. Available on-line at:
http://www-01.ibm.com/software/solutions/soa/Amy_Wohl_SOA_Governance_Analyst_White_Paper.

[2] R. Aversa, B. Di Martino, M. Rak, and S. Venticinque. Cloud Agency: A Mobile Agent Based Cloud System. In *Proceedings of the 2010 International Conference on Complex, Intelligent and Software Intensive Systems*, CISIS '10, pages 132–137, Washington, DC, USA, 2010. IEEE Computer Society.

[3] R. Barry. ESBs in the cloud: Tricky in the early going. News. June 2010. Available on-line at:
http://searchsoa.techtarget.com/news/1514427/ESBs-in-the-cloud-Tricky-in-the-early-going.

[4] S. Bennett, T. Erl, C. Gee, R. Laird, A. T. Manes, R. Schneider, L. Shuster, A. Tost, and C. Venable. SOA Governance: Governing Shared Services On-Premise & in the Cloud. Prentice Hall/PearsonPTR, 2011.

[5] T. Cecere. Five steps to creating a governance framework for cloud security. Cloud Computing Journal. November 2011. Available on-line at: http://cloudcomputing.sys-con.com/node/2073041.

[6] Cloud Computing Use Cases Group. Cloud computing use cases white paper. July 2010. Available on-line at:
http://opencloudmanifesto.org/Cloud_Computing_Use_Cases_Whitepaper-4_0.pdf.

[7] DMTF. Architecture for Managing Clouds. June 2010. Available on-line at: http://dmtf.org/sites/default/files/standards/documents/DSP-IS0102_1.0.0.pdf.

[8] DMTF. Use Cases and Interactions for Managing Clouds. June 2010. Available on-line at: http://www.dmtf.org/sites/default/files/standards/documents/DSP-IS0103_1.0.0.pdf.

[9] F.J. Garcia-Jimenez, M.A. Martinez-Carreras, and A.F. Gomez-Skarmeta. Evaluating open source enterprise service bus. In *e-Business Engineering (ICEBE), 2010 IEEE 7th International Conference on*, pages 284 –291, nov. 2010.

[10] C. Gong, J. Liu, Q. Zhang, H. Chen, and Z. Gong. The characteristics of cloud computing. In Wang-Chien Lee and Xin Yuan, editors, *ICPP Workshops*, pages 275–279. IEEE Computer Society, 2010.

[11] K. Kanetkar. A roadmap to building an ESB. May 2006. Available on-line at: http://www.saterisystems.com/Docs/whitepapers/Roadmap to Building an ESB.pdf.

[12] mOSAIC Consortium. The mOSAIC project. 2010. Available on-line at: http://mosaic-cloud.eu/.

[13] A. Mulholland, J. Pyke, and P. Fingar. Enterprise Cloud Computing: A Strategy Guide for Business and Technology Leaders. Meghan-Kiffer Press, Tampa, FL, USA, 2010.

[14] P. Mynampati. Soa governance: Examples of service life cycle management processes. November 2008. Available on-line at: http://www.ibm.com/developerworks/webservices/library/ws-soa-governance/index.html.

[15] D. Petcu, C. Crăciun, M. Neagul, I. Lazkanotegi, and M. Rak. Building an interoperability API for sky computing. In *High Performance Computing and Simulation (HPCS), 2011 International Conference on*, pages 405–411, july 2011.

[16] D. Petcu, S. Panica, and M. Neagul. From grid computing towards sky computing. case study for earth observation. Proceedings Cracow Grid Workshop 2010, pages 11–20. Academic Computer Center, Poland, 2010.

[17] S. Venticinque, R. Aversa, B. Di Martino, M. Rak, and D. Petcu. A cloud agency for SLA negotiation and management. In *Proceedings of the 2010 conference on Parallel processing*, Euro-Par 2010, pages 587–594, Berlin, Heidelberg, 2011. Springer-Verlag.

[18] S. Venticinque, R. Aversa, B. Di Martino, and D. Petcu. Agent based Cloud Provisioning and Management - Design and Prototypal Implementation. In *CLOSER 2010*, pages 184–191, 2011.

[19] P. Wainewright. Time to think about cloud governance. August 2011. Available on-line at: http://www.zdnet.com/blog/saas/time-to-think-about-cloud-governance/1376.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):128-136, December, 2012.

Implementing the Main Functionalities Required by Semantic Search in Decision-Support Systems

S.C. Necula

Sabina-Cristiana Necula
"Alexandru Ioan Cuza" University of Iasi
Romania, 700505 Iasi, 22 Carol I Blvd
E-mail: sabina.mihalache@gmail.com

Abstract:
This paper exploits different semantic web technologies and builds a prototype of semantic web mashup functionality based on an architecture proposed by us. The main scope is to improve decision-making processes. In this paper, we are focusing on querying different ontologies in order to improve decision-making semantic search. As a conclusion, we demonstrate that in order to improve decision-making semantic search there is a need of special constructions that a query language must support.

Keywords: semantic web, decision-making semantic search, ontology, SPARQL, SQWRL, RDF, OWL.

1 Introduction

The main problem treated by the present article is: improving Decision Support Systems (DSS) means improving search which leads to the semantic interoperability problem.

There is a main problem and, in the same time, a controversy here. When discussing enterprise decision support systems data exists, data comes from different sources (internal and external to enterprise) but they are differently described. There is an open world (the Web) and a closed world (the enterprise). If in the open world we discuss search engines and queries made by Internet users on existent data from www space, in the closed world we discuss databases and answers to queries that are priory represented. In the first case the answers to queries seems to be poorly represented in the aspect of semantics, in the second case the answers seems to be perfectly represented but not relevant. In this closed world of enterprise, usually decision-makers need answers to queries formulated by them and these queries are ad-hoc, very often cannot be anticipated. Their answers need a mixture of information that comes from the both worlds.

Even if all computer-based solutions are well intended the practical use shows that user satisfaction is low due to many reasons analyzed by studies that have been realized. There are many identified reasons: poor maintainability, poor flexibility and less reusability (see [22]). Much of the implicit information remains undiscovered thereby resulting in sub-optimal business decisions.

Web-based technologies are having a major impact on design, development, and implementation processes for all types of decision support systems (see [1]).

Therefore it is our goal to solve this problem by this article.

We consider in this study two main important reasons: (1) the possibility to apply knowledge in the decision-making moment at the decision-making place by the decision-maker; (2) semantic integration of information.

As a solution to the problem identified by us we propose practical implementations of semantic web search functionalities with existent Decision Support Systems. We present scenarios, technical requirements, and the architecture of a semantic web based application in decision support systems.

The scope of this paper is to present a method to improve search at the decision-making level. The main idea consists in using ontologies and semantic search technologies. The motivation is given by lack of interoperability and semantic consistency of different formats for the same content.

2 Related Work

In order to discuss differences of our approach we present the related achieved work in the field of adopting semantic web standards and the related achieved work in the field of developing applications in the context of decision-making.

Decision-making has been considered as one of the main concern in practice as in theory (see [19], [3], [13]and [10]).

A reference paper in the field belongs to H.K. Bhargava, D.J. Power, and D. Sun (see [2]) which presents the technological challenges that web technologies create for Decision Support Systems. They show that DSS generally require repeated interactions with a model, while the basic Web architectural model was designed for random jumps in hyperspace (see [2]). Integration is a major challenge in the context of Web-based decision computation.

Most of all previous studies focus on success factors of information systems (see [4], [5], [12]and [17]).

Describing big data sets according to schema(s) and accessing data by overlapping schemas is the big problem that Semantic Web is trying to address. Its main applications are in the field of managing the currently linked data, learning how to extract information from currently linked data (Linked Data Browsers), sense-making of events (there is a lot of life data streams like tweets) in order to provide a solution able to gather, collect and analyze in real time a large number of live data streams (e.g. twits), to extract the information contained and to map any reference to both a) geographical location/point of interest, etc. and b) domain specific facts (e.g. music events or violence/ demonstration). The goal is to identify events happening in a specific area (e.g. a specific city) in a short time (e.g. some hours). Sense making of the events is provided via mapping events over location and time. There is a lot of use for social purposes like emergency operators, governmental bodies. We must say that Semantic Web is intended to be used by people in the way which web intended to be at its origins.

Semantic web technologies are by far most often used for data integration and for improving the search (see [9]). The use of ontologies for knowledge sharing, heterogeneous database integration, and semantic interoperability has been long realized (see [8], [11]and [18]).

Metadata (data about data)/ ontology promises to overcome the problem of interoperability. An accepted standard for representing metadata should provide information about the syntax, the structure and the semantic context of data. Many standards have been proposed. For the moment the World Wide Web Consortium recommends RDF (S) (Resource Description Framework - Schema), and OWL (Ontology Web Language) for representing linked web data.

A number of query languages have been developed to query RDF and OWL. SPARQL (SPARQL Protocol and RDF Query Language) is currently the de facto standard RDF query language. Since OWL can be serialized as RDF, SPARQL can be used to query it. There is thus a need for an expressive OWL query language that supports comprehensive querying of OWL (see [14]). These authors proposed SQWRL (Semantic Query-enhanced Web Rule Language) built on SWRL (Semantic Web Rule Language).

A mashup is a Web page or application that uses and combines data, presentation or functionality from two or more sources to create new services. Mashup techniques retrieve content from several sources to create a new service or application.

3 Problem Statement

The limits of information integration means for decisions modeling the following issues: (1) input data in the decision-making model is changing. The set of models uses to change over time due to changing conditions and priorities. Thus, it becomes necessary that the conceptual structure can change as the data impose. A new proposed solution is to use ontologies; (2) at the same time, data come from multiple data sources (internal and external to organization) (see [6]and [7]). Thus, it becomes necessary to ensure semantic interoperability.

Two expensive stages appear in the management of decision-making models: 1) each data source has to have a schema and 2) there is a need to overlap different schemas.

Our research tries first to identify the main factors that are important in Decision Support usefulness through an empirical study and then to discuss how to overcome the limitations by applying Semantic Web to Decision Support Systems.

In order to clarify what we are talking about, we need to discuss what we mean when treating the problem of decision-making semantic search: (1) we don't discuss methods for representing knowledge or data structures or search algorithms in order to prove that one method or another technique is better and improves the semantics of content. Instead we the problem of decision-making semantic search as a search need to provide answers to queries formulated by enterprises decision-makers from two main worlds: the web and the internal world of enterprise; (2) we present the problem of decision-making semantic search in the light of the main technologies involved, the architecture and the main resources involved. We demonstrate it by presenting practical examples. These examples prove that by using semantic web technologies search results improves.

4 Our Approach

We propose the research model shown in Figure 1 by considering the characteristics of Decision Support Systems and also by referring to DeLone and McLeans Information Systems success model (see [4]).

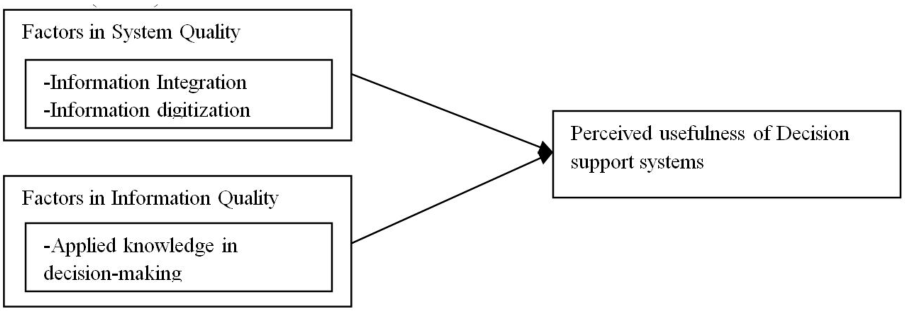
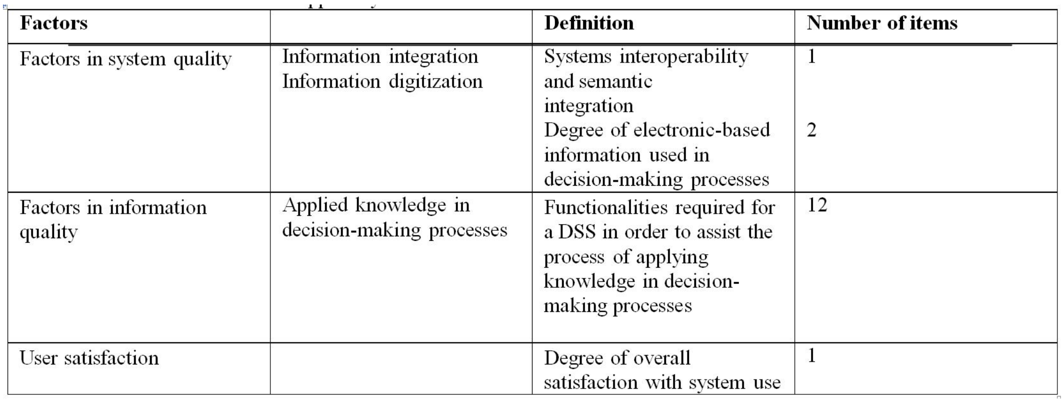

Factors in System Quality
-Information Integration
-Information digitization
Factors in Information Quality
-Applied knowledge in decision-making
Perceived usefulness of Decision support systems

Figure 1: The research model

We developed a number of items to measure each construct from the research model by making references to previous work in the research field. We present a survey and the analysis realized on the results obtained by conducting the survey.

Table 1 provides the definition of the factors listed in Fig. 1, and indicates the number of items used to measure each construct. We developed the measures by referring to the following previous work: 1) system quality and information quality of the IS success model (see 4, 5); 2) application of the Semantic Web to KM and (3) the possibilities to apply knowledge by the decision-makers (see [15], [16]and [21]).



Factors		Definition	Number of items
Factors in system quality	Information integration Information digitization	Systems interoperability and semantic integration Degree of electronic-based information used in decision-making processes	1 2
Factors in information quality	Applied knowledge in decision-making processes	Functionalities required for a DSS in order to assist the process of applying knowledge in decision-making processes	12
User satisfaction		Degree of overall satisfaction with system use	1

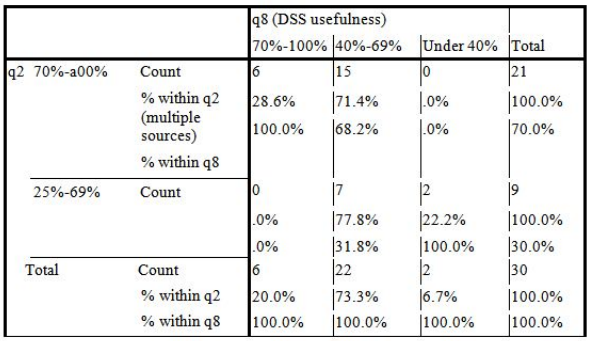
Table 1: Definition of Decision Support Systems success factors

We derived two main hypotheses from the research model.

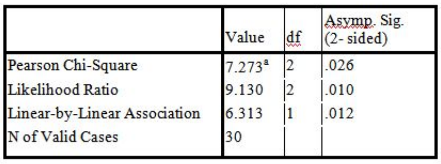
H1: applying knowledge in the decision-making moment has a positive impact on user satisfaction with the DSS H2: information integration has a positive impact on user satisfaction with DSS.

We interviewed 18 decision-makers from local public authorities that adopted and used a Decision Support System for three years to pretest the survey questionnaire. The primary goal of the pretest was to check content validity of the questionnaire. We revised a few question items. We then addressed the questionnaire to 34 decision-makers that own a business and that have invested in IS solutions. A total of 30 responses were used for statistical analysis (the software package SPSS version 17.0).

We used CrossTab correlation (Table 2) to see if is any direct relation between the degree in which it is considered that information needed in decision-making process comes from multiple sources and the degree in which the user is satisfied with DSS use (direct relation the main majority of the subjects responded affirmatively).



			q8 (DSS usefulness)			
			70%-100%	40%-69%	Under 40%	Total
q2	70%-a00%	Count	6	15	0	21
		% within q2 (multiple sources)	28.6%	71.4%	.0%	100.0%
		% within q8	100.0%	68.2%	.0%	70.0%
	25%-69%	Count	0	7	2	9
			.0%	77.8%	22.2%	100.0%
			.0%	31.8%	100.0%	30.0%
	Total	Count	6	22	2	30
		% within q2	20.0%	73.3%	6.7%	100.0%
		% within q8	100.0%	100.0%	100.0%	100.0%



	Value	df	Asymp. Sig. (2- sided)
Pearson Chi-Square	7.273^a	2	.026
Likelihood Ratio	9.130	2	.010
Linear-by-Linear Association	6.313	1	.012
N of Valid Cases	30		

Table 2: CrossTab correlation and Chi-Square Tests

Semantic web applications can be classified into two categories: generic applications (Linked Data browsers and Linked Data Search Engines) and domain-specific applications (ESW LOD wiki; Semantic Web Use Case Studies and Use Cases).

A mashup should be developed following three steps:

1) Discover data sources that provide data by following RDF links from an initial seed URI into other data sources.

2) Download data from the discovered data sources and store the data together with provenance meta-information in a local RDF store.

3) Retrieve information from the local store using SPARQL query language.

While there are not many implementations in the field of decision-making semantic search we benefit from the existing grounding technologies like ontology, RDF/OWL descriptions, SPARQL language. Our solution consists on all of the above technologies and the proposed architecture is presented in Figure 2.

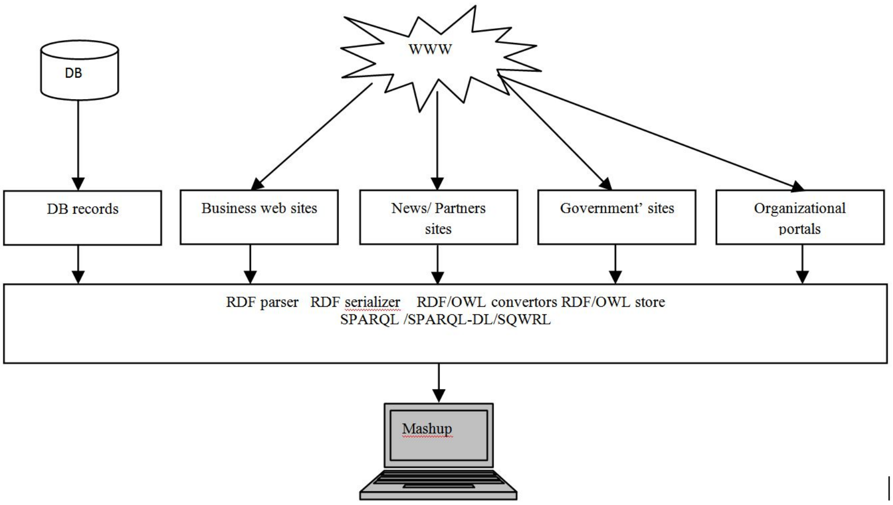
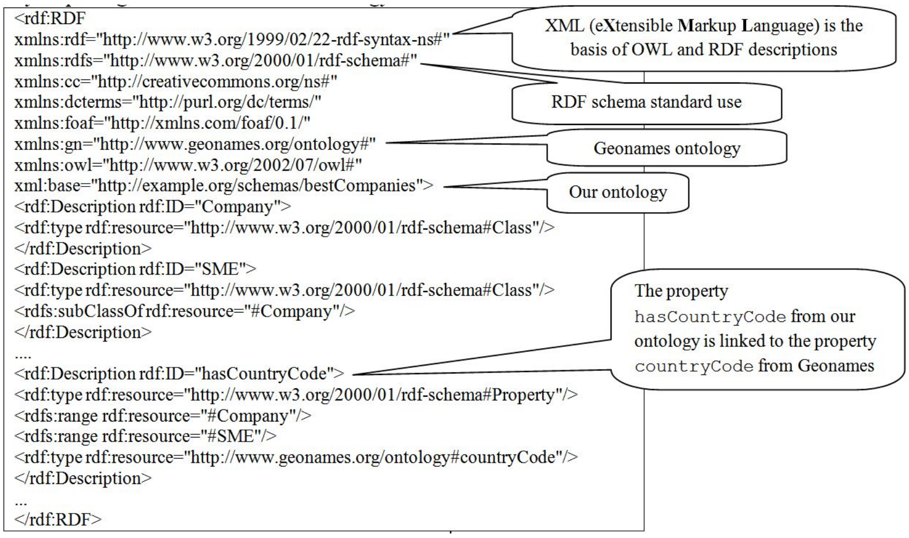


Figure 2: Semantic Web applications architecture

Our work is built upon five specifications: RDF, RDFS, OWL, SPARQL, and SWRL (SQWRL).

In Figure 3 we present a schema for describing Companies and an example of data described by respecting the defined schema/ ontology.



```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:cc="http://creativecommons.org/ns#"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:gn="http://www.geonames.org/ontology#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xml:base="http://example.org/schemas/bestCompanies">
<rdf:Description rdf:ID="Company">
<rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
</rdf:Description>
<rdf:Description rdf:ID="SME">
<rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
<rdfs:subClassOf rdf:resource="#Company"/>
</rdf:Description>
....
<rdf:Description rdf:ID="hasCountryCode">
<rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Property"/>
<rdfs:range rdf:resource="#Company"/>
<rdfs:range rdf:resource="#SME"/>
<rdf:type rdf:resource="http://www.geonames.org/ontology#countryCode"/>
</rdf:Description>
...
</rdf:RDF>
```

XML (e**X**tensible **M**arkup **L**anguage) is the basis of OWL and RDF descriptions

RDF schema standard use

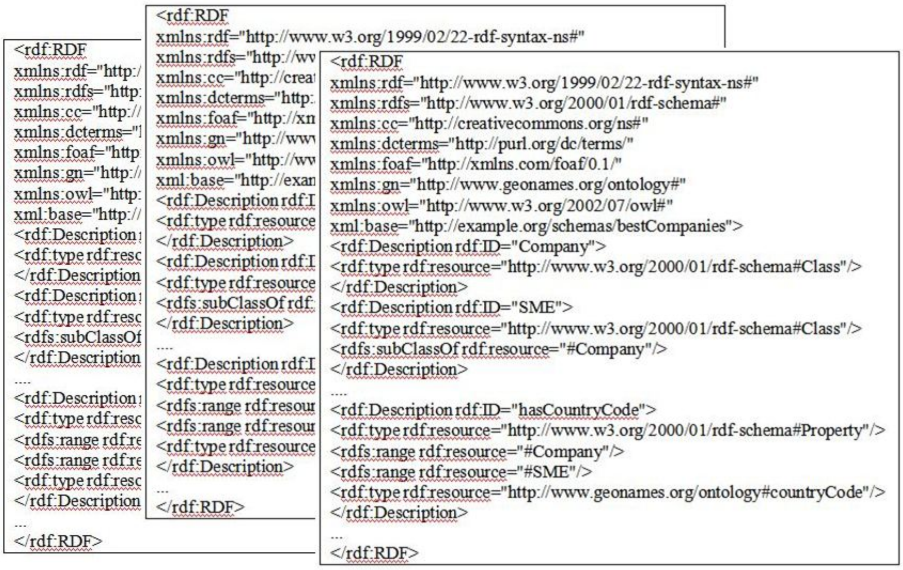
Geonames ontology

Our ontology

The property `hasCountryCode` from our ontology is linked to the property `countryCode` from Geonames

Figure 3: Companies RDF schema/ ontology

Web Data that has been cached locally is usually either accessed via SPARQL queries or via an RDF API. Using these specifications we developed a prototype Web based application to semantically improve a mashup that advises decision-makers of an organization. The Web-based application allows decision-makers to receive advice concerning their compared assets value to the ones of competitors. We present an example of different schemas/ ontologies needed to represent data stored in the RDF store (Figure 4).



```
<rdf:RDF
xmlns:rdf="http:/
xmlns:rdfs="http:
xmlns:cc="http://
xmlns:dcterms="l
xmlns:foaf="http
xmlns:gn="http://
xmlns:owl="http:
xml:base="http://
<rdf:Description
<rdf:type rdf:resc
</rdf:Description
<rdf:Description
<rdf:type rdf:resc
<rdfs:subClassOf
</rdf:Description
....
<rdf:Description
<rdf:type rdf:resc
<rdfs:range rdf:re
<rdfs:range rdf:re
<rdf:type rdf:resc
</rdf:Description
...
</rdf:RDF>
```

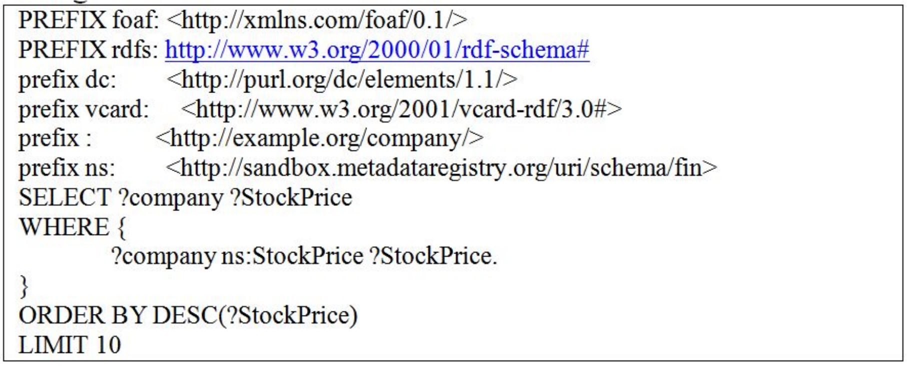
```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://wv
xmlns:cc="http://crea
xmlns:dcterms="http:
xmlns:foaf="http://xn
xmlns:gn="http://ww
xmlns:owl="http://wv
xml:base="http://exan
<rdf:Description rdf:I
<rdf:type rdf:resource
</rdf:Description>
<rdf:Description rdf:I
<rdf:type rdf:resource
<rdfs:subClassOf rdf:
</rdf:Description>
....
<rdf:Description rdf:I
<rdf:type rdf:resource
<rdfs:range rdf:resour
<rdfs:range rdf:resour
<rdf:type rdf:resource
</rdf:Description>
...
</rdf:RDF>
```

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:cc="http://creativecommons.org/ns#"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:gn="http://www.geonames.org/ontology#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xml:base="http://example.org/schemas/bestCompanies">
<rdf:Description rdf:ID="Company">
<rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
</rdf:Description>
<rdf:Description rdf:ID="SME">
<rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
<rdfs:subClassOf rdf:resource="#Company"/>
</rdf:Description>
....
<rdf:Description rdf:ID="hasCountryCode">
<rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Property"/>
<rdfs:range rdf:resource="#Company"/>
<rdfs:range rdf:resource="#SME"/>
<rdf:type rdf:resource="http://www.geonames.org/ontology#countryCode"/>
</rdf:Description>
...
</rdf:RDF>
```

Figure 4: Different schemas and data stored in the RDF store

SPARQL makes it possible to send queries and receive results, e.g., through Hypertext Transfer Protocol (HTTP) or Service Oriented Architecture Protocol (SOAP).

Using SPARQL consumers of the Web of Data can extract possibly complex information (i.e., existing resource references and their relationships). If we want to query what are the uri and the StockPrice of highest SockPriced companies we will write a query that look like query depicted in Figure 5.



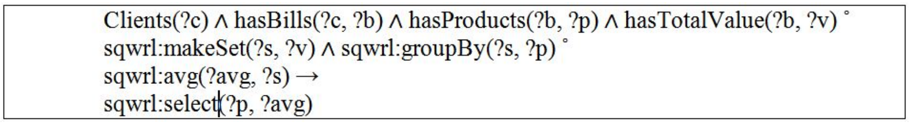
```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rdfs: http://www.w3.org/2000/01/rdf-schema#
prefix dc:      <http://purl.org/dc/elements/1.1/>
prefix vcard:     <http://www.w3.org/2001/vcard-rdf/3.0#>
prefix :       <http://example.org/company/>
prefix ns:      <http://sandbox.metadataregistry.org/uri/schema/fin>
SELECT ?company ?StockPrice
WHERE {
          ?company ns:StockPrice ?StockPrice.
}
ORDER BY DESC(?StockPrice)
LIMIT 10
```

Figure 5: A SPARQL query that returns uri and StockPrice of the ten companies that have the highest Stock Price

SPARQL is particularly adequate for extracting data from ontology and, through its CONSTRUCT statement, for generating new data.

But when it comes to aggregated functions there is no aggregated function implemented by SPARQL yet. So, because of an evident need in the decision-making processes for aggregated functions we were making use of SQWRL a powerful tool to express queries with aggregated functions. Figure 6 presents a query with aggregated function.

When executing SQWRL query, first SWRL inference rules are taken into account/executed. Running SQWRL requires a rule engine, currently Jess. We were able to derive answers for



