INTERNATIONAL JOURNAL

of

COMPUTERS COMMUNICATIONS & CONTROL

ISSN 1841-9836

A Bimonthly Journal
With Emphasis on the Integration of Three Technologies

Year: 2019 Volume: 14 Issue: 1 Month: February

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

Member since 2012
JM08090

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

CCC Publications

BRIEF DESCRIPTION OF JOURNAL

Publication Name: International Journal of Computers Communications & Control.
Acronym: IJCCC; **Starting year of IJCCC:** 2006.
ISO: Int. J. Comput. Commun. Control; **JCR Abbrev:** INT J COMPUT COMMUN.
International Standard Serial Number: ISSN 1841-9836.
Publisher: CCC Publications - Agora University of Oradea.
Publication frequency: Bimonthly: Issue 1 (February); Issue 2 (April); Issue 3 (June); Issue 4 (August); Issue 5 (October); Issue 6 (December).
Founders of IJCCC: Ioan DZITAC, Florin Gheorghe FILIP and Misu-Jan MANOLESCU.

Indexing/Coverage:

- Since 2006, Vol. 1 (S), IJCCC is covered by Clarivate Analytics and is indexed in ISI Web of Science/Knowledge: Science Citation Index Expanded.

  2018 Journal Citation Reports® Science Edition (Clarivate Analytics, 2017):
  Subject Category: (1) Automation & Control Systems: Q4(2009, 2011, 2012, 2013, 2014, 2015), Q3(2010, 2016, 2017); (2) Computer Science, Information Systems: Q4(2009, 2010, 2011, 2012, 2015), Q3(2013, 2014, 2016, 2017).

  Impact Factor/3 years in JCR: 0.373(2009), 0.650 (2010), 0.438(2011); 0.441(2012), 0.694(2013), 0.746(2014), 0.627(2015), 1.374(2016) **1.29 (2017)**.
  Impact Factor/5 years in JCR: 0.436(2012), 0.622(2013), 0.739(2014), 0.635(2015), 1.193(2016), **1.179(2017)**.

- Since 2008 IJCCC is indexed by Scopus: **CiteScore 2017 = 1.04**.
  Subject Category:
  (1) Computational Theory and Mathematics: Q4(2009, 2010, 2012, 2015), Q3(2011, 2013, 2014, 2016, 2017);
  (2) Computer Networks and Communications: Q4(2009), Q3(2010, 2012, 2013, 2015), Q2(2011, 2014, 2016, 2017);
  (3) Computer Science Applications: Q4(2009), Q3(2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017).

  SJR: 0.178(2009), 0.339(2010), 0.369(2011), 0.292(2012), 0.378(2013), 0.420(2014), 0.263(2015), 0.319(2016), 0.326 (2017).

- Since 2007, 2(1), IJCCC is indexed in EBSCO.

Focus & **Scope:** International Journal of Computers Communications & Control is directed to the international communities of scientific researchers in computers, communications and control, from the universities, research units and industry. To differentiate from other similar journals, the editorial policy of IJCCC encourages the submission of original scientific papers that focus on the integration of the 3 "C" (Computing, Communications, 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 & Soft computing (with particular emphasis on fuzzy logic-based methods, computing with words, ANN, evolutionary computing, collective/swarm intelligence, membrane computing, quantum computing);

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

EDITORIAL ADDRESS:
Agora University, Cercetare Dezvoltare Agora, Tineretului 8, 410526 Oradea, Bihor, Romania,
Tel./ Fax: +40 359101032, E-mail: ijccc@univagora.ro, rd.agora@univagora.ro
URL: http://univagora.ro/jour/index.php/ijccc/

EDITORIAL BOARD OF IJCCC (MEMBERS, 2018):

Contents

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 14(1), 7-20, February 2019.

Sentiment Analysis Method based on Piecewise Convolutional Neural Network and Generative Adversarial Network

C. Du, L. Huang

Changshun Du*, Lei Huang
School of Economics and Management
Beijing Jiaotong University
Beijing 100044, China
*Corresponding author: summer2015@bjtu.edu.cn

Abstract: Text sentiment analysis is one of the most important tasks in the field of public opinion monitoring, service evaluation and satisfaction analysis under network environments. Compared with the traditional Natural Language Processing analysis tools, convolution neural networks can automatically learn useful features from sentences and improve the performance of the affective analysis model. However, the original convolution neural network model ignores sentence structure information which is very important for text sentiment analysis. In this paper, we add piecewise pooling to the convolution neural network, which allows the model to obtain the sentence structure. And the main features of different sentences are extracted to analyze the emotional tendencies of the text. At the same time, the user's feedback involves many different fields, and there is less labeled data. In order to alleviate the sparsity of the data, this paper also uses the generative adversarial network to make common feature extractions, so that the model can obtain the common features associated with emotions in different fields, and improves the model's Generalization ability with less training data. Experiments on different datasets demonstrate the effectiveness of this method.
Keywords: Sentiment analysis, Piecewise Convolution Neural Network, Generative Adversarial Network.

1 Introduction

As one of the most important tasks in the field of public opinion monitoring, service evaluation and satisfaction analysis in network environments, text sentiment analysis needs to determine the opinions and preferences of customers in the text. In this article, these subjective texts with emotional polarity mainly refer to users' comments on products or services, which can provide the decision-making bases of potential consumers when purchasing products and services. Analysis of these comments is extremely helpful in mining the potential needs of users and improving products and services, but these comments are growing in large numbers every day. The analysis by humans is not only costly, but also lagging in time. Therefore, it is necessary to analyze the emotional polarity of these texts using appropriate intelligent algorithms. Although there have been many studies on text sentiment analysis, it is still a huge challenge.

At present, text sentiment orientation analysis can be mainly divided into two methods: the first category is the sentiment orientation analysis method based on an emotion dictionary, and the second category is the statistical machine learning based analysis method. The first type of dictionary-based approach first requires the construction of an emotional dictionary. The Chinese Emotional Dictionary primarily includes two emotional dictionaries, HowNet and NTUSD. HowNet has been released by China Knowledge Network and NTUSD was released by Taiwan University. Such methods include Chen Xiaodong's [1] application of a sentiment lexicon to analyze the emotional tendencies of microblog texts; Li Chun et al [4] chose HowNet's strong tendencies as seed words, combined with contextual influence. It calculates the emotional propensity score of a word by calculating the similarity between the common word and the emotional seed word, and then determines the emotional tendency of the sentence. This kind of sentiment analysis method is used to analyze the emotional tendency of sentences by examining the emotional polarity of words. It is a shallow analysis method that doesn't analyze or model the overall semantics of sentences. In the machine learning-based method, traditional methods usually use adjective features, word frequency features and N-Gram features as the characteristics of the sentiment orientation analysis of the text. These features only consider the meaning of the words themselves, or model adjacent words. The words are coded as vectors (word vectors) in vector space. With the development of Internet applications, the number of corresponding network data increases sharply, and text data is analyzed simply by artificial design features or traditional natural language processing grammar analysis tools. This method is not only noneffective but also inefficient. The rise of neural networks enables them to be used on emotional analysis of texts for semantic synthesis to automatically extract the features of sentences. Finally, classifiers are used to classify their emotional polarity. There are many works of this kind of methods, such as Socher et al. [5–7, 9, 28] using a cyclic neural network to extract the semantic features of sentences; Zeng Daojian et al. [12] used Convolutional neural networks to extract sentence features of specified tasks. The two neural network structures, recurrent neural networks and convolutional neural networks, are the two most effective methods for automatically learning sentence features in deep learning. This method of self-learning features has also made great progress in other fields of artificial intelligence. In terms of natural language processing tasks, it does not need to rely on tools such as traditional grammar analysis. It can automatically learn features from sentences, and thus it has received extensive attention from scholars. Note that except of text classification, in the field of pattern classification, classification and application of other methods also have a wide range of trying. [33, 34]

The research by Hang Cui et al. [2] shows that the sentiment analysis of texts has domain dependence. In different fields, traditional sentiment analysis methods are difficult to maintain at optimal levels at all times. Different methods have different adaptabilities in different fields. In a neural network approach, a convolutional neural network is capable of modeling the combination of word features with a small number of parameters. At the same time, convolutional neural networks have the following advantages over traditional natural language processing analysis tools:(1) As an effective method for automatically acquiring sentence features in deep learning, it is possible to automatically learn from the sentences the features that are most relevant to a sentiment analysis task. This improves the performance of the sentiment analysis model by extracting important features related to it in different fields.(2) Emotional analysis using convolutional neural network does not require the use of additional resources such as a sentiment lexicon, avoiding the problem of constructing an emotional lexicon and low emotional dictionary coverage. However, the original convolutional neural network model ignores the sentence structure information that is important for text sentiment analysis, and it is easy to overfit. In view of the above deficiencies, this paper adopts the piecewise pooling strategy, which enables the deep learning-based convolutional neural network model to model sentence structures and segment the main features of different structures. The Segmented Convolutional Neural Network (PCNN) analyzes the emotional propensity of the text by combining the structural information and the domain information of the text; and uses the Dropout algorithm to enhance the generic ability of the model.