```
Clients(?c) ∧ hasBills(?c, ?b) ∧ hasProducts(?b, ?p) ∧ hasTotalValue(?b, ?v) ˚
sqwrl:makeSet(?s, ?v) ∧ sqwrl:groupBy(?s, ?p) ˚
sqwrl:avg(?avg, ?s) →
sqwrl:select(?p, ?avg)
```

Figure 6: SQWRL query that returns the average value for each sold product

questions like: what is the average value for each sold product. In order to obtain an answer we used SQWRL, because SPARQL doesnt have the necessary specifications.

5 Conclusions and Future Works

We have two contributions in this paper: (1) We propose that a combination of RDF/OWL with SQWRL to add semantic markup is useful for DSS; (2) We show how to combine them technically.

The need today is for a distributed evolution of ontologies. The overall problem for ontology engineering is that the number of ontologies which is available is currently very limited, and it is hard to validate the approaches using real ontologies.

Open issues that remains to discuss are: (1) consistency in order to meet the requirements of future real-life applications; (2) evolution of ontologies and metadata; (3) method and tools that scale up to handle a large number of networked ontologies and related metadata.

The web is effective at bringing any resource to the web user, but if the information the user needs is not represented in a single place, the job of integration belongs to the user.

How much from the intended message could be provided by using semantic web technologies? Pretty much how much it is intended to be represented. The power of represented linked data should be in discovering relations in existing represented data. The whole idea of linked data consists in consuming and publishing data. There is no intent in representing a standard/ de facto ontology that is the best in the modeled domain, the whole intent is to link our data with others data.

The main problem remains that of scrapping data from the Web. In order to not scrap for data, every web source provider should have his/her data represented in a standard semantic web format. In this way, semantic web application could gather data semantically described and sharing it to the user.

For the future work, (1) currently we are building software to improve mashup; (2) We plan to improve the existing ontology mediation.

Acknowledgment

This work was supported by CNCSIS-UEFISCSU, project number PN II-RU code 188/2010.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):128-136, December, 2012.

Bibliography

[1] Bharati P., Chaudhury A., An empirical investigation of decision-making satisfaction in web-based decision support systems, *DECISION SUPPORT SYSTEMS*, 37: 187-197, 2004.

[2] Bhargava H.K., Power D.J., Sun D., Progress in Web-based decision support technologies, *DECISION SUPPORT SYSTEMS*, 43: 1083-1095, 2007.

[3] Chu P.C., Spires E.E., The joint effects of effort and quality on decision strategy choice with computerized decision aids, *DECISION SCIENCES*, 31 (2): 259-292, 2000.

[4] Delone W.H., McLean E.R., Information system success: The quest for the dependent variable, *INFORMATION SYSTEM RESEARCH*, 3(1): 60-95, 1992.

[5] Delone W.H., McLean E.R., The DeLone and McLean model of information systems success: A ten-year update, *JOURNAL OF MANAGEMENT INFORMATION SYSTEMS*, 19(4): 9-30, 2003.

[6] Eppler M., Mengis J., The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines, *THE INFORMATION SOCIETY*, 20(5): 325-344, 2004.

[7] Farhoomand A.F., Drury D.H., Managerial information overload, *COMMUNICATION OF THE ACM*, 45(10): 127-131, 2002.

[8] Gruber T., The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases, in *Principles of Knowledge Representation and Reasoning*, J. Allen, R. Fikes, and E. Sandewall, eds., Morgan Kaufman, San Mateo, CA, pp. 601-602, 1991.

[9] Janev V., Vrane S., *Applicability assessment of Semantic Web technologies*, *INFORMATION PROCESSING AND MANAGEMENT*, Elsevier, 2010.

[10] Janjua N. K., Hussain F. K., Web@IDSS Argumentation-enabled Web-based IDSS for reasoning over incomplete and conflicting information, *KNOWLEDGE-BASED SYTEMS*, 2011, Article in press.

[11] Kashyap V., Sheth A., Semantics-based Information Brokering, in *Proceedings of the Third International Conference on Information and Knowledge Management (CIKM)*, pp. 363-370, 1994.

[12] Kulkarni U., Ravindran S., Freeze R., A knowledge management success model: Theoretical development and empirical validation, *JOURNAL OF MANAGEMENT INFORMATION SYSTEMS*, 23(3): 309-347, 2007.

[13] Lau H.C.W., Tsui W.T., An iterative heuristics expert system for enhancing consolidation shipment process in logistics operations, in: Z. Shi, K. Shimohara, D. Feng (Eds.), *INTELLIGENT INFORMATION PROCESSING*, Springer, Boston, pp. 279-289, 2006.

[14] Connor M., Das A., SQWRL: a Query Language for OWL, OWL: Experiences and Directions (OWLED 2009), *Fifth International Workshop*, Chantilly, VA, 2009.

[15] Pomerol J.-C., Scenario development and practical decision making under uncertainty, *DECISION SUPPORT SYSTEMS*, 31 (2):197-204, 2001.

[16] Ramirez R., Melville N., Lawler E., Information technology infrastructure, organizational process redesign, and business value: An empirical analysis, *DECISION SUPPORT SYSTEMS*, 49(4): 417-429, 2010.

[17] Seddon P.B., A respecification and extension of the DeLone and McLean Model of IS Success, *INFORMATION SYSTEMS RESEARCH*, 8(3): 240-253, 1997.

[18] Sheth A., Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics, in *INTEROPERATING GEOGRAPHIC INFORMATION SYSTEMS*, M. Goodchild, M. Egenhofer, R. Fegeas, and C. Kottman, Eds., Kluwer Publishers, 1998.

[19] Simon H.A., Administrative Behavior. A Study of Decision-Making Processes in Administrative Organization, *The Free Press*, New York, 1976.

[20] Turban E., Aronson J.E., Decision Support Systems and Intelligent Systems, 6th ed, Prentice Hall, Upper Saddle River, NJ, 2001.

[21] Wagner E.L., Scott S.V., Galliers R.D., The creation of 'best practice' software: Myth, reality and ethics, *INFORMATION AND ORGANIZATION*, 16(3): 251-275, 2006.

[22] Xie Y., Wang H., Efstathiou J., A research framework for Web-based open decision support systems, *KNOWLEDGE-BASED SYSTEMS*, 18 (7): 303-319, 2005.

[23] The W3C Semantic Web Development Tools website http://www.w3.org/wiki/SemanticWebTools

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):137-144, December, 2012.

Evolutionary Algorithm based on the Automata Theory for the Multi-objective Optimization of Combinatorial Problems

E. Niño-Ruiz

Elias D. Niño-Ruiz
Universidad del Norte
KM5 Via Puerto Colombia
Barranquilla, Colombia
Web: http://combinatorialoptimization.blogspot.com/
E-mail: enino@uninorte.edu.co

Abstract:
This paper states a novel, Evolutionary Metaheuristic Based on the Automata Theory (EMODS) for the multiobjective optimization of combinatorial problems. The proposed algorithm uses the natural selection theory in order to explore the feasible solutions space of a combinatorial problem. Due to this, local optimums are often avoided. Also, EMODS exploits the optimization process from the Metaheuristic of Deterministic Swapping to avoid finding unfeasible solutions. The proposed algorithm was tested using well known multi-objective TSP instances from the TSPLIB. Its results were compared against others Automata Theory inspired Algorithms using metrics from the specialized literature. In every case, the EMODS results on the metrics were always better and in some of those cases, the distance from the true solutions was 0.89%.

Keywords: Combinatorial Optimization, Multi-objective Optimization, Automata Theory, Metaheuristic of Swapping.

1 Introduction

As well known, Combinatorial Optimization is a branch of the Optimization. Its domain is optimization problems where the set of feasible solutions is discrete or can be reduced to a discrete one, and the goal is to find the best possible solution [8]. In this field it is possible to find a large number of problems denominated NP-Hard, that is mean that the problem does not have a solution in polynomial time. One of the most classical problems in the combinatorial optimization field is the Traveling Salesman Problem (TSP), it has been analyzed for years [6] either in a mono or multi-objective manner. Formally, TSP is defined as follows:

$$min \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \cdot X_{ij}, \tag{1}$$

subject to:

$$\sum_{j=1}^{n} X_{ij} = 1, \forall i = 1, \ldots, n, \tag{2a}$$

$$\sum_{j=1}^{n} X_{ij} = 1, \forall j = 1, \ldots, n, \tag{2b}$$

$$\sum_{i \in \kappa} \sum_{j \in \kappa} X_{ij} \leq |\kappa| - 1, \forall \kappa \subset \{1, \ldots, n\}, \tag{2c}$$

$$X_{ij} = 0, 1 \forall i, j, \tag{2d}$$

where C_{ij} is the cost of the path X_{ij} and κ is any nonempty proper subset of the cities $1, \ldots, m$. (1) is the objective function. The goal is the optimization of the overall cost of the tour. (2a), (2b) and (2d) fulfill the constrain of visiting each city only once. Lastly, Equation (2c) set the subsets of solutions, avoiding cycles in the tour.

TSP has an important impact on different sciences and fields, for instance in Operations Research and Theoretical Computer Science. Most problems related to those fields, are based in the TSP definition. For instance, The Hard Scheduling Optimization [5] had been derived from TSP. Although several algorithms have been proposed for the solution of TSP, there is not one that optimal solves it. For this reason, this paper discuss novel metaheuristics based on the Automata Theory in order to approach the solution of the Multi-objective Traveling Salesman Problem. This paper is structured as follows: in section 2 important definitions about the multi-objective combinatorial optimization and the metaheuristics based on the automata theory are given, section 3 discusses an evolutionary metaheuritic based on the automata theory for the multi-objective optimization of combinatorial problems, lastly, in section 4 and 5 experimental results are given for each algorithm in order to estimate their performance using multi-objective metrics from the specialized literature.

2 Preliminaries

2.1 Multi-objective Optimization

The multi-objective optimization consists in two or more objectives functions to optimize and a set of constraints. Mathematically, the multi-objective optimization model is defined as follows:

$$\textit{optimize} \quad F(X) = \{f_1(X), f_2(X), \ldots, f_n(X)\}, \tag{3}$$

subject to:

$$H(X) = 0, \tag{4a}$$

$$G(X) \leq 0, \tag{4b}$$

$$X_l \leq X \leq X_u, \tag{4c}$$

where $F(X)$ is the set of objective functions, $H(X)$ and $G(X)$ are the constraints of the problem. Lastly, X_l and X_u are the bounds for the set of variables X.

2.2 Metaheuristic of Deterministic Swapping (MODS)

Metaheuristic Of Deterministic Swapping (MODS) [4] is a local search strategy that explores the feasible solution space of combinatorial problems based on a data structure named Multi Objective Deterministic Finite Automata (MDFA) [3]. A MDFA is a Deterministic Finite Automata that allows the representation of the feasible solution space of combinatorial problems. Formally, a MDFA is defined as follows:

$$M = (Q, \Sigma, \delta, Q_0, F(X)), \tag{5}$$

where Q represents all the set of states of the automata (feasible solution space), Σ is the input alphabet that is used for δ (transition function) to explore the feasible solution space of a combinatorial problem, Q_0 contains the initial set of states (initial solutions) and $F(X)$ are the objectives to optimize. MODS explores the feasible solution space represented through a MDFA using a search direction given by an elitist set of solutions (Q_*). The elitist solution are states that, when were visited, their solution dominated at least one solution in Q_ϕ. Q_ϕ contains all the states with non-dominated solutions.

Lastly, the template algorithm of MODS is defined as follows:

1. Create the initial set of solutions Q_0 using a heuristic relative to the problem to solve.

2. Set Q_ϕ as Q_0 and Q_* as ϕ.

3. Select a random state $q \in Q_\phi$ or $q \in Q_*$

4. Explore the neighborhood of q using δ and Σ. Add to Q_ϕ the solutions found that are not dominated by elements of Q_f. In addition, add to Q_* those solutions found that dominated at least one element from Q_ϕ.

5. Check stop condition, go to 3.

2.3 Simulated Annealing Metaheuristic of Deterministic Swapping (SAMODS)

Simulated Annealing & Metaheuristic Of Deterministic Swapping [2] (SAMODS) is a hybrid local search strategy based on the MODS theory and Simulated Annealing algorithm for the multi-objective optimization of combinatorial problems. Its main propose consists in optimizing a combinatorial problem using a Search Direction and an Angle Improvement. SAMODS is based in the next Automata:

$$M = (Q, Q_0, P(q), F(X), A(n)), \tag{6}$$

Alike MODS, Q_0 is the set of initial solutions, Q is the feasible solution space, $F(X)$ are the functions of the combinatorial problem, $P(q)$ is the permutation function ($P(q) : Q \to Q$) and $A(n)$ is the weighted function ($A(n) : \mathbb{N} \to \Re^n$). n represents the number of objective for the combinatorial problem.

SAMODS exploits the search directions given by MODS and it proposed an angle direction given by the function $A(n)$. Due to this, SAMODS template is defined as follows:

1. Setting sets. Set Q_0 as the set of Initial Solutions. Set Q_ϕ and Q_* as Q_0.

2. Settings parameters. Set T as the initial temperature, n as the number of objectives of the problem and ρ as the cooler factor.

3. Setting Angle. If T is equal to 0 then got to 8, else set $T_{i+1} = \rho \times T_i$, randomly select $s \in Q_\phi$, set $W = A(n) = \{w_1, w_2, \cdots, w_n\}$ and go to 4.

4. Perturbing Solutions. Set $s' = P(s)$, add to Q_ϕ and Q_* according to the next rules:

$$Q_\phi = Q_\phi \cup \left\{s'\right\} \Leftrightarrow (\nexists r \in Q_\phi)(r \text{ is better than } s'), \tag{7a}$$

$$Q_* = Q_* \cup \left\{s'\right\} \Leftrightarrow (\exists r \in Q_*)(s' \text{ is better than } r), \tag{7b}$$

then, if Q_ϕ has at least one element that dominated to s' go to step 5, otherwise go to 7.

5. Guess with dominated solutions. Randomly generated a number $n \in [0,1]$. Set z as follows:
$$z = e^{(-(\gamma/T_i))}, \tag{8}$$
where T_i is the temperature value in moment i and γ is defined as follows:
$$\gamma = \sum_{i=1}^{n} w_i \cdot f_i(s_X) - \sum_{i=1}^{n} w_i \cdot f_i(s'_X), \tag{9}$$
where s_X is the vector X of solution s, s'_X is the vector X of solution s', w_i is the weight assigned to the function i and n is the number of objectives of the problem. If $n < z$ then set s as s' and go to 4 else go to 6.

6. Change the search direction. Randomly select a solution $s \in Q_*$ and go to 4.

7. Removing dominated solutions. Remove the dominated solutions for each set (Q_* and Q_ϕ). Go to 3.

8. Finishing. Q_ϕ has the non-dominated solutions.

2.4 Genetic Simulated Annealing Metaheuristic of Deterministic Swapping (SAGAMODS)

Simulated Annealing, Genetic Algorithm & Metaheuristic Of Deterministic Swapping [2] (SAGAMODS) is a hybrid search strategy based on the Automata Theory, Simulated Annealing and Genetics Algorithms. SAGAMODS is an extension of the SAMODS theory. It comes up as result of the next question: could SAMODS quickly avoid local optimums? Although, SAMODS avoids local optimums guessing, it can take a lot of time accepting dominated solutions for finding non-dominated. Thus, the answer to this question is based on the Evolutionary Theory. SAGAMODS proposes crossover step before SAMODS template is executed. Due to this, SAGAMODS supports to SAMODS for exploring distant regions of the solution space. Formally, SAGAMODS is based on the next automata:
$$M = (Q, Q_S, C(q,r,k), F(X)), \tag{10}$$
where Q is the feasible solutions space, Q_S is the initial solutions and $F(X)$ are the objectives of the problem. $C(q,r,k)$ is defined as follows:
$$C(q,r,k) : Q \to Q, \tag{11}$$
where $q, r \in Q$ and $k \in N$. q and r are named parents solutions and k is the cross point. Lastly, SAGAMODS template is defined as follows:

1. Setting parameters. Set Q_S as the solution set, x as the number of solutions to cross for each iteration.

2. Set Q_C (crossover set) as selection of x solutions in Q_S, Q_M (mutation set) as ϕ and k as a random value.

3. *Crossover.* For each $s_i, s_{i+1} \in Q_C / 1 \leq i < \|Q_C\| : \quad Q_M = Q_M \cup \{C(s_i, s_{i+1}, k)\}$

4. *Mutation.* Set Q_0 as Q_M. Execute SAMODS as a local search strategy.

5. Check stop conditions. Go to 2.

3 Evolutionary Metaheuristic of Deterministic Swapping (EMODS)

Evolutionary Metaheuristic of Deterministic Swapping (EMODS), is a novel framework that allows the Multiobjective Optimization of Combinatorial Problems. Its framework is based on MODS template therefore its steps are the same: create initial solutions, improve the solutions (optional) and execute the core algorithm. Unlike SAMODS and SAGAMODS, EMODS avoids the slowly convergence of Simulated Annealing's method. EMODS explores different regions from the feasible solution space and search for non-dominated solution using Tabu Search. The core algorithm is defined as follows:

1. Set θ as the maximum number of iterations, β as the maximum number of state selected in each iteration, ρ as the maximum number of perturbations by state and Q_ϕ as Q_0.

2. Randomly select a state $q \in Q_\phi$ or $q \in Q_*$.

3. *Mutation - Tabu Search* Set N as the new solutions found as result of perturbing q. Add to Q_ϕ and Q_* according to the next equations:

$$(Q_\phi = Q_\phi \cup \{q\}) \iff (\nexists r \in Q_\phi / r \text{ is better than } q) \tag{12a}$$

$$(Q_* = Q_* \cup \{q\}) \iff (\exists r \in Q_\phi / q \text{ is better than } r) \tag{12b}$$

   and then, the states with dominated solutions for each set are removed.

4. *Crossover.* Randomly, select states from Q_ϕ and Q_*. Generate a random point of cross.

5. Check stop condition, go to 3.

Step 2 and 3 support the algorithm in removing dominated solutions from the set of solutions Q_ϕ as can be seen in figure 3. However, one of the most important steps in the EMODS algorithm is 4 where new solutions are found after the crossover step.

4 Experimental Analysis

4.1 Experimental Settings

The algorithms were tested using well-known instances from the multi-objective TSP taken from TSPLIB [1]. The test of the algorithms was conducted using a dual core computer with 2 Gb RAM. The optimal solutions were constructed based on the best non-dominated solutions of all algorithms in comparison for each instance used. The instances were constructed using the combination of the mono-objective instances KROA100, KROB100, KROC100, KROD100 and KROE100. For instance, KROAB100 is a bi-objective instance whose matrices of distance are given by the instance KROA100 and KROB100. We full combine the instances (KROAB100, KROAC100, ..., KROABCDE100) and then we run the experiments. The metrics used for the measurement of the different algorithms are described below, most of them use two Pareto fronts. The first one is PF_{true} and it refers to the real optimal solutions of a combinatorial problem. The second is PF_{know} and it represents the optimal solutions found by an algorithm. In all the cases $\|\cdot\|$ represents the number of elements.

$$\begin{aligned} GNDV &= \|PF_{know}\|, & (13) \\ ReGNDV &= \|\{y | y \in PF_{know} \wedge y \in PF_{true}\}\|, & (14) \end{aligned}$$

where Generation of Non-dominated Vectors (GNDV) and Real Generation of Non-dominated Vectors (ReGNDV) measure the number of solutions and the number of true solutions found by an algorithm respectively. On the other measures the number of true solutions generated. On the other hand, Generational Distance (GD) and Inverse Generational Distance (IGD) measure the distance between FP_{know} and FP_{true}:

$$GD = \left(\frac{1}{\|PF_{know}\|}\right) \cdot \left(\sum_{i=1}^{\|PF_{know}\|} d_i\right)^{(1/p)}, IGD = \left(\frac{1}{\|PF_{true}\|}\right) \cdot \left(\sum_{i=1}^{\|PF_{know}\|} d_i\right), \tag{15}$$

where d_i is the smallest Euclidean distance between the solution i of FP_{know} and the solutions of FP_{true} and p is the dimension of the combinatorial problem. For the measurement of the range variance of neighboring solutions in PF_{know} the Spacing (S) is proposed:

$$S = \left(\frac{1}{\|PF_{know}\| - 1}\right)^2 \cdot \left(\sum_{i=1}^{\|PF_{know}\|} \left(\overline{d} - d_i\right)^2\right)^{(1/p)} \tag{16}$$

where d_i is the smallest Euclidean distance between the solution i and the rest of solutions in PF_{know}. $\overline{d} = \frac{1}{\|PF_{true}\|} \sum_{i=1}^{\|PF_{true}\|} d_i$. The Error Rate ($\varepsilon$) depicts the error rate respect to the precision of the solutions as follows:

$$\varepsilon = \left(\left|\frac{\|PF_{true}\| - \|ReGNDV\|}{\|PF_{true}\|}\right|\right) \cdot 100\% \tag{17}$$

4.2 Experimental Results

The average of the metrics applied to each algorithm are shown in table 1. Furthermore, a graphical comparison for tri-objectives instances is shown in figure 1.

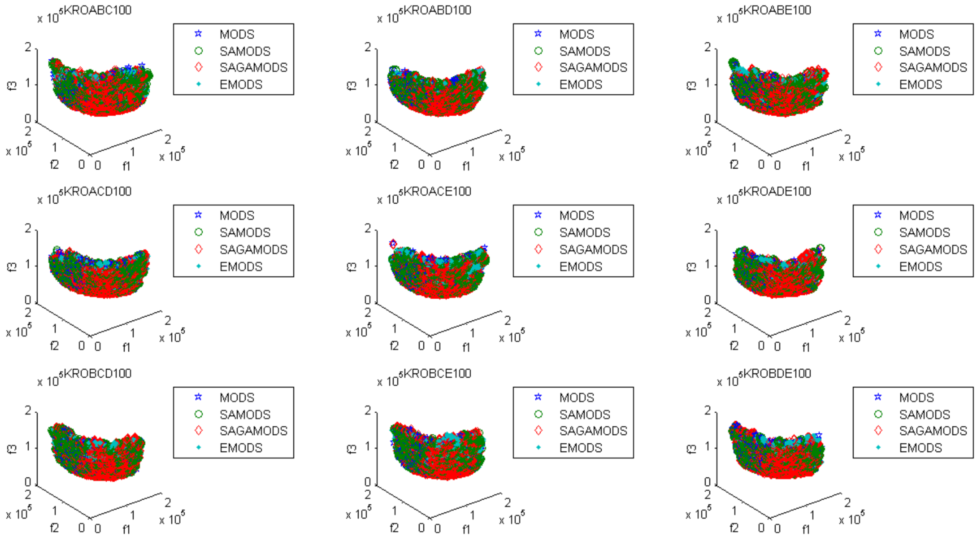


Figure 1: Graphical comparison between MODS, SAMODS, SAGAMODS and EMODS for tri-objective TSP instances.

Table 1: Average performance for the algorithms in comparison using multi-objective instances of TSP with multi-objective optimization metrics.

$INSTANCE$	$ALGORITHM$	$GNDV$	$ReGNDV$	$\left(\frac{ReGNDV}{GNDV}\right)\%$	S	GD	IGD	ε
Bi-objective TSP	MODS	262.7	0	0%	0.0286	21.6672	2329.4338	100%
	SAMODS	6487.2	1425.7	22.03%	0.0016	0.2936	265.8974	89.47%
	SAGAMODS	6554.5	1581.8	23.97%	0.0015	0.3062	286.547	88.3%
	EMODS	19758.6	10671.8	54.89%	0.0003	0.0492	75.2773	22.23%
Tri-objective TSP	MODS	1992.5	63.9	3.21%	0.1508	0.302	3206.7459	99.91%
	SAMODS	12444.2	269.3	2.16%	0.0727	0.0434	2321.5258	99.6%
	SAGAMODS	12332.5	271.1	2.2%	0.0743	0.0437	2312.3389	99.6%
	EMODS	68969.1	67097	97.3%	0.0468	0.0011	6.3914	0.89%
Quad-objective TSP	MODS	5364.8	3273.2	60.99%	0.3468	0.0252	5810.4824	94.31%
	SAMODS	27639.6	11594.2	41.94%	0.2325	0.0043	3397.7495	79.87%
	SAGAMODS	35649.6	14754.8	41.4%	0.2231	0.0032	3013.1894	74.39%
	EMODS	200420.6	27991.6	13.97%	0.176	0.0005	1891.9864	51.43%
Quint-objective TSP	MODS	7517	7517	100%	0.5728	0.0125	15705.6864	98.41%
	SAMODS	26140	26140	100%	0.4101	0.0033	10801.6382	94.46%
	SAGAMODS	26611	26611	100%	0.4097	0.0033	10544.8901	94.36%
	EMODS	411822	411822	100%	0.3136	0.0001	950.4252	12.77%

5 Conclusion

SAMODS, SAGAMODS and EMODS are algorithms based on the Automata Theory for the multi-objective optimization of combinatorial problems. All of them are derived from the MODS metaheuristic, which is inspired in the Theory of Deterministic Finite Swapping. SAMODS is a Simulated Annealing inspired Algorithm. It uses a search direction in order to optimize a set of solution (Pareto Front) through a linear combination of the objective functions. On the other hand, SAGAMODS, in addition to the advantages of SAMODS, is an Evolutionary inspired Algorithm. It implements a crossover step for exploring far regions of a solution space. Due to this, SAGAMODS tries to avoid local optimums owing to it takes a general look of the solution space. Lastly, in order to avoid slow convergence, EMODS is proposed. Unlike SAMODS and SAGAMODS, EMODS does not explore the neighborhood of a solution using Simulated Annealing, this step is done using Tabu Search. Thus, EMODS gets optimal solution faster than SAGAMODS and SAMODS. Lastly, the algorithms were tested using well known instances from TSPLIB and metrics from the specialized literature. The results shows that for instances of two, three and four objectives, the proposed algorithm has the best performance as the metrics values corroborate. For the last instance worked, quint-objective, the behavior of MODS, SAMODS and SAGAMODS tend to be the same, them have similar error rate but, EMODS has a the best performance. In all the cases, EMODS shows the best performance. However, for the last test, all the algorithms have different solutions sets of non-dominated solutions, and those form the optimal solution set.

Acknowledgment

First of all, I want to thank to God for being with me in my entire life, He made this possible. Secondly, I want to thank to my parents Elias Niño and Arely Ruiz and my sister Carmen Niño for their enormous love and support. Finally, and not less important, to thank to my beautiful wife Maria Padron and our baby Maria Gabriela for being my inspiration.

Bibliography

[1] University Of Heidelberg. Tsplib - office research group discrete optimization - university of heidelberg. http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95/.

[2] Elias D. Niño. Samods and sagamods: Novel algorithms based on the automata theory for the multi-objective optimization of combinatorial problems. *Int. J. of Artificial Intelligence - Special issue of IJAI on Metaheuristics in Artificial Intelligence*, accepted, 2012.

[3] Elias D. Niño, Carlos Ardila, Yezid Donoso, and Daladier Jabba. A novel algorithm based on deterministic finite automaton for solving the mono-objective symmetric traveling salesman problem. *Int. J. of Artificial Intelligence*, 5(A10):101-108, 2010.

[4] Elias D. Niño, Carlos Ardila, Yezid Donoso, Daladier Jabba, and Agustin Barrios. Mods: A novel metaheuristic of deterministic swapping for the multi objective optimization of combinatorials problems. *Computer Technology and Application*, 2(4):280-292, 2011.

[5] Elias D. Niño, Carlos Ardila, Adolfo Perez, and Yezid Donoso. A genetic algorithm for multiobjective hard scheduling optimization. *INT J COMPUT COMMUN*, 5(5):825-836, 2010.

[6] J.G. Sauer and L. Coelho. Discrete differential evolution with local search to solve the traveling salesman problem: Fundamentals and case studies. In *Cybernetic Intelligent Systems, 2008. CIS 2008. 7th IEEE International Conference on*, pages 1-6, 2008.

[7] Yang Xiawen and Shi Yu. A real-coded quantum clone multi-objective evolutionary algorithm. In *Consumer Electronics, Communications and Networks (CECNet), 2011 International Conference on*, 4683-4687, 2011.

[8] Qin Yong-Fa and Zhao Ming-Yang. Research on a new multiobjective combinatorial optimization algorithm. In *Robotics and Biomimetics, 2004. ROBIO 2004. IEEE International Conference on*, 187-191, 2004.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):145-153, December, 2012.

Decision Support for Healthcare ICT Network System Appraisal

A.M. Oddershede, L.E. Quezada, F.M. Cordova, R.A. Carrasco

Astrid M. Oddershede, Luis E. Quezada,
Felisa M. Cordova
University of Santiago of Chile,
Industrial Engineering Department
Av. Ecuador 3769, Santiago, Chile
E-mail: astrid.oddershede@usach.cl,
luis.quezada@usach.cl, felisa.cordova@usach.cl

Rolando A. Carrasco
Newcastle University, UK, School of IEEE
Newcastle Upon Tyne, UK
E-mail: r.carrasco@ncl.ac.uk

Abstract:
A framework to support the appraisal process to improve the quality of service (QoS) of an Information and Communication Technology (ICT) network system in health care service is presented. Most of health-related activities stand to benefit from ICT endorsement; however, technical problems may appear, as an inadequate physical infrastructure, insufficient access by the user to the hardware/software communication infrastructure and QoS issues. The aim is to develop a prototype assessment model based on data collected from the main users of a health network system An evaluation process is carried out to analyze and assess the support of QoS of ICT, its infrastructure and user interface perception of the QoS offered through case study for hospitals in Chile. Performance has been evaluated by simulation and modelling network Architecture. The Optimization Network Engineering Tool (OPNET) simulation platform is used to examine the network behaviour and performance to ensure consistency and reliability for thousands of staff across the hospital network.
Keywords: ICT, Healthcare, OPNET, MCDM.

1 Introduction

Nowadays, Healthcare Institutions are universally urged to improve quality of their services and Hospital units are concentrated on how to develop better services, to allocate methods, to provide resources to satisfy professional aspirations and to comply with citizen necessities. Most of the actions oriented to improve the operation and the quality of healthcare service depends, to a great extent, on the level of information available and the communications system. A poor ICT network system implementation may generate a negative effect on the service, patients and health care providers.

Furthermore, the ICT system from a health centre may be endowed by attributes that not often matches with the user's requirement s (quality level, performance, cost and others). The ICT system may be inappropriate for satisfying user requirements or may be inefficient in doing it. ICT system should be a facilitator for health care users since they need to access to all types of data existing on all types of systems.

The challenge of providing quality of service (QoS) in a health environment is further complicated by the extremely variable QoS needs of individual health organizations over time. The kinds of information exchanges in which an organization engages typically vary considerably in the course of a day, from simple exchanges of information regarding a patient's coverage by a health plan, through transfers of medical records with affiliated organizations, to the exchange of large medical images for interpretation and diagnosis. For example, the bandwidth needs of a small medical clinic could, accordingly, vary enormously during the course of a day, ranging from near nothing one minute to several megabits per second the next. Finding ways to satisfy such variable demand for bandwidth economically represents a significant challenge.

The agents (patient, clinical doctors, hospital staff, government, etc) involved should attend contradictory petitions: on one hand, to respond the increasing demand and, by another, to attend to the budgetary restrictions.

The answer to this complexity is to improve QoS and the efficiency [1] [2] . User needs and expectations are indispensable to count on better information that lead to an improvement in operation services, welfare and clinical information. Reorienting the Health care system towards the needs of the client, medicine based on the evidence, clinical management and the continuous improvement of the processes (QoS), Total quality, contemplating the human factor as the main assets of the business) are key elements in future.

The hospital management challenge is based on the achievement of the articulation and the convergence of values among the agents, the staff members and the patients; the Information Technology and Communications Systems needed are not just mere data agents, instead, they are of information, knowledge and intellectual capital [3] Some general questions to be answered are: What are user expectations about the ICT needs? Is the current ICT infrastructure appropriate? What are the main QoS parameters? How can be evaluated the ICT network for Healthcare service? What technical capabilities do health applications demand of the Internet? How do these capabilities differ from those needed by applications in other sectors, such as banking, defense, and entertainment?.

To deal with this complexity, a two stages methodology involving user perception is proposed. The first stage corresponds to the identification of the user types, the activities they are involved in, the ICT network system requirements, applications and priorities attributes. This is achieved through Multicriteria Decision Making approach (MCDM) [4], using the Analytical Hierarchy Process (AHP) in author previous work [8] obtaining key QoS parameters.

This paper focuses on the second stage, which is the development of a model for evaluating ICT healthcare network quality of service (QoS) integrating user perception. This stage is concerned with modelling and simulation to examine the network performance on applications. The QoS technical metrics related to each attribute has to be defined together with the applications profile. QoS offered by a particular network could be established by technical parameters that can be measured objectively. However, user perception depends upon their needs, their precise applications and their expectations. It is a difficult task to find a set of universal parameters for every type of service because there are many and dissimilar parameters involved in the performance evaluation.

A typical public hospital zone is considered as a case study to analyze technology infrastructure and network performance. Further, profile applications must be set up according to the main role health care users perform. A new model based on the results obtained on the first stage is designed and OPNET simulation platform is used to examine the network behaviour and performance .The model incorporates communications resources of LAN architecture from typical Chilean hospitals where. OPNET model has demonstrated to give a good representation to real world to analyze traffic flows, and network performance providing a tool to demonstrate different type of networks and protocols.

This pilot case study revealed that the model shows to be useful by evaluating quality, user pertinent criteria and to connect higher objectives with lower performance metrics and conclude that through the proposed assessment model is possible to detect whether there is connection between human perception of QoS and the technical metrics associated to the ICT network. This analysis helps decision makers, network planners and operations engineers to manage complex and constantly changing networks, using predictive planning for reacting to significant network issues using real-time network visualization and troubleshooting. This is critical in a healthcare institution.

Section 2 provides a description of the conceptual model, in section 3 the case study description is presented .The simulation results generate new information, and section 5 provides the conclusions.

2 Conceptual Model for Healthcare Service ICT Network

A conceptual model is proposed for health ICT network system evaluation connecting QoS user perception (qualitative) with QoS technical aspects, (parameters, and network performance). Modeling and simulation provides a good tool, to plan network infrastructure and manage application performance using predictive planning. The model scheme is depicted through Figure 1 illustrating the diverse factors involved in the evaluation of QoS for a particular service starting from the end user. Initially, the main users, the type of user and its applications must be categorized , then the ICT support needs to deliver a better service, the attributes indispensable to meet the requirement in concordance to their expectation, are to be characterized , next , technical metrics should be defined and analyzed to check the performance on an application [5].

The perceived QoS cluster includes the parameters related to each service the user perceive and determine the satisfaction of the service received. For example: success in the connection, accessibility, velocity, etc. The technical ICT parameters refers to the basic metrics that would guarantee a service, i.e., end to end delay, packet loss and should be detected by the network operators in charge.

QoS offered by a particular network could be established by technical parameters that could be measured objectively. However the user perception depends upon their needs, their directly application and their expectations. It is a difficult task to find a set of universal parameters for every type of service for the reason that there are many and dissimilar parameters involved in the performance evaluation. Then, it is useful to analyze which of the parameters is relevant when considering the user perception for a determined service. The organization must then define a service level agreement (SLA) for their main applications.

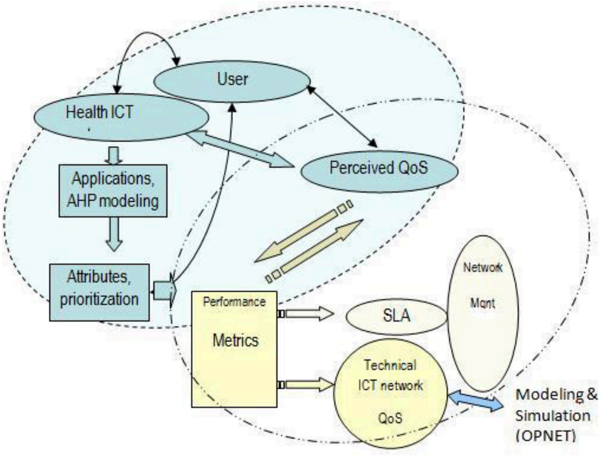


Figure 1: Proposed Conceptual model for health ICT network evaluation

For the case in study, we suggest modeling and simulation using Optimization Network Engineering Tool (OPNET) simulation platform, (www.opnet.com). [6], [7] to examine the network behavior and performance when it is not possible to compute through the operators the basic metrics. OPNET supports simulation technologies and it is well suited to examining network behavior.

The OPNET models execute the protocol in much the same way as a production environment. The modeling it is based on Cisco network which is 99% accurate and good representation of real world. [8] It is possible to study protocol behavior under different network conditions and application performance.

3 Hospital ICT Network Scenarios Simulations with OPNET

A pilot case study is pursued following the methodology proposed, where the results can be used for analyzing the network performance en ICT healthcare system, when it is not possible to measure directly, the performance metrics, and parameters.

According to OPNET methodology, the initial phase for analyzing the effect of QoS on an ICT network and/or on application performance is to select a network topology, followed by traffic configuration for each application before running simulations. Then QoS has to be configured defining SLAs to analyze the results.

3.1 Hospital Zone Case Study

The authors have collected information from a Public Community hospital/medical centre/-clinic Hospital Dr. Luis Tisne Brousse in Santiago Chile. A Local Ethernet network from a zone of the medical centre is considered as an initial pilot study case. In view of the fact that the purpose is to examine the backbone utilization, the study zone is represented as a partial network topology along with the methodology.

Hospital Zone Description

The backbone Giga Bit Ethernet in study is located at a building next to the hospital where they perform administrative functions to assist the hospital and some health service functions.

A diagram showing the zone in study is given in Figure 2.

The section in study is a two floor building, plus a basement named: "Zócalo". Within the first floor, there are rooms used for providing diagnosis for minor ailments and medical consultation. This first floor includes 3 switches:

- WS2950G-24, switch, located at the Electric room, with five workstations,
- WS2950-24, switch located at the medicine/Physiatrist room, with six work stations,
- WS2950G-12 located at the X ray room, with 3 workstations.
- The basement floor includes a WS2950G-24, Switch with 24 ports and there are 6 workstations connected,
- The first floor switches are connected to the second floor to a WS2950G-24 switch located at the Director office by fibber optical cable. The switches are connected to and between each other by Optic fiber, Except for the X ray room switch which is a cable UTP cat. 5E.

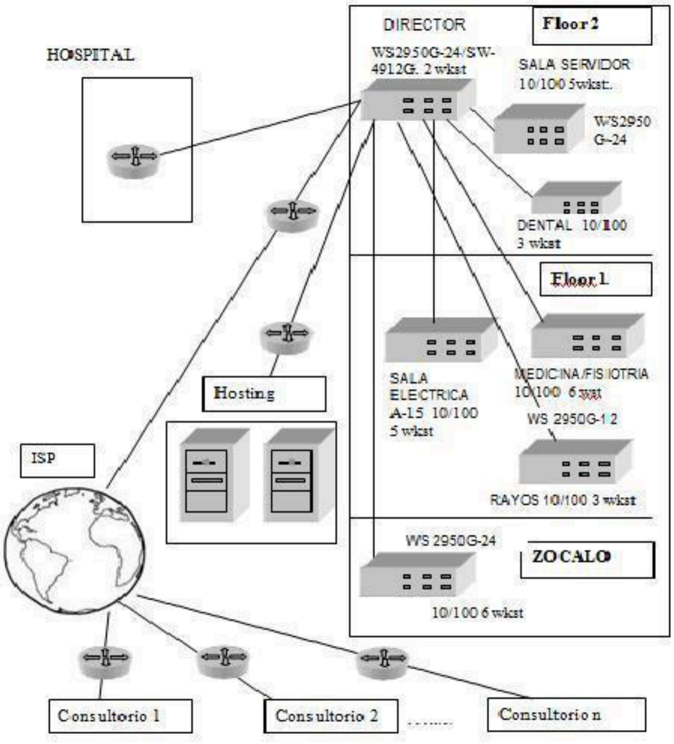


Figure 2: Hospital zone diagram

3.2 OPNET Network Modeling for Public Hospital Zone Case Study

The authors have designed network modeling and implementation for the public Hospital stated in conformity with the diagram described on figure 4. The partial network topology is represented following OPNET methodology in Figure 3.

The partial network characterizes the initial situation and first scenario to be explored. A total of 30 workstations are initially linked to the switches. The different services the users deliver are expected to be supported by the network through different applications that will help to accomplish their assignment. The applications regard as sustained by the server are based on the results obtained on the previous study.

3.3 Health ICT Network Applications

Communications networks enable applications to exchange data. Popular applications that use data networks include virtual terminal services, file transfer utilities, database transactions, and e-mail. Each of these applications generates its own sort of traffic: virtual terminals slowly generate many small packets, while file transfer utilities send long streams of large packets.

Each type of traffic causes and experiences a different set of problems in the underlying network, so you may want to accurately model the traffic patterns generated by a variety of applications. OPNET uses a generic network application model to generate typical application traffic patterns. This is the applications model, also called the standard network application model.. Depending on their underlying networks, application architectures may differ [7].

The applications are modeled explicitly, and end-to-end delays or response times are studied in detail. The factors that contribute to application response time include,:

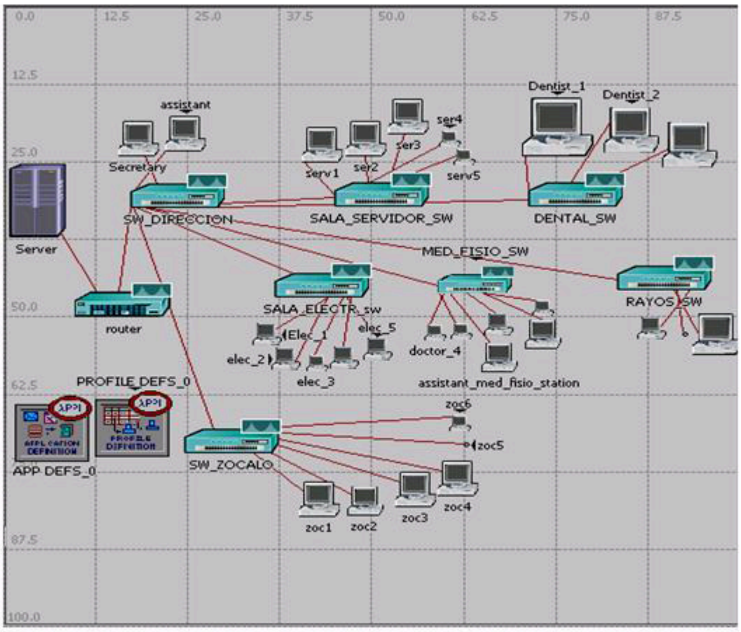


Figure 3: Partial network topology

- Delays due to contention on servers (server processing time),or/and
- Delays due to contention on the client (client processing time),
- Delays due to contention with "other" traffic at the various intermediate devices (queuing delays),
- Delays due to contention with traffic of the same application type from other users at the intermediate devices (queuing delays),
- Network delays (transmission and propagation),
- Delays due to protocol effects (TCP retransmissions, windowing etc.)

In the course of the investigation and survey developed on a previous work [9], [10], the findings revealed that the relative **usage** of ICT applications in health care differs to some extent depending on the institution. Indicating that, so far, the real application that has more usage is e- mail and there is very little **usage** of the others. Web browsing is used mainly on research activities.

However, through the multicriterial analysis by the use of AHP, to find out the importance of ICT provision in quality of service in healthcare institutions developed in[10], one of the important new information, conclusions and contribution are the relative importance users assigned to the applications in performing a health care service. User assigned the highest priority in importance to data base access and the least relative importance to e-mail. This shows the existence of a gap between what healthcare participants actively use and what they consider important for the development of their daily work as a service for the patients.

Then, considering these results the applications to be supported by the server for every scenario are: Email, Web Browsing (Http 1.1), File transfer, Database Access, File Print, Video Conferencing, and Voice.

3.4 Applications Configuration

A profile is applied to each workstation, server, or LAN. It specifies the applications used by a particular group of users. An application may be any of the common applications, email, file transfer, etc., that may be defined. The next step is to set up profile applications according to each user type. Every workstation will have a profile application consistent with the users' main role.

After the participants have studied the nature of the system, then profiles are designed with respect to the functions a user makes use of in/at a health service. This profile name is assigned for classification. These applications profiles were defined using the results from the analytic hierarchy process.

Once applications and profiles are defined, it is possible to characterize different scenarios for each study case intended for visualizing how sensible is a networks performance with respect to changes.

4 Simulation Results

Different scenarios for the study case are characterized to determine the sensitivity of main key parameters (as throughput and utilization). Initially, the applications described are modeled explicitly and all background traffic is disabled. Simulations are ran for each scenario, increasing the number of users and their respective applications. Through Figure 4 it is possible to visualize overlaid run results for delay average (in Ethernet. delay (secs) for four scenarios. As the number of users is increased, the average delay time increases at different rates, until it tends to be stabilized.

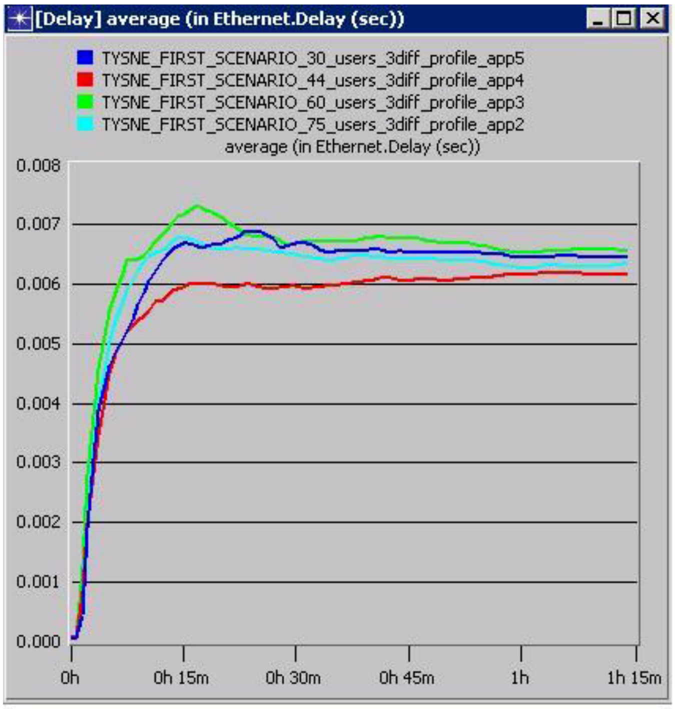


Figure 4: [Delay] average [in Ethernet. Delay (sec)] for four overlaid scenarios.

The simulations scenarios are run allowing the same probability distribution for the start time. The difference in average delay for the scenarios is due to contention among the network users on all the intermediate devices and links. Figure 5 gives a picture of the delay variation in seconds as the number of users is augmented.

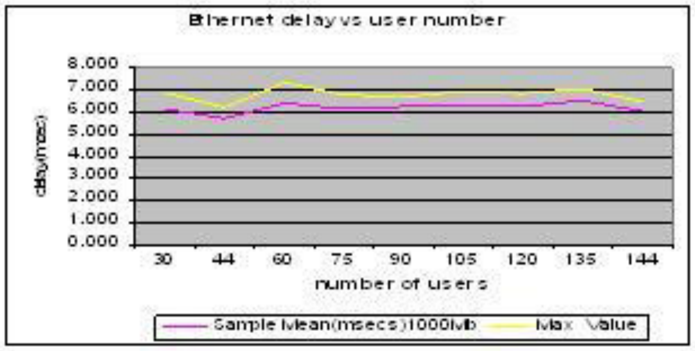


Figure 5: Ethernet average delay vs number of users.

After running simulations for different scenarios and for all the applications considered, increasing number of users, incorporating new number of applications and devices, varying traffic and others we could observe that for this hospital zone ICT network infrastructure there are no problems concerning channel utilization.

Ethernet delay showed to be low with a high throughput. Then, the network prepared to support a greater number of users, more and new applications, as voice, images, video and others.

5 Conclusions

Through the case studied, we observe for data transmission that there is no connection with human perception. Even though from human perspective showed discontent about the access and ubiquity. QoS parameters are good.

However, users complain about the straightforward access, new applications and number of computers. In this respect, it is possible to conclude that if the delay is low and throughput is high, there should not be any availability problem, but one of the reasons could be caused of unbalanced resources distribution, some resource policy distribution, or other.

From the various simulations results we conclude that the model is able to detect whether there is connection between human perception of QoS and the technical metrics associated to the ICT network.

From the technology aspects perspective there is a big potential forthcoming for gaining benefit from of ICT support in healthcare service that has to be analyzed also from financial perspective to optimise the available resources. An increase in the number of workstations, increasing the number of new applications, preparing and training people in informatics and technology is a tactic that contributes and leads to increase the usage of technology in the Healthcare sector, mainly in rural communities.

Bibliography

[1] Elske Ammenwertha,*, Stefan Gräberb, Gabriele Herrmannc, Thomas Bürkled, Jochem Königb, Evaluation of health information systems-problems and challenges , INT JOURNAL OF MEDICAL INFORMATICS 71, 125-135, 2010.

[2] E. Babulak, Quality of service provision assessment in the healthcare information and telecommunications infrastructures, *INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS*, 75(3):246-252, 2006.

[3] Networking Health: *Prescriptions for the Internet, Computer Science and Telecommunications Board*, National Academic press, ISBN-10: 0-309-06843 , 2000.

[4] Triantaphyllou, E.. " Multi-Criteria Decision Making Methods: A comparative Study" (Applied Optimization, Volume 44) Nov 2000

[5] Sun, L., and Ousmanou, K., Articulation of Information Requirements for Personalized Knowledge Construction, JOURNAL OF REQUIREMENTS ENGINEERING, 11(4):279-293, 2006.

[6] Optimized Network Engineering Tool (OPNET). www.opnet.com, Opnet: User's Manual, http://www.opnet.com/university_program/teaching_with_opnet
/textbooks_and_materials/materials/OPNET_Modeler_Manual.pdf, 2004.

[7] OPNET documentation V.11.O.A, OPNET Technologies, Inc., Bethesda, MD, 2004.

[8] Heath A., Carrasco R.(2001), Access techniques for 3G multimedia wireless packet switched networks: simulation using OPNET, *IEE/IEEE/BCS 6th International Symposium on Communication Theory and Applications (ISCTA01)*, Lancaster University, 15-20 Jul 2001.

[9] Oddershede, A.M, Carrasco, R.A E. Barham. Multicriteria Decision Model for Assessing Health Service Information Technology Network Support using AHP, *IBEROAMERICAN JOURNAL OF COMPUTING*, Computacion y Sistemas, ISSN 1405-5546, 12 (2): 173-182, 2008,

[10] Oddershede, A.M, Carrasco, R.A, Information and Communications Technology Significance in Health Care: User Perception, *MEDITERRANEAN JOURNAL OF ELECTRONICS AND COMMUNICATIONS*, 2(2): 82-89, 2006.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):154-167, December, 2012.

Mining Association Rules from Empirical Data in the Domain of Education

D. Radosav, E. Brtka, V. Brtka

Dragica Radosav, Eleonora Brtka,
Vladimir Brtka
University of Novi Sad
Technical Faculty "Mihajlo Pupin"
Serbia, 23000 Zrenjanin, Djure Djakovica bb
E-mail: radosav@tfzr.uns.ac.rs,
brtka@sbb.rs, vbrtka@tfzr.uns.ac.rs

Abstract:
The data mining techniques and their applications are widely recognized as powerful tools in various domains. In the domain of education there is a variety of data of various types that are collected. The important question is: Is it possible to process collected data with the data mining technique and what are main advantages of data mining and e-learning interaction? If an e-learning system accumulates a huge volume of data, then it is possible to deploy techniques and tools from the domain of data mining in order to gain valuable information. The research presented in this paper is conducted on real-life data that originates from the Balkan region. The software system Weka is used to generate association rules. The main result of this research is the assessment of the parameters that are associated with the opinion that computer skills will be helpful in the future, from the students point of view. This result is very important because it gives the exact insight to computer technology usage in the Balkans schools. Furthermore, some advantages of the usage of data mining techniques in the domain of education are determined.

Keywords: association rules, education, data mining.

1 Introduction

In the past few years e-learning techniques have significantly improved as the result of progress and increased use of the Internet. The "desktop" e-learning systems, in many cases, have been replaced by systems that operate using the Internet. Some of the web-based systems allow the determination of preferences for each participant in the process and adjustment of activities in accordance with the profile of participants [1]. Recently, techniques from the domain of data mining have been incorporated into systems for e-learning. The changes that are constantly taking place in terms of rapid technical and technological developments affect society as a whole. The educational system is experiencing changes in terms of modernization and globalization. However, the educational system that is "inert" is not suitable for rapid change and modernization. The educational processes in Serbia and some other countries in the region of western Balkans are changing so that the "reproduction" style of learning is replaced with the style that prefers "understanding" of the learning content and usage of the acquired knowledge. The theories of learning that are used are no longer associative and behavioral, but have become constructive and cognitive. Students are required to improve the style of self-training and their skills. In these processes the information capacity is an important factor in development, especially the Internet and the resources that are available through World Wide Web. The usage of the Internet by some Course Management System (CMS) is not unusual occurrence in Serbia, but cannot be said that such systems are widely present.

Efforts that have been invested in the integration of CMS and Data Mining (DM) systems are evident. This integration often means adding DM modules to the existing CMS [1, 2], but it is possible to approach to CMS and DM system integration through serial connection [3, 4]. Serial connection, in this case, means collecting data with CMS, and then processing collected data by DM system. The results of DM analysis are fed back to CMS in order to improve their effectiveness.

However, this paper lists the basic DM techniques and some tools that allow the application of these techniques in domain of education, but CMS and their application is not the topic of this paper, although the importance of CMS is evident. This paper deals with some special DM techniques when applied to data collected in the domain of education. The application of DM techniques results in some rules or patterns that can be used as feedback to CMS. Rather than investigation of the connection between CMS and DM system, this paper deals with DM techniques when applied to data in the domain of education and gives some conclusions and remarks about the importance of inferred knowledge. Special contribution is the analysis of the results of DM technique when applied to real-life data collected from the region of Serbia and Bosnia and Herzegovina.

The paper is organized as follows: section two gives the short description of various DM techniques used in the domain of education. One of DM techniques is chosen to be used for the analysis of the real-life data. It is explained why this DM technique is the most suitable in this particular case. Section three contains data description, as well as the methodology description. Section four lists the results obtained by application of the DM technique, as well as the interpretation of these results. Finally, section five contains conclusions and remarks about applicability of DM techniques in the domain of education.

2 Previous Work and DM Techniques

The well-known CMS that are in use are: Blackboard, WebCT, ANGEL and Moodle. In previous papers DM techniques are used as a functional element (module) of CMS [2, 5]. The DM module can be an integral part of CMS but, in some cases, can be used separately. The field of data mining, like statistics, concerns itself with "learning from data" or "turning data into information" [6]. According to [5, 7], DM can be defined as the intersection of the domain of statistics, computer science, artificial intelligence, machine learning, database management and data visualization. Data mining is the process of identifying valid, novel, potentially useful, and ultimately comprehensible and understandable patterns or models.

The basic techniques of data mining are [7]:

- *Classification* - examining the feature of a newly presented object and assigning it to a predefined set of classes.
- *Affinity grouping or association rules* - determining which things go together, also known as dependency modeling.
- *Clustering* - segmenting a population into a number of subgroups or clusters. Description and visualization - exploratory or visual data mining.

2.1 The Application of the DM

In practice, there are a lot of general and specific data mining tools [2]. The commercial mining tools are: DBMiner [8], SPSS Clementine: [9] and DB2 Intelligent Miner [10], etc. Some public domain mining tools are: Weka [11] and Keel [12]. There are also specific educational data mining tools such as: Mining tool [13] for association and pattern mining, MultiStar [14] for association and classification, KAON [15] for clustering and text mining and CIECoF [16] for association rule mining. The application of these systems in the field of education includes the selection of suitable DM techniques. It is not unusual that multiple DM techniques are applied to the same data sample.

As in [2, 7] the main steps of application of DM techniques are:

1. Collection of the data. The CMS system is used and the collected data are stored in database. This step can be executed by a questionnaire or some other data collection technique instead of CMS usage.
2. Preprocessing the data. The data is "cleaned" and transformed into an appropriate format to be mined.
3. Application of suitable DM technique. The DM technique is applied to build the model that discovers new rules, patterns and knowledge. To execute this step, either a general or a specific data mining tool, or a commercial or a free data mining tool can be used.
4. The interpretation, evaluation and deployment of the results.

In particular, it is necessary to apply and elaborate in detail each of these steps depending on the data to be analyzed. Some of the systems for data mining that are used to analyze data from different domains are:

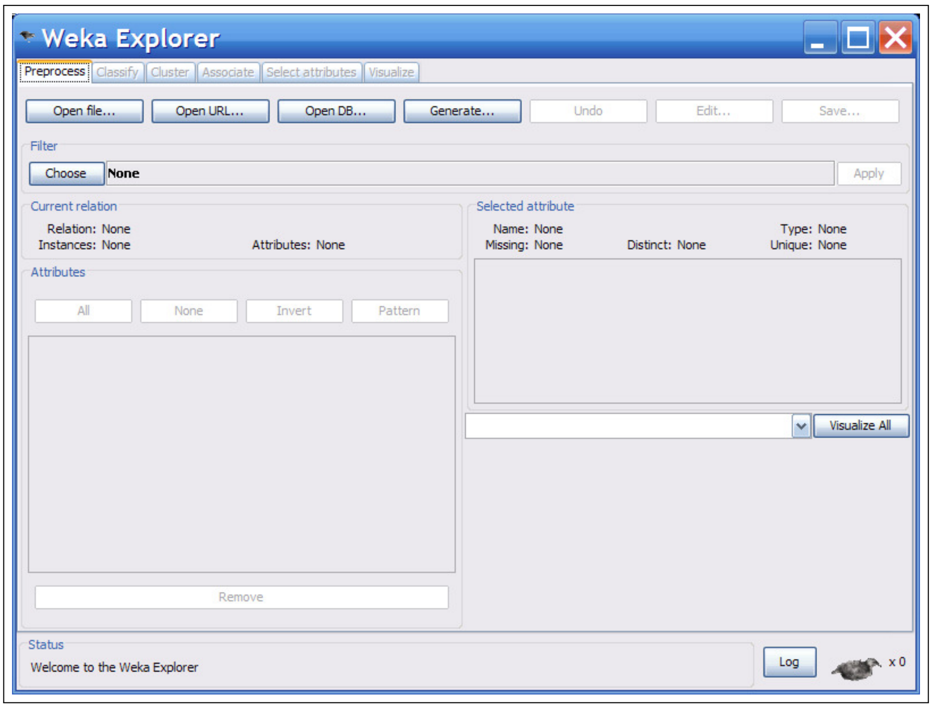
- **Rosetta** system, developed by researchers from the University of Warsaw and the University of Trondheim [17, 18]. Rosetta is capable of synthesis of the IF ... THEN rules by usage of the Rough sets theory. This system is based on classification, reduction of data and decision rules synthesis.
- **Weka** system was developed by researchers from the University of Waikato, New Zealand [19].

Rosetta system allows data to be loaded from MS Excel table; the format of the loaded data can also be CSV (Comma Separated Values). Rosetta system performs the extraction of the IF...THEN rules. The data can be collected by various methods; the format of the collected data does not have to be specially adapted to DM techniques implemented in Rosetta.

On the other hand, the Weka GUI Chooser provides a starting point for launching Wekas main GUI applications and supporting tools. The Weka system can be used to start the particular DM applications:

- Explorer - An environment for exploring data with Weka.
- Experimenter - An environment for performing experiments and conducting statistical tests between learning schemes.
- KnowledgeFlow - This application supports essentially the same functions as the Explorer but with a drag-and-drop interface. One advantage is that it supports incremental learning.
- SimpleCLI - Provides a simple command-line interface that allows direct execution of Weka commands for operating systems that do not provide their own command line interface.

System Weka allows (see Figure 1):



Filter: None

Relation: None
Instances: None
Attributes: None

Name: None
Type: None
Missing: None
Distinct: None
Unique: None

Welcome to the Weka Explorer

Figure 1: The main menu of Weka system

- **Classification** (*Classify*) - A classifier is a mapping from a (discrete or continuous) feature space X to a discrete set of labels Y [20]. Classification or discriminant analysis predicts class labels. This is supervised classification which provides a collection of labeled pre-classified patterns; the problem being to label a newly encountered, still unlabeled, pattern. In e-learning, classification has been used for: discovering potential student groups with similar characteristics and reactions to a specific pedagogical strategy [21]; predicting students performance and their final grade [22]; detecting students misuse or students playing around [23]; predicting the students performance, as well as assessing the relevance of the attributes involved [24]; grouping students as hint-driven or failure-driven and finding students common misconceptions [25]; identifying learners with little motivation and finding remedial actions in order to lower drop-out rates [26]; for predicting course success [27].

- **Clustering** (*Cluster*) - Clustering is a process of grouping objects into classes of similar objects [28]. It is an unsupervised classification or partitioning of patterns into groups or subsets (clusters) based on their locality and connectivity within an n-dimensional space. In e-learning, clustering has been used for: finding clusters of students with similar learning characteristics, and for promoting group-based collaborative learning, as well as for providing incremental learner diagnosis [29]; grouping students and personalized itineraries for courses based on learning objects [30]; grouping students in order to give them differentiated guiding according to their skills and other characteristics [31]; grouping tests and questions into related groups based on the data in the score matrix [32].

- **Association rule mining** (*Associate*) - Association rule mining discovers relationships among attributes in databases, producing if-then statements concerning attribute-values [33]. An association rule expresses a close correlation between items (attribute-value) in a database with values of support and confidence. The confidence of the rule is the percentage of transactions that contains the consequence in transactions that contain the antecedent. The support of the rule is the percentage of transactions that contains both antecedent and consequence in all transactions in the database. Association rule mining has been applied to web-based educational systems for: building recommender agents that could recommend on-line learning activities or shortcuts [34]; diagnosing student learning problems and offering students advice [35]; guiding the learners activities automatically and recommending learning materials [36]; determining which learning materials are the most suitable to be recommended to the user [37]; identifying attributes characterizing patterns of performance disparity between various groups of students [38]; discovering interesting relationships from students usage information in order to provide feedback to the author of the course [39]; finding out relationships in learners behavior patterns [40]; finding students mistakes that often accompany each other [41]; guiding the search for the best fitting transfer models of student learning [42]; and optimizing the content of the e-learning portal by determining what most interests the user [43].

- **Selecting Attributes** (*Select attributes*) - Attribute selection involves searching through all possible combinations of attributes in the data to find which subset of attributes works best for prediction. To do this, two objects must be set up: an attribute evaluator and a search method. The evaluator determines what method is used to assign a value to each subset of attributes. The search method determines what style of search is performed.

- **Visualization** (*Visualize*) - Information visualization [44] is a branch of computer graphics and user interface which is concerned with the presentation of interactive or animated digital images so that users can understand data. These techniques facilitate analysis of large amounts of information by representing the data in some visual display. Weka visualization section allows visualizing 2D plots of the current relation.

Rosetta system and Weka are particularly suitable for data analysis in the field of education because they offer a selection of DM techniques and are relatively easy to use.

3 Methodology

The data sample, in form of MS Excel document, consists of a total of 256 instances (students). It is important to mention that the data are not collected with the aim to be analyzed by DM techniques. The survey was conducted on students from the territory of the Republic of Serbia and the territory of Bosnia and Herzegovina. The computer technology that is used in this region is mostly out of date but is sufficient for elementary usage in education. Description of the data, presented in Table 1, shows the names of attributes and their possible values. Attribute names and associated values comply with the form of survey.

Various techniques can be used on this small data set: First of all, there are statistical techniques, for example students distribution that is used when estimating the mean of a normally distributed population when the data set is small; then there are techniques for inferring decision rules (based on Pawlaks rough sets theory, decision trees, etc.), even neural networks can be trained on small data set [45]. In addition, the data sample which is described in Table 1 can be analyzed by various DM techniques or DM systems. The association rule mining is adopted as the most suitable DM technique.

As defined by Agrawal et al. [33] the problem of association rule mining is defined as:

Let $U = \{u_1, u_2, ..., u_m\}$ be a discrete universe, a finite set of objects. Let $A = \{a_1, a_2, ..., a_n\}$ be a finite set of attributes with binary values. Each object of universe U is described by attributes $a_i,\ i = 1, 2, ..., n$ thus generating a data set. An associative rule is defined as an implication of the form $X \Rightarrow Y$ where $X, Y \in A$ and $X \cap Y \neq \oslash$. The set of attributes X is

Table 1: Description of used data, attribute names and their possible values

Name	Value, number of objects and distribution	Description of the attribute
A1	1. yes, 210 , 82.03125% 2. no, 46, 17.96875%	Does the student has a computer at home
A2	1. do not know, 31, 12.109375% 2. other, 7, 2.734375% 3. PII, 7, 2.734376% 4. PIII, 60, 23.4375% 5. PIV, 144, 56.25% 6. laptop, 7, 2.734376%	What type of computer do the student have
A3	1. yes, 123, 48.046875% 2. no, 133, 51.953125%	Does the student use the Internet
A4	1. 1 hour per day, 98, 38.28125% 2. 2 hours per day, 85, 33.203125% 3. 3 hours a day, 23, 8.984375% 4. 4 hours a day, 24, 9.375% 5. 5 hours a day, 26, 10.15625%	How many hours per day does the student use the computer
A5	1. 0 hours a day, 67, 26.171875% 2. 1 hour per day, 74, 28.90625% 3. 2 hours per day, 64, 25.0% 4. 3 hours a day, 18, 7.03125% 5. 4 hours a day, 12, 4.6875% 6. 5 hours a day, 21, 8.203125%	How many hours per day the student use the Internet
A6	1. yes, 89, 34.765625% 2. no, 167, 65.234375%	Does the student use e-mail
A7	1. web sites on Serbian, 142, 55.46875% 2. web sites on other lenguages, 114, 44.53125%	What web sites does the student visit most frequently
A8	1. educational, 27, 10.546875% 2. entertainment, 116, 45.3125% 3. other, 113, 44.140625%	What kind of web sites is the most visited by student
A9	1. yes, 185, 72.265625% 2. no, 71, 27.734375%	Does the student use his/her home computer for learning
A10	1. educational, 42, 16.40625% 2. film, 18, 7.03125% 3. music, 130, 50.78125% 4. other, 66, 25.78225%	What type of media does the student use most frequently at home
A11	1. yes, 159, 62.109375% 2. no, 5, 1.953125% 3. do not know, 92, 35.9375%	Does the student want more educational computer software to be used in school in order to improve teaching
A12	1. yes, 192, 75.0% 2. no, 30, 11.71875% 3. do not know, 34, 13.28125%	Does the student want to review learning materials used in school, by Distant Learning System at home
A13	1. games, 78, 30.46875% 2. educational, 11, 4.296875% 3. games, educational, 167, 65.234375%	What type of software is most frequently used by student
A14	1. yes, 221, 86.328125% 2. no, 35, 13.671875%	Does the student like the subject of informatics
A15	1. yes, 219, 85.546875% 2. no, 37, 14.453125%	Does the student believe that he/she gained enough knowledge to work independently on a computer
A16	1. independently, 82, 32.03125% 2. from others, 36, 14.0625% 3. at school, 138, 53.90625%	From who or where has the student learned most about computer usage
A17	1. yes, 235, 91.796875% 2. no, 21, 8.203125%	Does the student think that his/her computer skills will help him/her in the future

Table 2: Simple example of data set

Object	The student has a computer at home (A1)	The student uses the Internet (A2)	The student thinks that his/her computer skills will help him/her in the future (A3)
1.	yes	yes	no
2.	no	yes	yes
3.	no	no	no
4.	yes	yes	yes
5.	no	yes	no

called antecedent (left-hand-side or LHS) of the rule; the set of attributes Y is called consequent (right-hand-side or RHS) of the rule.

There are many rules of the form $X \Rightarrow Y$, but to select interesting rules from the set of all possible rules, various measures of significance can be used; the best-known are minimum thresholds on support and confidence. The support $supp(X)$ is defined as the proportion of objects in the data set which contains the attributes from X. The confidence of a rule is defined as:

$$conf(X \Rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)}$$

The previous concepts are explained in next simple example. For a data set given in Table 2 it is possible to infer some association rules, as well as the confidence and support parameters.

For $X = \{A1, A2, A3\}$ $supp(X) = \frac{1}{5} = 0.2$ because there is one object (number four) for which there is a "yes" value for every attribute.

For example, the confidence of the rule $X \Rightarrow Y$, where $X = \{A1, A2\}$ and $Y = \{A3\}$ is:

$$conf(X \Rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)} = \frac{0.2}{0.4} = 0.5$$

The previous rule $X \Rightarrow Y$ is interpreted as follows: The student who has a computer at home and uses the Internet is associated with the opinion that his/her computer skills will help him/her in the future.

There are many algorithms for association rule computation, but so called apriori algorithm [46] is the best-known algorithm to mine association rules. It is based on breadth-first search strategy [47]. In the data set described by Table 2 attribute selection is more complicated by the fact that some attribute values are not binary so this involves more extensive search.

The association rule generation is chosen to be performed due to multiple reasons:

1. This is an exact method and therefore excludes any subjective influence while analyzing data set.

2. The result is presented in a readable and easy-to-understand form.

3. It is expected that number of generated association rules would not be high (data set contains 256 instances) due to computation of confidence for each rule and selection of rules with the highest confidence.

4. It is expected that association rules have a great value when inferred from data set in education domain because association rules can be treated as a hypothesis.

The experiment was conducted on data set by software system Weka in order to generate association rules. After loading, the data are ready for pre-processing and application of DM techniques. Weka system requires that the attributes with numerical values do not participate in the association rule mining, so they are ignored. Association rules generated by Weka system (apriori algorithm is used) are shown in Table 3.

Table 3: Association rules generated by apriori algorithm

Rule	IF	THEN	Rule confidence
1	A11=yes, 159	A17=yes, 154	0.97
2	A9=yes AND A14=yes, 159	A17=yes, 154	0.97
3	A13=games, educational, 167	A17=yes, 159	0.95
4	A9=yes AND A15=yes, 166	A17=yes, 158	0.95
5	A12=yes AND A14=yes, 169	A17=yes, 160	0.95
6	A9=yes, 185	A17=yes, 174	0.94
7	A14=yes AND A15=yes, 195	A17=yes, 183	0.94
8	A12=yes, 192	A17=yes, 180	0.94
9	A6=no, 167	A17=yes, 156	0.93
10	A14=yes, 221	A17=yes, 206	0.93

There are 10 rules generated. The IF part of every rule is followed by support measure, as is the case with the THEN part of each rule. The confidence for each rule is given in the separate column.

4 Results

The Analysis of generated association rules provides insight into the dependence of the monitored parameters. Each rule is accompanied by a factor of confidence that takes a value in the range [0, 1]. Ten association rules have been generated, see Table 3.

By association rule 1 attribute A11 (Does the student want more educational computer software to be used in school in order to improve learning) is associated with attribute A17 (Does the student think that his/her computer skills will help him/her in the future). The factor of confidence for this rule is 0.97, that rule is guaranteed to a great extent. If the value of attribute A11 is "yes" then, by this association rule, the value of attribute A17 is also "yes". A possible conclusion which can be drawn is that most of the students think that usage of computer technology in school generates skills and knowledge that could be used in the future. This students opinion is a good indicator of the importance of the usage of computer technology and educational software in teaching process. It is evident that technology usage in future is closely related to computer technology software and methods which are used at present. Other association rules can be interpreted in analogous way.

The rule 2 can be interpreted as follows. Two facts: The fact that student uses home computer for learning and the fact that student likes the subjects of computer science, are associated with opinion that computer skills will be helpful in the future. Rule number 3 associates the type of software that is most frequently used by student (games, educational software) with the opinion that computer skills will be helpful in the future.

The opinion that computer skills will be helpful in the future is mostly associated with: believing that student gained enough knowledge to work independently on a computer, the usage of the computer for learning and aspiration to use distant learning system at home to review learning materials previously used in school. Attribute A14 (The student likes the subject of computer science) is most frequently associated with opinion that computer skills will be helpful in the future.

However, rule number nine associate the students that are not satisfied with the usage of e-mail with the opinion that computer skills will be helpful in the future. In fact, this may be in accordance with rule number eight: The usage of distant learning system at home to review learning materials previously used in school is associated with the opinion that computer skills will be helpful in the future. So, in students opinion, distant learning system makes e-mail service obsolete in a way.

5 Conclusions and Future Works

The application of association rule mining allows automatic generation of hypotheses and factors of confidence that are related to them. Considering rule number one (see Table 3) the hypothesis is: If a student wants more educational computer software to be used in school in order to improve learning then the student thinks that his/her computer skills will help him/her in the future. Other rules may also be interpreted as a hypothesis. Obviously the factor of confidence has a great impact on confirmation of hypotheses. At the Technical Faculty "Mihajlo Pupin" in Zrenjanin, Serbia, extensive research is underway, investigating the possibilities of applying DM techniques to data from the domain of education, extracted from a survey conducted in the wider Balkan region. The fact that it is not obligatory that the data are collected with the aim to be analyzed by DM techniques, offers an excellent chance to assess real possibilities in the actual practice. In future, this leads to identification of the advantages of DM techniques over standard statistical techniques. So far, there have been identified the following advantages of application of the DM techniques to data from the domain of education:

- It is possible to extract the "special" cases of the statistical rarity that are often discarded as "noise". The discovery of such "isolated" cases is conducted by automated IF ... THEN rule synthesis. As in [48] the association data mining based on a uniform support misses some patterns of low support. Although it is possible to exploit support constrains, which specifies some minimum supports, we opted to use automated IF ... THEN rules synthesis by Rosetta system or similar.

- Clustering technique allows the exact definition of average student and special groups (clusters) of students, so that further action of e-learning can be implemented while respecting the characteristics of each cluster.

- Analysis of generated association rules provides insight into the dependence of the monitored parameters.

- Ranking of attributes (parameters) in order of importance of influence on a selected attribute allows the rejection of the parameters of lesser importance.

- Data visualization provides a figurative description of the data, so we can actually see patterns which may exist.

The usage of Weka system, or a similar system, is of great help because it generates association rules (hypotheses) automatically. Furthermore, Weka system generates association rules for which the factor of confidence is high. This method can be used in any case of data, but Weka system requires that attributes with numerical values do not participate in generating the association rule mining, so they are ignored. This is not a great handicap because linguistic terms are frequently used in queries and surveys. The final conclusion is: the usage of Weka system in order to generated association rules automatically is of great help because the hypotheses of little importance are avoided.

Future work will be practical and it will refer to the selection of a particular CMS and integration with DM system.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):154-167, December, 2012.

Bibliography

[1] C. Chih-Ming, Personalized E-learning system with self-regulated assisted mechanisms for promoting learning performance", *An Int. J. of Expert Systems with Applications* 36: 8816-8829, 2009.

[2] C. Romero, S. Ventura, E. Garcia, Data mining in course management system: Moodle case study and tutorial, *An Int. J. of Computers and Education* 51, pp. 368-384, 2008.

[3] E. Brtka, D. Radosav, V. Brtka, The Data Mining module as a part of the e-learning system, (In Serbian), *In Proc. of InfoTech Conference*, Vrnjacka Banja, Serbia, 2009.

[4] E. Brtka, The data mining analysis approach in pedagogical research, Master Thesis, (In Serbian), Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia, 2009.

[5] E. Gaudioso, L. Talavera, Data mining to suport tutoring in virtual learning communities: Experiences and challenges. *In C. Romero and S. Ventura (Eds.), Data mining in e-learning*, Southampton UK, Wit Press, pp. 207-226, 2006.

[6] K. Diego, Data Mining and Statistics: What is the Connection?, *The Data Administration Newsletter*, LLC - www.TDAN.com

[7] D. Hand, H. Mannila, P. Smyth, *Principles of Data Mining*, MIT Press, Cambridge, MA. ISBN 0-262-08290-X. OCLC 226126187, 2001.

[8] DBMiner (2007), http://www.dbminer.com

[9] Clementine (2007), http://www.spss.com/clementine/

[10] Miner (2007), http://www-306.ibm.com/software/data/iminer/

[11] Weka (2007), http://www.cs.waikato.ac.nz/ml/weka/

[12] Keel (2007), http://www.keel.es/

[13] O. Zaiane, J. Luo, Web usage mining for a better web-based learning environment, *In Proc. of conf. on advanced technology for education*, Banff, Alberta, pp. 60-64, 2001.

[14] D. Silva, M. Vieira, Using data warehouse and data mining resources for ongoing assessment in distance learning, *In IEEE int. conf. on advanced learning technologies*, Kazan, Russia, pp. 40-45, 2002.

[15] J. Tane, C. Schmitz, G. Stumme, Semantic resource management for the web: An elearning application, *In Proc. of the WWW conference*, New York, USA, pp. 1-10, 2004.

[16] E. Garcia, C. Romero, S. Ventura, C. Castro, Using rules discovery for the continuous improvement of e-learning courses, *In Int. conf. intelligent data engineering and automated learning*, Burgos, Spain, pp. 887-895, 2006.

[17] Z. Pawlak, A. Skowron, Rudiments of rough sets, *An Int. J. of Information Sciences* 177:3-27, 2007.

[18] A. Ohrn, *Discernibility and Rough Sets in Medicine: Tools and Applications*, PhD thesis, Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway, 1999.

[19] I. H. Witten, E. Frank, *Data Mining: Practical machine learning tools and techniques*, 2nd Edition, Morgan Kaufman, San Francisco, 2005.

[20] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern classification*, Wiley Interscience, 2000.

[21] G. Chen, C. Liu, K. Ou, B. Liu, Discovering decision knowledge from web log portfolio for managing classroom processes by applying decision tree and data cube technology, *Journal of Educational Computing Research* 23(3):305-332, 2000.

[22] B. Minaei-Bidgoli, W. Punch, Using genetic algorithms for data mining optimization in an educational web-based system, *In Genetic and evolutionary computation conference*, Chicago, USA, pp. 2252-2263, 2003.

[23] R. Baker, A. Corbett, K. Koedinger, Detecting student misuse of intelligent tutoring systems, *In Intelligent tutoring systems*, Alagoas, Brazil, pp. 531-540, 2004.

[24] S. B. Kotsiantis, C. J. Pierrakeas, P. E. Pintelas, Predicting students performance in distance learning using machine learning techniques", *Applied Artificial Intelligence* 18(5):411-426, 2004.

[25] M. V. Yudelson, O. Medvedeva, E. Legowski, M. Castine, D. Jukic, C. Rebecca, Mining student learning data to develop high level pedagogic strategy in a medical ITS, *In Proceedings of AAAI workshop on educational data mining*, Boston, pp. 1-8, 2006.

[26] M. Cocea, S. Weibelzahl, Can log files analysis estimate learners level of motivation? *In Proceedings of the workshop week Lernen - Wissensentdeckung - Adaptivitat*, Hildesheim, pp. 32-35, 2006.

[27] W. Hamalainen, M. Vinni, Comparison of machine learning methods for intelligent tutoring systems, *In Proceedings of the eighth international conference in intelligent tutoring systems*, Taiwan, pp. 525-534, 2006.

[28] A. K. Jain, M. N. Murty, P. J. Flynn, Data clustering: A review, *ACM Computing Surveys* 31(3):264-323, 1999.

[29] T. Tang, G. McCalla, Smart recommendation for an evolving e-learning system, *International Journal on E-Learning* 4(1):105-129, 2005.

[30] E. Mor, J. Minguillon, E-learning personalization based on itineraries and long-term navigational behavior, *In Proceedings of the 13th international world wide web conference*, pp. 264-265, 2004.

[31] W. Hamalainen, J. Suhonen, E. Sutinen, H. Toivonen, "Data mining in personalizing distance education courses", *In World conference on open learning and distance education*, Hong Kong, pp. 1-11, 2004.

[32] J. Spacco, T. Winters, T. Payne, T. Inferring use cases from unit testing, *In AAAI workshop on educational data mining*, New York, pp. 1-7, 2006.

[33] R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large databases, *In Proc.of the ACM SIGMOD international conference on management of data*, Washington DC, USA, pp. 1-22, 1993.

[34] O. Zaiane, Building a recommender agent for e-learning systems, *In Proc.of the int. conference in education*, Auckland, New Zealand, pp. 55-59, 2002.

[35] G. J. Hwang, C. L. Hsiao, C. R. Tseng, A computer-assisted approach to diagnosing student learning problems in science courses, *Journal of Information Science and Engineering* 19: 229-248, 2003.

[36] J. Lu, Personalized e-learning material recommender system, *In International conference on information technology for application*, Utah, USA, pp. 374-379, 2004.

[37] P. Markellou, I. Mousourouli, S. Spiros, A. Tsakalidis, Using semantic web mining technologies for personalized e-learning experiences, *In Proc. of the web-based education*, Grindelwald, Switzerland, pp. 461-826, 2005.

[38] B. Minaei-Bidgoli, P. Tan, W. Punch, Mining interesting contrast rules for a web-based educational system, *In Int. conf.on machine learning applications*, Los Angeles, California, pp. 1-8, 2004.

[39] C. Romero, S. Ventura, P.D. Bra, Knowledge discovery with genetic programming for providing feedback to courseware author, *User Modeling and User-Adapted Interaction: The Journal of Personalization Research* 14(5):425-464, 2004.

[40] P. Yu, C. Own, L. Lin, On learning behavior analysis of web based interactive environment, *In Proc. of the implementing curricular change in engineering education*, Oslo, Norway, pp. 1-10, 2001.

[41] A. Merceron, K. Yacef,Mining student data captured from a web-based tutoring tool: Initial exploration and results, *Journal of Interactive Learning Research* 15(4):319-346, 2004.

[42] J. Freyberger, N. Heffernan, C. Ruiz, Using association rules to guide a search for best fitting transfer models of student learning, *In Workshop on analyzing studenttutor interactions logs to improve educational outcomes at ITS conference*, Alagoas, Brazil, pp. 1-10, 2004.

[43] A. A. Ramli, Web usage mining using apriori algorithm: UUM learning care portal case, *In Int. conf. on knowledge management*, Malaysia, pp. 1-19, 2005.

[44] R. Spence, *Information visualization*, Addison-Wesley, 2001.

[45] R. Andonie, Extreme Data Mining: Inference from Small Datasets, *INT J COMPUT COMMUN*, ISSN 1841-9836, Vol. 5(3):280-291, 2010.

[46] R. Agrawal, R. Srikant, Fast algorithms for mining association rules in large databases, *In Jorge B. Bocca, Matthias Jarke, and Carlo Zaniolo (eds.), Proc. of the 20th International Conference on Very Large Data Bases*, VLDB, Santiago, Chile, pp. 487-499, 1994.

[47] G. Luger, W. Stubblefield, Artificial Intelligence - structures and strategies for complex problem solving, University of New Mexico, Albuquerque, The Benjamin/Cummings Publishing Company Inc, 1993.

[48] M. Pater, D.E. Popescu, Multi-Level Database Mining Using AFOPT Data Structure and Adaptive Support Constrains, *INT J COMPUT COMMUN*, ISSN 1841-9836, 3(S):437-441, 2008.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):168-180, December, 2012.

Cooperative Robot Structures Modeled After Whale Behavior and Social Structure

I.C. Reșceanu, G.C. Călugăru, C.F. Reșceanu, N.G. Bîzdoacă

Ionuț Cristian Reșceanu
Cristina Floriana Reșceanu
Nicu-George Bîzdoacă
University of Craiova
Faculty of Automation, Computers and Electronics
Romania, 200440 Craiova, Decebal Blvd., no. 107
E-mail: $\{resceanu, cristina, nicu\}$@robotics.ucv.ro

George-Cristian Călugăru
University of Craiova
Faculty of Automation, Computers and Electronics
Romania, 200440 Craiova, Decebal Blvd., no. 107
E-mail: calugaru.george.nds@gmail.ro

Abstract:
This paper analyses the communications and social structure of whale pods and tries to apply their principles on cooperative robot structures which can be guided to perform a certain task. The communication patterns and social structure are presented at first in order to define real, natural phenomena which can then be translated through cooperative robotic structures. Whale communication has suffered modifications in the latter years mostly because of the increase of noise in the ocean. Problems regarding communications between the cooperative structures chosen for modelling whale behavior are solved by data fusion techniques. In the last part of the paper the problem of dynamic compensation of disturbances is studied with regard to cooperative structures.

Keywords: cooperative structures, data fusion, manipulative robots.

1 Introduction

The order Cetacea (Cetus, whale, from Greek) includes the marine mammals commonly known as whales, dolphins and porpoises. Cetus is Latin and is used in biological names to mean "whale"; its original meaning, "large sea animal", was more general. It comes from Ancient Greek (ketos), meaning "whale" or "any huge fish or sea monster". In Greek mythology the monster Perseus defeated was called Ceto, which is depicted by the constellation of Cetus. Cetology is the branch of marine science associated with the study of cetaceans. Suborders of the Cetacea order are Mysticeti, Odontoceti and Archaeoceti (extinct ancient whales, depict the evolution of whales throughout time). Cetaceans are mammals that evolved throughout the ages for aquatic environments. Their body is fusiform (spindle-shaped) and the forelimbs are modified into flippers as a result of this evolution. The tiny hindlimbs are vestigial; they do not attach to the backbone and are hidden within the body. The tail has horizontal flukes. Cetaceans are nearly hairless, and are insulated from the cooler water they inhabit by a thick layer of blubber. Some species are noted for their high intelligence.

1.1 Whale Social Structure