At the same time, the user's feedback on the service involves many different fields, and there is less labeled data in each field. The lack of data makes the training of the convolutional neural network more difficult. The parameters cannot be fully optimized and cause the model to be under-fitting. Therefore, in the case that data volume expansion is difficult, in order to alleviate the sparseness of the data, this paper uses the generated adversarial network to extract the common features of texts in different fields, so that the model can acquire the common features related to emotions in the feedback of different fields and enhance the generalization ability of the model in cases where there is less training data. Experiments on different data demonstrate the effectiveness of this method.

2 An emotional analysis model that combines PCNN and GAN

Convolutional neural networks are widely used in natural language processing tasks such as relation extraction, information retrieval, and sentiment analysis, and have achieved remarkable results. In this section, this paper will use a convolutional neural network to extract the features of the text, and automatically select the important features of the sentiment tendency that are suitable for the text field to be classified from the input text, and alleviate the dependence of the sentiment analysis model on the text field. Furthermore, the deficiencies of the convolutional neural network ignoring the structural information of text sentences are improved. The pooling layer is used to extract the structural features of the text and improve the performance of the sentiment analyzer. We also introduce the generative adversarial network to model, which enables the model to extract common features related to sentiment analysis between different domains of text, and improve the effectiveness of the model in the case of sparse data.

2.1 Emotional analysis model based on convolutional neural network

The model mainly consists of two parts. The first part is the feature extraction operation, which consists of two steps: convolution and pooling. The second part is the emotion classifier. The following is a description of each component separately.

Convolution operation

A standard convolutional neural network usually consists of a Convolution Layer and a Pooling Layer. Unlike the input of image processing, which is composed of image pixels, in natural language processing tasks, a matrix is usually used to represent a sentence or a paragraph as an input. Each line of the matrix represents a language token, which can be a character or a word. In the model presented in this paper, each row of the input matrix is a vectorized representation of a linguistic symbol. In image processing, the Convolution Kernel of the convolutional layer usually slides over all parts of the image, while in natural language processing the convolution kernel slides only in the direction of text expansion. That is, the width of the input matrix (the dimension of the word vector) coincides with the width of the convolution kernel. Assuming that the height of the convolution kernel is w , and the width and word vector dimensions are both d , the convolution kernel can be represented as a matrix $W \in R^{w \times d}$. Let the vectorization of the i th language symbol in the input be represented as s_i , and the input text can be represented by the matrix formula $S = (s_1^T, s_2^T, \cdots, s_{|s|}^T)$. Then the convolution operation can be expressed as follows:

$$c_j = W \bigotimes S_{j:j+w-1} \tag{1}$$

$1 \leq j \leq |S| - w + 1$, c_j is a feature value extracted by the convolution operation between the word j and the word in the window of convolution kernel.

The effective features obtained by feature extraction of text by a convolution kernel are not comprehensive. In order to be able to extract more abundant information from text, a number of different convolution kernels are usually used, which can be expressed as a three-dimensional tensor $\hat{W} = \{W_1, W_2, \cdots, W_n\}$. The convolution operation of the convolutional layer can be expressed as follows:

$$c_{i,j} = W_i \bigotimes S_{j:j+w-1} \tag{2}$$

$1 \leq i \leq n$, The input text is subjected to the $i-th$ convolution kernel convolution operation to obtain the feature vector $c_i = \{c_{i,1}, c_{i,2}, \cdots, c_{i,|S|-w+1}\}$, Then all convolution kernels can get a total of n feature vectors, also known as Feature Map.

Sentiment classifier based on dropout algorithm and softmax classifier

The feature vector extracted by the convolution kernel still has many eigenvalues. If the feature analyzer is used directly to train the sentiment analyzer, the number of parameters that need to be optimized is still very large, and the model training is difficult and easily susceptible to overfitting. The pooling operation can further select features, effectively reduce parameters, and select some features that are most suitable for sentiment analysis as the final features of the sentiment analyzer.

The pooling layer compresses the input feature map, which can simplify the network computation complexity; on the other hand, the pooling operation can extract the main features. The essence of the pooling operation is sampling; which extracts features of the text from a sentence-level or higher-level text representation consisting of a vector representation of words. For different input texts, even when some of the expressions associated with the sentiment analysis of the current domain change, the output after the pooling operation is constant, which can enhance the robustness of the convolutional neural network. There is a certain anti-disturbance effect. Because the model uses a maximum pooling operation in the local neighborhood, the model can obtain the maximum degree of translational invariance of the text features. This feature is very important for sentiment analysis models because the model can effectively extract strong emotional features from the text. These features are in different positions in the text, and the pooling layer can select the most relevant features from the feature map through the pooling operation [10] . For the feature vector c_i obtained by the input text S through the i th convolution kernel, the maximum pooling operation can be expressed by the following formula:

$$p_i = \max(c_i) \tag{3}$$

Before the pooling layer, the feature vector obtained by the convolution kernel is connected to the pooling layer through a nonlinear mapping layer. Then Formula 3 can be rewritten as follows:

$$\begin{aligned} \hat{c_l} &= f(c_i) \\ p_i &= \max(\hat{c_l}) \end{aligned} \tag{4}$$

f represents a nonlinear mapping function.

At this time, the neural network of the feature extraction operation section can be represented as shown by the convolution and pooling layers in Figure 1 . The resulting feature vector is $p_S = [p_1, p_2, \cdots, p_n]$.

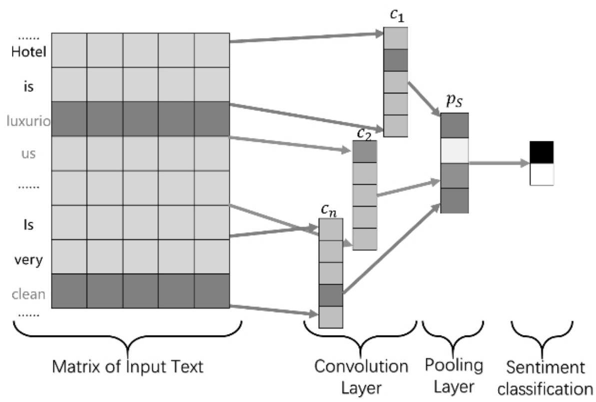


Figure 1: Sentiment analysis model based on convolution neural network

Pooling operations

The output obtained by the convolution operation and the pooling operation represents the advanced features of the original input text. In order to get the final sentiment analysis result, these features need to be input into the classifier through the fully connected layer for emotional tendency classification.

As shown in Figure 1 , after the features of the text are extracted by the convolutional neural network, the feature vectors are input to the softmax classifier for classification. This article uses the dropout method to connect feature vectors with the softmax classifier. The input data of the neural network is randomly connected to the neurons of the next layer according to a preset proportional parameter ρ . In the convolutional neural network of this paper, the pooled feature vector is set to 0. Therefore, in the subsequent calculation process, only other elements that are not set to 0 participate in the operation to obtain the output of the network. The specific process of parameter update is as follows: First, the pooled feature vector is set to 0 according to the ratio ρ , and the element without 0 participates in the operation of the softmax classifier and obtains the gradient-optimized neural network outputting the neural network parameters. Then, the input vectors of all the samples in the sample set are sequentially accepted, and the connection of the input elements is randomly set to 0 in the same manner, and the elements participating in the training are selected until all the samples are used by the neural network to train the model parameters. Each time a sample is entered, the probability that the input element is set to 0 is ρ , so each time the input element of the participating operation is connected differently, the updated network weight parameters are also different. When using neural networks for prediction, it is necessary to multiply the parameters of the entire network by $1-\rho$ to obtain the final classifier network parameters. Assuming that the eigenvector obtained by the convolution and pooling operations of the input text S is p_S , the way the dropout algorithm sets its element to 0 can be represented by the Bernoulli distribution B . First, use the Bernoulli distribution to generate a binary vector of the same dimension as c (the element takes only 0 or 1) r , ie $r \sim B(\rho)$, and then do the Hadamard product with the eigenvector. The vector that is ultimately entered into the classifier can be expressed as:

$$c_d = p_S \times r \tag{5}$$

Set the network parameter W_e of the softmax classification layer, and the bias term is b_e ,then the neural network output can be expressed as:

$$o = f(W_e c_d + b_e) \tag{6}$$

Where f is the activation function. Then the probability that the entered text sentiment tends to i is:

$$p(i|\theta) = \frac{e^{o_i}}{\sum_{j=1}^{N} e^{o_j}} \tag{7}$$

where θ represents all parameters of the neural network, o_i represents the value of the i -th item of the output vector, and N represents the number of categories of the text. Let the sample set be expressed as Ω , then the model's optimization objective function can be calculated by the following formula.

$$L_{sen} = \sum_{i=1}^{|\Omega|} -logp(y_i|S_i,\theta) + \lambda \parallel \theta \parallel_2^2 \tag{8}$$

Where λ is the parameter of the regularization term. In the actual experiment, the traditional convolutional neural network uses the stochastic gradient descent method to optimize the objective function. The update method of the parameter θ is:

$$\theta = \theta - \alpha \frac{\partial L}{\partial \theta} \tag{9}$$

α is the learning rate.

2.2 Piecewise pooling based convolutional neural network sentiment classification model

There is a deficiency in the process of extracting text features using the sentiment analysis model based on a traditional convolutional neural network. Regardless of whether or not the texts or Chinese, English or other natural language texts, the sentences have a certain structure, and traditional convolutional neural networks ignore the structural features of these sentences. As shown in Figure 2 , both Chinese and English sentences can contain grammatical structures such as a subject, predicate and object.