Whale family structures are fascinating. One of the most important facts about whales is that they are particularly intelligent mammals and like humans, place much value on their families and the role that each member plays within the unit. Notably, the individual families also travel and migrate together in pods. Each family member continues to play a vital role within that pod, as a greater unit of the family. These groups demonstrate the sociable nature of whales and their unspoken cooperation with one another is evidence of the insight and sense of responsibility inherent to these animals.

Another interesting whale fact: whales tend to separate themselves into pods according to age and sex. The whale cows and their calves travel together in pods of up to 30 members at a time, accompanied by one dominant bull. Cows without their calves, or whose calves are mature enough, act as midwives to pregnant and nursing mothers. They assist them with their birth by ensuring that the newborn reaches the surface of the water for air. Cows are also babysitters to the other mother's calves in her absence, and assist her with the care of her new baby in a general sense.

The calves stick close to their mothers for an average of three to six years. But this period can be even longer, depending on the individual calf and their species. Even after they have left their mother's side, they may still return to the main pod to visit her. Females are also known for returning to their first pod when they become pregnant with their own calf.

The youths of the group eventually branch off into a smaller juvenile pod. They will move into larger pods once they reach sexual maturity and begin to calve. In some cases, the juvenile pod is either replaced or broken away from by a 'bachelor pod', consisting of only the young bulls. Whales are considered 'juveniles' from about three years of age to approximately thirteen.

There is a dominant bull in each core pod, and he is responsible for the pod in which he resides. He is sexually mature and cares for his harem of cows and calves. The other males tend to stick to themselves, traveling separately from the rest of the main pod, as to respect the 'property' of the dominant bull. This family- and pod structure is designed to protect the weak and the young of the group. Because whale calves do not mature as quickly as some other mammals do, they require time to grow and develop within a protected environment. The organization of the dominant male and the group of mothering cows ensures that calves are isolated from the dangers of the deep.

Traveling in this way also ensures that whale migrations remain orderly and safe for all involved, preventing smaller family units from drifting off course or facing the dangers that come with isolation. Of course, whales are also social creatures and benefit from the close interaction with others of their sort. This level of mutual understanding and cooperation is another indication of the brilliance of the whale creation and instinct. [12]

1.2 Whale Communication

Whales can communicate through a very intriguing method called echolocation. Echolocation, also known as biosonar, is the biological sonar used by several animals such as shrews, most bats and most cetaceans. The term was first used by Donald Griffin, whose work with Robert Galambos was the first to conclusively demonstrate its existence in bats. Two bird groups also employ this system for navigating through caves, the so called cave swiftlets in the genus Aerodramus (formerly Collocalia) and the unrelated Oilbird Steatornis caripensis. [13]

Echolocating animals emit calls out to the environment and listen to the echoes of those calls that return from various objects in the environment. They use these echoes to locate, range and identify the objects. Echolocation is used for navigation and for foraging (or hunting) in various environments.

Echolocation makes use of active sonar, using sounds made by an animal. Ranging is done by measuring the time delay between the animal's own sound emission and any echoes that return from the environment. The relative intensity of sound received at each ear provides information about the horizontal angle (azimuth) from which the reflected sound waves arrive. Unlike some sonar that relies on an extremely narrow beam to localize a target, animal echolocation relies on multiple receivers. Echolocating animals have two ears positioned slightly apart. The echoes returning to the two ears arrive at different times and at different loudness levels, depending on the position of the object generating the echoes. The time and loudness differences are used by the animals to perceive distance and direction. With echolocation, the bat or other animal can see not only where it is going but also how big another animal is, what kind of animal it is, and other features.

The sounds produced by whales can travel for miles as objects found in the water can help amplify them. The sounds will echo back to the whale that emitted them. This form of communication has an estimated speed of 1 mile/second.

Sounds made by whales are in fact very unique. The clicks are a basic part of this communication. The clicks also help the whales navigate through the waters as well as being a language feature. Different species of whales communicate in various ways. This is due to the fact that they don't have an inner ear.

The humpback whales have a very particular way of communication called singing. This signing mainly consists of periodical sound patterns. These songs can be up to 30 minutes in length and can travel up to 100 miles. Sperm whales have only been heard making clicks, while toothed whales use echolocation that can generate sounds of 30000 Watts at 163 decibels. Whales create small pods with various communication patterns within. Researchers established that whales from different locations in the world have different terminology. This terminology is also different at whales in captivity in comparison with free whales.

It is believed that whales have amplified their songs in the past few decades. Whale songs can travel thousands of miles, but new research shows that their ability to communicate is severely affected by the increase of the amount of noise in an ocean. Thus the whales adapted by making their songs louder. [11]

2 Using Data Fusion for Process Surveillance and Diagnosis In the Modelling of Whale Communication

Terms like data fusion, multi-sensor data fusion, sensor fusion and information fusion or multi-sensor integration are used frequently in literature to depict a variety of techniques, technologies, systems and applications which use data provided by multiple sources. The data fusion applications are various - from real-time sensor fusion applied to mobile robots navigation to the fusion of strategic intelligence developed for military purposes. [6]

Data fusion is used in the development of modern applications regarding process surveillance and diagnosis.

Using specific algorithms, damaged sensors can be determined and in the same time signals can be synthesized in order to replace the erroneous information.

Data fusion techniques combine data from a multitude of sensors and related information to obtain more specific results than in the case of one sensor.

The data fusion concept can find its equivalent in nature. Throughout their evolution, people and animals developed their ability to use multiple senses in order to survive. For example, establishing if a food is eatable or not cannot be determined by using only the sight sense; the combination of sight, touch, smell and taste is far more effective. In a similar manner, when the ability to see is blocked by built structures or vegetation, the hearing sense can offer advanced warnings in case of imminent danger. Thus, multi-sensor data fusion is performed naturally by animals and people in order to evaluate the configuration of the surrounding environment and in order to detect possible threats.

Even though the concept of data fusion is well established, the development of new sensors, new techniques of processing and hardware made real-time data fusion viable.

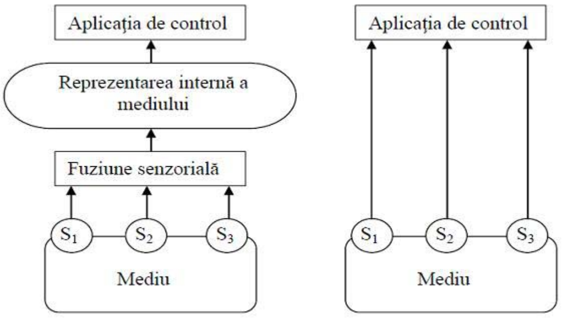
Recent progresses made in the development of computers and sensors offer the ability to emulate the natural abilities regarding data fusion for both people and animals through hardware and software means. Nowadays, data fusion systems are used in military applications such as: target recognition (e.g. intelligent weapons), vehicle guidance and remote detection like IFFN systems (Identification Friend-Foe-Neutral Systems). Non-military applications include monitoring of manufacturing processes, condition based maintenance for complex machinery, robotics and medical applications.

2.1 Multi-sensor Advantages

Another frequent term is that of multi-sensor integration which signifies the synergic use of sensor data to perform a specific task.

Sensor fusion is different from multi-sensor integration because the first includes the current combination of sensor information in a single representation format.

Multiple sensor data fusion offers multiple advantages in comparison with information prelevated through one sensor. First, if multiple sensors are used (e.g. identical radars following an object in motion) combining all the determinations will provide a better estimate of the motion and position of the object. A statistical advantage is obtained by adding N independent determinations (e.g. estimation of the position and speed of the target is improved by a factor of), with the assumption that data are combined in an optimal manner. The same result could be obtained by adding N independent determinations to a single sensor.



a) Fuziune senzorială b) Integrare multisenzorială

Figure 1: Data fusion and multiple sensor integration configurations

A second advantage implies the use of relative placement and motion of sensors to improve the process of observation. For example, 2 sensors that measure the angular motion of an object can be coordinated in order to determine the position of an object through triangularization. This technique is used for surveillance and commercial navigation.

A third advantage is the improved observability. The initial increase in physical observability may lead to significant improvements.

2.2 Control Architectures

For system control, the following types of data fusion can be defined with regard to sensors, [1]:

- *Complementary data fusion* - the fusion of multiple sensors spread across an area provides partial information about the environment; e.g. the fusion of multiple sensors which measure distances or video cameras pointing in different directions. This type of fusion solves the problem of incomplete information.

- *Competitive fusion* - the fusion of uncertain data obtained from multiple sensors - e.g. a radar and a video camera which detect the same object. Through data fusion, the distance to the object can be obtained with a higher accuracy. This type of fusion is used to reduce the effect of uncertainties and erroneous measurements.
- *Cooperative fusion* - the fusion of different sensors in which one of the sensors is based on determinations of another sensor in order to obtain its own set of data. For example, when a tactile sensor supplies information about the shape of an object previously estimated by a proximity sensor. This type of fusion is used to diminish the uncertainty effect, measurement errors as well as the incomplete state of the data.

Three basic alternatives can be used for multi-sensor data:

1. Direct fusion of data provided by sensors
2. Representation of sensor data through characteristic vectors with subsequent fusions of them.
3. Processing of each sensor in order to obtain high level control decisions which are later combined. Each of these approaches will use different data fusion techniques.

If multi-sensor data are proportional (if sensors measure the same physical phenomenon) then data from the sensors can be directly combined. The techniques for fusion of sheer data imply classical estimation methods like Kalman filtering. However, if sensor data is not proportional the data must be merged at the level of the characteristic/state vector or at a decision level.

Data fusion at the characteristics level implies the extraction of the representative characteristics from the sensor. It has been demonstrated that people use a cognitive function based on characteristics in order to determine objects. In case of data fusion at the characteristics level, the characteristics are extracted from multiple sensor observers and combined in a single vector which is subjected to pattern recognition techniques like neural networks or grouping algorithms.

Decision level data fusion combines sensor information after every sensor performed a preliminary determination of the entity location, attributes and identification.

2.3 A Model for Processing Data Fusion

In order to improve communications between military researchers and system developers, JDL (Joint Directors of Laboratories), established in 1986, began an effort to define the specific terms regarding data fusion. The result of the effort led to the creation of a process model for data fusion and specific terminology. The JDL process model is set to be very general and useful in multiple application domains, identifies processes and categories of techniques applicable to data fusion. The model is hierarchical with 2 levels. At the upper level, the data fusion process is described by the sensor inputs, the interaction between man and computer, database management, source preprocessing and 4 key sub-processes:

Level 1 processing (object refining) combines sensor data to obtain the most reliable and accurate estimation of the position of the object, its velocity, attributes and identity.

Level 2 processing (scenario refining) tries dynamically to develop a description of the current relationship between entities and events taking place in their surrounding environment.

Level 3 processing (threat refining) projects the current scenario into the future in order to draw the inferences about for threats, about friends and the foe vulnerabilities and opportunities of action.

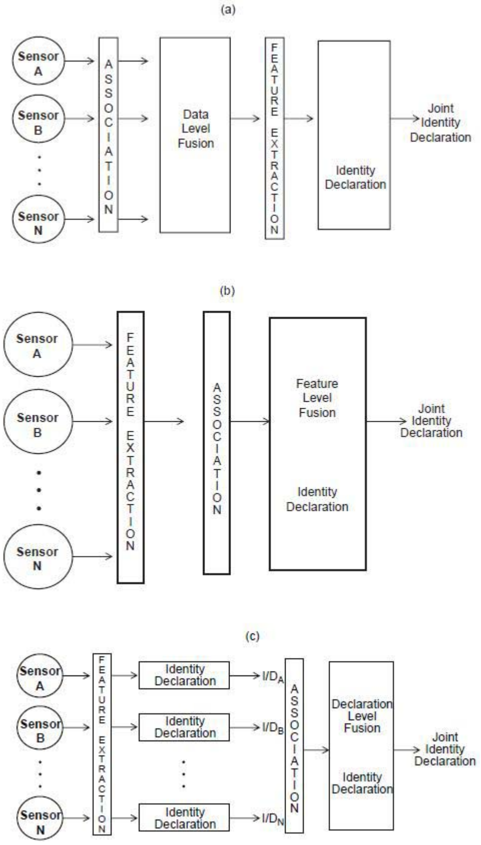


Figure 2: a. Direct data fusion of sensors; b. Representation of sensor data through characteristic vectors and subsequent fusions of them; c. The processing of each sensor in order to obtain high level inferences and decisions which are then combined

Level 4 processing (process refining) is a meta-process which monitors the global process of data fusion in order to evaluate and improve the real-time system.

For each of these sub-processes, the JDL hierarchical model identifies specific function and technique categories (in the second level of the model) and specific techniques (in the lower layer of the model). The implementation of the data fusion systems integrates and correlates those functions in a general work flow.

3 The Dynamic Model of a Manipulative Robot Used In Modelling Whale Behavior

The cooperative structure used to model whale behavior contains manipulative robots. The dynamic model of a such a robot is presented below. For more details on the principles of bio-insipired computing and robots see [10]. Also for more details on cooperative robot structures see [3].

The dynamic model is given by:

$$J\prime_1\ddot{\theta_1} + J^{*}cos(\theta_1 - \theta_2)\ddot{\theta_2} + J^{*}sin(\theta_1 - \theta_2)\dot{\theta}_2^2 + M\prime_1 cos\theta_1 = M_1 \qquad (1)$$

$$J\prime_2\ddot{\theta_2} + J^{*}cos(\theta_1 - \theta_2)\ddot{\theta_1} - J^{*}sin(\theta_1 - \theta_2)\dot{\theta}_1^2 + M\prime_2 cos\theta_2 = M_2 \qquad (2)$$

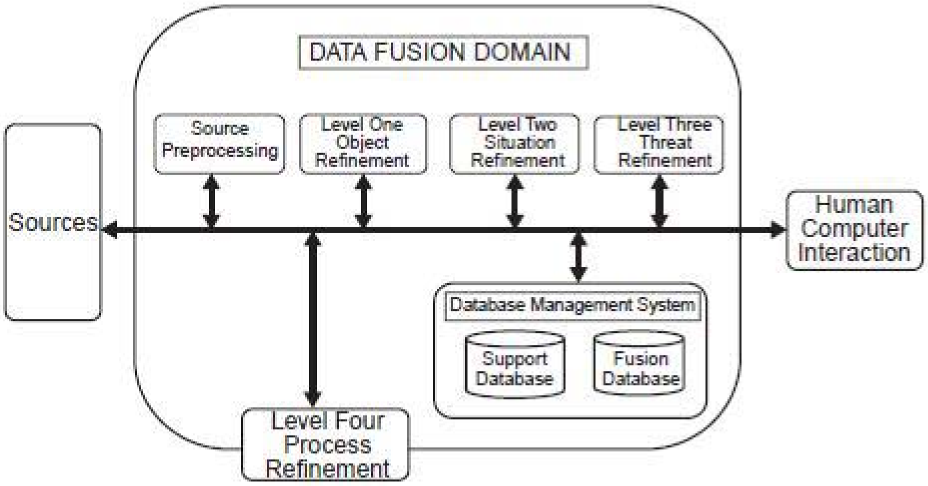

DATA FUSION DOMAIN
Sources
Source Preprocessing
Level One Object Refinement
Level Two Situation Refinement
Level Three Threat Refinement
Human Computer Interaction
Database Management System
Support Database
Fusion Database
Level Four Process Refinement

Figure 3: The JDL process model for data fusion

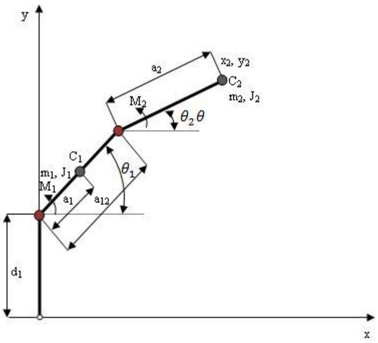


Figure 4: The kinematical model of the manipulative robot

3.1 Determining the Operation Span In Fault Conditions

The present day migration courses of whale pods are affected by naval traffic. For instance, in 2007 a mother whale and her calf were trying to return from the Sacramento- San Joaquin Delta to the Pacific Ocean. The two whales had deviated off course a week earlier. The change in course is believed to be influenced by the sounds of tug boats met along the way. The two whales appeared to have been wounded by a ship's propeller. In case the whale swimmers are affected, the whale will try to maintain its course using the wounded swimmers with a smaller operation span. In this paragraph the operation span in fault conditions is determined.

According to the kinematical configuration of the robot system considered, it has an area of operation in the shape of an annulus sector as depicted in the figure below:

In the case of a fabrication line which contains a number of robots with the same kinematical configuration, trespassing over these admissible operation areas can lead to interference problems. In order to avoid this, all areas that can be trespassed are eliminated apriori and thus each robot can have its own distinct area. So, for every robot a rectangular operation area (called an operation cell) will be defined.

Fig. 5.a) depicts the three parameters that describe the area of operation: maximum angle φ , minimum range r_{min} and respectively, the maximum range r_{max}.

In figure 6.b) a rectangular operation cell is defined where x and y represent the height and width of the cell. Between the 2 regions the following relations are defined:

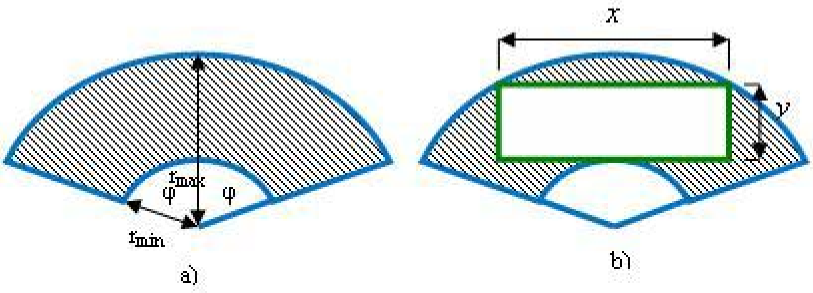


Figure 5: Definition of the area of operation for the manipulative robot: a) area of operation b) definition of the operation cell

$$\frac{y}{2} < r_{min} tg\varphi \tag{3}$$

$$r_{max}^2 = (r_{min} + y)^2 + \left(\frac{y}{2}\right)^2 \tag{4}$$

3.2 Manipulator Fault Due to a Jammed Articulation

It is assumed that the fault which appears at the manipulator is due to a jammed articulation. It is also assumed that the value of the angle of the jammed articulation is known. Even though the blocking position is not available from the sensor attached to every articulation, this can be calculated based on the position of the griper of the faulted manipulator by solving backward kinematics problems.

An Articulation Fault

When the first articulation is jammed, the faulted manipulator can only move its third joint by means of the second joint.

Depending on the configuration of the system at the moment the fault occurred, the faulted manipulator can be placed in a position similar to the one below:

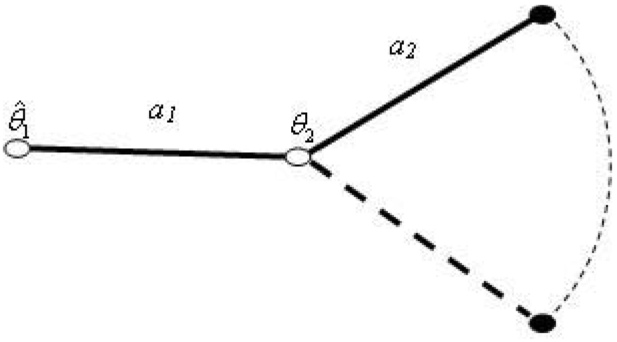


Figure 6: Aerial view of the manipulator with the first articulation blocked

The admissible region from the area of operation is, in this case, projected on a line of the arc as can be seen in figure 8.

In this case, the terminal of the faulted manipulator can be placed in one of the two possible positions, A and A', from its trajectory. The kinematical constraints that can guarantee the existence of these two placement positions are defined like this:

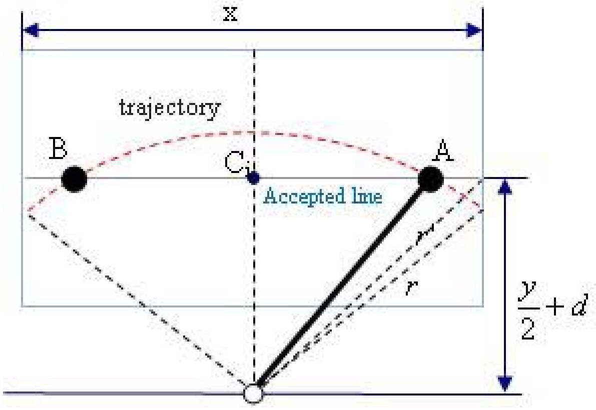


Figure 7: The kinematical constraints on case of first articulation failure

$$\frac{y}{2}+d \leq r \leq r\prime \tag{5}$$

Where r is the radius of the radius of the arc and r' is the distance between the point in which the arm attaches itself to the support and the limits of terminal of the manipulator.

$$r\prime = \frac{1}{2}\sqrt{x^2+(y+2d)^2} \tag{6}$$

$$r = a_1 cos\hat{\theta_1} + a_2 cos\theta_2 \tag{7}$$

where $\hat{\theta_1}$ is the angle of the jammed articulation. Note the fact that r is identical to the length of the projection of the manipulator on the area of operation.

$$\frac{y}{2}+d \leq a_1 cos\hat{\theta_1} + a_2 cos\theta_2 \leq \frac{1}{2}\sqrt{x^2+(y+2d)^2} \tag{8}$$

4 Dynamic Delay Compensation Modelled After Whale Communication Habits

This part of the paper deals with dynamic compensation of disturbances, [7] In the last decades, the amount of noise present in the ocean has increased significantly in the ocean. This is largely due to the increase of naval traffic, the exploitation of certain resources(e.g. oil through oil platforms) and the intensive use of sonars. The whales themselves have adapted by increasing the intensity of their sounds. With regard to robotic structure communication a dynamic delay compensation structure is necessary. Thus, alternative Smith predictor structures were used for the dynamic delay compensation, [2], [4], [5], [9], [8]. The following simulations were performed using Smith predictor structures on a robotic arm acted on by a Quanser SRV-02 servo-motor.

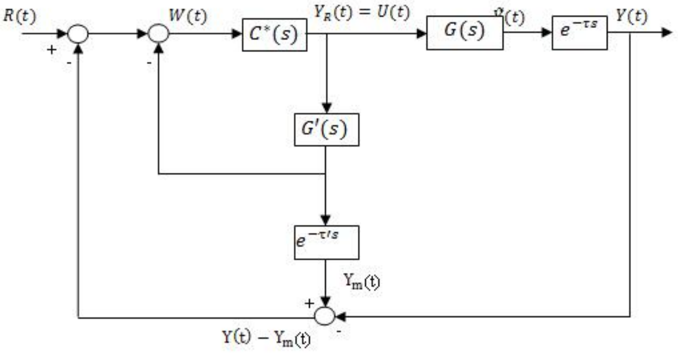


Figure 8: Modified Smith predictor structure

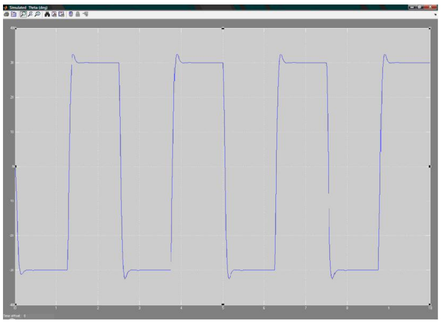


Figure 9: System response to a signal generator

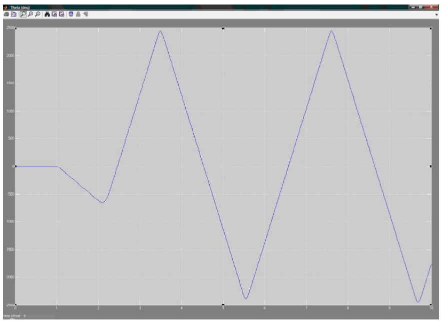


Figure 10: System response in case of delay occurence

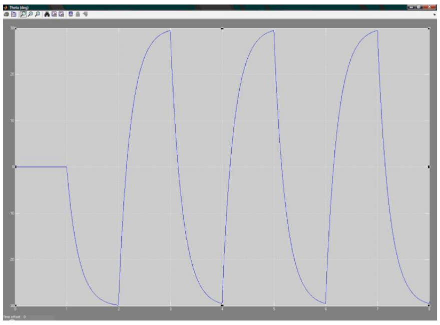


Figure 11: System response when using a modified Smith predictor

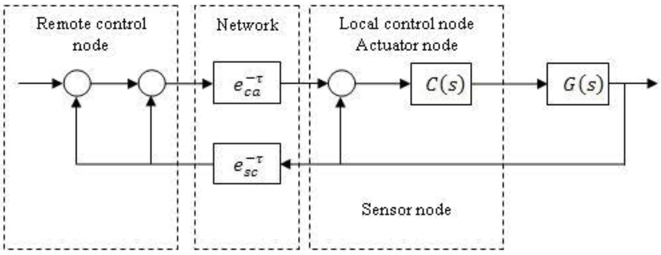


Figure 12: Improved Smith predictor structure

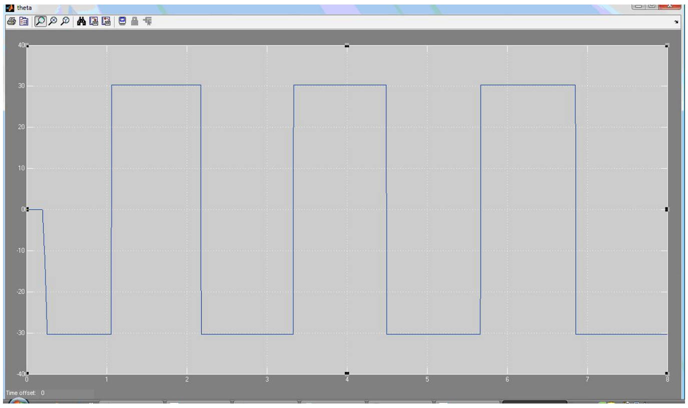


Figure 13: Position control of the Quanser SRV-02 plant using an improved Smith predictor

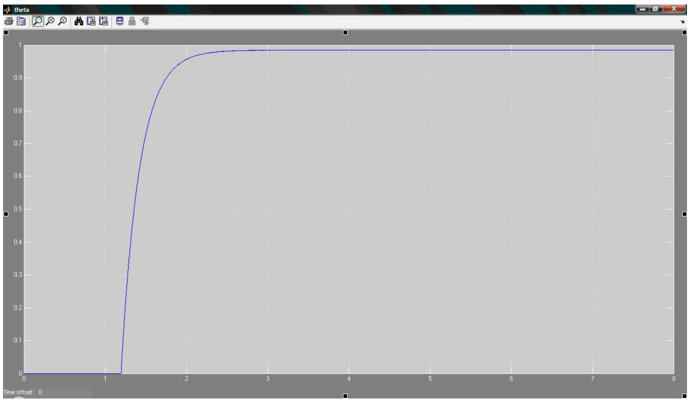


Figure 14: Step response of the system using the improved Smith predictor and a PI controller

5 Conclusions

This paper analyses the possibility of modelling cooperative robot structures after the behavior and communication habits of whale pods. Certain key aspects are studied like the modelling of whale communication using data fusion techniques, determining the operation span in fault conditions and the dynamic delay compensation modelled also after whale communication. Aspects like communication and social structure of whale pods are well documented in literature. What this paper is trying to establish is that there is the possibility of implementing this complex behavior in a cooperative robot structure with a well determined purpose. Further research and the study of other key behavior aspects should lead to the appearance of another class of problems based on real-life phenomena-the evolution of whale pods in matters of communication.

Acknowledgement. This work was supported by the strategic grant POSDRU/89/1.5/S/61968, Project ID61968 (2009), co-financed by the European Social Fund within the Sectorial Operational Program Human Resources Development 2007-2013

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):168-180, December, 2012.

Bibliography

[1] K.J. Astrom, B. Wittenmark, *Computer Controlled Systems: Theory and Design*, Prentice-Hall, Englewoods Cliffs, NJ, 1984

[2] K.J. Astrom, C.C. Hang, B.C. Lin, *A New Smith Predictor for Controlling a Process With an Integrator and Long Dead-Time*, IEEE Transactions on Automatic Control, 39(2), pp. 343-345, 1994

[3] Y.U. Cao, A.S. Fukanaga, A.B. KAHNG, *Cooperative Mobile Robotics: Antecedents and Directions*, Autonomous Robots, 4, 1-23, Kluwer Academic Publishers, Boston, 1997

[4] D. Feng, D.,Wencai, L. Zhi, *New Smith Predictor and Nonlinear Control for Networked Control Systems*, Proceedings Of The International MultiConference of Engineers and Computer Scientists, March 18-20, Hong Kong, Vol. 2, pp. 1184-1188, 2009

[5] D. Feng, Q. Qian, *Study of NCS with Improvement Smith Predictor and Fuzzy Immune Control*, Proceedings of the International Conference on Intelligent Systems and Knowledge Engineering, Chengdu, China, 2007

[6] D.L. Hall, J. Llinas,*Handbook of Multisensor Data Fusion*, CRC Press, New York, USA, 2001

[7] J.G. Truxal, M.J. Shooman, W.R. Blesser, J.W. Clark, *Remote Control, in the Handbook of Telemetry and Remote Control*, Mc-Graw Hill, New-York, pp 1-169, 1967

[8] J. Velagic, *Design of Smith-like Predictive Controller with Communication Delay Adaptation*, World Academy of Science, Engineering and Technology, no. 47, pp 199-203, 2008

[9] M. Veronesi, *Performance Improvement of Smith Predictor through Automatic Computation of Dead Time*, Yokogawa Technical Report English Edition, pp 25-30, 2003

[10] X. Yang, *Bio-Inspired Computing and Networking*, CRC Press, 2011.

[11] http://www.eoearth.org/article/whale_communication_and_culture

[12] http://www.napali.com/na_pali_coast/hawaii_whale_watching/social_communication.html

[13] http://www.whales.org.za/facts_communication.aspx

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):181-191, December, 2012.

Function Approximation with ARTMAP Architectures

L.M. Sasu, R. Andonie

Lucian M. Sasu
1. Transilvania University of Braşov
Mathematics and Computers Department
Romania, 500091 Braşov, Iuliu Maniu, 50
lmsasu@unitbv.ro
2. Siemens Corporate Technology
Romania, 500096 Braşov, 15 Noiembrie, 46
E-mail: lucian.sasu@siemens.com

Răzvan Andonie
1. Computer Science Department
USA, Central Washington University, Ellensburg
400 East University Way
Ellensburg, WA 98926, USA
2. Transilvania University of Braşov
Electronics and Computers Department
Romania, 500024 Braşov, Politehnicii, 1
E-mail: andonie@cwu.edu

Abstract: We analyze function approximation (regression) capability of Fuzzy ARTMAP (FAM) architectures - well-known incremental learning neural networks. We focus especially on the universal approximation property. In our experiments, we compare the regression performance of FAM networks with other standard neural models. It is the first time that ARTMAP regression is overviewed, both from theoretical and practical points of view.

Keywords: fuzzy ARTMAP, universal approximation, regression.

1 Introduction

The approximation of functions that are known only at a certain number of discrete points is a classical application of neural networks. Almost all approximation schemes can be mapped into some kind of network that can be dubbed as a "neural network" [1]. A neural network has the *universal approximation property* if it can approximate with arbitrary accuracy an arbitrary function of a certain set of functions (usually the set of continuous function) on a compact domain. The drawback is that such an approximation may need an unbounded number of "building blocks" (i.e., fuzzy sets or hidden neurons) to achieve the prescribed accuracy. Therefore it is reasonable to make a trade-off between accuracy and the number of the building blocks, by determining the functional relationship between them.

Historically, of fundamental importance was the discovery [2] that a classical mathematical result of Kolmogorov (1957) was actually a statement that for any continuous mapping $f : [0,1]^n \subset \Re^n \longrightarrow \Re^m$ there must exist a three layered feedforward neural network of continuous type neurons that implements f exactly. This existence result was the first step. Cybenko [3] showed that any continuous function defined on a compact subset of $\Re^n$ can be approximated to any desired degree of accuracy by a feedforward neural network with one hidden layer using sigmoidal nonlinearities. Many other papers have investigated the approximation capability of three layered networks in various ways. In addition to sigmoid functions, more general functions can be used as activation functions of universal approximator feedforward networks [4].

Girosi and Poggio proved that radial basis function (RBF) networks also have universal approximation property [1]. Hartman and Kowalski [5] proved that a one hidden layer neural network with Gaussian hidden nodes is a universal approximator for real-valued maps defined on convex, compact sets of $\Re^n$. Additional related papers are [6] and [7].

The Fuzzy ARTMAP (FAM) family of neural networks is one of the best known incremental learning systems. There are many variations of Carpenter's *et al.* [8] initial FAM model, including Gaussian ARTMAP (GAM) [9], PROBART [10], FAMR [11], GART [12], [13], and AppART [14]. Compared to FAM classification, the function approximation (regression) capability of FAM was less frequently addressed. It is our goal here to discuss FAM regression capability for different FAM architectures.

The FAM maps subsets of $\Re^n$ to $\Re^m$, accepting both binary and analogue inputs in the form of pattern pairs. The initial FAM, PROBART, and the FAMR architectures have been used for incremental regression estimation. Since the initial FAM was proved to be universal approximator [15], it is reasonable to believe that members of the FAM family may also have the universal approximation capability. However, since some of the FAM variations are quite different than the initial FAM, each model should be considered individually.

The Bayesian theory allows for elaboration of general neural network training methods [16].Recently, Vigdor and Lerner have combined the Bayesian theory and the FAM introducing the Bayesian ARTMAP (BA) [17]. Like the GAM and the GART networks, during training, the BA uses Gaussian categories and FAM competitive learning. However, the BA prediction phase is very different than the FAM competitive algorithm, being a Bayesian approach. Vigdor and Lerner have compared the BA performance with respect to classification accuracy, learning curves, number of categories, sensitivity to class overlapping and risk with those of the FAM. Generally, the BA outperformed the FAM in classification tasks. Up to our contribution, the BA regression capability was not discussed or tested.

Our paper is the first overview of both theoretical and practical aspects of FAM regression, considering several major FAM architectures: the initial FAM of Carpenter *et al.*, PROBART, FAMR, BA, GAM, and AppART. We discuss universal approximation capabilities of these FAM models. In our experiments, we compare the regression performance of FAM networks with standard neural networks: Multi Layer Perceptron (MLP), RBF, General Regression Neural Network (GRNN), and FasBack. Section 2 reviews the main notations and paradigms of FAM. In Section 3, we discuss the universal approximation capability of the following FAM architectures: the original FAM, PROBART, FAMR, BA, and AppART. We synthesize our comparative experiments in Section 4. Section 5 contains the final remarks.

2 Fuzzy ARTMAP

A FAM consists of a pair of fuzzy ART modules, ART_a and ART_b, connected by an inter-ART module called Mapfield, F^{ab}. ART_a contains a preprocessing layer F_0^a, an input (or short-term memory) layer F_1^b and a competitive layer F_2^b. The following notations apply: M_a is the number of nodes in F_1^a, N_a is the number of nodes in F_2^a, and $\mathbf{w}^a$ is the weight vector between F_1^a and F_2^a. We say that a node – also called a category – from F_2^a is *uncommitted* if it has not learned yet an input pattern, and *committed* otherwise. Analogous layers and notations are used in ART_b. Each node j from F_2^a is linked to each node from F_2^b via a weight vector $\mathbf{w}_j^{ab}$ from F^{ab}, the jth row of the matrix $\mathbf{w}^{ab}$, $1 \leq j \leq N_a$. All weights are initialized to 1.

All input vectors are complement-coded by the F_0^a layer in order to avoid category proliferation [8], [18], [19]: the input vector $\mathbf{a} = (a_1, \ldots, a_n) \in [0,1]^n$ produces the normalized vector $\mathbf{A} = (a_1, \ldots, a_n, 1-a_1, \ldots, 1-a_n)$. During pattern processing, the operator $\wedge$ used is the fuzzy

AND operator defined as $(\mathbf{p} \wedge \mathbf{q})_i = \min(p_i, q_i)$, where $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$. $|\cdot|$ denotes the L_1 norm.

Before learning a normalized input vector $\mathbf{A}$, the vigilance parameter factor ρ_a is reset to its baseline value $\overline{\rho}_a$ and each input category is considered as not inhibited, competing for the current input pattern. A fuzzy choice function is computed for every ART_a category: $T_j(\mathbf{A}) = \frac{|\mathbf{A} \wedge \mathbf{w}_j^a|}{\alpha_a + |\mathbf{w}_j^a|}$, for $1 \leq j \leq N_a$. The non-inhibited node of index J having the maximum fuzzy choice function value is further checked whether it passes the resonance condition, i.e. if the input is similar enough to the winner's prototype: $|\mathbf{A} \wedge \mathbf{w}_J^a|/|\mathbf{A}| \geq \rho_a$. If this condition is not fulfilled, then the node having index J is inhibited and another non-inhibited node maximizing the fuzzy choice function is considered as above. If no such node exists, a new node with index J is created to represent the input vector. In parallel, a similar step is performed in the ART_b module; we obtain output vector $\mathbf{y}^b = (\delta_{iK})_{1 \leq i \leq N_b}$, where K is the index of the output winner node ($1 \leq K \leq N_b$) and δ_{ij} is Kronecker's delta. If input node J is newly added, then we associate it with the current output: $\mathbf{w}_{Jk}^{ab} = \delta_{kK}$ and this association becomes permanent. Each time input node J is activated, it predicts as output value the only index k for which $w_{Jk}^{ab} = 1$. If node J is not new, then we check whether its predicted value is K. If the prediction is incorrect, a new activity (called *match tracking*) is triggered in ART_a solely. Otherwise, learning occurs in both ART_a and ART_b:

$$\mathbf{w}_J^{a(new)} = \beta_a \left(\mathbf{A} \wedge \mathbf{w}_J^{a(old)} \right) + (1 - \beta_a)\mathbf{w}_J^{a(old)} \tag{1}$$

where $\beta_a \in (0, 1]$ is the learning rate parameter. A similar learning step takes place in ART_b.

The match tracking raises the ρ_a threshold for the current input pattern: $\rho_a = \delta + |\mathbf{A} \wedge \mathbf{w}_J^a|/|\mathbf{A}|$. If $\rho_a > 1$ then the current input pattern is rejected; otherwise, the search for an appropriate input category is continued, as described above.

For each F_2^a category we have the following geometrical interpretation. Node w_j^a is a hyperrectangle R_j inside the n-dimensional hypercube, having size $n - |w_j|$ [8]. Learning, as in equation (1), is equivalent to expanding the hyperrectangle towards the current input pattern, unless this pattern is not already in R_j. If $\beta_a = 1$, then R_j expands to $R_j \oplus \mathbf{a}$, the minimal hyperrectangle containing both R_j and input pattern $\mathbf{a}$. A similar geometrical interpretation applies to ART_b.

3 FAM Architectures used in Regression

3.1 The initial FAM for regression

The FAM regression capability was first tested by Carpenter *et al.* for univariate real functions [8]. Input categories were considered to predict not real values, but real intervals. The experiments targeted the study of predicted output intervals' geometry and the number of resulted categories for various values of ρ_b. For the test set, the authors counted the matchings between predicted output categories and actual output values. A matching between $f(a)$ and the predicted output category (a rectangle) R_K^b was established if the size of $R_K^b \oplus f(a)$ did not exceed $(1 - \rho_b)$. As expected, the number of matchings increased with ρ_b.

Verzi *et al.* [15] proved that a slightly modified FAM version can be used to universally approximate any measurable function in $L^p([0, 1])$. More specifically, given $1 \leq p < \infty$, for every $f \in L^p([0, 1])$, $f \geq 0$, a series of FAM computable functions s_n with the following property were determined: functions s_n approximate f in the limit and s_n are dense in $L^p([0, 1])$. One can extend this result to the initial FAM.

3.2 PROBART for function approximation

PROBART is a modification of FAM motivated by empirical findings on the operational characteristics of FAM under certain conditions [10]. The authors replaced the Mapfield update rule FAM by

$$\mathbf{w}_J^{ab} = \begin{cases} \mathbf{y}^b + \mathbf{w}_J^{ab} & \text{if the } J\text{-th } F_2^a \text{ node is active and } F_2^b \text{ is active} \\ \mathbf{w}_J^{ab} & \text{if the } J\text{-th } F_2^a \text{ node is active and } F_2^b \text{ is inactive} \end{cases} \tag{2}$$

Thus, w_{jk}^{ab} indicates the number of associations between the j-th ART_a node and k-th ART_b node. Initially, $w_{jk}^{ab} = 0$, i.e. no association has been made yet.

There is no match tracking phase. The predicted value for an input pattern activating the Jth ART_a category is

$$\mu_{Jl} = \frac{1}{|\mathbf{w}_J^{ab}|} \sum_{k=1}^{N_b} \epsilon_{kl} w_{Jk}^{ab}, \quad 1 \leq l \leq M_b \tag{3}$$

where μ_{Jl} is the expected value of the l-th component of the predicted output pattern associated with the current input pattern, $|\mathbf{w}_J^{ab}|$ is the total number of associations of the J-th ART_a category and each category from ART_b, and ϵ_{kl} represents the kth ART_b category. Specifically, for PROBART the authors considered ϵ_{kl} as the lth component of the kth ART_b category exemplar. Only the first m components of each output category w_k^b are meaningful for computing the prediction corresponding to the current input pattern.

Equation (3) can be written as $\mu_{Jl} = \sum_{k=1}^{N_b} \epsilon_{kl} p_{Jk}$, where p_{Jk} is the empirically estimated association probability between the Jth ART_a category and the kth ART_b category: $p_{Jk} = w_{Jk}^{ab}/|\mathbf{w}_J^{ab}|$.

3.3 The FAMR Model for Function Approximation

The FAMR (Fuzzy ARTMAP with Relevance factor), a version of the FAM, has a novel learning mechanism. We will review here the FAMR basic notations (details in [11]) and discuss its function approximation capabilities.

The main difference between the FAMR and the initial FAM is the update method of the w_{jk}^{ab} weights. The FAMR uses the following updating formula [11]:

$$w_{jk}^{ab(new)} = \begin{cases} w_{jk}^{ab(old)} & \text{if } j \neq J \\ w_{JK}^{ab(old)} + \frac{q_t}{Q_J^{new}}\left(1 - w_{JK}^{ab(old)}\right) & \\ w_{Jk}^{ab(old)}\left(1 - \frac{q_t}{Q_J^{new}}\right) & \text{if } k \neq K \end{cases} \tag{4}$$

where q_t is the relevance assigned to the t-th input pattern ($t = 1, 2, \dots$) and $Q_J^{new} = Q_J^{old} + q_t$. The *relevance* q_t is a real positive finite number directly proportional to the importance of the experiment considered at step t. Initially, each Q_j ($1 \leq j \leq N_a$) has the same initial value q_0.

To maintain the stochastic nature of each $\mathbf{w}_j^{ab}$ row in Mapfield, we modified the Mapfield dynamics: when a new input category is created, a new row filled with $1/N_b$ is added to $\mathbf{w}^{ab}$; when a new ART_b category indexed by K is added, each existing input category is linked to it by $w_{jK}^{ab} = \frac{q_0}{N_b Q_j}$, and the rest of elements w_{jk}^{ab} are decreased by $\frac{w_{jK}^{ab}}{N_b - 1}$, for $1 \leq j \leq N_a$, $1 \leq k \leq N_b$, $k \neq K$. The update in eq. (4) preserves the stochastic property of each row. Finally, the vigilance test is changed to: $N_b\, w_{JK}^{ab} \geq \rho_{ab}$.

According to [11], this w_{jk}^{ab} approximation is a correct biased estimator of posterior probability $P(k|j)$, the probability of selecting the k-th ART_b category after having selected the j-th ART_a.

To estimate the corresponding output value for a given input pattern, FAMR uses the same formula as in eq. (3), but in this case ϵ_k contains the coordinates of the kth ART_b category centroid. During the FAMR training process, the l-th component of the centroid can be updated by Kohonen's learning rule: $\epsilon_{kl}^{b(new)} = \epsilon_{kl}^{b(old)} + (b_l - \epsilon_{kl}^{b(old)})/size_J^b$.

This rule incorporates an idea from [20]. The value $size_J^b$ is the number of output vectors of the k-th ART_b category and b_l is the l-th component of $\mathbf{b}$, the output vector of the current training pair $(\mathbf{a}, \mathbf{b})$.

3.4 The Bayesian ARTMAP Function Approximation Algorithm

In BA, in contrast to FAM, w_j^a is not a weight vector (a prototype), but simply a category label. Also, the ART categories are Gaussians, similar to the GAM. Each BA category j is characterized by the n-dimensional vector $\hat{\boldsymbol{\mu}}_j^a$ (mean), the $n \times n$ covariance matrix $\hat{\boldsymbol{\Sigma}}_j^a$, and the count number of training patterns clustered to category j, n_j^a. Analogous notations appear in ART_b, where one provides m-dimensional vectors.

The associations between input and output categories are stored inside the Mapfield module, as PROBART does, and one can approximate the conditional probability $P(w_k^b|w_j^a)$ as $\hat{P}(w_k^b|w_j^a) = w_{jk}^{ab}/\sum_{l=1}^{N_b} w_{jl}^{ab}$.

The following description uses ART_a notations; analogous notations are used for ART_b. All existent ART_a categories compete to represent the current input pattern. The posterior probability of category j given input $\mathbf{a}$ is estimated according to Bayes' theorem:

$$\hat{P}(w_j^a|\mathbf{a}) = \frac{\hat{p}(\mathbf{a}|w_j^a)\hat{P}(w_j^a)}{\sum\limits_{i=1}^{N_a} \hat{p}(\mathbf{a}|w_i^a)\hat{P}(w_i^a)} \tag{5}$$

where $\hat{P}(w_j^a)$ is the estimated prior probability of the j-th ART_a category, $\hat{P}(w_j^a) = n_j^a/\sum_{i=1}^{N_a} n_i^a$.

The conditional probability $p(\mathbf{a}|w_j^a)$ is estimated using all patterns already associated with Gaussian category w_j^a:

$$\hat{p}(\mathbf{a}|w_j^a) = \frac{1}{(2\pi)^{n/2}\left|\hat{\boldsymbol{\Sigma}}_j^a\right|^{1/2}} \cdot \exp\left\{-\frac{1}{2}(\mathbf{a} - \hat{\boldsymbol{\mu}}_j^a)^t(\hat{\boldsymbol{\Sigma}}_j^a)^{-1}(\mathbf{a} - \hat{\boldsymbol{\mu}}_j^a)\right\} \tag{6}$$

During the category choice step in ART_a, the winning category J is the one maximizing the posterior probability $\hat{P}(w_j^a|\mathbf{a})$.

The following vigilance test is performed: $S_J^a \leq S_{MAX}^a$, where $S_J^a = \left|\hat{\boldsymbol{\Sigma}}_J^a\right|$ is the hyper-volume of the winning category, and S_{MAX}^a is an upper bound threshold. During processing a training pattern, S_{MAX}^a may decrease from its initial value $\overline{S_{MAX}^a}$. In contrast, S_{MAX}^b remains unchanged. Every newly recruited category inside an ART_a (ART_b) module is centered in the current pattern and has the initial covariance matrix set to $\lambda(S_{MAX}^b)^{1/m} \cdot \mathbf{I}_m$ (and $\lambda(\overline{S_{MAX}^b})^{1/n} \cdot \mathbf{I}_n$, respectively), where λ is a small positive constant. This is done when none of the categories fulfills the vigilance test. Adding a new input (output) category triggers the addition of a new zero-filled line (column) to the association matrix $\mathbf{w}^{ab}$.

If the connection strength $\hat{P}(w_K^b|w_J^a)$ between winning categories w_J^a and w_K^b is below a fixed threshold P_{min}, then S_{MAX}^a is slightly decreased under the current winner input category's S_J, and the quest for another input category is continued. Otherwise, if the current winner input category was not newly added during processing the current pattern, ART_a learns the current pattern:

$$\hat{\boldsymbol{\mu}}_J^a(new) = \frac{n_J^a}{n_J^a + 1}\hat{\boldsymbol{\mu}}_J^a(old) + \frac{1}{n_J^a + 1}\mathbf{a}\ , \tag{7}$$

$$\hat{\boldsymbol{\Sigma}}_J^a(new) = \frac{n_J^a}{n_J^a + 1}\hat{\boldsymbol{\Sigma}}_J^a(old) + \frac{1}{n_J^a + 1}(\mathbf{a} - \hat{\boldsymbol{\mu}}_J^a(new))(\mathbf{a} - \hat{\boldsymbol{\mu}}_J^a(new))^t * I_n \tag{8}$$

$$n_J^a = n_J^a + 1 \tag{9}$$

Unless w_K^b is a newly added category for the current training pattern, an analogous learning process in ART_b takes place. Finally, the Mapfield association counter w_{JK}^{ab} is updated.

After learning, the BA can be used for prediction. We estimate the probabilistic association of an output category w_k^b with input test pattern $\mathbf{a}$:

$$\hat{P}(w_k^b|\mathbf{a}) = \frac{\sum\limits_{j=1}^{N_a} \hat{P}(w_k^b|w_j^a)\hat{p}(\mathbf{a}|w_j^a)\hat{P}(w_j^a)}{\sum\limits_{l=1}^{N_b}\sum\limits_{j=1}^{N_a} \hat{P}(w_l^b|w_j^a)\hat{p}(\mathbf{a}|w_j^a)\hat{P}(w_j^a)} \tag{10}$$

As in [21], we assume the conditional independence of activating categories w_k^b and w_j^a, given input pattern $\mathbf{a}$. For function approximation the following average formula is used:

$$\hat{f}(\mathbf{a}) = \sum_{k=1}^{N_b} \hat{P}(w_k^b|\mathbf{a}) \cdot \hat{\boldsymbol{\mu}}_k^b \tag{11}$$

Since, under certain mild conditions on the kernel function, RBF networks are universal approximators [1], [5], [6], [7], and the FAM also has universal approximation capability [15], it looks natural for the BA, which is essentially a FAM architecture with Gaussian categories, to be universal approximator. However, this statement can not be directly deducted from the RBF and FAM results. This is was a good reason for us to proof the following theoretical result [22]:

Theorem 1. *BA is a universal approximator on a compact set $X \subset \Re^n$.*

3.5 AppART: Hybrid Stable Learning for Universal Function Approximation

AppART [14] is an ART-based neural network model that incrementally approximates continuous-valued multidimensional functions through a higher-order Nadaraya–Watson regression.

An input pattern $\mathbf{x}$ is feedforwarded from input layer $F1$ to the $F2$ layer. The $F2$ layer consists of N categories, modeling a local density of the input space using Gaussian receptive fields with mean $\boldsymbol{\mu_j}$ and standard deviation $\boldsymbol{\sigma_j}$. A match criterion is used to detect whether the current leaning pattern activates an existing $F2$ category or a new one should be added. The match function is:

$$G_j = \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \mu_{ji}}{\sigma_{ji}}\right)^2\right),\ 1 \le j \le N \tag{12}$$

If all G_j values are below threshold ρ_{F2}, a new node is recruited to represent the current input pattern. Otherwise, the input strength of each $F2$ node is computed as $g_j = I(G_j > \rho_{F2}) \cdot (\eta_j G_j / \prod_{i=1}^{n}\sigma_{ji})$, where η_j is a measure of the prior activation probability of the jth category, and I is the binary indicator function: $I(P) = 1$ iff P is true. The activation values v_j of the $F2$ nodes are obtained by normalizing g_j. One can use v_j as an approximation of the posterior probability $P(j|\mathbf{x})$ of category j given input pattern $\mathbf{x}$.

The P and O layers together compute the prediction of the network. In the P layer, there are $m+1$ nodes whose corresponding values are computed as: $a_k = \sum_{j=1}^{N} \alpha_{kj} v_j \ (1 \leq k \leq m), \ \ b = \sum_{j=1}^{N} \beta_j v_j$ where α_{kj} and β_j are weights connecting each $F2$ category to the each node in the P layer. Each α_{kj} is the sum of values of output feature k, learned when the jth $F2$ node was active. β_j counts how many patterns the jth $F2$ category has learned. Output layer O has m output nodes, whose predictions are $o_k = I(b > 0) \cdot a_k/b$.

Incorrect predictions are detected by comparing a threshold ρ_O with the degree of closeness between the prediction of the network and the desired output. If an incorrect prediction is produced, a match tracking mechanism (similar to the one in FAM) is triggered. This might produce a new $F2$ node or find a more suitable node for the current input pattern.

The learning process takes place for μ_j, σ_j, η_j, α_j and β_j:

$$\eta_j(t+1) = \eta_j(t) + v_j, \ \mu_{ji}(t+1) = (1 - \eta_j^{-1} v_j)\mu_{ji}(t) + \eta_j^{-1} v_j x_i$$

$$\lambda_{ji}(t+1) = (1 - \eta_j^{-1} v_j)\lambda_{ji}(t) + \eta_j^{-1} v_j x_i^2, \ \sigma_{ji}(t+1) = \sqrt{\lambda_{ji}(t+1) - \mu_{ji}(t+1)^2}$$

$$\alpha_{kj}(t+1) = \alpha_{kj}(t) + \epsilon^{-1} v_j y_k, \ \beta_j(t+1) = \beta_j(t) + \epsilon^{-1} v_j$$

A common value $\gamma_i = \gamma_{common}$ may be used for the standard deviation in case of all input features.

An important theoretical result of AppART is [14]:

Theorem 2. *AppART with $\rho_{F2} = 0$, $\rho_O = 0$ and $\gamma_i = \gamma_{common}$, $1 \leq i \leq n$ behaves as GRNN.*

Since the GRNN can be viewed as a normalized RBF expansion, one can transitively apply to AppART two important properties of RBF networks: the universal approximation and the best approximation properties [1].

4 Experimental Results

For the first test, we consider function [10] $f : [0,1] \to [0,1]$ defined by $f(x) = (10 + \sum_{t=1}^{7} \sin(10tx))/20$.

We use independent, randomly generated datasets for training, validation and testing, consisting of 800, 200, and 1000 patterns, respectively. Each training pattern is a $(x, f(x))$ input-output pair. The testing set was not used in the training phase, but only to assess generalization performance.

The BA parameters $\overline{S^a_{MAX}}$, S^b_{MAX}, and P_{min} are optimized on the validation set by trial and error, for $\overline{S^a_{MAX}}, S^b_{MAX} \in \{10^{-3}, 5{\cdot}10^{-4}, 10^{-4}, 5{\cdot}10^{-5}, 10^{-5}, 5{\cdot}10^{-6}\}$ and $P_{min} \in \{0, 0.1, \ldots, 0.9\}$.

The BA with optimized parameters (i.e., generating the lowest RMSE on the validation set) was trained on the training+validation dataset. The generalization performance of the trained BA was assessed on the testing set in two ways (see Table 4):

1. The “BA(1)”, corresponds to a BA network with unbounded number of categories.

2. For “BA(2)”, we considered only BA models with similar number of input categories as for PROBART.

	ART_{a} categories no.	ART_{b} categories no.	$RMSE$
FAM	312	53	0.0074
PROBART	110	53	0.0169
BA(1)	185.6	57.8	0.0076
BA(2)	111.0	35.8	0.0106

Table 1: FAM, PROBART, and BA performance for regression on data generated by function f.

The RMSE for BA(1) and BA(2) were each averaged for five different runs, using each time randomly generated training, validation, and test sets. The results for PROBART and FAM are from [10]. FAM in our experiments is Carpenter's initial FAM version.

The BA(1) results are very similar to the FAM results, but for a considerably smaller number of input categories. On average, BA(2) produced one more input category than PROBART, while improving the RMSE by 40.23%. It is quite difficult to directly compare the resulted BA(2) and FAM, since BA(2) has 64.42% less input categories than the FAM.

Considering both the RMSE score and the number of input categories, we may conclude that, for this experiment, the BA performs better than the FAM and PROBART.

In the second test, we use the fifth-order chirp function [14]: $g(x) = 0.5+0.5\sin(40\pi x^5)$. Marti *et al.* have experimentally compared the function approximation performance of the following neural models [14]: AppART, Multi Layer Perceptron (MLP), RBF, General Regression Neural Network (GRNN), FAM, GAM, PROBART, and FasBack [23]. The reported score was the mean squared error (MSE). The authors run the training algorithms for several epochs. The data set consisted of 10000 points $x \in [0, 1]$, of which 70% were used for training and the rest for testing. The cited paper does not fully describe the parameter values used for each of the networks.

In our experiment, we partition a dataset of 10000 patterns into a 4000 patterns training set, a 3000 validation set, and a 3000 patterns testing set. We perform a trial and error search for $\overline{S^a_{MAX}}, S^b_{MAX} \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$, $P_{min} \in \{0, 0.1, \ldots, 0.9\}$. The values producing the best MSE on the validation set are used to train the BA on the train+validation dataset, and the testing set MSE was reported. The above procedure are repeated five times, for randomly generated datasets. We only use single epoch training.

Table 2 contains the results for MLP, RBF, GRNN, FAM, GAM, PROBART, FasBack, AppART and BA. For the first eight neural networks the results are from [14]. BA produces a very good MSE score for this regression task, most likely due to the optimized parameter values obtained by trial and error.

Comparing the MSE BA score, obtained by single epoch training, and those reported in [14], where multi-epoch training was used, we can state that the BA clearly performs better.

5 Conclusions

Theoretical universal approximation results were obtained for several FAM architectures:

- Explicit results were obtained for a slight variation of the initial FAM and for the BA.
- Implicit results, derived by association with other networks: FAMR, PROBART, and AppART.

Model	MSE	Training epochs
MLP	0.4362	30000+
RBF	0.2701	10000
GRNN	0.1540	150
FAM	0.1802	140
GAM	0.1521	45
PROBART	0.1435	50
FasBack	0.0915	10000
AppART	0.0803	30
BA	0.0086	1

Table 2: BA vs. other neural networks generalization performance for data generated by function g.

The result showing FAM networks to be universal approximators is an important fact in establishing the utility of FAM architectures. A learning algorithm which is known to be a universal approximator can he applied to a large class of interesting problems with the confidence that a solution is at least theoretically available. Experimentally, FAM architectures performed well compared to other neural function approximators.

The FAM model, as well as other universal approximators, suffer from the curse of dimensionality, as defined by Bellman [24]: an exponentially large number of ART categories may be required to reach a final solution. Therefore, the universal approximation capability of a network is an generally an existential result, not a constructive procedure to obtain a guaranteed compact network approximation of an arbitrary function. An important problem we have not addressed here is that of determining the network parameters so that a prescribed degree of approximation is achieved (see [25]).

The FAM and its offsprings are incremental learning models. Therefore, they may be used for fast approximation of massive streaming input data. This may be a serious plus when compared to other neural predictors.

How could a neural posterior probability estimator, like the BA, be used in risk assessment and decision theory? One possibility would be to combine the inferred posterior probabilities with a loss function, as suggested for a more general framework in [26]. This way, we could obtain an incremental learning risk assessment tool capable of processing fast large amounts of data.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):181-191, December, 2012.

Bibliography

[1] Girosi, F.; Poggio, T. (1989); Networks and the Best Approximation Property, *Biological Cybernetics*, 63: 169-176.

[2] Hecht-Nielsen, R. (1987); Kolmogorov's mapping neural network existence theorem, *Proceedings of IEEE First Annual International Conference on Neural Networks*, 3: III-11–III-14.

[3] Cybenko, G. (1992); Approximation by superpositions of a sigmoidal function, *Mathematics of Control, Signals, and Systems*, 5(4): 455-455.

[4] Chen, T.; Chen, H. (1995); Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems, *IEEE Transactions on Neural Networks*, 6(4): 911-917.

[5] Hartman, E; Keeler, J.D.; Kowalski, J.M. (1990); Layered neural networks with Gaussian hidden units as universal approximations, *Neural Computations*, 2(2): 210-215.

[6] Park, J.; Sandberg, I.W. (1991); *Neural Computations*, 3(2): 246-257.

[7] Park, J.; Sandberg, I.W. (1993); *Neural Computations*, Approximation and radial-basis-function networks, 5(2): 305-316.

[8] Carpenter, G.A.; Grossberg, S.; Markuzon, N.; Reynolds, J.H.; Rosen, D.B. (1992); *IEEE Transactions on Neural Networks*, Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps, 3(5): 698-713.

[9] Williamson, J. (1996); *Neural Networks*, Gaussian ARTMAP: A neural network for fast incremental learning of noisy multidimensional maps, 9:881–897.

[10] Marriott, S.; Harrison, R.F. (1995); *Neural Networks*, A modified fuzzy ARTMAP architecture for the approximation of noisy mappings, 8(4): 619-641.

[11] Andonie, R.; Sasu, L. (2006); *IEEE Transactions on Neural Networks*, Fuzzy ARTMAP with Input Relevances, 17: 929-941.

[12] Yap, K.S.; Lim, C.P. Abidi, I.Z. (2008); *IEEE Transactions on Neural Networks*, A Hybrid ART-GRNN Online Learning Neural Network With a ε-Insensitive Loss Function, 19: 1641–1646.

[13] Yap, K.S.; Lim, C.P. Junita, M.S. (2010); *Journal of Intelligen & Fuzzy Systems*, An enhanced generalized adaptive resonance theory neural network and its application to medical pattern classification, 21: 65-78.

[14] Marti, L.; Policriti, A.; Garcia, L. (2002); *Hybrid Information Systems, First International Workshop on Hybrid Intelligent Systems, Adelaide, Australia, December 11-12, 2001, Proceedings*, AppART: An ART Hybrid Stable Learning Neural Network for Universal Function Approximation, 93-119.

[15] Verzi, S.J.; Heileman, G.L.; Georgiopoulos, M.; Anagnostopoulos, G.C. (2003); *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN 2003)*, Universal Approximation with Fuzzy ART and Fuzzy ARTMAP, (3): 1987-1992.

[16] MacKay, D.J.C. (1996); *Computation in Neural Systems*, Probable networks and plausible predictions - a review of practical Bayesian methods for supervised neural networks, 6: 469 - 505.

[17] Vigdor, B.; Lerner, B. (2007); *IEEE Transactions on Neural Networks*, The Bayesian ARTMAP, 18: 1628-1644.

[18] Moore, B. (1988); *Proceedings of the 1988 Connectionist Model Summer School*, ART1 and Pattern Clustering, 174-185.

[19] Carpenter, G.A.; Grossberg, S.; Reynolds, J.H. (1991); *Neural Networks*, Fuzzy ART: fast stable learning and categorization of analog patterns by an adaptive resonance system, (4): 759-771.

[20] Lim, C.P.; Harrison, R.F. (1997); *Neural Networks*, 10(5), An Incremental Adaptive Network for On-line Supervised Learning and Probability Estimation, 925-939.

[21] Lerner, B.; Guterman, H. (2008); *Computational Intelligence Paradigms - Studies in Computational Intelligence, Springer*, Advanced Developments and Applications of the Fuzzy ARTMAP Neural Network in Pattern Classification, 137: 77-107.

[22] Sasu, L; Andonie, R. (2012); The Bayesian ARTMAP for Regression, *under review.*

[23] Izquierdo, J.M.C.; Dimitriadis, Y.A.; Coronado, J.L. (1997); *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems*, FasBack: matching-error based learning for automatic generation of fuzzy logic systems, 3: 1561 -1566.

[24] Bellman, R.E. (1961), *Rand Corporation Research studies*, Adaptive control processes: a guided tour.

[25] Andonie, R. (1997); *Dealing with Complexity: A Neural Network Approach*, The Psychological Limits of Neural Computation, 252-263.

[26] Duda, R.O.; Hart, P.E.; David G.S (2000); Pattern Classification, 2nd edition.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):192-200, December, 2012.

The Specification of ETL Transformation Operations based on Weaving Models

M. Vučković, M. Petrović, N. Turajlić, M. Stanojević

Milica Vučković, Marko Petrović,
Nina Turajlić, Milan Stanojević
University of Belgrade, Faculty of Organizational Sciences
Serbia, 11000 Belgrade, Jove Ilića 154
E-mail: {milica.vuckovic, marko.petrovic,
nina.turajlic, milan.stanojevic}@fon.bg.ac.rs

Abstract: In the ETL process the transformation of data is achieved through the execution of a set of transformation operations. The realization of this process (the order in which the transformation operations must be executed) should be preceded by a specification of the transformation process at a higher level of abstraction. The specification is given through mappings representing abstract operations specific to the transformation process. These mappings are defined through weaving models and metamodels. A generated weaving metamodel (GWMM) is proposed giving the complete mapping semantics through specific link types (representing the abstract operations) and appropriate OCL constraints. Weaving models specifying the actual mappings must be in accordance with this proposed GWMM.

Keywords: ETL process, MDD, Weaving models.

1 Introduction

In crisis management the tracking of a large amount of information (regarding people, material, financial, medical and other resources) is crucial. The establishment of a data warehouse, in which all of the relevant information could be easily stored, processed and analyzed, would enable the crisis management coordinators to make efficient decisions.

One of the most demanding phases in the data warehouse design process is the design of the process for transforming the source data into a form suitable for its further analysis (the Extract-Transform-Load process). Mistakes made during this phase may lead the whole project to failure. Since it is usually very complex and time-consuming it is necessary to provide data warehouse designers with adequate techniques to aid them in overcoming this complexity.

The ETL process consists of the Extract, Transform and Load phases. The focus of this paper is the Transform phase of the ETL process. This transformation process involves the execution of a set of operations through which the actual transformations are achieved. Most of the existing approaches directly define the order in which the transformation operations must be executed during the transformation process. We consider this to be too complex because it involves the definition of the process realization at a low level of abstraction and propose that this phase should be preceded by a specification of the transformation process at a higher level of abstraction. The main focus of this paper is the specification of the key abstract operations specific to the transformation process. These abstract transformation (AT) operations denote the semantics related to the different possible types of correspondences that exist between the source models and the target model and are the basis for the specification of mappings.

We propose an approach in accordance with Model Driven Development (MDD) which is based on the premise that the most important product of software development is not the source code itself but rather the models representing knowledge about the system that is being developed. The main goal of MDD is to automate software development through the successive application of model transformations, starting from the model representing the specification of the system and ending in a model representing the detailed description of the physical realization, from which the executable code can ultimately be generated.

In accordance with MDD a special kind of model i.e. a weaving model (WM) is used for the specification of mappings between heterogeneous models [1–3]. Through these weaving models the correspondences between individual elements of different models (called woven models) are defined. In compliance with the OMG MDA an appropriate metamodel (i.e. a weaving-metamodel WMM) is defined and weaving models must conform to it. The WMM actually defines the types of correspondences which may occur between particular concepts of concrete models. However, metamodel concepts (e.g. relational metamodel concepts) cannot be used in the definition of the WMM, hence only new link types can be defined. This implies that the corresponding WM i.e. the mappings between concepts of concrete models (e.g. a concrete relational model) cannot be semantically controlled.

In this paper mapping models and metamodels are used for specifying the AT operations. In the proposed solution we provide for the explicit introduction of metamodels into the weaving model approach as well as an appropriate metamodel for the semantic mapping between these metamodels (the generated WMM or GWMM). The AT operations are represented through the concepts of this GWMM. These concepts allow the establishing of semantic correspondences only between those metamodel concepts on which the respective abstract operations can be applied. Since weaving models must conform to the GWMM this implies that both the syntax and the semantics of the correspondences between concepts of concrete models can be controlled.

The paper is organized as follows. The next section describes the main issues of the design of the ETL processes and briefly presents the related work regarding the different approaches to its design. Section 3 describes the existing MDD approach used in the Eclipse Modeling Framework for the specification of model mappings. The proposed solution for the specification of the GWMM as well as several examples demonstrating its application, are given in Section 4. Finally, a conclusion is given detailing the main benefits of the proposed approach.

2 Design Issues

One of the main issues in data warehouse is the problem of data integration, i.e. the integration of heterogeneous data sources into a single source. To this end first a global model i.e. a reconciled model should be created giving a unified view of the relevant source data. The main benefit of this approach is that it creates a common reference data model for the whole organization [7]. After the reconciled model has been created the next step is the population of the data warehouse (DW) with the actual data. If the chosen architecture for DW design presumes that the reconciled model is materialized, first the reconciled data layer will be populated from the sources and then the DW we be populated from the reconciled data layer. We have adopted such an approach in this paper since it enables the clear separation of source data extraction and integration from DW population [7]. Therefore, the data extracted from the sources first needs to be transformed into a format compliant with the defined reconciled model.

Assuming that a single reconciled model has previously been created on the basis of the source models (e.g. by overlapping the source models) we propose that the first step in the transformation process design should be the specification of the correspondences between the source model concepts and the reconciled model concepts at the highest level of abstraction. At this stage conflicts can occur both at the structure and the instance levels (i.e. the data level). Structure level conflicts are caused by the fact that different structures, and relationships among them, are used to represent the same real world concepts in different source models. Instance level conflicts are more complex and may be the result of different granularity levels (e.g. events recorded on a daily or weekly level) or different formats in which the data is recorded.

At the structure level we can differentiate between mappings at two different granularity levels i.e. the element and attribute level. Element level mappings are used to define the correspondences between source model and reconciled model concepts by which same real world concepts are modeled. These mappings represent different abstract transformation (AT) operations (e.g. *Join*, *Equivalence*, etc.) by which the semantics of the correspondences between the related elements are defined and which are easily understandable from the end-user viewpoint. The AT operations are not executable and will be transformed, in the subsequent phases of the ETL process design, into one or more actual operations (e.g. *SQL JOIN*, *UNION*, etc.), though this transformation is not in the scope of this paper.The attribute level mappings are at a lower granularity level and give the details of the established element level mappings. Attribute mappings represent AT operations that transform the values of one or more source model attributes into one or more reconciled model attribute values (e.g. *Equals*, *Concatenate*, *Split*, *Add*, etc.).

Most of the existing approaches to ETL design proposed in literature or implemented in commercial tools do not provide concepts which allow the explicit and formal definition of the semantics of the element or attribute mappings. In [11] mechanisms for the specification of the most common ETL process operations (e.g. *Join*, *Filter*, etc.) are provided and a set of corresponding UML stereotypes is defined. These mechanisms are related through UML dependencies, and attribute mappings are defined by notes attached to the dependencies. *Data mapping diagrams* based on the Data Mapping UML Profile are introduced in [10] to trace the flow of data and are organized into four levels (*Database*, *Dataflow*, *Table* and *Attribute levels*) through the use of UML packages. At the table level only data relationships are specified and not the actual processes therefore, these mappings do not carry any semantics. At the attribute level the semantics of the mappings are given either as UML notes attached to the target attributes if the relationship is represented as an association, or, if the relationship is represented as a mapping object, by the tag definition of the mapping object. In [12] a static conceptual model of the ETL process is proposed for identifying the transformations in the ETL process and it includes the transformation entity (an abstraction of modules of code executing a single task related to filtering, cleaning or transformation operations), ETL constraints (regarding the necessary data requirements), notes (explaining the semantics of the applied functions: the type or expression/condition). In these approaches most of the transformations semantics are given through notes, often in a natural language, which does not represent a formal specification.

On the other hand, most approaches directly define the order in which the transformation operations must be executed during the transformation process i.e. the dataflow, as in [10, 11]. Also in [9] ETL process models are designed in accordance with the introduced BPMN4ETL metamodel (based on the Business Process Modeling Notation) which defines the dataflow. We consider that the specification of the transformation process realization should be preceded by a specification of the process at a higher level of abstraction. These specifications should be formal enough to enable the designer to move to lower specification levels resulting in the dataflow specification, through certain model transformations (in accordance with MDD). To this end, in [12] a static conceptual model of the ETL process is proposed which can be mapped into a logical model representing the workflow. In [8] a UML profile is proposed which introduces the *MappingOperator* stereotype (among others) to tag classes defining the functions that can be used for specifying the mapping expression of a particular mapping.

Therefore, we propose that the specification of the dynamics of the ETL process should be preceded by a static specification in which the mappings are defined through concepts which explicitly represent the semantics of the transformation operations. This specification should be formal enough to allow its transformation into the dataflow specification (in accordance with MDD), it should be extensible to allow the designers to add their own concepts, and its concepts should be easily understandable from the end-user viewpoint.

3 The AMW Approach

In accordance with the MDD motto that everything should be treated as models, the weaving model approach treats mappings between heterogeneous models also as models [1, 2] and is supported by the ATLAS Model Weaver (AMW) toolkit [3] within the Eclipse Modeling Framework (EMF) environment. In the AMW approach weaving models (WM) are used for defining the correspondences between individual concepts of different models (called woven models).

In the context of the AMW approach mapping specifications can be regarded at different abstraction levels (illustrated in Figure 1.), which actually correspond to the different levels in the OMG MDA standard [4]. Taking into account the relationships which must exist between models from different abstraction levels in the OMG MDA, a mapping model at a given level of abstraction serves as a metamodel for mapping models from the lower abstraction level. Or, in other words, every mapping model must conform to a mapping metamodel from the higher abstraction level. It should be noted that in this paper we concentrate only on the metamodel (M2) and model (M1) levels.

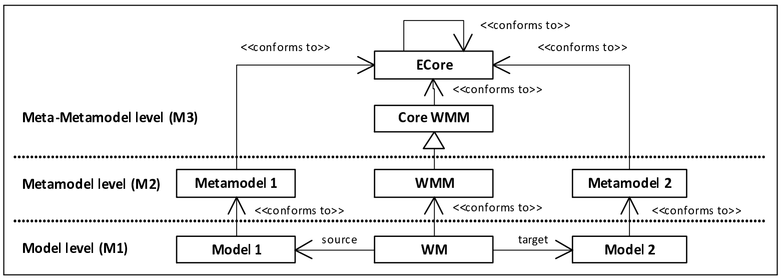


Figure 1: The AMW Approach

In accordance with the previous statements, a WM must also conform to a particular weaving metamodel (WMM). This WMM is actually an extension of the generic WMM i.e. the Core WMM, which is based on the meta-metamodel namely, the ECore meta-metamodel in the EMF environment. Since the concepts of the Core WMM do not give the definition of the specific semantics of the correspondences, it should be extended [1] to provide the semantics relevant for the specific context in which the models are used (i.e. it should be extended to include concepts specific to a certain domain). Therefore, the WMM will actually define the types of correspondences which may exist between concepts of woven models, and consequently correspondences specified in a WM must only be instances of the types defined in the WMM.

However, a WMM cannot specify the mappings between metamodel concepts (e.g. the relational metamodel, the XSD metamodel concepts etc.), it merely defines new correspondence types. More precisely, the semantics of an introduced correspondence type are given only through the definition of its name without specifying the type of metamodel concepts that may participate in that type of correspondence. Thus, one of the main drawbacks of the AMW approach is the fact that a WMM does not include the necessary semantic rules for establishing mappings between metamodels so mappings between meta-concepts cannot be defined. Instead it is only possible to define mappings between concepts of concrete models. Consequently, only the syntax of these mappings can be controlled and not their semantics. In other words, it is up to the designer to know these semantic rules and ensure they are fulfilled when defining the mappings.

In [5,6] a solution is proposed for overcoming this problem in the context of the specification of mappings between heterogeneous schemas. This solution is based on the explicit introduction of meta schemas into the WM approach as well as an appropriate WMM for the semantic mapping between the concepts of these meta schemas. The details of this approach and it application for the specification of AT operations are given in the following section.

4 The Proposed Solution

To take into account the assertions made in the previous section an extension of the Core WMM is proposed which includes the concepts necessary for providing the semantics of the correspondences in the context of the transformation process. In Figure 2. this extended WMM is shown using the UML class diagram. New mapping link types representing the identified AT operations (*Join*, *Equivalence*, *Equals*, *Concatenate*, etc.) are defined as specializations of the WLink concept. It should be emphasized that every identified AT operation is represented by a separate *MappingLink* type in our extended WMM though we have only depicted those relevant for the following examples. These new *MappingLink* types are used for the specification of mapping between the concepts of different metamodels (the actual concepts are represented by the *MappingLinkEnd* subclass of the *WLinkEnd* class). References *source* and *target* enable the defining of many-to-many mappings between the concepts of different metamodels. *MappingModelRef* and *MappingElementRef* represent references to concrete UML models and their elements.

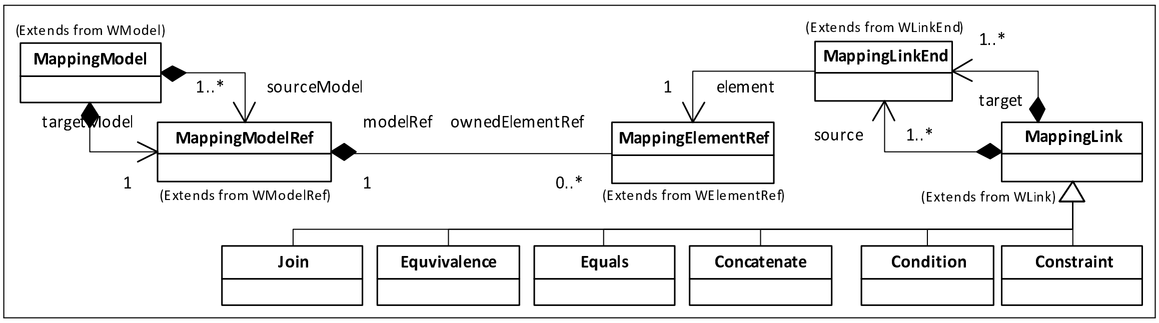


Figure 2: Extended Weaving Metamodel (simplified)

As explained in the previous section the extended WMM merely defines new mapping link types (whose semantics are only given through their names). Since each specific type of mapping assumes certain constraints regarding the number and type of model concepts that may participate in that type of mapping (e.g. the *Join* operation may only be used for mapping elements and must map at least two source model elements to a single reconciled model element) these constraints should be included in a WMM to prevent structural and semantic mistakes in WMs.

The procedure for generating the GWMM is based on the transformation given in [5, 6]. The transformation is accomplished by defining an individual WM for the mappings between metamodel concepts (the MM-WM) and then transforming it into a generated WMM with appropriate OCL constraints (e.g. for checking whether the type of model concepts involved in each mapping are valid). Specifically, a WM conforming to the proposed extended WMM is created in which the mappings are defined through a set of mapping links which are instances of the introduced mapping link types (e.g. *TableJoin* is an instance of the *Join* link type). This WM is then transformed into a new generated WMM (GWMM) in which the semantics of these mapping links are represented by appropriate OCL constraints. This GWMM now serves as a metamodel for weaving models at a lower level of abstraction (the M1 level in the OMG MDA). More precisely put, while the extended WMM defines the new mapping link types (representing the AT operations), it is the GWMM with OCL constraints that gives the complete semantics of the defined mappings (regarding the metamodel concepts on which those operations can be applied). This process is illustrated in Figure 3. It is assumed that all of the models (both source and reconciled) are expressed using the same formalism i.e. they are based on the same metamodel, which is common practice in DW design. For the purpose of this paper we have chosen the relational metamodel. The concrete models are represented as UML class diagrams using the appropriate UML stereotypes defined by the relational metamodel (the definition of these UML stereotypes is omitted from this paper due to space constraints).

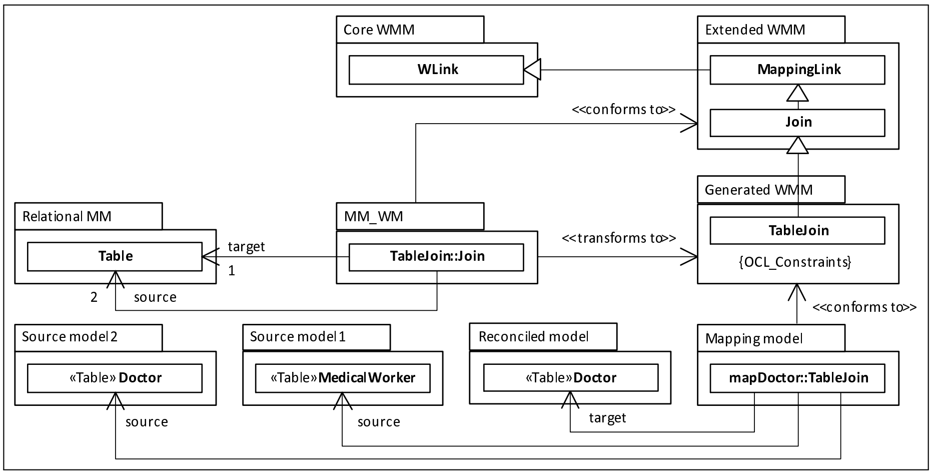


Figure 3: Proposed Solution

Next we give several examples of concrete weaving models conforming to the proposed GWMM. It should be emphasized that the first step in creating a particular WM is to define the element mappings. Subsequently, the attribute mappings are defined detailing the established element mappings. All mappings are defined by using the concepts of the proposed GWMM.

In Figure 4. an example is given illustrating the abstract *TableEquivalence* operation. The *MapDoctor* mapping is an instance of the *TableEquivalence* mapping link which states that the mapped elements *MedicalWorker* and *Doctor* are equivalent in the sense that they describe the same real world concept at the same abstraction level. The details of the *mapDoctor* element mapping are given through the child attribute mappings. The attribute mapping *mapFullName* is given by the *ColumnConcatenate* mapping link whose semantic indicates that the values of the *FirstName* and *LastName* attributes of the *MedicalWorker* element should be concatenated to obtain the value for the *FullName* attribute of the corresponding *Doctor* element. The *mapSSN* mapping is given by the *ColumnEquals* mapping link whose semantic indicates that attributes *SSN* of the *MedicalWorker* and *Doctor* elements exactly coincide.

Another example is given in Figure 4. (on the right) illustrating the situation in which the *MedicalWorker* element of the fist source model and the *Doctor* element of the second source model represent the same real world concept but record different information about it. Therefore they are represented by a single *Doctor* element in the reconciled model which includes all of the relevant information contained in both of the source models. For defining this type of mapping the abstract *TableJoin* operation is used whose semantic indicates that the *MedicalWorker* and *Doctor* elements of the two different source models should be joined to obtain all of the relevant information for the *Doctor* element in the reconciled model (the *mapDoctor* mapping link). In addition to defining the corresponding child attribute mappings it is also necessary to define the condition by which the elements should be joined. This is given by the *ColumnCondition* mapping link, also a child of the defined element mapping, which specifies that the join should be performed on the basis of the *SSN* attributes. The *ColumnConstraint* link, again, also a

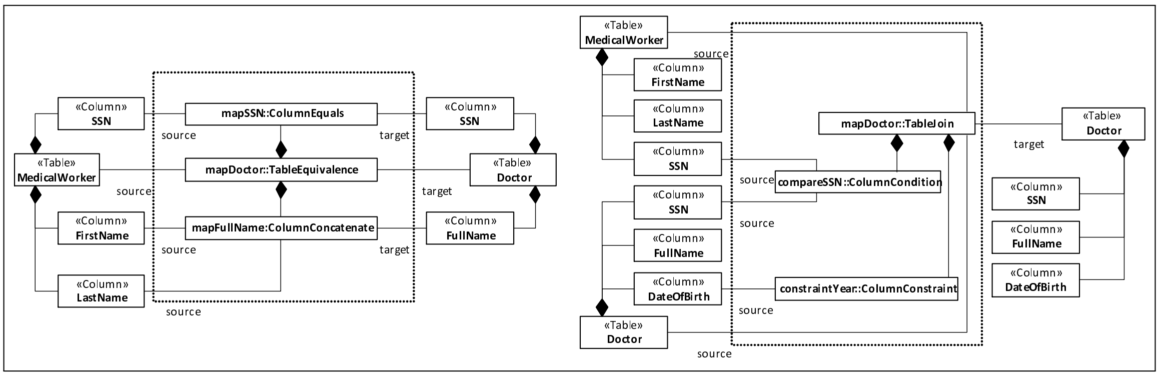


Figure 4: Illustrations of the proposed approach

child of the defined element mapping indicates that only those instances of the *Doctor* element which satisfy the defined *constraintYear* regarding the required *DateOfBirth* will be included in the transformation. The child attribute mappings giving the details of the *mapDoctor* element mapping have been omitted from because they would render the picture unnecessarly complex.

5 Conclusion

To overcome the compexity of ETL process design they should be designed gradually through the development of a series of models and the corresponding transformations between them (in accordance with MDD). The aim of this paper is to facilitate the defining of the transformation process at the highest level of abstraction. The specification of the transformation is given through mappings between the source models and the reconciled model which represent abstract operations specific to the transformation process. In the proposed solution the specification of the AT operations is based on the weaving model approch. To this end, first a description of the existing model weaving approach in the EMF environment is given and its drawbacks are discussed. Then, a solution is proposed which actually extends the existing AMW approach to overcome the identified drawbacks. The proposed solution is based on the introduction of a special generated WMM with OCL constraints through which the mapping links between concepts of concrete models are semanticaly controlled. Namely, the AT operations of the ETL process are actualy represented through appropriate semantic mapping links between metamodel concepts.Therefore, the main benefit of the proposed approach is the introduction of a formal specification of the semantics of the transformation operations giving a static view of the transformation process which is extensible and easily understandable from the end-user viewpoint. It is also platform independent so it can be used to complement other approaches dealing with the dynamic aspects of the transformation.

On the basis of these specifications the designer would, through certain model transformations (in accordance with MDD) move to lower specification levels which would result in the data flow specification i.e. the specification of the process dynamics (the AT operations would be transformed into one or more actual operations e.g. *SQL JOIN*, *UNION*, etc.). Therefore, further work in this area would be aimed at the realization of the proposed solution in the EMF environment. The specification of the AT operations given through weaving models would be used for obtaining the transformation models in a given transformation language (ATL, QVT or XSLT). These transformation models can be used for the automatic generation of code for any target platform.

Acknowledgment

The research presented in this paper was partially supported by the Ministry of Education and Science of the Republic of Serbia, Grant III-44010 and TR32013.

Bibliography

[1] Del Fabro, M.D., Bézivin, J., Jouault, F., Valduriez, P., *Applying Generic Model Management to Data Mapping.* In proc. of Base de Donnes Avances (BDA 2005), France, 2005.

[2] Del Fabro, M.D., Valduriez, P., *Semi-automatic model integration using matching transformations and weaving models.* In proc. of symposium on Applied computing, ACM, 2007.

[3] Del Fabro, M.D., Bézivin, J, and Valduriez, P., *Weaving Models with the Eclipse AMW plugin.* In: Eclipse Modeling Symposium, Eclipse Summit Europe 2006, Germany, 2006.

[4] Miller, J., Mukerji, J., *Model Driven Architecture (MDA).* http://www.omg.org; 2001.

[5] Nešković, S., Vučković, M., Aničić, N., *On Using Weaving Models to Specify Schema Mappings.* 2nd Intl Workshop on Future Trends of Model-Driven Development, Portugal, 2010.

[6] Aničić, N., Nešković, S., Vučković, M., Cvetković, R., *Specification of Data Schema Mappings using Weaving Models.* Paper accepted for publication in ComSiS Journal, March 2012.

[7] Golfarelli M., Rizzi, S., *Data Warehouse Design: Modern Principles and Methodologies.* McGraw-Hill, ISBN-13: 978-0071610391, 2009.

[8] Kurz, S., Guppenberger, M., Freitag, B., *A UML profile for modeling schema mappings.* In Proc. of the intl. conf. on Advances in Conceptual Modeling, (pp. 53–62), Springer, 2006.

[9] El Akkaoui, Z., Zimányi, E., Mazón, J.-N., Trujillo, J., *A model-driven framework for ETL process development.* In Proc. of 14th intl. workshop on DW and OLAP, ACM, 2011.

[10] Lujan-Mora, S., Vassiliadis, P., Trujillo, J., *Data Mapping Diagrams for Data Warehouse Design with UML.* In Proc. 23rd Intl. Conf. on Conceptual Modeling, Springer, 2004.

[11] Trujillo, J., Luján-Mora, S., *A UML Based Approach for Modeling ETL Processes in Data Warehouses.* Conceptual Modeling - ER 2003, (pp. 307–320), Springer-Verlag, 2003.

[12] Simitsis, A., *Mapping conceptual to logical models for ETL processes.* In Proc. of the 8th ACM international workshop on Data warehousing and OLAP, (pp. 67–76), ACM, 2005.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):201-214, December, 2012.

Error Correction Method in Classification by Using Multiple-Criteria and Multiple-Constraint Levels Linear Programming

B. Wang, Y. Shi

Bo Wang
School of Mathematical Sciences,
Graduate University of the Chinese Academy of Sciences,
Beijing 100190, China
E-mail: wangbo8014@126.com

Yong Shi (Corresponding author)
1. Research Center on Fictitious Economy & Data Science,
Chinese Academy of Sciences, Beijing 100190, China and
2. College of Information Science & Technology,
University of Nebraska at Omaha, Omaha, NE 68182, USA
E-mail: yshi@gucas.ac.cn

Abstract:
In classification based on multiple-criteria linear programming (MCLP), we need to find the optimal solution of the MCLP problem as a classifier. According to dual theory, multiple criteria can be switched to multiple constraint levels, and vice versa. A MCLP problem can be logically extended into a multiple-criteria and multiple-constraint levels linear programming (MC2LP) problem. In many applications, such as credit card account classification, how to handle two types of error is a key issue. The errors can be caused by a fixed cutoff between a "Good" group and a "Bad" group. Two types of error can be systematically corrected by using the structure of MC2LP, which allows two alterable cutoffs. In order to do so, a penalty (or cost) is imposed to find the potential solution for all possible trade-offs in solving MC2LP problem. Some correction strategies can be investigated by the solution procedure. Furthermore, a framework of decision supporting system can be illustrated for various real-life applications of the proposed method.

Keywords: Classification, Two Types of Error, Multiple-Criteria Linear Programming, Multiple-Criteria and Multiple-Constraint Levels Linear Programming, Decision Supporting System.

1 Introduction

Taking advantage of many optimization based classification algorithms, the transactions data collected by the bank can be analyzed in many ways. As a result, the profit of the new accounts can be predicted, that is, which accounts are inclined to be bankrupt and which ones have good credit. However, for many real-life problems, the accounts cannot be separated by the hyperplane in spite of using some kernel techniques [1]. Thus, how to decrease the number of misclassified samples becomes a big issue. In some applications, misclassifications have unequal importance. For example, in credit card account classification, it is essential to classify credit card customers precisely in order to provide effective services while avoiding losses due to bankruptcy from users' debt. Actually, even 0.01 percent increase in early detection of bad accounts can save millions, while losing a good account does not influence much [2], [3].

Linear programming (LP) is a useful tool for discriminating analysis of a problem given appropriate groups (e.g., "Good" and "Bad") [4]. And multiple-criteria linear programming

(MCLP) has improved the result by minimizing the sum external deviations and maximizing the sum of the internal deviations simultaneously. Mostly, the cutoff of MCLP is fixed to be a given number (e.g., 1), while this will cause some other problems. For instance, it cannot involve those possible cases that can achieve the ideal cutoff score to be zero. Formally, this means that the solutions obtained by linear programming are not invariant under linear transformations of the data [2], [3]. Particularly, it is not invariant under vector addition. Furthermore, if the classes of the samples exchange, i.e. "Good" and "Bad" classes swap to each other, the solutions are different.

Noticing these problems, some researchers made many efforts on this topic. Consequently, a new model based on multiple-criteria and multiple-constraint levels linear programming (MC2LP) was posed [5], [6]. However, these methods were domain-driven, which meant they needed some domain knowledge to help finding the best cutoff. Nevertheless, our new models are all automatically solving. As a result, it is useful to solve the problems that no domain knowledge is prepared. In particular, it solves classification problem twice. The maximal external deviation is found for the first time, while MC2LP is exploited to search for the optimal hyperplane based on minimizing two types of error for the second time.

In addition to this, we also provide another MC2LP based model for the purpose of errors correction. In this model, we fix the cutoff to be 1 whereas we also add two more hyperplanes to detect the misclassified points carefully. Accordingly, a subtle discussion is involved regarding the relationship between two types of error and the deviations . In fact, in the statistics theory, two types of error influence on each other oppositely. In other words, reducing of Type I error will cause the increasing of Type II error, and vice versa. Hence, we focus on different types of error respectively. Moreover, some more elaborate introductions are demonstrated in the next sections.

2 Multiple-criteria and Multiple-constraint Levels Linear Programming for Classification

2.1 The MCLP model for classification

Given a set of n variables about the records $X^T = (x_1, x_2, ..., x_l)$, and then let $x_i = (x_{i1}, x_{i2}, ..., x_{in})^T$ be one sample of data, where $i = 1, 2, ..., l$ and l is the sample size. In linear discriminant analysis, data separation can be achieved by two opposite objectives, that is, minimizing the sum of the deviations (MSD) and maximizing the minimum distances (MMD) of observations from the critical value [4]. That is to say, in order to solve classification problem, we need to minimize the overlapping of data, i.e. α, at the same time, to maximize the distances from the well classified points to the hyperplane, i.e. β.

However, it is difficult for traditional linear programming to optimize MMD and MSD simultaneously. According to the concept of Pareto optimality, we can check all the possible trade-offs between the objective functions by using multiple-criteria linear programming algorithm.

The MCLP model can be described by Figure 1:

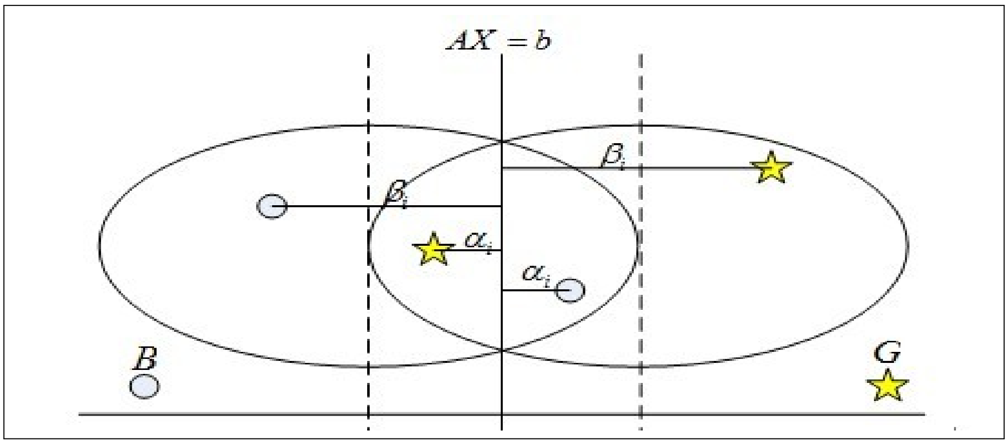


Figure 1: MCLP model

Moreover, the first Multiple Criteria Linear Programming (MCLP) model can be described as follows:

$$min \sum_i \alpha_i$$

$$max \sum_i \beta_i$$

$$\begin{aligned} s.t. \quad & A_iX = b + \alpha_i - \beta_i, \quad A_i \in Bad, \\ & A_iX = b - \alpha_i + \beta_i, \quad A_i \in Good, \\ & \alpha_i, \beta_i \geqslant 0, i = 1, 2, ..., l \end{aligned}$$

Here, α_i is the overlapping and β_i is the distance from the training sample x_i to the discriminator $(w \cdot x_i) = b$ (classification separating hyperplane).

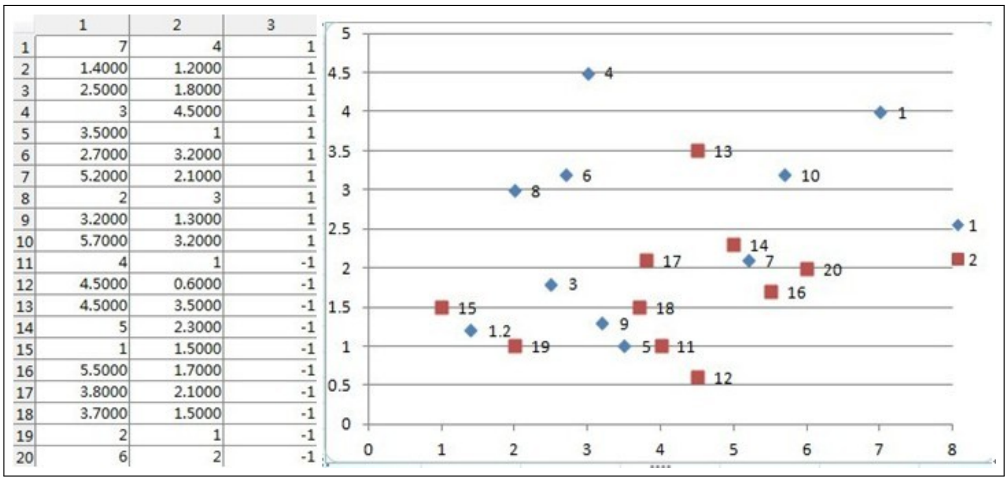
2.2 Discussion of the cutoff b

A boundary value b (cutoff) is often used to separate two groups, where b is unrestricted. A problem caused by treating b as a variable will meet many cases of no solution. For some applications, the user can choose a fixed value of b ($b = 1$) to get a solution as the classifier. As a result, efforts to improve the accuracy rate of classification have been greatly confined to the unrestricted characteristics of b (that is, a given b is put into calculation to find coefficients w) according to the user's experience facing the real time data set [3].

In such procedure, the goal of finding the optimal solution for classification question is replaced by the task of testing boundary b. That is to say, if b is given, we find a classifier using an optimal solution by solving the model above.

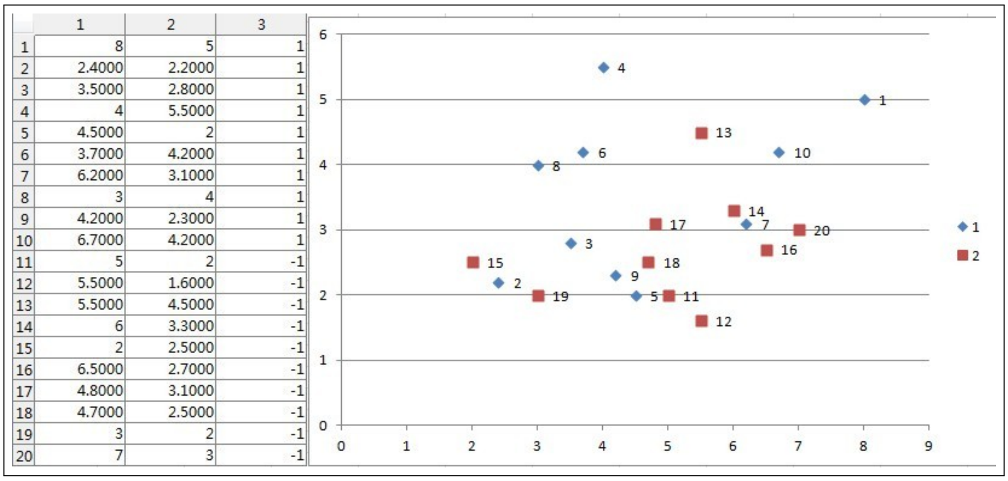
However, now we will point out the drawbacks of handling model in this way. At first, fixing $b = 1$ will make the solutions different under vector addition (one kind of linear transformations). Besides, the solution will change if we swap the classes of "Bad" and "Good", which seems to be illogical. Thus, we will introduce the examples below to illustrate these two issues in the rest of this part.

Example 1. *In figure 2, there are 20 points (10 points belong to class -1, 10 points belong to class 1). On the left, there are coordinates of the points. On the right, the distribution of the points is displayed. MCLP is applied to this data set and the solution is* $X = (0.58, -0.77)$.



	1	2	3
1	7	4	1
2	1.4000	1.2000	1
3	2.5000	1.8000	1
4	3	4.5000	1
5	3.5000	1	1
6	2.7000	3.2000	1
7	5.2000	2.1000	1
8	2	3	1
9	3.2000	1.3000	1
10	5.7000	3.2000	1
11	4	1	-1
12	4.5000	0.6000	-1
13	4.5000	3.5000	-1
14	5	2.3000	-1
15	1	1.5000	-1
16	5.5000	1.7000	-1
17	3.8000	2.1000	-1
18	3.7000	1.5000	-1
19	2	1	-1
20	6	2	-1

Figure 2: data1



	1	2	3
1	8	5	1
2	2.4000	2.2000	1
3	3.5000	2.8000	1
4	4	5.5000	1
5	4.5000	2	1
6	3.7000	4.2000	1
7	6.2000	3.1000	1
8	3	4	1
9	4.2000	2.3000	1
10	6.7000	4.2000	1
11	5	2	-1
12	5.5000	1.6000	-1
13	5.5000	4.5000	-1
14	6	3.3000	-1
15	2	2.5000	-1
16	6.5000	2.7000	-1
17	4.8000	3.1000	-1
18	4.7000	2.5000	-1
19	3	2	-1
20	7	3	-1

Figure 3: data2

And then, we transform the data by moving the points along the vector $(1, 1)$*, and the new data coordinates and the distribution are shown in figure 3. It is obvious that the samples' distribution is the same as the former ones', which means the optimal hyperplane will parallel to the former one. Conversely, the result obtained by MCLP is* $X = (0.4, -0.4)$*, which is quite different from the former one.*

In addition, if the classes are swapped, the result will change into $X = (-0.15, 0.77)$*, which should be the same as the first one. For simplicity, in these experiments, the weight between two criterions is always fixed to be* $(4, 1)$*.*

The discussion above shows that the result won't be invariant under linear transformation. This means when the data change, we cannot just keep b to be a fixed number, say 1, and then solve the MCLP problem for different data set. Moreover, we need to take advantage of MC2LP to trace a better b to decrease the number of errors.

2.3 The MC2LP model for classification

According to the discussion above, a non-fixed b is very important to our problem. At the same time, for the simplicity and existence of the solution, b should be fixed in some interval.

As a result, for different data, we fix b in different pairs of interval $[b_l, b_u]$, where b_l and b_u are two fixed numbers. Now our problem is to search the best cutoff between b_l and b_u at every level of their trade-offs, that is to say, to test every point in the interval $[b_l, b_u]$. We keep the multiple-criteria the same as MCLP, that is, MMD and MSD. And then, the following model is posed [3]:

$$
\begin{aligned}
& min \sum_i \alpha_i \\
& max \sum_i \beta_i \\
s.t. \quad & A_iX = [b_l, b_u] + \alpha_i - \beta_i, \quad A_i \in Bad, \\
& A_iX = [b_l, b_u] - \alpha_i + \beta_i, \quad A_i \in Good, \\
& \alpha_i, \beta_i \geqslant 0, i = 1, 2, ..., l
\end{aligned}
$$

where A_i, b_l and b_u are given, and X is unrestricted.

In the model, $[b_l, b_u]$ represents a certain trade-off in the interval. By virtue of the technical of Multiple-criteria and multiple-constraint levels linear programming (MC2LP), we can test each trade-off between the multiple-criteria and multiple-constraint levels as follows:

$$
\begin{aligned}
& min\ \lambda_1 \sum_i \alpha_i - \lambda_2 \sum_i \beta_i \\
s.t. \quad & A_iX = \gamma_1 b_l + \gamma_2 b_u + \alpha_i - \beta_i, \quad A_i \in Bad, \\
& A_iX = \gamma_1 b_l + \gamma_2 b_u - \alpha_i + \beta_i, \quad A_i \in Good, \\
& \alpha_i, \beta_i \geqslant 0, i = 1, 2, ..., l
\end{aligned}
$$

Here, the parameters of $\lambda \times \gamma$ are fixed for each programming problem. Moreover, the advantage of MC2LP is that it can find the potential solutions for all possible trade-offs in the parameter space systematically [7] [8], where the parameter space is

$$\{(\lambda, \gamma) | \lambda_1 + \lambda_2 = 1, \gamma_1 + \gamma_2 = 1\}.$$

Of course, in this model, choosing a suitable pair for the goal problem is a key issue and needs domain knowledge. Consequently, a non-parameter choosing MC2LP method should be posed.

3 A New Two Alterable Cutoffs Model based on MC2LP

3.1 A framework of the new MC2LP model

For the original MCLP model, one cutoff is used to predict a new sample's class, that is to say, there is only one hyperplane. The former MC2LP model points out that we can define two cutoffs instead of the original single cutoff. And then a systematical method can be used to solve this problem. Consequently, all potential solutions at each constrain level trade-off can be acquired. However, one problem is how to find the cutoffs, that is, b_l and b_u.

On one hand, we utilize two cutoffs to discover the solution of higher accuracy; on the other hand, we hope the cutoffs can be obtained from the system directly. Inspired by the idea above, we address our first MC2LP model, which solves the classification problem twice.

For the first step, MCLP model is used to find the vector of external deviations α. It is a function of λ. For simplicity, we set $b = 1$. And then, we fix the parameter of λ to get one potential solution. Now a non-parameter vector of external deviations α is acquired. The component ($\alpha_i > 0$) means the corresponding sample in the training set is misclassified. In other words, Type I and Type II errors occur. According to the idea of MC2LP, we can detect the result of every single MCLP by fixing the parameter of γ at each level in the interval $[b_l, b_u]$. Now, we find the maximal component of α:

$$\alpha_{max} = max\{\alpha_i, 1 \leq i \leq l\}. \tag{1}$$

Indeed, the smaller the weight of external deviations is, the bigger α_{max} is.

The misclassified samples are all projected into the interval $[1 - \alpha_{max}, 1 + \alpha_{max}]$ according to the weight vector X obtained from the MCLP model. In this way, we define b_l and b_u as $1 - \alpha_{max}$ and $1 + \alpha_{max}$, respectively. It is easy to see, if we want to lessen the number of two types of error, in effect, we just need to inspect the cutoffs by altering the cutoff in the interval $[1 - \alpha_{max}, 1 + \alpha_{max}]$.

Moreover, for the second step, a new MC2LP classification model can be stated as follows:

$$min\ \lambda_1 \sum_i \alpha_i - \lambda_2 \sum_i \beta_i$$

$$\begin{aligned} s.t. \quad & A_iX = [1 - \alpha_{max}, 1 + \alpha_{max}] + \alpha_i - \beta_i, \quad A_i \in Bad, \\ & A_iX = [1 - \alpha_{max}, 1 + \alpha_{max}] - \alpha_i + \beta_i, \quad A_i \in Good, \\ & \alpha_i, \beta_i \geqslant 0, i = 1, 2, ..., l \end{aligned}$$

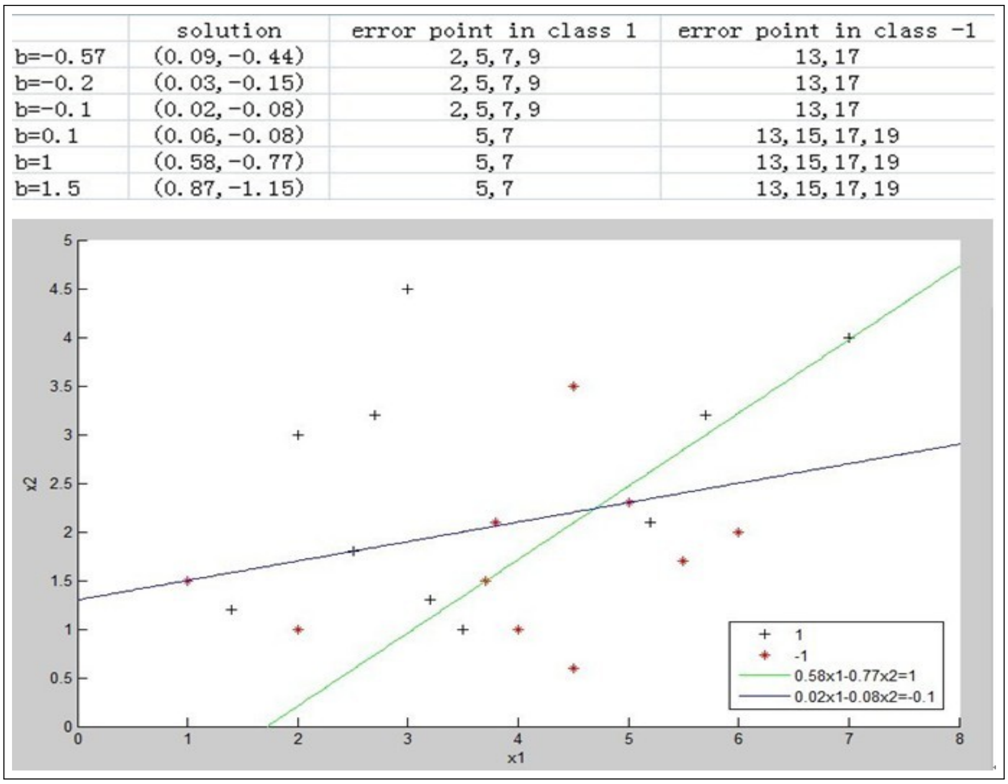
where A_i, α_{max} are given, and X is unrestricted, $[1 - \alpha_{max}, 1 + \alpha_{max}]$ means a certain trade-off in the interval. At the same time, $\lambda = (\lambda_1, \lambda_2)$ is the parameter chosen in the first step.

Claim 2. Furthermore, all the notations as $[a, b]$ in the models displayed in this paper mean a certain trade-off in the interval, where a and b are two real numbers.

3.2 Discussion of the new MC2LP model

The most direct modification of the new MC2LP model is to transfer the single objective function to be a multiple-criteria one. Because the vector of external deviations is a function of λ, it is easy to observe that if the weight between external deviations and internal deviations changes, α changes. Consequently, α_{max} alters. And the ideal α is the one that makes α_{max} not too huge. In other words, we do not hope to check the weight that satisfies λ_1 not too small. Actually, some papers have proved that only if $\lambda_1 > \lambda_2$, then $\alpha \cdot \beta = 0$, which makes the model meaningful [9]. As a result, we only need to check the parameters of objective functions that make α_{max} not too big, in short, not too far away from the original one.

On the other hand, we expect α_{max} not too small. That is to say, we hope the model has some generalization. Hence, two small positive numbers ϵ_1 and ϵ_2 are chosen manually. And then, the interval is builded as $[[1-\alpha_{max}-\epsilon_1, 1-\alpha_{max}+\epsilon_1], [1+\alpha_{max}-\epsilon_2, 1+\alpha_{max}+\epsilon_2]]$. This means that the lower and the upper bound of the interval should be trade-off of some intervals, i.e. the multiple-constrained levels are actually multiple-constrained intervals. Indeed, checking every trade-off of the intervals is the same as checking every trade-off of $1 - \alpha_{max} - \epsilon_1$ and $1 + \alpha_{max} + \epsilon_2$. In this case, we can consider the objective function as a multiple-criteria one. It



	solution	error point in class 1	error point in class -1
b=-0.57	(0.09,-0.44)	2,5,7,9	13,17
b=-0.2	(0.03,-0.15)	2,5,7,9	13,17
b=-0.1	(0.02,-0.08)	2,5,7,9	13,17
b=0.1	(0.06,-0.08)	5,7	13,15,17,19
b=1	(0.58,-0.77)	5,7	13,15,17,19
b=1.5	(0.87,-1.15)	5,7	13,15,17,19

Figure 4: the solutions for MC2LP

can be stated as follows:

$$
\begin{aligned}
& min \sum_i \alpha_i \\
& max \sum_i \beta_i \\
s.t.\ \ & A_iX = [1-\alpha_{max}-\epsilon_1, 1+\alpha_{max}+\epsilon_2] + \alpha_i - \beta_i, \quad A_i \in Bad, \\
& A_iX = [1-\alpha_{max}-\epsilon_1, 1+\alpha_{max}+\epsilon_2] - \alpha_i + \beta_i, \;\; A_i \in Good, \\
& \alpha_i, \beta_i \geqslant 0, i = 1, 2, ..., l
\end{aligned} \tag{2}
$$

where A_i, α_{max}, ϵ_1 and ϵ_2 are given, and X is unrestricted. Here, ϵ_1 and ϵ_2 are two nonnegative number.

Example 3. *A numeric experiment will be introduced to demonstrate that our new MC2LP model is effective. The data is the same as data1 that has been displayed in figure 2. As we have shown above, the solution for MCLP is* $X = (0.58, -0.77)$. *According to the result, we find there are 6 misclassified points in the data, that is, point 5, 7 for class 1 and point 13, 15, 17, 19 for class -1. Moreover, according to the solution,* α_{max} *is 1.57. Then, we keep* $\epsilon_1 = \epsilon_2 = 0$. *As a result,* $[b_l, b_u] = [-0.57, 2.57]$ *is obtained. Next, some trade-offs between multiple-constraint levels have been chosen and the solutions are listed below. For consistency, we keep the weight between the objective function to be* $(4, 1)$. *The result is laid out in Figure 4. The green hyperplane* $0.02 * x_1 - 0.08 * x_2 = 1$ *is obtained from negative cutoff. From the solutions, we can see that different trade-offs between the constraint levels yield different results, while the solutions will be the same when b is set to be the same sign. That is to say, for all the positive b, the solutions generate the same hyperplane, and then, for all the negative b, the solutions create the same hyperplane. Moreover, we want to prove this result as a lemma.*

Lemma 4. *For certain trade-off between the objective functions, if b is maintained to be the same sign, then hyperplanes, which are obtained in the MCLP model , keep the same. Furthermore, different signs result in different hyperplanes.*

Proof: Let's assume that the trade-off between the objective functions is $\lambda = (\lambda_1, \lambda_2)$ and X_1 is the solution obtained by fixing b to be 1. And then, set b_1 as an arbitrary positive number. The MCLP model can be transformed as follows:

$$min\ \lambda_1 \sum_i \alpha_i - \lambda_2 \sum_i \beta_i$$

$$\begin{aligned} s.t. \quad & A_i X = b_1 + \alpha_i - \beta_i, \quad A_i \in Bad, \\ & A_i X = b_1 - \alpha_i + \beta_i, \quad A_i \in Good, \\ & \alpha_i, \beta_i \geqslant 0, i = 1, 2, ..., l \end{aligned}$$

The problem above is the same as:

$$min\ \lambda_1 \frac{\sum_i \alpha_i}{b_1} - \lambda_2 \frac{\sum_i \beta_i}{b_1}$$

$$\begin{aligned} s.t. \quad & A_i \frac{X}{b_1} = 1 + \frac{\alpha_i}{b_1} - \frac{\beta_i}{b_1}, \quad A_i \in Bad, \\ & A_i \frac{X}{b_1} = 1 - \frac{\alpha_i}{b_1} + \frac{\beta_i}{b_1}, \quad A_i \in Good, \\ & \alpha_i, \beta_i \geqslant 0, i = 1, 2, ..., l \end{aligned}$$

And then, we let $\alpha'_i = \frac{\alpha_i}{b_1}, \beta'_i = \frac{\alpha_i}{b_1}, X' = \frac{X}{b_1}$. It is obvious that the solution is $X' = \frac{X_1}{b_1}$ and the hyperplane $AX' = b_1$ is the same as $AX_1 = 1$.

Similarly, we can prove that when b is a negative number, the solution is the same as the one that is obtained from $b = -1$.

As a result, we just need to compare the solutions (hyperplanes) resulted from $b = 1$ and $b = -1$. For this case, it is easy to see that the signs before α_i and β_i swap when we transform $b = -1$ into $b = 1$. If this happens, then the objective function changes into $-\lambda_1 \sum_i \alpha_i + \lambda_2 \sum_i \beta_i$. This means that the solutions will be different. □

According to the lemma, we have the theorem below:

Theorem 5. *For our MC2LP model (2) above, according to the solutions (hyperplanes), space γ is divided into two non-intersect parts.*

Remark 6. When $[1 - \alpha_{max}, 1 + \alpha_{max}]$ is achieved, ϵ_1 and ϵ_2 are chosen to satisfy that 0 is contained by the interval $[1 - \alpha_{max} - \epsilon_1, 1 + \alpha_{max} + \epsilon_2]$. In this case, for any λ, the solutions belong to the trade-offs with same sign will result in the same hyperplane. In other words, there are only two different hyperplanes corresponding to model (2). In short, the flexibility of model (2) is limited.

4 A New Model based on Correcting of Two Types of Error

In many classification models, including original MCLP model, two types of error is a big issue. In credit card account classification, to correct two types of error can not only improve the accuracy of classification but also help to find some important accounts.

Accordingly, many researchers have focused on this topic. Based on this consideration, more attention should be paid to the samples that locate between two hyperplanes acquired by the original MCLP model, that is, the points in the grey zone [10]. Consequently, we define the external deviations and internal deviations related to two different hyperplanes, the left one and the right one, that is, α^l, α^r, β^l and β^r.

Definition 7. The conditions the deviations should satisfy are stated as follows:

$$\alpha_i^l = \begin{cases} 0, & A_iX < 1-\alpha_{max} \ and \ A_i \in Bad; \\ A_iX-(1-\alpha_{max}), & A_iX \geq 1-\alpha_{max} \ and \ A_i \in Bad; \\ 0, & A_iX \geq 1-\alpha_{max} \ and \ A_i \in Good; \\ (1-\alpha_{max})-A_iX, & A_iX < 1-\alpha_{max} \ and \ A_i \in Good. \end{cases}$$

$$\alpha_i^r = \begin{cases} 0, & A_iX < 1+\alpha_{max} \ and \ A_i \in Bad; \\ A_iX-(1+\alpha_{max}), & A_iX \geq 1+\alpha_{max} \ and \ A_i \in Bad; \\ 0, & A_iX \geq 1+\alpha_{max} \ and \ A_i \in Good; \\ (1+\alpha_{max})-A_iX, & A_iX < 1+\alpha_{max} \ and \ A_i \in Good. \end{cases}$$

$$\beta_i^l = \begin{cases} (1-\alpha_{max})-A_iX, & A_iX < 1-\alpha_{max} \ and \ A_i \in Bad; \\ 0, & A_iX \geq 1-\alpha_{max} \ and \ A_i \in Bad; \\ A_iX-(1-\alpha_{max}), & A_iX \geq 1-\alpha_{max} \ and \ A_i \in Good; \\ 0, & A_iX < 1-\alpha_{max} \ and \ A_i \in Good. \end{cases}$$

$$\beta_i^r = \begin{cases} (1+\alpha_{max})-A_iX, & A_iX < 1+\alpha_{max} \ and \ A_i \in Bad; \\ 0, & A_iX \geq 1+\alpha_{max} \ and \ A_i \in Bad; \\ A_iX-(1+\alpha_{max}), & A_iX \geq 1+\alpha_{max} \ and \ A_i \in Good; \\ 0, & A_iX < 1+\alpha_{max} \ and \ A_i \in Good. \end{cases}$$

Figure 5 is a sketch for the model. In the graph, the green and the red lines are the left and right hyperplane, b^l and b^r respectively, which are some trade-offs in two intervals, i.e. $[1-\alpha_{max}-\epsilon_2, 1]$ and $[1, 1+\alpha_{max}+\epsilon_1]$. And all the deviations are measured according to them in different colors. For instance, if a sample in "Good" class is misclassified as "Bad" class, it means $\alpha_i^r > \beta_i^l \geq 0$ and $\alpha_i^l = \beta_i^r = 0$. And then, if a sample in "Bad" class is misclassified as "Good" class, it means $\alpha_i^l > \beta_i^r \geq 0$ and $\alpha_i^r = \beta_i^l = 0$. Thus, for the misclassified ones, $\alpha_i^r + \alpha_i^l - \beta_i^r - \beta_i^l$ should be minimized.

As a result, a more meticulous model could be stated as follows:

$$min \ \sum_i (\alpha_i^r + \alpha_i^l)$$

$$min \ \sum_i (\alpha_i^l - \beta_i^r)$$

$$min \ \sum_i (\alpha_i^r - \beta_i^l)$$

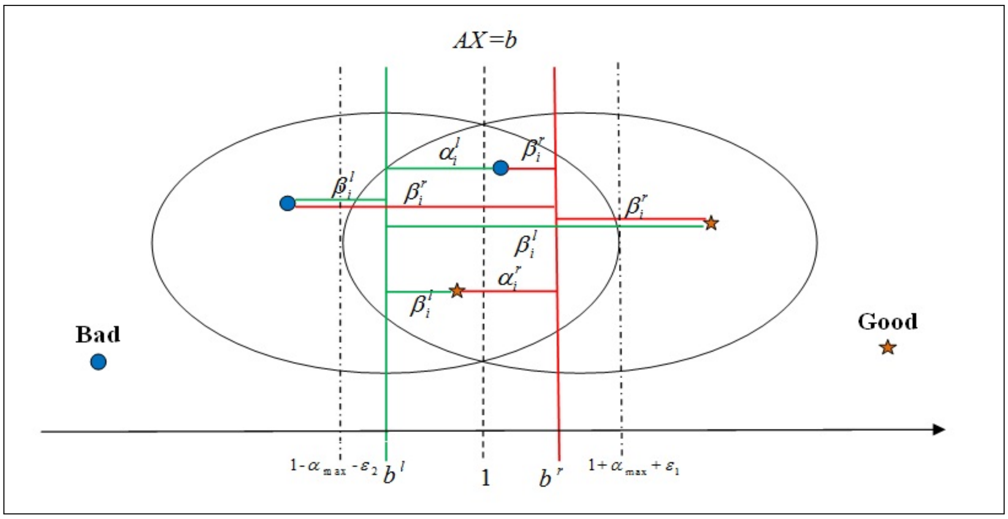


Figure 5: MC2LP model

$$
\begin{aligned}
& max \ \sum_i (\beta_i^r + \beta_i^l) \\
s.t. \quad & A_iX = 1 + [0, \alpha_{max} + \epsilon_1] + \alpha_i^r - \beta_i^r, \quad A_i \in Bad, \\
& A_iX = 1 - [0, \alpha_{max} + \epsilon_2] + \alpha_i^l - \beta_i^l, \quad A_i \in Bad, \\
& A_iX = 1 + [0, \alpha_{max} + \epsilon_1] - \alpha_i^r + \beta_i^r, \quad A_i \in Good, \\
& A_iX = 1 - [0, \alpha_{max} + \epsilon_2] - \alpha_i^l + \beta_i^l, \quad A_i \in Good, \\
& \alpha_i^r, \alpha_i^l, \beta_i^r, \beta_i^l \geqslant 0, i = 1, 2, ..., l.
\end{aligned}
$$

where A_i, α_{max}, $\epsilon_1 > 0$, $\epsilon_2 > 0$ are given, and X is unrestricted.

In Figure 5, for each point, there are at most two kinds of deviations nonzero. The objective functions appear to deal with the deviations according to the position shown in Figure 5, respectively, whereas they have their own special meaning. That is to say, it measures two types of error in some degree by means of the second and third objective functions. As a result, in this new version of MC2LP, we not only consider the deviations respectively, but also take the relationship of the deviations based on two types of error into account in the objective functions. By virtue of MC2LP method, each trade-off between $1 - \alpha_{max} - \epsilon_2$ and 1 for the left hyperplane as well as each trade-off between 1 and $1 + \alpha_{max} + \epsilon_1$ for the right hyperplane can be checked.

After obtaining the weight vector X of the hyperplane, $AX = 1$ is still used to be the classification hyperplane. However, in our new model, we minimize the distance between the left hyperplane and the right one. In other words, we discover the hyperplane that genders the smallest grey area.

Example 8. *We show how the new MC2LP model based on error correction works. The data is the same as data1 that has been displayed in figure 2. Two kinds of situations will be considered in this example, that is, the pairs of hyperplanes are $AX = 0.57, AX = 2.57$ and $AX = -6, AX = 10$. We choose ϵ_1, ϵ_2 bigger than 10, so that we can get the pair of hyperplanes $AX = -6, AX = 10$. From the new model, we obtain two hyperplanes, which are displayed in Figure 6. The blue hyperplane is obtained from $AX = 0.57, AX = 2.57$, while the green one is corresponding to $AX = -6, AX = 10$. In both experiments, we fix the weight between α_i^l, α_i^r and β_i^l, β_i^r to be (4,1). According to the result, we can see that the point marked is a misclassified sample corresponding to the former pair, while it is classified correctly based on the latter one.*

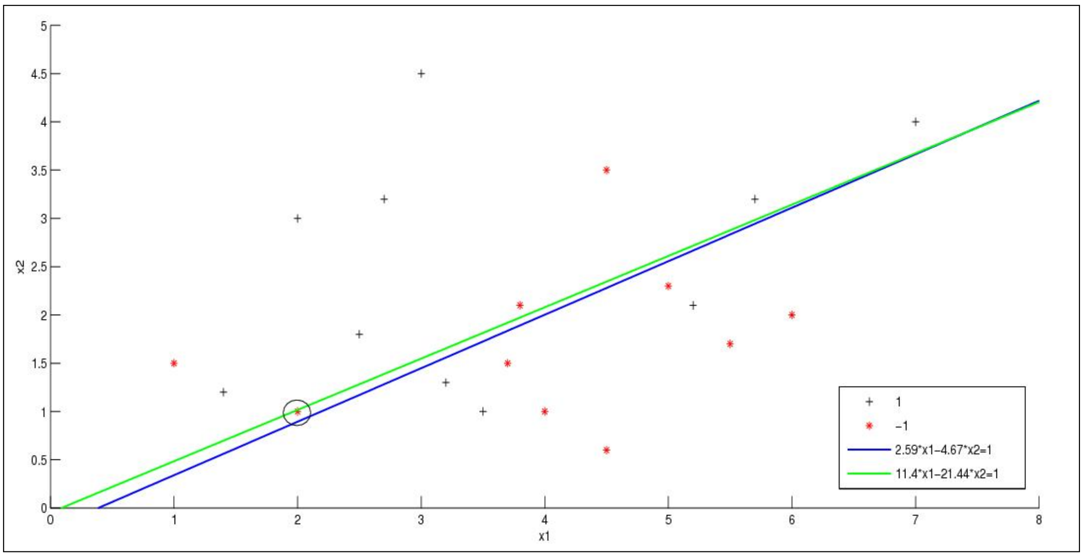


Figure 6: solutions for new MC2LP model

Actually, in statistics, Type I and Type II errors are two opposite objectives. That is to say, it is very hard to correct both of them at the same time. As a result, we modify the former model into two different models focusing on two types of error respectively as follows:

$$\begin{aligned}
&min \ \sum_i(\alpha_i^r+\alpha_i^l)\\
&min \sum_i(\alpha_i^l-\beta_i^r)\\
&max \ \sum_i(\beta_i^r+\beta_i^r)\\
s.t. \quad & A_iX=1+[0,\alpha_{max}+\epsilon]+\alpha_i^r-\beta_i^r, \quad A_i\in Bad,\\
& A_iX=1+\alpha_i^l-\beta_i^l, \qquad\qquad\qquad A_i\in Bad,\\
& A_iX=1+[0,\alpha_{max}+\epsilon]-\alpha_i^r+\beta_i^r, \quad A_i\in Good,\\
& A_iX=1-\alpha_i^l+\beta_i^l, \qquad\qquad\qquad A_i\in Good,\\
& \alpha_i^r,\alpha_i^l,\beta_i^r,\beta_i^l\geqslant 0, i=1,2,...,l.
\end{aligned} \tag{3}$$

where A_i, α_{max} and $\epsilon>0$ are given, and X is unrestricted. In this model, $\sum_i(\alpha_i^r-\beta_i^l)$ is not contained in the objective functions. This model can deal with Type II error, that is, classifying a "Good" point to be a "Bad" one. Now we provide an example to illustrate the effect of model (3).

Example 9. *The data set is the same as data1 that has been displayed in figure 2. Here, the label "-1" is regarded as "Good" class, and the label "1" is regarded as "Bad" class. Two kinds of situations will be considered in this example, that is, the pairs of hyperplanes are $AX=1, AX=2.57$ and $AX=1, AX=8$. The left hyperplane is always kept to be $AX=1$, which is also the hyperplane that is used to be the classification hyperplane at last. We set ϵ as 10. It is obvious that more remarkable result will be obtained from bigger ϵ. After fixing the weight of the objective functions as (4,0.1,1), the outcome is shown in Figure 7. The blue hyperplane is obtained from $AX=1, AX=2.57$, while the green one is corresponding to $AX=1, AX=8$. Two marked points are the ones corrected by model (3) after the right hyperplane moving to the right.*

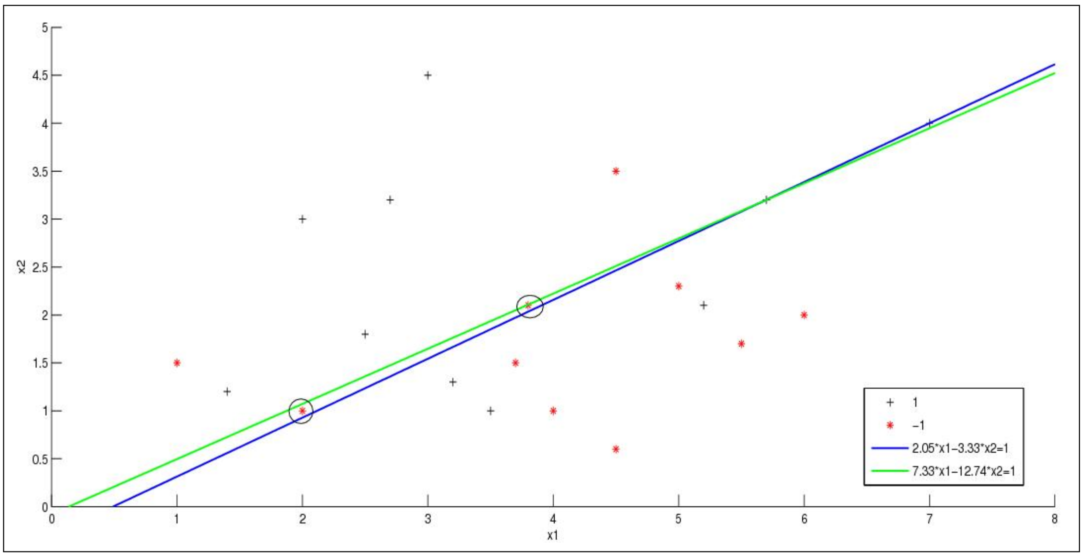


Figure 7: solutions for new MC2LP model (3)

As the result shown above, model (3) can correct Type II error in some degree. We conclude this in the proposition below.

Proposition 10. *Model (3) can correct Type II error by moving the right hyperplane to the right based on the concept of multiple-constraint levels.*

Remark 11. The second objective function in model (3) is nonzero for the samples in class "Bad" and getting negative when the right hyperplane moving to the right. That is to say, we tolerate some Type I errors. At the same time, the first objective function in model (3) renders Type II errors an increasing punishment with moving the right hyperplane to the right. As a result, it can correct Type II error in some degree.

Similarly to model (3), we pose model (4) which can deal with Type I error as follows:

$$
\begin{aligned}
&min \sum_i (\alpha_i^r + \alpha_i^l) \\
&min \sum_i (\alpha_i^r - \beta_i^l) \\
&max \sum_i (\beta_i^r + \beta_i^r) \\
s.t.\quad & A_i X = 1 + \alpha_i^r - \beta_i^r, && A_i \in Bad, \\
& A_i X = 1 - [0, \alpha_{max} + \epsilon] + \alpha_i^l - \beta_i^l, && A_i \in Bad, \\
& A_i X = 1 - \alpha_i^r + \beta_i^r, && A_i \in Good, \\
& A_i X = 1 - [0, \alpha_{max} + \epsilon] - \alpha_i^l + \beta_i^l, && A_i \in Good, \\
& \alpha_i^r, \alpha_i^l, \beta_i^r, \beta_i^l \geqslant 0, i = 1, 2, ..., l.
\end{aligned}
\tag{4}
$$

where A_i, α_{max} and $\epsilon > 0$ are given, and X is unrestricted. In this model, $\sum_i (\alpha_i^l - \beta_i^r)$ is not contained in the objective functions. This model focuses on Type I error, that is, classifying a "Bad" point to be a "Good" one.

Example 12. *In order to compare model (3) and model (4), we draw two hyperplanes obtained from these two models in the same graph. Figure 8 is the result of using pair $AX = 1, AX = 8$ for model (3) and pair $AX = -0.57, AX = 1$ for model (4).*

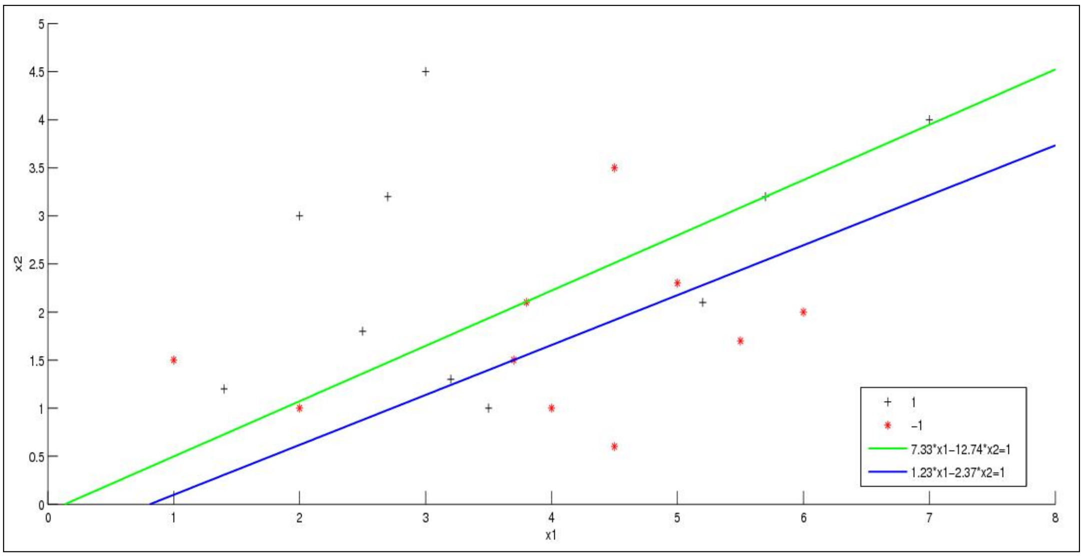


Figure 8: comparison of model (3) (4)

In figure 8, the green hyperplane is obtained from model (3) with less Type II errors, while the blue one is acquired from model (4) with less Type I errors.

In the last two models introduced above, the first one defines $AX = 1$ as the left hyperplane, at the same time, $AX = 1 + [0, \alpha_{max} + \epsilon]$ as the right one. Oppositely, the second model defines $AX = 1 - [0, \alpha_{max} + \epsilon]$ as the left hyperplane, and then, $AX = 1$ as the right one. For each trade-off, our model can deal with the misclassifications. In this way, two types of error can be corrected respectively.

5 Conclusion

This paper focuses on how to correct two types of error. At first, the disadvantage of MCLP by fixing the cutoff is addressed. And then, a MC2LP based model is introduced. According to the topic of error correction, we develop MC2LP by measuring the deviations of the misclassified points more sophisticatedly. In this way, we obtain a brand new hyperplane that is different from the one obtained by MCLP model. Moreover, focusing on correcting two types of error, a result with less errors is gained.

Acknowledgement

This work has been partially supported by grants from National Natural Science Foundation of China (No. 70921061, No. 71110107026) and the CAS/SAFEA International Partnership Program for Creative Research Teams.

Bibliography

[1] Zhang Z., Zhang D., Tian Y., Shi Y., *Kernel-based Multiple Criteria Linear Programming Classifier*, Procedia CS 1(1): 2407-2415, 2010.

[2] Thomas L.C., Edelman D.B., Crook J.N., *Credit Scoring and Its Applications*, SIAM, 2002.

[3] He J., Zhang Y., Shi Y., Huang G., *Domain-Driven Classification Based on Multiple Criteria and Multiple Constraint-Level Programming for Intelligent Credit Scoring*, IEEE Transactions on Knowledge and Data Engineering, vol. 22, No. 6: 826-838, 2010.

[4] Freed N., Glover F., *Simple but Powerful Goal Programming Models for Discriminant Problems*, European J. Operational Research, vol. 7: 44-60, 1981.

[5] Shi Y., Tian Y., Kou G., Peng Y., Li J., *Optimization Based Data Mining: Theory and Applications, Advanced Information and Knowledge Processing*, Springer, 2011.

[6] Shi Y., *Multiple Criteria Optimization-based Data Mining Methods and Applications: A Systematic Survey*, Knowl Inf Syst, 24: 369-391, 2010.

[7] Shi Y., *Multiple Criteria and Multiple Constraint Levels Linear Programming: Concepts, Techniques and Applications*, World Scientific, 2001.

[8] Shi Y., He J., Wang L., Fan W., *Computer-based Algorithms for Multiple Criteria and Multiple Constraint Level Integer Linear Programming*, Comput. Math. Appl. 49(5): 903-921, 2005.

[9] Nakayama H., Yun Y., *Generating Support Vector Machines using Multiobjective Optimization and Goal Programming*, Studies in Computational Intelligence, vol. 16: 173-198, 2006.

[10] Chen Y., Zhang L., Shi Y., *Post Mining of Multiple Criteria Linear Programming Classification Model for Actionable Knowledge in Credit Card Churning Management*, ICDMW, IEEE Computer Society: 204-211, 2011.

INT J COMPUT COMMUN, ISSN 1841-9836
7(5):215-225, December, 2012.

A Fault-Tolerant Scheduling Algorithm using Hybrid Overloading Technology for Dynamic Grouping based Multiprocessor Systems

X.-B. Yu, J.-S. Zhao, C.-W. Zheng, X.-H. Hu

Yu Xing-biao
1.Institute of Software Chinese Academy of Sciences
4# South Fourth Street, Zhongguancun, Beijing, P.R. China
2.Graduate University of Chinese Academy of Sciences
80# Zhongguancun East Road, Haidian District, Beijing, P.R. China
E-mail: eliteman5317@hotmail.com

Zhao Jun-suo, Zheng Chang-wen, Hu Xiao-hui
Institute of Software Chinese Academy of Sciences
4# South Fourth Street, Zhongguancun, Beijing, P.R. China

Abstract:
In order to extend the application area of fault-tolerant scheduling algorithm based on hybrid overloading for multiprocessor and increase the fault-tolerant number of processors, we propose a new fault-tolerant scheduling algorithm, which is based on hybrid overloading and dynamic grouping for multiprocessor by combining logic grouping strategy for processors in primary backup overloading and backup backup overloading.This algorithm presents the formalization of the dynamic grouping for processors in fault-tolerant scheduling based on hybrid overloading and enlarges the task number included in overloading task link. In the process of fault-tolerant scheduling the processors are dynamically divided into some groups based on overloading task link, so as to keep good scheduling success ratio and enhance the fault-tolerant performance of processors. Both theoretical analysis and simulation experiment prove this algorithm's effectiveness respectively.

Keywords: dynamic grouping; fault-tolerant; overloading; primary backup; backup backup

1 Introduction

Real-time embedded system has been applied in many fields such as military, aeronautics, astronautics and communication, and the relevant research also has made important progress. The result of task scheduling in real-time system lies on not only scheduling correctness but also time restriction. The multiprocessor platform of real-time system takes advantage of the technology of resource redundancy and time redundancy in order to meet the demand of scheduling correctness and system reliability, among them primary-backup overloading(PB) [1, 2] and backup-backup overloading(BB) [2–4]approach is most important one. Supposing that there is just a processor fault in same period, the scheduling time of different versions of different tasks is overloaded in scheduling period of same processor for PB overloading , and the scheduling time of backup-backup versions of different tasks is overloaded in scheduling period of same processor for BB overloading. The fault-tolerant scheduling technology of hybrid overloading [2, 6–8] is the combination of PB overloading and BB overloading with better scheduling efficiency and fault-tolerant performance of processor, but still supposing there is only a processor fault in same period. Logic grouping strategy [1] for processor based on PB overloading and BB overloading dynamically divides processors into groups in the process of task scheduling so as to tolerate a processor fault in every group, which is more adaptable to apply practically. But this strategy yet supposes the task number of overloading task chain is not more than two tasks, therefore the assigning of grouping size and group number is not enough flexible to be suited for hybrid overloading. In order to simplify the application of overloading method in scheduling algorithm, some basic application principles of overloading technology were introduced. [2]

In this paper Dynamic Grouping PB-BB Algorithm(DG_PB-BB Algorithm) combines the advantage of two overloading scheduling technology and logic grouping in the precondition of application principle for overloading technology, not only efficiently improving the efficiency of task scheduling but also enlarging tolerance area of the process. DG_PB-BB Algorithm has better application value than no grouping algorithm and logic grouping algorithm.

2 System Model

System is composed of m same processor nodes based on real time multiprocessor platform with shared memory, centralized dispatcher and processor communication medium without communication cost. Processor responsible for scheduling is dispatcher and responsible for execution is executer, which runs in parallel with dispatcher. New real time task is received by dispatcher and centralized dispatched to form assignment queue to execute on each processor.

Real time tasks to be scheduled are independent aperiodic and nonpreemptalbe scheduling with primary-backup copy technology. Each task t_i has two versions, namely primary copy(pr_i) and backup copy(bk_i) with identical attributes and resource requirements. Task sets $\tau = \{t_i | i = 1, 2, \cdots, n\}$, t_i=(a_i,r_i ,c_i,d_i), a_i is the start time of task t_i, r_i is the ready time of task t_i, c_i is the maximal execution time of task t_i, d_i is the deadline time of task t_i. Processor sets $\omega = \{p_i$ $| i = 1, 2, \cdots, m\}$. $pr_i \rightarrow bk_i$ is task chain, if other tasks are overloading scheduled on task t_i, then it is overloading task chain. The parameters of system model are defined as follows:

Definition 1. $st(t_i)$ is the start scheduling time of task t_i. $ft(t_i)$ is the finish scheduling time of task t_i. $r_i \leq st(pr_i) \leq ft(pr_i) \leq st(bk_i) \leq ft(bk_i) \leq d_i$.

Definition 2. $proc(pr_i)$ is the processor on which the primary task is scheduled, $proc(bk_i)$ is the processor on which the backup task is scheduled, $proc(pr_i) \neq proc(bk_i)$.

Definition 3. $s(t_i)$ is the time interval on which the primary or backup task is scheduled. $s(pr_i) \cap s(bk_i) = \phi$.

Definition 4. $n_{cascade}$ is the cascade number of overloaded tasks within a processor group and a time slot. $groupsize(g_i)$ is the processor number of a processor group g_i. $n_{cascade}$=1 represents the scheduling assignment of no task overloading, $n_{cascade}$= groupsize(g_i) represents the task can not be overloading scheduled.$1 \leq n_{cascade} \leq groupsize(g_i)$.

Definition 5. $t_{overload}$ is the part time of a task overloaded on other tasks. $0 \leq t_{overload} \leq c_i$.

3 A Fault-tolerant Scheduling Algorithm Using Hybrid Overloading Technology for Dynamic Grouping Based Multiprocessor Systems(DG_PB-BB)

This paper states the strategy of dynamic grouping of fault-tolerant processor based on fault-tolerant scheduling technology of hybrid overloading and the finishing of task scheduling with the help of AP(Allocation Parameter) [5,6]algorithm. In DG_PB-BB Algorithm the system can tolerate more than two processor faults at same moment and the tasks of overloading chain can not again be limited as two numbers.

3.1 Validity checking

1. Backups can be overloaded on any task. Primaries can be overloaded only on backups.

2. If a primary pr_i is overloaded on a backup bk_j, then st(pr_i)≥ft(pr_j).

3. A overloading task chain should not be looped. A overloading task chain is relating to some processors and the maximal primary number of a overloading task chain is groupsize(g_i)-1 in the group g_i.

4. No more than one processor is belong to different processor groups. Every processor group has at least two processors.

5. The size of every processor group is possibly unequal and dynamically changes in the process of task scheduling.

6. The backups and primaries of same task are scheduled on same processor group. Overloading scheduling of task copy can only happen within same processor group.

7. If proc(pr_i)=proc(pr_j) within same processor group, then s(bk_i)∩s(bk_j)=ϕ, stating the scheduling time of task bk_i and bk_j can not be overloading.

3.2 Scheduling algorithm

Next specifically describing DG_PB-BB algorithm. In DG_PB-BB algorithm processor grouping is concluded as four conditions, formalizably described as following:

1. If the copy of task t_i, pr_i and bk_i are not overloaded on other tasks,then the processors occupied by two copies form a new processor group g_e.
   g_e=form_group(proc(pr_i),proc(bk_i)).

2. If pr_i is not overloaded on other tasks, but bk_i is, then the task chain in which task t_k overloaded on bk_i is situated links with the task chain in which $pr_i \rightarrow bk_i$ is situated to extend as a new processor group g_e.
   (proc(pr_k),proc(bk_k))∈group(g_e),
   expand_group(g_e,proc(pr_i)).

3. If bk_i is not overloaded on other tasks, but pr_i is, then the task chain in which task t_k overloaded on pr_i is situated links with the task chain in which $pr_i \rightarrow bk_i$ is situated to extend as a new processor group g_e.
   (proc(pr_k),proc(bk_k))∈group(g_e),
   expand_group(g_e,proc(bk_i)).

4. If both of pr_i and bk_i is overloaded on other tasks, then the task chain in which task t_k overloaded on bk_i is situated links with the task chain in which task t_j overloaded on pr_i is situated to extend as a new processor group g_e.
   (proc(pr_j),proc(bk_j))∈group(g_e),
   (proc(pr_k),proc(bk_k))∈group(g_f),
   g_e=expand_group(g_e,g_f).

Fig.1 states four conditions of dynamic processor grouping. OTC is overloading task chain. OTC(A) represents task chain $pr_1 \rightarrow bk_1$ and the processors occupied by it form a new processor group, which is the first condition of dynamic processor grouping. Within OTC(B_1), the task chain t_4 is situated in links with $pr_3 \rightarrow bk_3$ task chain and the processors occupied by it form

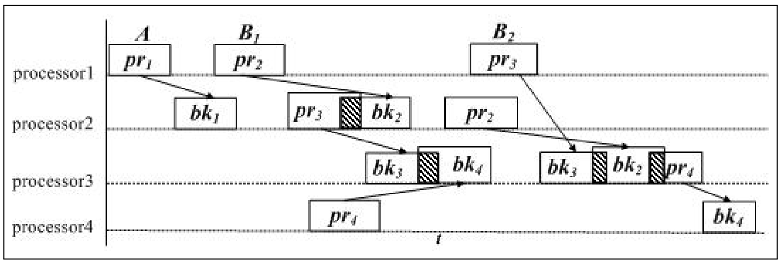


Figure 1: dynamic processor grouping

a new processor group, showing the second condition of dynamic processor grouping. Within OTC(B_2), the task chain t_4 is situated in links with task chain $pr_2 \rightarrow bk_2$ and the processors occupied by it form a new processor group, expressing the third condition of dynamic processor grouping. Within OTC(B_1), task chain $pr_3 \rightarrow bk_3$ links with both task chain t_2 and t_4, and the processors occupied by it form a new processor group, which is the fourth condition of dynamic processor grouping.

DG_PB-BB algorithm is based on AP scheduling strategy of hybrid overloading. New tasks form ready scheduling queue according to FCFS(First Come First Server) method. Bigger The distance of start scheduling time between primaries and backups, smaller the effect of backups to task receiving ratio, and it means start scheduling time of primaries is as early as possible, st(pr_i)$\rightarrow r_i$, finish scheduling time of backups is as late as possible, ft(bk_i)$\rightarrow d_i$. In order to increase task receiving ratio, tasks contained in task sets of hybrid overloading should be as much as possible and scheduling time occupied by them should be as few as possible. In the process of task scheduling the strategy of dynamic processor grouping is adopted to enhance tolerant level of the processor. The principle of processor grouping is the processors occupied by every overloading task chain constitute one group, if new overloading task chain has no relationship with old overloading task chain, then it should add a new processor to schedule new overloading task chain again to form a new processor group.

If e=groupsize(g_e),then:

$$AP[pr_i, p_j, ft(pr_i)] = \begin{cases} \frac{d_i - ft(pr_i)}{d_i - r_i}\frac{1}{e} & n_{cascade} = 1 \\ \frac{d_i - ft(pr_i)_i}{d_i - r_i}\frac{n_{cascade}}{e}\frac{t_{overload}}{c_i} & 1 < n_{cascade} \leq e \end{cases}$$

$$AP[pr_i, p_j] = \max\{AP[pr_i, p_j, ft(pr_i)]\}$$

$$AP[bk_i, p_j, st(bk_i)] = \begin{cases} \frac{st(bk_i) - r_i}{d_i - r_i}\frac{1}{e} & n_{cascade} = 1 \\ \frac{st(bk_i) - r_i}{d_i - r_i}\frac{n_{cascade}}{e}\frac{t_{overload}}{c_i} & 1 < n_{cascade} \leq e \end{cases}$$

$$AP[bk_i, p_j] = \max\{AP[bk_i, p_j, st(bk_i)]\}$$

AP[pr_i,p_j,ft(pr_i)] is evaluation factor of scheduling scheme that a new primary pr_i is assigned to schedule on processor j. similarly,AP[bk_i,p_j,ft(bk_i)] is evaluation factor of scheduling scheme that a new backup bk_i is assigned to schedule on processor j. AP is bigger, showing that the scheme of scheduling assignment is more optimal and scheduling efficiency is better.

DG_PB-BB Algorithm

1. The copy of first task, pr_i and bk_i are assigned respectively processor p_1 and p_2 to form a new group g_e. g_e is present processor group, G is assigned processor group.
   schedule(pr_i)$\rightarrow p_1$,schedule(bk_i)$\rightarrow p_2$,g_e=form_group(p_1,p_2),validity(),G=g_e.

2. **(1)** Within assigned processor group G, next task t_i finishes scheduling with the adoption

of scheduling technology of hybrid overloading and allocation parameter.