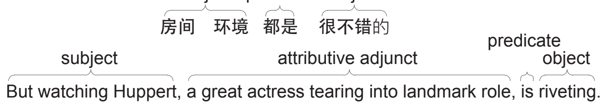


Figure 2: Structures in sentences

While existing deep learning method has difficulty parsing sentences, modeling the sentence structures will make the calculation process more complicated. However, if the simulation of grammatical structures can be added to network structures, the learning of sentence features will be significantly improved. Traditional maximum pooling extracts a maximum value from the features of a convolution kernel convolution, and does not make any distinguish between the grammatical elements of sentences. In response to this problem, this paper proposes the use of

Piecewise Pooling strategy for sentiment analysis. The piecewise pooling strategy divides the feature vector of a sentence into segments and performs maximum pooling operations on each segment. By segmenting the different grammatical structural components in a sentence, the segmented pooled feature vector can extract the features of the corresponding components in the sentence. The difference between segmented pooling and non-segmented pooling operations can be seen from the comparison in Figure 3 .

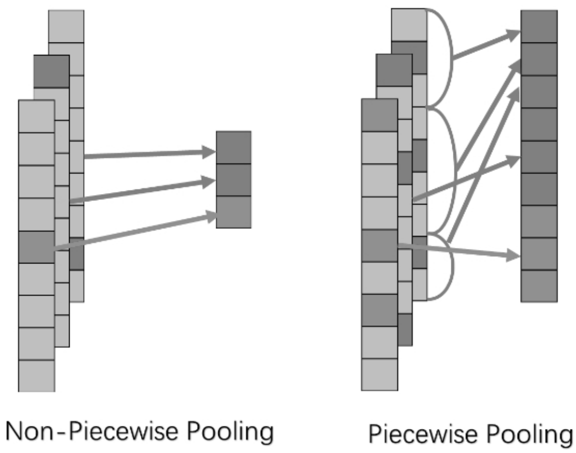


Figure 3: Piecewise pooling and non-piecewise pooling

The rest of the model is identical to the traditional convolutional neural network model described above, with the main difference being the pooling part. The feature vector c_i obtained after the convolution kernel convolution operation is different from the original pooled layer pair by Single Max-Pooling. The model in this section uses Piecewise Max-Pooling. The convoluted feature vector is divided into m segments, that is, $c_i = (c_{i,1}, c_{i,2}, \cdots, c_{i,m})$. For different segments, the maximum pooling operation is performed separately, and the piecewise pooling operation may be expressed as Equation 10 .

$$p_i = [\max(c_{i,1}), \max(c_{i,2}), \cdots, \max(c_{i,m})] \tag{10}$$

The piecewise pooling operation is performed on the feature vectors obtained by all convolution kernels, and then the feature vectors are nonlinearly transformed by the fully connected layer to obtain the feature vector of the input text, namely:

$p_S = [f(p_1), f(p_2), \cdots, f(p_i)]$

$1 \leq i \leq n$, f represents a nonlinear activation function.

2.3 Multi-domain common sentiment feature extraction based on generative adversarial networks

Although the piecewise convolution neural network can effectively analyze the emotional tendency of the text, in order to make the neural network fully trained, the generalization of the network parameters is good enough, and more annotation data is needed. However, in each segment, there is relatively little data to be labeled, and data sparseness makes network training more difficult. Therefore, this paper proposes to use the generated adversarial network to extract common features related to emotions in different fields to alleviate the above problems.

The shared private model separates the feature space into shared and private spaces, but there is no guarantee that shared features cannot exist in the private feature space, and vice versa. Therefore, some useful sharing features can be ignored in the shared private model, and the shared feature space is also vulnerable to certain task-specific information. Therefore, a simple principle can be applied to multitasking learning. A good shared feature space should contain more public information and no task-specific information. To solve this problem, we introduce adversarial training into the multitasking framework.

Introduction for generative adversarial network

First, We give a brief introduction to the generative adversarial network (GAN), generated by the Goodfellow et al. [3] was proposed in 2014, and quickly received extensive attention from the academic and industrial circles. The goal of GAN is to learn a generator distribution $P_G(x)$ that is fully matched to the real data distribution $P_{data}(X)$. Specifically, the GAN learns $P_G(x)$ by training a generation network G and a discrimination network D, wherein G is a sample generated from the distribution $P_G(x)$ to generate forged samples, and attempts to make the discrimination network unable to distinguish the forged samples from the real samples. ;D determines whether the sample is from $P_G(x)$ or $P_{data}(X)$, trying to distinguish the forged sample from the real sample. The two games are mutually optimized and alternately optimized. When the discriminator reaches the Nash equilibrium, that is, the discriminator cannot distinguish the forged sample from the real sample, the obtained $P_G(x)$ is consistent with $P_{data}(X)$. This minimum-maximum game can be optimized by the following risks:

$$\max_D \min_G V(D,G) = E_{x\sim P_{data}}[logD(x)] + E_{z\sim p(z)}[log(1 - D(G(x))] \tag{11}$$

Due to the role of the discriminator, the sample generated by the generator is forced to approach the true distribution of the data progressively and unbiased. If the generator produces a sample distribution that perfectly matches the real data distribution, the discriminator will not be able to tell whether the input is from real data or a sample generated by the generator network, giving a probability value of 0.5 for all inputs, which is the Nash equilibrium. Although GAN was originally proposed for the generation of random samples, GAN can be used as a general tool for measuring the equivalence between distributions [11] , which can be used for feature extraction.

Multi-domain common sentiment feature extraction GAN

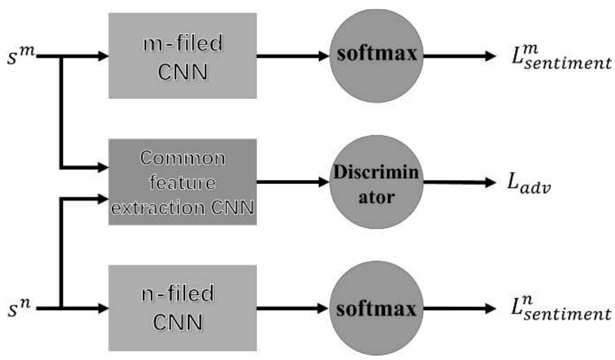


Figure 4: Multi-domain common sentiment feature extraction with generative adversarial network

Inspired by GAN, we propose a adversarial-sharing-private model for multi-domain common sentiment feature extraction. The shared convolutional neural layer first, and then the private convolutional neural network extract features from the input text. The features extracted by the shared convolutional neural network are input into the discriminator network, trying to make the discriminator unable to distinguish which subdivision from which the feature comes from, and the discriminator tries to distinguish the features of different domains. Then, the extracted common features and the private features extracted by the piecewise domain are spliced as a feature to the accurate prediction of the text sentiment tendency in the sentiment orientation analyzer. This kind of adversarial training encourages the sharing space to be purer, ensures that the shared representation is not polluted, and can well preserve the common emotional features of different fields and eliminate the private emotional characteristics of the segment.

The domain discriminator network is used to map the shared representation of the sentence into a probability distribution and estimate which fields the encoded sentence comes from.

$$D(p_s, \theta_D) = Softmax(b + Up_s) \tag{12}$$

U is a learnable parameter and b is the bias.

In order to train the discriminator network, the parametric loss L_{adv} is used to optimize the parameters of the network to prevent the private sentiment features related to the emotional tendency of the specific domain from spreading to the shared sentiment orientation feature space. The counter-loss is used to train the model to extract common sentiment features in different domain texts so that the discriminator network cannot reliably predict which category of the currently entered text belongs to base on these features. The original form discriminator of GAN can only be used in the case of two classifications. In order to overcome this problem, this paper extends its s to multiple types of forms, so that the model of this paper can train sentiment analyzers in multiple fields together. The objective function against the network can be expressed as follows:

$$L_{adv} = \min_{\theta_s}(\lambda \max_{\theta_D}(\sum_{k=1}^{K}\sum_{i=1}^{N} d_i^k log[D(E(S^k))])) \tag{13}$$

d_i^k indicates the domain tag to which the currently input text belongs. The basic idea is that, given a sentence, the shared convolutional neural network extracts a representation of an eigenvector misleading discriminator network. At the same time, the discriminator tries to classify the field of the input text correctly. After training, the shared sentiment feature extractor network and the discriminator network reach the Nash balance, and the discriminator cannot distinguish which domain these common features come from. The common feature extractor model can represent Figure 4.

Finally, the loss function of the sentiment analysis method model of PCNN and GAN proposed in this paper can be expressed as follows:

$$L = L_{sen} + \lambda L_{adv} \tag{14}$$

λ is a hyperparameter for balance sentiment analysis loss and adversarial loss.