```
for task $t_i$,
while validity()$\neq$ failed
AP($pr_i$,$p_{j1}$)=next[max(AP($pr_i$,$p_j$))]or AP($bk_i$,$p_{j2}$)=next[max(AP($bk_i$,$p_j$))],
$p_j \in$ group($G$),$j = 1, 2, \cdots , m$,$j_1 \neq j_2$,
schedule($pr_i$)$\rightarrow p_{j1}$,schedule($bk_i$)$\rightarrow p_{j2}$,validity().
```

(2) If task t_i can not be scheduled within assigned processor group, then it should be to add a new processor into assigned processor group G to schedule task t_i again and judge whether new and old overloading task chain can be united according to four conditions. Otherwise to judge whether task chain $pr_i \rightarrow bk_i$ and task chain included in processor group G are separated each other, and whether processor group g_e formed by task chain in which task t_i is situated is really contained in group G. If it is true, then group g_e should be decomposed, otherwise task t_i is dealt with according four conditions. After processors are assigned completely, if new task can not be scheduled, then fault-tolerant scheduling algorithm of dynamic grouping is started again according to principle of processor grouping.

```
if schedule($t_i$)=failed in $G$
then {for new processor $p_k$, $G$=$p_k \cup G$,go to (1) (constraint $p_k$=$p_{j1} \vee p_k$=$p_{j2}$)
    if (combine(link($pr_i \rightarrow bk_i$),(group($G$) $\rightarrow$ $link$))=failed)