3 Experiment settings

3.1 Data sets

Three data sets were used primarily in the experiment. The first is the Chinese Hotel Data Collection (Ctrip Hotel for short), which was compiled by Tan Songbo from the Institute of Computing Technology of the Chinese Academy of Sciences and is a commentary on the service of large-scale customers. The corpus was automatically collected from the Ctrip (http://hotels.ctrip.com/) and then organized. The corpus size is 10,000, including 7000 positive evaluation samples and 3,000 negative evaluation samples. The second data set is the English data set, from Stanford University's emotional tree library, which labels the emotional polarity of each phrase in the sentence and the entire sentence. In this paper, only the emotional polarity of the sentence is extracted and classified. The data set contains a total of 11,855 sentences, including training sets, validation sets, and test sets containing 8544, 1101, and 2210 samples, respectively. The emotional polarity of each sentence is in the range of [0,1]. The smaller the score, the more the emotion tends to be negative. Otherwise, the emotion tends to be positive. The emotional scores of all sentences in the data set are manually labeled and then averaged. It has a good reliability. According to the distribution of the emotional scores of the dataset and its data description, the sample with the sample score between [0, 0.5) is first divided into negative samples, and the sample score is divided into positive emotions between [0.5, 1]. This is a coarser-grained sentiment analysis. Further divided into five levels of sentiment, including negative [0, 0.2), negative [0.2, 0.4), neutral [0.4, 0.6), positive [0.6, 0.8), positive [0.8, 1].In order to make the experiment closer to a real production environment, and to verify the effectiveness of the sentiment analysis method of PCNN and GAN proposed in this paper, this paper uses the third data set. It is the review text (Dianping for short) from different fields that we crawled from the public service website (http://www.dianping.com/). Its data includes six segments of food, hotels, movies, entertainment, marriage, home improvement, and the comments are divided into different emotional tendencies based on the scoring information in the comments. This paper selects 30,000 reviews as the training set and 10,000 as the test set.

3.2 Data pre-processing

For the Chinese data set, the Chinese word piecewise package NLPIR developed by the Chinese Academy of Sciences is first used for Chinese word piecewise. Since the experiment in this article has a Chinese data set, you need to call the package participle. The English data itself contains independent words, so there is no need for word piecewise. Since the minibatch training model is used during training (multiple samples are learned at a time, and the text length of multiple samples may be different), so the length of the text needs to be fixed-length. Since the length of the natural language text is inconsistent, the longest sentence length l _ max is first calculated. For sentences with a sentence length less than l _ max, the text is uniformly filled with the $< \backslash s >$ symbol to the length l _ max (the vector of $< \backslash s >$ is always set to 0), this will unify the text length. The purpose of this approach is to improve computational efficiency, and when the length of the data is uniform, the computational time overhead can be effectively reduced. At the same time, in order to ensure the features are extracted at the beginning and the end of the text during the convolution process, a certain number of $< \backslash s >$ corresponding to the convolution kernel should be added at the beginning and end of the longest text as Padding.

3.3 Pre-training of word embedding

Word embedding is required before training of the model. Word embedding acts as a distributed representation of words as an input suitable for neural networks. Many current studies have shown that executing word embedding pre-training on a large-scale corpus, and then applying the obtained word embedding to subsequent training, can speed up the convergence of neural network models and achieve a better local optimal solution. In this paper, the word2vec algorithm is used to pre-train word embedding. The word embedding of this algorithm shows better performance in many natural language processing tasks, and has higher efficiency. This paper chooses the Skip-gram model and the Negative Sampling model to pre-train the word embedding of Chinese and English words. The pre-training of Chinese word embedding uses the text content crawled on Baidu Encyclopedia. The English word vector is pre-trained using New York Times.

3.4 Setting of experimental parameters

In the training optimization process of the model, we use the Adam optimizer to train and optimize the model. The parameter settings of the optimizer adopt the recommended parameter values. In this paper, the model mainly has the following hyperparameters: the dimension d of the word vector, the number n of convolution kernels, the ratio ρ of the Dropout algorithm, and the piecewise number t of the piecewise pooling in the piecewise convolutional neural network. In order to obtain the optimal hyperparameter setting, this paper uses grid-search to determine the value of some hyperparameters. The word embedding dimension N selects a value from { 50, 100, 200, 300} ; the number n of convolution kernels takes a value in { 100, 150, 200 } . The ratio ρ of the Dropout algorithm is 0.5 based on experience. In the experiments in this paper, multiple experiments were performed using these parameters, and then the average of the results was obtained. The number of segments of the piecewise pooling was performed in this paper. The results were given in the analysis.

4 Results and analysis

In order to verify the effectiveness of the sentiment analysis method of the fusion piecewise convolution neural network and the generated adversarial network, the method of this paper is compared with some mainstream emotional analysis baseline methods. The parameter sensitivity experiment is carried out on the piecewise pooling segment number hyperparameters which are important to the performance of the model in the hyperparameters proposed in this paper, so as to find the more suitable parameter settings in the application scenario.

4.1 Comparison method

In order to verify the validity and correctness of the proposed model, this paper selects a model based on traditional methods and a neural network method such as RNTN proposed by Richard Socher et al. as a baseline method. The first method of comparison is the naive Bayesian method (NB for short) using the word bag feature. The second method is to use the word bag feature as input to perform emotion classification using a support vector machine (SVM for short) classifier. The third method is the naive Bayesian method (BiNB for short) of the bag feature obtained using the binary grammar language model. The fourth method is to use the average word vector of the sentence as the input feature and use the fully connected network as the classifier (VecAvg for short). The fifth method is the Recurrent Neural Network (RNN). The sixth method is a recurrent neural network (MV-RNN) with a semantic transformation matrix [7] . The seventh method is based on a tensor-based cyclic neural network (RNTN). The last comparison method is the traditional convolutional neural network.

4.2 Analysis of experimental results

First, because there is no multi-domain information on the Ctrip Hotel dataset and the Stanford English Emotional Tree Database dataset, the performance of the segmented convolutional neural network model and the baseline model is first compared. It can be seen from the results in Table 1 that the neural network method generally has higher performance than

Table 1: Text classification precision on different data sets(%)

Model	**Stanford**		**Ctrip**
	Two levels sentiment	**Five levels sentiment**	
NB	81.8	41.0	80.2
SVM	79.4	40.7	86.7
BiNB	83.1	41.9	85.9
VecAvg	80.1	32.7	82.1
RNN	82.4	43.2	87.8
MV-RNN	82.9	44.4	87.6
RNTN	**85.4**	45.7	89.3
CNN	81.9	45.6	88.5
PCNN	**85.4**	**45.9**	**89.7**

Table 2: Text classification recall on different data sets(%)

Model	**Two levels sentiment**	**Five levels sentiment**	**Two levels sentiment(Reduced)**
NB	72.1	35.7	66.1
SVM	72.6	33.2	61.2
BiNB	75.2	36.6	62.4
VecAvg	74.1	36.3	62.3
RNN	77.3	40.2	69.7
MV-RNN	78.0	42.2	73.1
RNTN	80.3	42.6	75.7
CNN	79.1	41.5	73.5
PCNN	80.8	43.2	74.9
PCNN+GAN	**83.3**	**47.9**	**80.1**

the conventional method. Especially in the five-level sentiment analysis with finer granularity, the neural network method can obtain the key features of the text well. An important reason for the BiNB method to achieve better results is that the binary grammar model considers a certain combination of semantics, but with it is a large computational overhead. At the same time, unlike the method of cyclic neural networks, the use of maximum pooling in convolutional neural networks can automatically extract the features most relevant to sentiment analysis tasks, and has positive significance for text sentiment analysis tasks, so it has achieved good results. Traditional methods such as BiNB can only extract combined features from adjacent words, and traditional convolutional neural networks cannot model grammatical structures. The segmented convolutional neural network method has the best effect on both datasets because it simulates the grammatical structure information of the text, supplements the original emotional words, and extracts the combined semantic features of different positions.

In the Dianping data, the data comes from different fields. We compare the performance of different methods in this data to fully demonstrate the effectiveness of the proposed PCNN and GAN methods. It can be seen from Table 2 that the fusion segmented convolutional neural network and the sentiment analysis method of the generated adversarial network proposed in this paper have achieved the best results. The comments crawled from the Dianping contain different subdivisions. The comment text in each field is relatively small, and the comment text itself is relatively short. Therefore, samples in different fields have a certain degree of sample sparseness.

This paper also reduces the number of samples in each field, making sparse data a more serious side-effect. It can be seen that all methods have performance degradation problems to varying degrees, but the method proposed in this paper still achieves good results. The stability of the proposed method in the case of multi-domain data sparseness is fully explained.

In the piecewise convolutional neural network, a very important parameter is the piecewise pooling segment number. Since the piecewise is a simulation of the text's grammatical structure, the piecewise number determines the validity of the text structure information extraction. In this paper, the parameter sensitivity test is carried out on the three data sets for the number of segments. The results are shown in Figure 5. It can be seen that a reasonable piecewise can better obtain the grammatical structure information of the text, and an excessively large number of segments destroys the structure of the sentence itself. The pooling operation cannot effectively extract the features most relevant to sentiment analysis, making the sentiment analysis effect worse. Therefore, it is very important to select the appropriate segment number to simulate the grammatical structure of the sentence.