  then condition 1,validity()
      else  condition $i$,validity(),go to (1) until $j_1 \vee$ $j_2$=$m$}
else
    {if ((combine(link($pr_i \rightarrow bk_i$),(group($G$) $\rightarrow$ $link$))=failed)  and
    $form\_group$($pr_i$,$bk_i$)$\in$group(G))

  then decompose(form_group($pr_i$, $bk_i$))
      else  condition $i$,validity(),go to (1) until $j_1 \vee j_2$=$m$}
```

Next giving an example of DG_PB-BB algorithm to express the thinking of algorithm.T_1=(2, 2,2,7.5),T_2=(4,4,2,8),T_3=(4.5,4.5,2,9),T_4=(5,5,2,10),T_5=(5,5,4,13), T_6=(6,6,3.5,14.5),m=6. Fig.2 describes DG_PB- BB algorithm scheduling example of above queue.

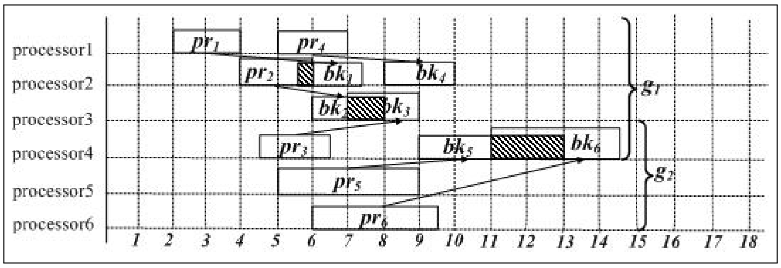


Figure 2: DG_PB-BB algorithm

Firstly m=2, processor p_1 and p_2 form group g_1.St(pr_2)=4 ,m=3 ,proc(bk_2)=p_3, pr_2 and bk_1 can be PB overloading ,p_3 is added into group g_1.st(pr_3)=4.5,if m=3, pr_3 and bk_3 can not be scheduled within group g_1, then m=4, bk_3 and bk_2 are BB overloading, proc(pr_3)=p_4, p_4 is added into group g_1. st(pr_4)=5,m=4, pr_4 and bk_4 can be scheduled within group g_1,proc(pr_4)=p_1, proc(bk_4)=p_2,and task chain $pr_4 \rightarrow bk_4$ is separated from old overloading task chain within group g_1 and really included in group g_1, so that task chain $pr_4 \rightarrow bk_4$ is combined into group g_1. st(pr_5)=5,m=4,pr_5 and bk_5 can not be scheduled within group g_1, so it is to extend a new processor p_5 and rescheduling pr_5 and bk_5 within processor scheduling sets composed of p_5 and group g_1, proc(pr_5)=p_5, proc(bk_5)=p_4.As a result, task chain $pr_5 \rightarrow bk_5$ is separated from old overloading task chain within group g_1 and not really included in group g_1, therefore it is to extend task chain $pr_5 \rightarrow bk_5$ as group g_2. st(pr_6)=6, task t_i can not be scheduled neither in group g_1 nor group g_2, so that a new processor p_6 is added into group g_2 and task t_i is rescheduled within extended group g_2, proc(pr_6)=p_6, proc(bk_6)=p_4, bk_5 and bk_6 is BB overloaded on processor p_4. All processors have been assigned completely so that DG_PB-BB algorithm is restarted from processor p_1 while there are new tasks arriving into scheduling queue.

3.3 Algorithm analysis

If task sets showed in Fig.2 are scheduled by AP scheduling algorithm of hybrid overloading without grouping(PB-BB_AP Algorithm),although it only needs four processors to finish scheduling, but can just tolerate a processor fault and is too strict to apply widely. DG_PB-BB algorithm needs to increase two processors to finish scheduling, but new processor group g_2 in system makes fault-tolerant number of processor extend as two processors so as to strengthen the reliability of system. Time complexity of PB-BB_AP algorithm is $O[N^2 \cdot m \cdot (m-1)]$, N is average task number of task sets on which a processor has ever scheduled, m is processor number of the system. If in DG_PB-BB algorithm average processor number of group g_e is k, then time complexity of DG_PB-BB algorithm is $O[N^2 \cdot k \cdot (k-1)]$.In regard to DG_PB-BB algorithm, comparing with PB-BB_AP algorithm, the algorithm cost decreases and the system reliability increases, but the guarantee ratio decreases, because PB-BB_AP algorithm is an ideal method.

3.4 Theory testification

The theory testification follows the methodology used in [8] and is based on the following assumptions:

- All tasks have unit worst execution time, for example: c_i=1.
- Backup slots are preallocated in the schedule.
- FIFO scheduling strategy is used.
- Task deadlines follow uniform distribution [W_{min},W_{max}],called deadline window. If $P_{win}(w)$ is the probability that an arriving task has a relative deadline w, then $P_{win}(w) = 1/(W_{max} - W_{min} + 1), W_{min} \leq w \leq W_{max}$.
- Task arrivals follow uniform distribution[0,A_{max}], with mean A_{av}=$A_{max}/2$. If $P_{ar}(k)$ is the probability of k tasks arriving at a given time, then $P_{ar}(k) = 1/(A_{max}+1), 0 \leq k \leq A_{max}$.

A simple pre-allocation policy for BB overloading is to reserve a slot for backups every n time slots on each processor. Backup slots on the three processors can be staggered. For a task t_i, bk_i is scheduled immediately after pr_i with probability 0.5 and is scheduled two slots later than pr_i with probability 0.5.

In PB overloading there are three different types of time(0,1 and 2), if (t-1)mod 3=i, any time t has a type of i. At any time t, the number of primaries that can be scheduled to start at that time is s_0 if t is of type 0, s_1 if t is of type 1, s_2 if t is of type 2.

Using FIFO scheduling is equal to maintaining a task queue, to which arriving tasks are appended. Given that the number of task that can be scheduled on each time unit is known, then the position of a task in the Q indicates its scheduled start time. In BB overloading two tasks can be scheduled on each time (one slot is reserved for backups). If at the beginning of time slot t, a task t_i is the qth task in Q, then t_i is scheduled to execute at time slot $t+g_q^{BB}$. g_q^{BB} is the time at which a task, whose position in the Q is $q\,(q = 1, 2, \cdots, 2W_{\max})$, will be executed and is defined as $g_q^{BB} = \lfloor \frac{q}{2} \rfloor$. In PB overloading s_0,s_1 and s_2 tasks can be scheduled on a given time slot t depending on whether t is of type 0,1,or 2 respectively. The time g_q^{PB} is defined as $g_q^{PB} =(i+j+\mathrm{l}), \sum_{c=1}^{i} s_0 + \sum_{c=1}^{j} s_1 + \sum_{c=1}^{l} s_2 \le q-1, |i-j| \le 1, |j-l| \le 1, |l-i| \le 1$. where $i \ge j \ge l$ if t is of type 0, $j \ge l \ge i$ if t is of type 1, and $l \ge i \ge j$ if t is of type 2. When a task t_i arrives at time t, its schedulability depends on the length of Q and on the relative deadline w_i of the task. In BB overloading, if t_i is appended at position q of Q and $w_i \ge g_q^{BB}$, then the primary task pr_i is guaranteed to execute before $t+w_i$, Moreover, if $w_i \ge g_q^{BB} + 2$, then bk_i is also guaranteed to execute before time $t+w_i$. In PB overloading, if t_i is appended at position q of Q and $w_i \ge g_q^{PB}$, then the primary task pr_i is guaranteed to execute before $t+w_i$, Moreover, if $w_i \ge g_q^{PB} + 2$, then bk_i is also guaranteed to execute before time $t+w_i$. Let $p_{q,k}$ be the probability that one of the k tasks is rejected when the queue size is q, and its value is the probability that the relative deadline of the task is smaller than $g_b^{*}+\delta$,*=PB or BB, δ=1 or 2.

g_b^{logic} is the time at which a task, whose position in the Q is b, will be executed in the PB-BB overloading scheduling strategy of logic grouping, $b = q + k/2$. Showing as t_0-t_{12} in Fig.3, processor p_1-p_{12} are divided into 4 groups and every group has 3 processors, adopting with scheduling strategy of BB overloading, processor p_{13}-p_{24} are also divided into 4 group and every group has 3 processors, adopting with scheduling strategy of PB overloading. Above method is described as PB-BB overloading scheduling strategy of logic grouping.

$$g_b^{logic} = \frac{q+k/2}{\left\lfloor (i+j+l)(n-n/3)+\sum_{c=1}^{i} s_0+\sum_{c=1}^{j} s_1+\sum_{c=1}^{l} s_2 \right\rfloor},$$

$$\sum_{c=1}^{i} s_0 + \sum_{c=1}^{j} s_1 + \sum_{c=1}^{l} s_2 \le q + k/2 - 1, |i-j| \le 1, |j-l| \le 1, |l-i| \le 1.$$

$g_b^{dynamic}$ is the time at which a task, whose position in the Q is b, will be executed in the PB-BB overloading scheduling strategy of dynamic grouping, $b = q + k/2$. Different with logic grouping, in dynamic grouping s_0=3, s_1=4 and s_2=2. Describe as t_{12}-t_{24} in Fig.3, processor p_1-p_{12} is divided into 3 groups and every group has 4 processors, adopting with scheduling strategy of BB overloading, processor p_{13}-p_{24} are also divided 3 groups and every group has 4 processors, adopting with scheduling strategy of PB overloading, Above method is described as PB-BB overloading scheduling strategy of dynamic grouping.

$$g_b^{dynamic} = \frac{q+k/2}{\left\lfloor (i+j+l)(n-n/4)+\sum_{c=1}^{i} s_0+\sum_{c=1}^{j} s_1+\sum_{c=1}^{l} s_2 \right\rfloor}, \sum_{c=1}^{i} s_0 + \sum_{c=1}^{j} s_1 + \sum_{c=1}^{l} s_2 \le q + k/2 - 1, |i-j| \le 1, |j-l| \le 1, |l-i| \le 1.$$

Obviously, $g_b^{dynamic} < g_b^{logic}$, $g_b^{dynamic}$ decreases more quickly than g_b^{logic} with increasing n, therefore DG_PB-BB algorithm is more efficient than LG_PB-BB algorithm with increasing n.

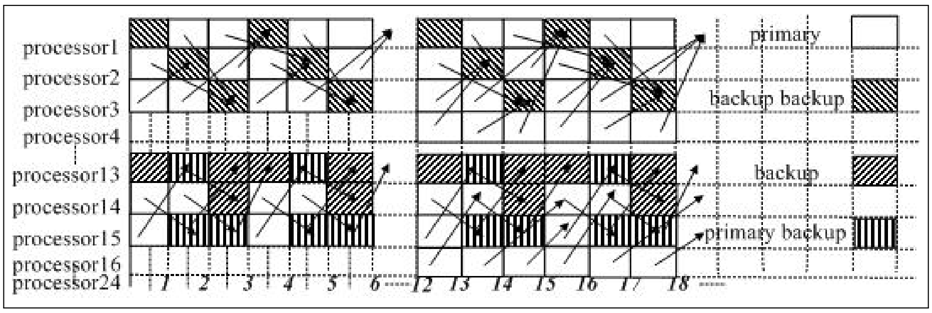


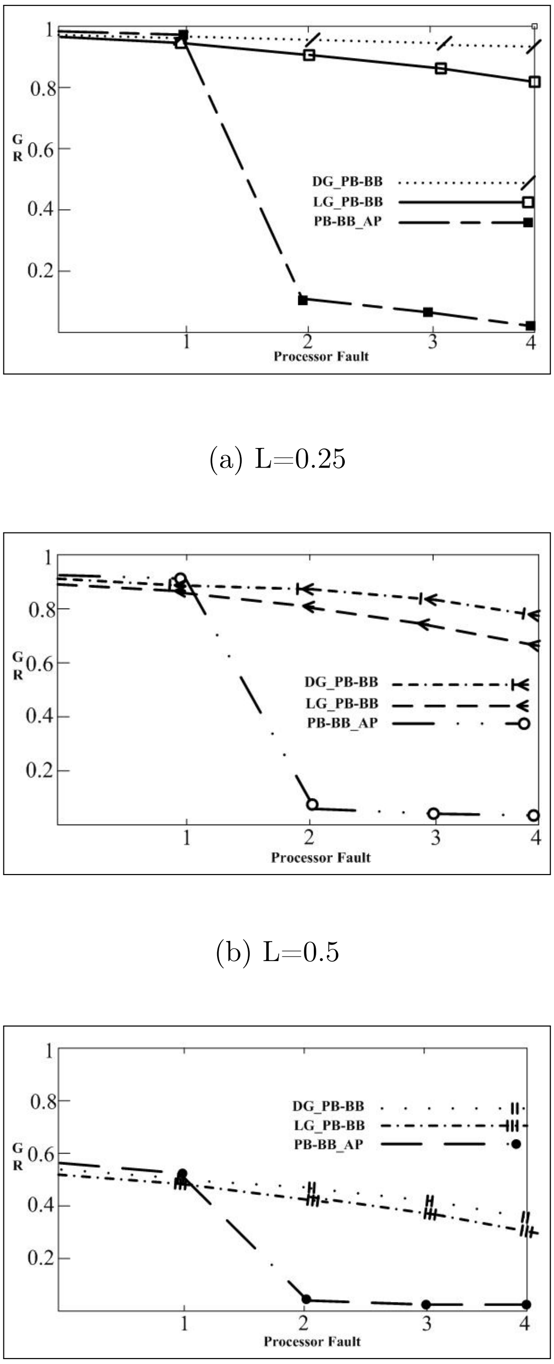
Figure 3: theroy testification of LG_PB-BB and DG_PB-BB

4 Simulation Experiment

Supposing LG_PB-BB represents fault-tolerant scheduling algorithm of hybrid overloading based on the strategy of logic grouping. Simulation experiment mainly compares with change situation of the guarantee ratio(GR) in variety of fault-tolerant number of processor in circumstance of different task load for DG_PB-BB, LG_PB-BB and PB-BB_AP algorithm respectively. The guarantee ratio=the number of tasks guaranteed/the number of tasks arrived. Simulation parameters as following:

- The inter-arrival time of tasks follows exponential distribution with mean θ. θ=8.
- The inter-arrival time of faults follows exponential distribution with mean λ. λ=200.
- The execution time of a task is chosen uniformly [2,8].
- The deadline of a task is chosen uniformly $[r_i + c_i, r_i + R * c_i]$, where$R \geq 1$.
- Processor number m=10, task number n=50.
- R(task laxity) represents flexibility time task t_i can stay in ready queue in precondition of finishing scheduling before deadline. R=3.
- L (task load) is the expected number of task arrivals per mean service time. $L = C/\theta$, C is the mean execution time, θ is the inter-arrival average time of tasks t_i. Bigger L is, more the average load of processor is and lower the guarantee ratio is.

In Fig.4 (a),(b) and (c) respectively show the relationship of processor fault and guarantee ratio in DG_PB-BB,LG_PB-BB and PB-BB_AP algorithm for the system of L=0.25,L=0.5 and L=1. Experiment results prove GR decreases along with L and processor faults increase. When there is only one processor fault in DG_PB-BB,LG_PB-BB and PB-BB_AP algorithm, the differences of GR for the system of L=0.25,L=0.5 and L=1 are small. When processor fault increases to more than two faults, GR for three kinds of task load in PB-BB_AP algorithm are all low and failure possibility of task scheduling is high. When L=0.25 and L=0.5 in DG_PB-BB algorithm GR enhances significantly comparing with LG_PB-BB algorithm, and for the system of L=1 the difference of GR between two algorithm is small, stating in the system of not full load task DG_PB-BB algorithm can tolerate processor fault better than LG_PB-BB and PB-BB_AP algorithm.



(a) L=0.25

(b) L=0.5

(c) L=1

Fig.4. comparison with dynamic grouping, logic grouping and no group algorithm

5 Conclusions

This paper improves task number and grouping strategy included in overloading task chain and proposes DG_PB-BB algorithm based on PB-BB_AP algorithm according to logic grouping strategy of PB overloading and BB overloading. Simulation experiment shows DG_PB-BB algorithm not only has good guarantee ratio of task scheduling, but also improves fault-tolerant level of processor, with better application valuation.

Main creative achievements include 1)introducing the formalization of processor dynamic grouping in fault-tolerant scheduling technology of hybrid overloading, 2)proposing the method of processor dynamic grouping based on overloading task chain,and 3)extending task number included in overloading task chain and increasing fault-tolerant level of processor by the adoption of processor dynamic grouping.

Bibliography

[1] R.Al-Omari,Arun K.Somani,G.Manimarna,Efficient overloading techniques for primary-backup scheduling in real-time systems, *J.Parallel and Distributed Computing*,64:629-648,2004.

[2] Wei Sun,Naixue Xiong,Laurence T.Yang,Chunming Rong, Towards free task overloading in passive replication based real-time multiprocessors, *10th IEEE International Conference on Computer and Information Technology*, 1735-1742, 2010.

[3] Bindu Mirle,Albert M.K.Cheng, *Simulation fault-tolerant scheduling on real-time multiprocessor systems using primary backup overloading*, University of Houston, 1-10,2006.

[4] R.Al-Omari,Arun K.Somani,G.Manimarna, An adaptive scheme for fault-tolerant scheduling of soft real-time tasks in multiprocessor systems, *J.Parallel and Distributed Computing*, 65:595-608, 2005.

[5] W.Sun,Y.Zhang,C.Yu,X.Defago,Y.Inoguchi, Dynamic scheduling real-time task using primary-backup overloading strategy for multiprocessor systems, *IEICE Transactions on Information and Systems*, 796-806, 2008.

[6] W.Sun,Y.Zhang,C.Yu,X.Defago,Y.Inoguchi,Real-time task scheduling using extended overloading technique for multiprocessor system, *11th IEEE Symposium on Distributed Simulation and Real-time Applications*, 95-102, 2007.

[7] W.Sun,Y.Zhang,C.Yu,X.Defago,Y.Inoguchi,Hybrid overloading and stochastic analysis for redundant scheduling in real-time multiprocessor systems, *26th IEEE International Symposium on Reliable Distributed Systems*, 265-274, 2007.

[8] G. Manimaran, C. Siva Ram Murthy,A fault-tolerant dynamic scheduling algorithm for multiprocessor real-time systems and its analysis, *IEEE Trans. Parallel Distributed System* , 9(11):1137-1152, 1998.

Author index