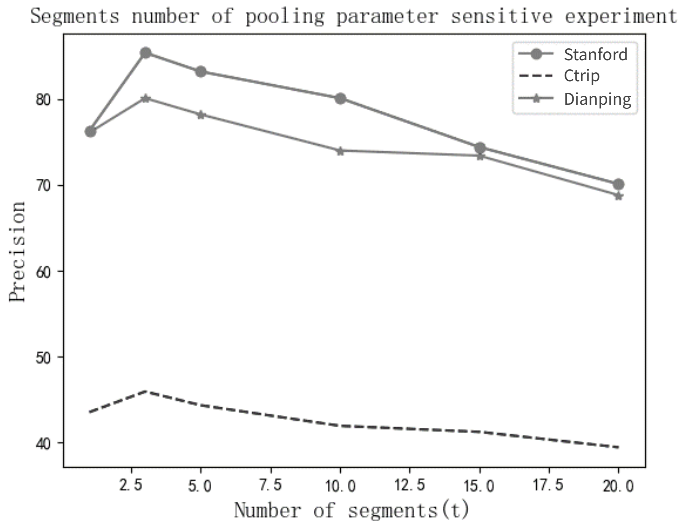


Figure 5: Segments number of pooling parameter sensitive experiment

5 Conclusion

In this paper, we show the difficulty of maintaining optimality in different fields for traditional text sentiment analysis methods. Different methods have different applicability problems in different fields. The original convolutional neural network model ignores the sentence structure information that is very important for text sentiment analysis, and it is susceptible to over-fitting. In this paper, the piecewise pooling strategy is adopted to enable the deep learning-based convolutional neural network model to model the sentence structure and segment the main features of different structures. It combines the structural information and domain information of the text to analyze the emotional tendency of the text; and uses the Dropout algorithm to enhance the generic ability of the model.

At the same time, user feedback on services involves many different areas, with less data in each field. Less data makes the training of convolutional neural networks more difficult, and the parameters cannot be fully optimized, resulting in under-fitting of the model. Therefore, in the case that data volume expansion is more difficult, in order to alleviate the sparseness of data, this paper uses the generated adversarial network to extract common features of texts in different fields. This enables the model to acquire common features related to emotions in feedback from different fields, and to enhance the generalization ability of the model with less training data. Experiments on different data demonstrate the effectiveness of this method.

Bibliography

[1] Chen, X. (2012); *Research on Sentiment Dictionary based Emotional Tendency Analysls of Chinese MicroBlog*, Huazhong University of Science & Technology, 2012.(in Chinese)

[2] Cui, H.; Mittal, V.; Datar, M. (2006); Comparative experiments on sentiment classification for online product reviews, *AAAI*, 6, 1265-1270, 2006.

[3] Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M. et al. (2014); Generative adversarial nets, *Intl Conf. on Neural Information Processing Systems*, MIT Press, 2672-2680, 2014.

[4] Li, D.; Qiao, B.; Yuanda, Y.; et al. (2008); Word Orientation Recognition Based on Semantic Analysis, *Pattern Recognition & Artificial Intelligence*, 2008.(in Chinese)

[5] Luong, T.; Socher, R.; Manning, C.(2013); Better word representations with recursive neural networks for morphology, *Proc. of the Seventeenth Conf. on Computational Natural Language Learning*, 104-113,2013.

[6] Socher, R. (2014); Recursive deep learning for natural language processing and computer vision, *Stanford University*, 2014.

[7] Socher, R.; Huval, B.; Manning, C.D. et al.(2012); Semantic compositionality through recursive matrix-vector spaces, *Proc. of the 2012 joint conf. on empirical methods in natural language processing and computational natural language learning*, Association for Computational Linguistics, 1201-1211, 2012.

[8] Socher, R.; Chen, D.; Manning, C.D. et al (2013); Reasoning with neural tensor networks for knowledge base completion, *Advances in neural information processing systems*, 926-934, 2013.

[9] Socher, R.; Perelygin, A.; Wu, J. et al.(2013); Recursive deep models for semantic compositionality over a sentiment treebank, *Proc. of the 2013 conf. on empirical methods in natural language processing*, 1631-1642, 2013.

[10] Srinivas, S.; Sarvadevabhatla, R.K.; Mopuri, K.R. et al. (2016); A taxonomy of deep convolutional neural nets for computer vision, *Frontiers in Robotics and AI*, 2:36, 2016.

[11] Taigman, Y.; Polyak, A.; Wolf, L. (2016); Unsupervised Cross-Domain Image Generation, *arXiv preprint arXiv*, 1611.02200, 2016.

[12] Zeng, D.; Liu. K.; Lai, S. et al. (2014); Relation classification via convolutional deep neural network, *Proc, of COLING 2014, the 25th Intl Conf. on Computational Linguistics: Technical Papers*, 2335-2344, 2014.

[13] Zhang, W., Zhang, Z., Chao, H.C. et al. (2018); Kernel mixture model for probability density estimation in Bayesian classifiers, *Data Mining and Knowledge Discovery*, 32(3), 675-707, 2018.

[14] Zhang, W.; Zhang, Z.; Qi, D. et al. (2014); Automatic crack detection and classification method for subway tunnel safety monitoring, *Sensors*, 14(10), 19307-19328, 2018.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 14(1), 21-38, February 2019.

RESP: Relay Suitability-based Routing Protocol for Video Streaming in Vehicular Ad Hoc Networks

O.A. Hammood, N. Nizam, M. Nafaa, W.A. Hammood

Omar A. Hammood*, Mohd Nizam,
Muamer Nafaa, Waleed A. Hammood
Department of Computer System and Software Engineering
University Malaysia Pahang (UMP) , Malaysia
Kuantan-26300, Pahang, Malaysia
*Corresponding author: omer_almajeed@yahoo.com
mnizam@ump.edu.my, Muamer.scis@gmail.com, engwaleed54@yahoo.com

Abstract: Video streaming in Vehicular Ad Hoc Networks (VANETs) is a fundamental requirement for a roadside emergency and smart video surveillance services. However, vehicles moving at a high speed usually create unstable wireless links that drop video frames qualities. In a high-density network, network collision between vehicles is another obstacle in improving the scalability of unicast routing protocols. In this paper, the RElay Suitability-based Routing Protocol (RESP) which makes a routing decision based on the link stability measurement was proposed for an uninterrupted video streaming. The RESP estimates the geographic advancement and link stability of a vehicle towards its destination only in the small region. To ensure the reliability while extending the scalability of routing, the relay suitability metric integrates the packet delay, collision dropping, link stability, and the Expected Transmission Count (ETX) in the weighted division algorithm, and selects a high-quality forwarding node for video streaming. The experimental results demonstrated that the proposed RESP outperformed the link Lifetime-aware Beacon-less Routing Protocol (LBRP) and other traditional geographical streaming protocols in providing a high packet delivery ratio and less packet delay with various network densities, and proved the scalability support of RESP for video streaming.
Keywords: Video streaming, scalability, geographic unicast routing, link stability, reliability, QoS.

1 Introduction

Vehicular Ad Hoc Networks (VANETs) are widely applied in information and communication technology to ease transportation problems [8,23]. Video streaming in VANETs is a central part towards the realization of intelligent transportation systems [2,31], and it has many promising potential applications, ranging from safety to entertainment. However, video streaming in VANETs faces several challenges due to the stringent video quality level requirements [7, 9] . The major problems associated with scalable video streaming over VANETs are frequent link failures [33] due to greedy vehicle mobility, lack of link quality measurement, and link lifetime measurement in terms of the vehicle velocity rather than the relative velocity of vehicles. The network condition is dynamic with respect to vehicle densities and the distance between the source and the destination vehicle [22]. The existing geographic routing protocols based on greedy or improved greedy routing determine the shortest routing paths in such a way that it can be scaled up to a large VANET environment; however, it degrades the performance of the routing approach under high mobility scenarios. The integration of different routing aspects such as relative velocity, density-based collision probability, distance to reach the destination, and Expected Transmission Count (ETX) [15] in the measurement of routing relay suitability can improve the performance of video streaming over the dynamic VANET environment. An improper integration of these parameters tends to compensate each other for their influence on the scalable and efficient router selection for video streaming [4]. To cope with these challenges, the geographic unicast routing protocols have been employed as a suitable solution for the distribution of video flows in VANETs [24].

Recently, several geographic unicast routing protocols have evolved to achieve scalability in video streaming, however, a scalable video streaming that meets the stringent video quality requirements without video distortions under a highly dynamic network topology is a challenging task in unicast routing. Greedy Perimeter Stateless Routing (GPSR) is a commonly used geographic unicast routing protocol in wireless networks. The conventional location-based routing protocols depend on the distance metric in which the nearest vehicle is selected as a forwarder. In highly dynamic networks, the greedy node selection can lead to frequent link failures and retransmissions. Beside the dynamic network topology, video streaming solution has to comply with the Quality of Service (QoS) requirements [25]. To decide the reliable and QoS forwarding vehicle in the network, the recent geographic routing protocols must accumulate multiple routing factors.

Geographic unicast routing is more effective for a wireless communication [10, 14], however, there are many issues in implementing a scalable and reliable geographic unicast routing over VANETs. For example, unicast geographic routing selects a greedy node in the communication range with the aim of advancing the packet forwarding [3]. Despite distance information, a common aspect of the geographic routing protocol is that it must select and relay data packet through a stable node. Two key facts that limit greedy routing performance in video streaming are node mobility and collision due to the unreliability of the greedy node. When a packet is dropped due to node mobility or collision, the sender node either retransmit the packet to the same forwarder or through the currently available greedy node. Consequently, it triggers unnecessary retransmissions, an end-to-end packet delay, and even video buffering and distortion. It is essential to focus on the trade-off between the suitability of a node in advancing packets toward the destination and the stability of its connectivity to the sender node.

In this paper, the RElay Suitability-based unicast routing Protocol (RESP) for an uninterrupted video streaming in the VANET topology is proposed with the aim of extending the scalability of geographic unicast routing protocols. The proposed RESP builds a small region at the boundary of the communication range using an offset angle and integrates the compensating factors using a weighted division algorithm to improve the performance of a VANET router selection process. It significantly reduces packet retransmission and propagation delay in video streaming.

The main contributions of the proposed RESP protocol are summarized as follows:

- The proposed RESP protocol considers multiple routing aspects for an efficient unicast routing and scalable video delivery with minimized distortions and buffering.

- Node density-aware data collision and relative velocity-aware link stability measurements facilitate the scaling up of RESP to a larger vehicular network.

- The geographical advancement of a near-by vehicle towards the destination is considered in the router selection to avoid the involvement of several intermediate vehicles in the routing. Further, the consideration of ETX avoids packet loss, resulting in a successful timely video frame delivery.

- The weighted division algorithm integrates multiple routing aspects and ensures an efficient and stable link selection for video delivery.

- The performance of the RESP routing protocol was evaluated using a network simulator and compared to the routing policy centralized in terms of packet delivery ratio (PDR), delay, overhead, and frame setback ratio (FSR).

The remaining part of this paper is organized as follows: Section 2 presents a discussion of the existing related works while section 3 presents the idea of the proposed system model, including the positive forwarding region formation, the degree of mobility, different routing aspects of relay suitability factor, forwarding probability and delay metrics, and scalable and QoS data forwarding by applying weighted division algorithm. In section 4, the performance of the proposed system was evaluated and presented while section 5 presents a conclusion and a brief introduction to the intended future work.

2 Related works

Several unicast routing protocols have been proposed for video streaming in VANETs [20]. The geographic unicast routing protocols are classified into three types, which are greedy routing, opportunistic routing, and trajectory-based routing protocols [5, 19]. In VANETs, a commonly used routing strategy is the greedy forwarding strategy in which a node selects a neighbor close to the destination for data forwarding [26]. The opportunistic routing scheme follows the store-carry-and-forward technique and avoids unnecessary packet dropping especially when no forwarding node is available. However, it is prone to long delays and inadequate for video-streaming. The trajectory-based strategy provides more importance to a vehicle to act as a forwarder when it is moving towards the destination vehicle. The GPSR is a greedy-based unicast routing protocol in VANETs which assumes that the nodes are aware of their own position and the location of the destination [18]. The GPSR follow two different modes for data routing, which are the greedy mode and the perimeter mode. The perimeter mode is applied when there is no greedy router in the communication range. However, the greedy mode can fail since there is a high chance of the greedy router to move out of the communication range. Conventionally, several works have been presented on improving the basic GPSR [5, 28]. The movement prediction routing (MOPR) considers the link stability concept and attempts to solve the issues in greedy routing mode while the improvement GPSR (I-GPSR) consolidates different metrics such as the distance to the destination, vehicle density, moving direction, and vehicle speed to make routing decisions. Likewise, the multimedia multimetric map-aware routing protocol (3MRP) [17] exploits the distance to the destination, vehicle density, trajectory, and available bandwidth to select the next hop for routing while the available bandwidth estimation (ABE) [1] calculates the available bandwidth on a link between two nodes. The MMMR exploits the roadmap considering the possible obstacles; however, it assigns an equal importance or weight to all the metrics in the score estimation. A better scheme must provide variable weights to the routing metrics depending on the current network conditions. However, the acquisition of motion prediction information is a difficult task. Most of the conventional routing techniques focus only on better packet delivery rates with low data delivery delays. However, they do not address the quality of video delivery when users are watching video sequences. Video dissemination techniques have not considered video qualities [6,20] which is a crucial concept for the success of safety and traffic-warning video streaming in VANETs [18]. Thus, the multi-hop routing service is essential to the awareness of Quality of Experience (QoE) requirements and network conditions. It assists the video streaming approach to avoid the loss of video sequences and quality. A real-time video dissemination with a QoE-supportive solution in VANET [6] imposes the use of a distortion threshold. QoE represents the experience of the end users on a delivered service and their satisfaction rate about it. It considers the average delay in routing decisions, however, average delay reduction is not equivalent to QoE maximization [25]. The network conditions include collision warning, traffic congestion, and sharing parking information. When a collision or network congestion occurs, video streaming might be affected. The main reason behind network congestions and collision is a high routing overhead and to solve this issue, the beaconless routing protocols have been proposed [20, 21] Rather than a continuous information exchange with the neighboring vehicles, each vehicle considers the speed and position information before making any forwarding decision [11]. From the information, the receivers set a timer and the vehicle to first reach count zero will be selected as the next hop relay. For a high traffic collision environment, taking a routing decision based on the distance and vehicle stability is insufficient. The reactive routing protocols have a better mean opinion score in a VANET environment [27] whereas the location-based VANET routing protocols have a good result due to location accuracy [12, 24]. However, they face a degradation in average mean opinion score [32] and affect the quality of video streaming over VANETs. Several works have focused on the performance evaluation of the position-based routing protocols for video streaming in VANETS based on the peak signal-to-noise ratio, video quality metric, and QoE measures [25, 32]. To avoid the issue of retransmission, ETX has been employed in VANET protocols, however, ETX alone tends to present a long routing path and high delays. Thus, neither the location-based distance measure nor the ETX metric alone is efficient for video streaming over VANETs. To accomplish this, the proposed RESP protocol exploits relay suitability as a metric of communication quality and considered different routing aspects in relay suitability estimation.

3 The proposed system model idea

The proposed work considered the proximity of the vehicles to the destination. Expected Transmission Count (ETX), collision probability, and relative velocity based link stability in relay suitability measurement. The weighted division algorithm is used for an efficient integration of factors in the scalable router selection for unicast geographic routing. There are two different routing aspects considered in relay suitability measurement - data forwarding probability and data forwarding delay. The first aspect includes the link stability and collision probability. Figure 1 demonstrates the RESP architecture of the proposed work.

Considering the vehicle velocity alone is not feasible for scalable router selection because the velocity difference between two vehicles is dependent on the proximity of the vehicle. A considerable number of neighbors increase the possibility of packet losses due to collisions. A high traffic, to a greedy vehicle, results in collision and packet loss and limit the quality of a greedy link since it increases video distortion. Thus, the proposed work considered the collision probability in a scalable router selection. The second routing aspect handles the problem of data forwarding delay with the support of geographical advancement towards the destination and ETX. The proposed work only considered a few vehicles located in a preferable forwarding region (near to the destination vehicle). It avoids the selection of the number of hops to reach the destination and a high computational time. The sending vehicle applies the relay suitability measurement only on the vehicles in a preferable forwarding region. The distance routing metric alone enables the routing protocol to select the shortest routing path, however, it is prone to transmission failures due to poor link quality since the transmission capability of a node is not considered in the decision-making process.

FThe selection of routing vehicles with low ETX increases the reliability of a network. However, ETX alone is prone to long routing paths and high latency on routing video frames. Thus, the distance or ETX metric alone is inefficient for video streaming over VANETs. To accomplish this, the proposed work takes into account different routing aspects in relay suitability estimation. Moreover, simply combining the routing factors is not an adequate way, since these are compensating with each other. Thus, the proposed work designs an efficient weighted division algorithm and efficiently delivers the video stream over selected scalable routing vehicles.

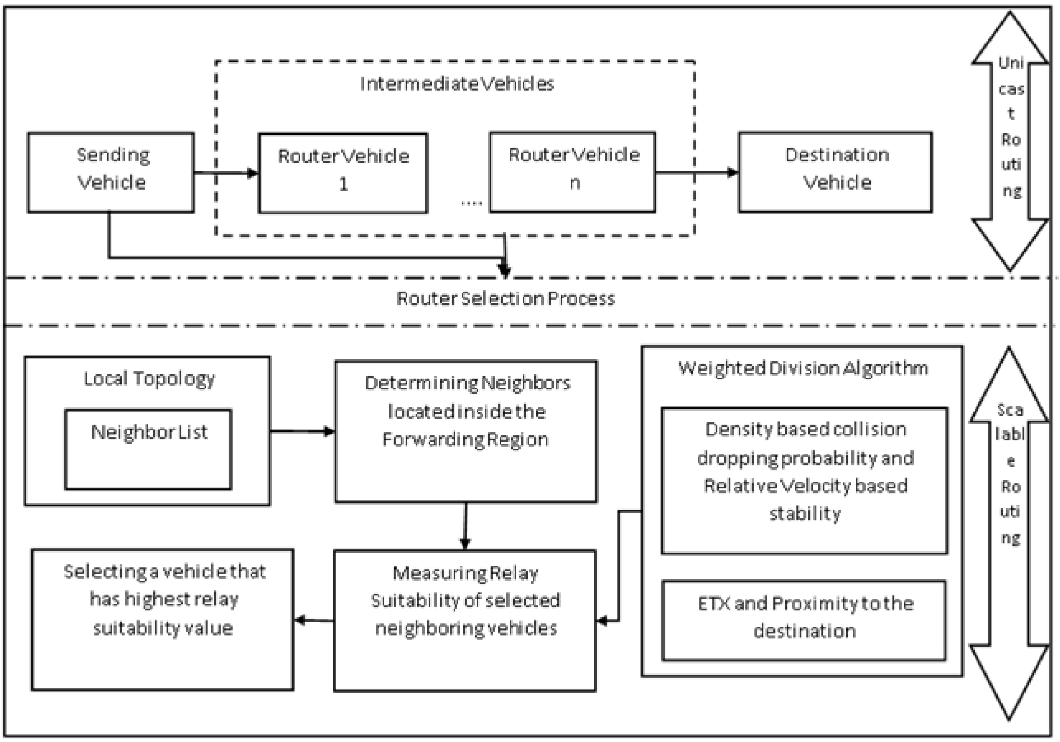


Figure 1: RESP architecture

4 Proposed routing mechanism

This section is intended to explain the mechanism of the proposed routing protocol and how this new mechanism can enhance the performance of the proposed protocol.

4.1 Positive forwarding region formation

The proposed work focused on the consolidation of multi-objective metrics which combine delay and stability with a successful forwarding probability. Delay and packet loss should not exceed their predefined limits since both are the deciding factors of video quality. The transmission of a video demands the use of high-quality links within a short time, thus, high-quality positive forwarding nodes must be selected from the neighbor list. Every node must broadcast a neighbor's discovery message to identify its neighboring nodes (Nn) in the communication range R. A node which receives the request message must respond by sending its identification number (ID) and the information relevant to the multi-objective metric. Every node receives several messages from its neighbors and builds its routing table. The nodes follow the periodic alive message propagation to maintain their presence in the communication range as a neighbor. If a node does not receive an alive message from the registered neighboring node until a predefined period, it erases the entry related to such neighbor from its routing table. Before selecting the router node, the source node (Sn) executes the multi-objective function. The proposed work employs no execution control, the sender node would be executed for the neighboring nodes located in all the directions ad this results in an imperfect router selection and affects the video quality. Since there are compensating factors for distance metric, the sender selects and transmits the packet to a node in the backward direction to the destination. Therefore, to minimize the unnecessary transmission hop increment and quality degradation, each node that transmits the data packet must generate a "forwarders list" based on the distance information and apply the link quality measurement on them using a weighted division algorithm. The notations used in the work are listed in table 4.

Table 1: Notations used in RESP protocol

Notations	**Description**
Nn	Neighboring node
R	Communication range
SN	Source node
Lsx, Lsy	Location Coordinates of node SN
DN	Destination node
DistND	Distance between the destination and neighboring node
DistSD	Distance between the source and destination node
POSFN	Positive forwarding region
ETX	Expected Transmission count
$\Delta\ v_j$	Relative velocity between node $_j$ and $_i$
CP_j	Collision probability of node $_j$
NNM	The neighbors which are added in the MPFN of a node
Link $\mathrm{Quality}_{i,j}$	Quality of a link $_{i,j}$
α and β	Weighting factors
$P_{i,j}$ and $P_{j,i}$	Weighting The packet delivery ratio of links $_{i,j}$ and $_{j,i}$

Every forwarder of the data packets must generate a forwarders list. Note that the SN and intermediate nodes exploit the same process. To route towards the destination nodes (DN), the SN obtains the location information of the DN using the external location server. By applying Equation 1, the sender node estimates the distance between the source (Lsx, Lsy) and destination with the location coordinates (Ldx , Ldy). The DistND represents the distance between the destination and the neighboring nodes while DistSD represents the distance between the source and the destination nodes. From Equation 2, the sender node generates the set of forwarding nodes (FN), which include the nodes that satisfy the condition of DistSD > DistND.

$$Dist\_SD = [(L\_dx - L\_sx)^2 + (L\_dy - L\_sy)^2]^{(1/2)} \tag{1}$$

$$PSO_{FN} = \{F_N |\, N_N \in Dist_{SD} > Dist_{ND}\} \tag{2}$$

The process of forwarding node selection from the POSitive FN (POSFN) set provides the direction to deliver the data packets to the destination DN. In this way, although the data packets can approach the DN, the reachability cannot be guaranteed due to the negligence of the communication quality between the nodes. The proposed RESP needs to identify the quality routers that forward the data packets by considering the offset angle-based degree of mobility.

4.2 Degree of mobility

The proposed RESP balances the router suitability to forward the video packets between the node geographic advancement towards the destination and the link stability. To do this, the proposed RESP protocol considered a small region at the boundary of the communication range formed by the offset angle and the source node as the center of the circle. The restricted small region has been referred to as the "most preferable forwarding area" (MPFN) (Figure 2). Since the vehicles present at the edge of the communication range can move out of the communication range at any time, the vehicles present in the MPFN area were selected for the measurement of relay suitability. The preferable routers have a small degree of mobility. Also, there is a chance of experiencing a bad signal reception which can negatively impact the video quality. The forwarding region (FN) and the MPFN are shown in Figure 2

The area of the segment Area$_{TUV}$ in a communication range of a node is named as the forwarding region, and the nodes located within the Area$_{TUV}$ are denoted as a set of FN. The restricted segment area of Area$_{TUV}$ is denoted as MPFN (Area$_{OPQR}$) because the nodes within the segment of the Area$_{VOP}$ can move out of the communication range of a sender node at any time. This can increase the uncertainty of a forwarding node, thus, the RESP restricts the forwarding region segment area and selects a preferable forwarding region for routing. Notably, the nodes within the segments ofArea$_{SQR}$, Area$_{STO}$, and Area$_{SUP}$ increased the delay due to the long traveling path of a packet towards the destination. TheArea$_{OPQR}$ is estimated using the following equations:

$$Area_{OPQR} = Area_{SOP} - Area_{SQR} \tag{3}$$

The points 'O' and 'p' lie on the circle of communication range, with an angle of α. Initially, the value of α is assigned as 450. If there are no nodes within the MPFN, all the nodes in the FN that satisfy the conditions of Equation 2 are involved in the relay suitability measurement. Moreover, points 'Q' and 'Z' lie at the distance of half of the communication range with an angle of α.

$$Area_{SOP} = (\frac{Dist_{op}}{4})\ x(4R^2 - {Dist_{op}}^2)^{1/2} \tag{4}$$

$$Area_{SQR} = (\frac{Dist_{QR}}{4})\ x(4R^2 - {Dist_{QR}}^2)^{1/2} \tag{5}$$

The Area$_{OPQR}$ is estimated by substituting Equations 4 and 4 in 3. From the measured area, the nodes present inside the area are called the preferable forwarding nodes. The sender node applies the relay suitability measurement to these nodes and improves the efficiency of routing. The proper selection of routers facilitates the scaling of RESP routing to a large network without reducing its routing reliability.

4.3 Different routing aspects of relay suitability factor

The main aim of the proposed methodology is to design a scalable and an efficient geographic routing protocol by considering different aspects of routing strategy and efficiently integrating the compensating routing factors. The source selects the neighbors from the preferable forwarding region before sending the data packet. To select the router, the RESP protocol exploits the RElay Suitability (RES) as a metric of communication quality. A sender node selects a neighbor node $\in$ MPFN with high RES to enhance the reliability. If the sender node simply selects a neighbor with a low PER as per the distance metric, the routing delay increases because of the increase in the number of hops to the DN. It limits the scalability of the network to support video streaming in VANETs. Hence, the proposed RESP applies the delay, collision impact, stability, and ETX metrics in weighted division algorithm and selects a high-quality router for video streaming.

4.4 Forwarding probability and delay metrics

The most deciding factors of forwarding probability are collision-free network and link stability. The proposed RESP protocol suggests the estimation of link stability according to the

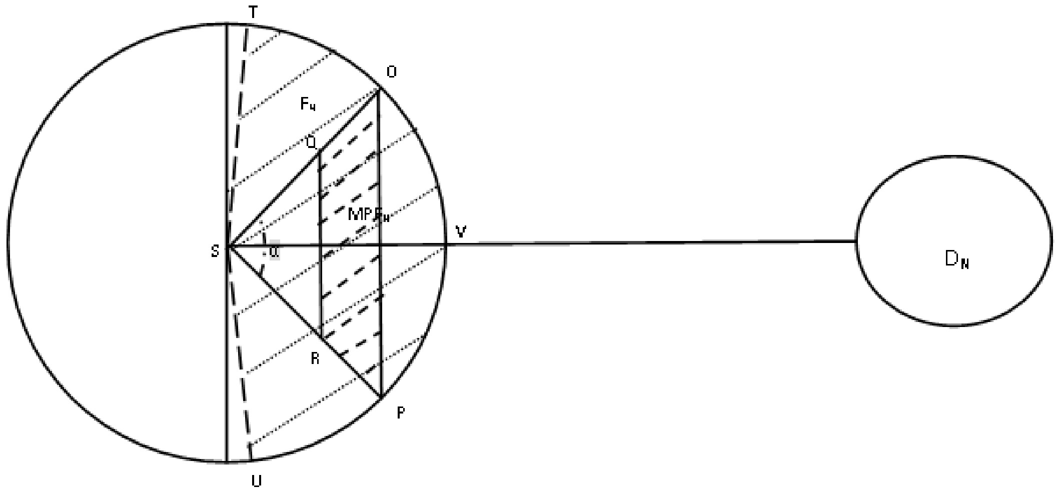


Figure 2: Most preferable forwarding regions

relative velocity of two vehicles and data collision probability using node density. The current works considered node velocity but the use of node velocity alone cannot guarantee the estimation of the future network topology from the relative velocity of two vehicles since their velocity does not depend on their proximity. Based on this, the relative velocity of a node with the neighboring nodes located in the restricted forwarding region was considered in the proposed protocol. If the vehicles move in the same direction at the same speed, their relative velocity is considered as zero, but if they are moving in the same direction with a significant difference in their speed, the relative velocity is considered high. The relative velocity (ΔV) is used as a primary factor in the determination of relay suitability.

$$\Delta V_i = V_{SN} - V_i\,,\ Where\, i\, \text{varies}\, from\, 1\, to\, |MPF_N| \tag{6}$$

Moreover, the packet collision increases with the number of active nodes (the nodes transmitting the packets simultaneously), even as a high traffic to a selected intermediate vehicle can cause both collision and packet loss. Thus, the RESP estimates the probability of packet dropping due to collisions. The packet level lossy network collision model for each link $N_{i,j}$ is taken as the input number of nodes that add the node j to its corresponding MPFN (NNM) over the total neighbors. Equation 7 showed the estimation of collision probability of node j.

$$CP_j = \frac{(\,Area_{network}\, x\, NN_M)}{(Total\, Nodes\, x\, 3.14\, xR^2)} \tag{7}$$

It considered the network area, total number of nodes, and the communication range of the nodes. Assume that all the nodes know the network area and the number of nodes; when a node is added in the MPFN of most of its neighboring nodes, there is a chance for packet collision. The delay and ETX are essential parameters for video streaming over VANETs and to deliver video files to their destination on time, it is essential to select efficient routing nodes. Considering ETX alone expose the routing protocol to long routing path determination and high latency on routing video frames.

$$ETX_{i,j} = \frac{1}{(P_{i,j}\, x\, P_{j,i})} \tag{8}$$

By applying Equation 8, the ETX value for a communication link ij can be estimated, where Pij and Pji represent the packet delivery ratio of links ij and ji, respectively. By using all the factors, the proposed RESP estimates the RES value for all the nodes ∈ MPFN. Moreover, the distance metric-based selection router or the greedy node is not much reliables since the can move out of the communication range anytime. Thus, the distance or ETX metric alone is inefficient for video streaming over VANETs and in the RESP, both metrics were considered as forwarding delay factors.

4.5 Scalable and QoS data forwarding by applying weighted division algorithm

The proposed RESP considered multiple routing aspects such as the probability of packet dropping due to the collision, ETX, relative velocity, and distance factor. However, an improper integration of these factors tends to compensate each other for their influence on the scalable and efficient router selection for video streaming since the greedy node drops the packet due to mobility and enables the sender node to retransmit the same packet frequently. Moreover, the selection of the shortest link for routing increases video frame delivery delay and video buffering. Thus, the proposed RESP integrates the compensating factors using a weighted division algorithm to improve the performance of a VANET router selection process.

$$Link\,Quality_{i,j} = \left\{ \frac{[(1-\alpha)\,(1-\frac{\Delta V_{i,j}}{R})\,] + (\alpha)\,(1-CP_j)}{(1-\beta)\,(ETX) + (\beta)\,(1 - [\frac{Dist_{i,j}}{R}]} \right\} \tag{9}$$

In Equations 8 and 9, α and β are the weighting factors, and moreover, both values are in the range of 0 to 1. The stability and collision probability is the forwarding probability relevant metrics. If the ratio of $\frac{\Delta V_{i,j}}{R}$ is nearly 1, the link will be disconnected. The factor CP_j also creates a negative impact. The weighted addition of factors 1- ($\frac{\Delta V_{i,j}}{R}$) and (1-CP_j) are taken in the denominator whereas the weighted summation of ETX and geographic advancement are taken in the numerator. The result of a weighted division algorithm represents the quality of the communication link. The node with a high link quality in the MPFN is selected in the packet forwarding. The collision-free links with a better stability significantly mitigate the packet delay and dropping. Thus, the selection of high-quality links ensure a successful delivery of video frames and a significant reduction in video distortion and buffering.

The algorithm 1 illustrates the process of RESP protocol in detail. The processes in RESP protocol are divided into three procedures, such as Most Preferable forwarding region selection, Relay suitability (Link quality) Measurement, and Scalable router selection. Every node that has data packets to forward to the destination executes the aforesaid three procedures to select the scalable router for video streaming. The following flow chart in Figure 3 provides a detailed picture of processes in RESP protocol.

5 Performance evaluation

The performance of the proposed RESP was evaluated in an NS2 simulator on a 2000 m road scenario where the vehicles move in two independent directions. The simulation model imports the roadmap from the SUMO. The maximum velocity of vehicles in each lane was 130 km/h. In this scenario, the vehicles follow MOVE mobility model for movement. To evaluate the performance of the RESP, it was compared to LBRP [11], LSGR [13] and GPSR [18] under various scenarios by varying the number of vehicles. The agents such as MyTrafficTrace, MyUDP, and MyUDPSink were built inside the NS2 to generate video traffic. The sender forwards the video format of YUV QCIF or YUV CIF to the destination. To support the digitized video transfer, this work utilized the MPEG4 codec to compress the file. The MyTrafficTrace agent fragments

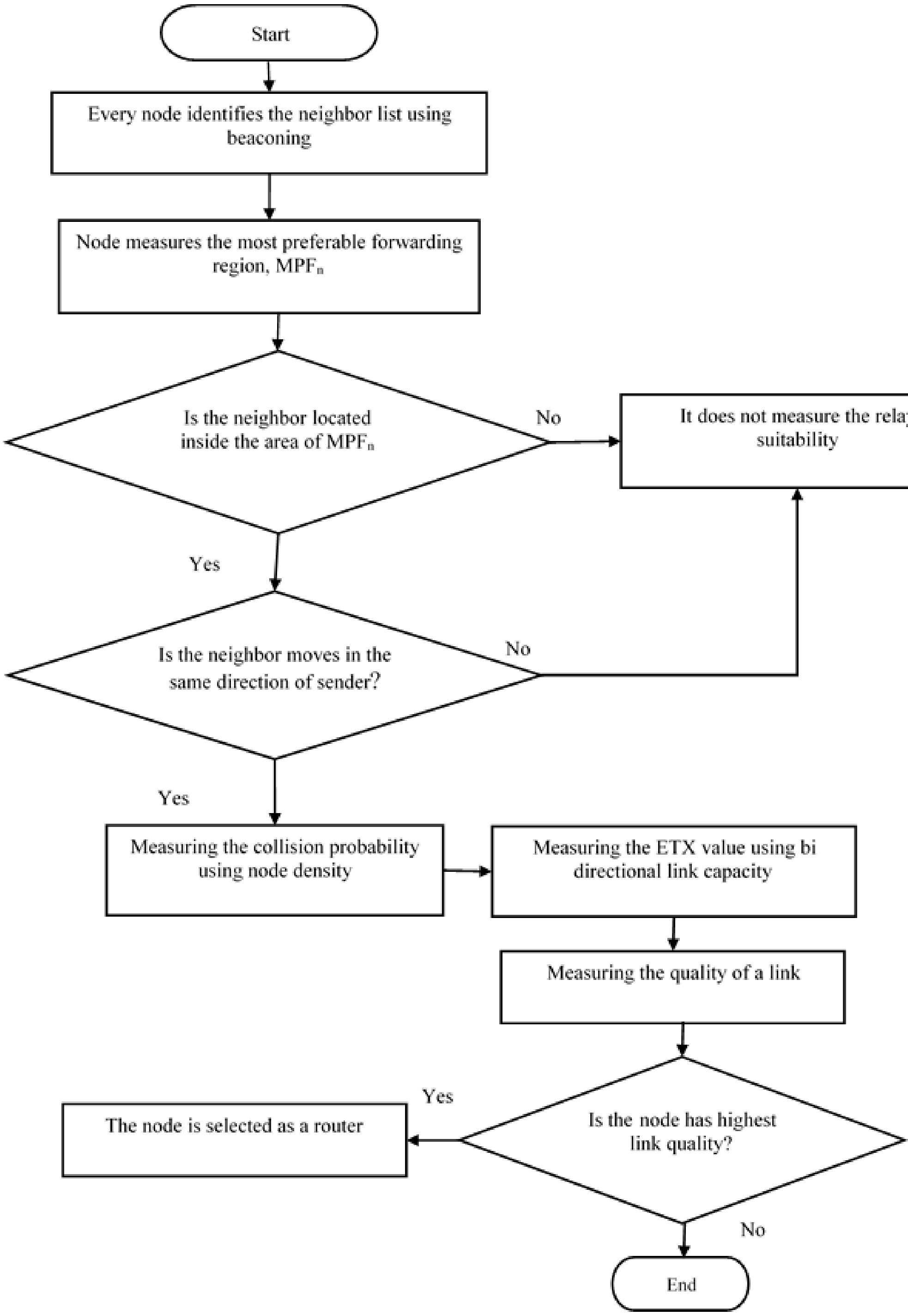


Figure 3: Flow chart of process of RESP protocol

Algorithm 1 Scalable router selection for video streaming

Input: Neighbor List of a sender node
Output: Scalable video delivery
Procedure 1: Most Preferable forwarding region selection
Procedure 2: Relay suitability (Link quality) Measurement
Procedure 3: Scalable router selection
Every sender node ()
{
Procedure 1 ()
if $Dist_{SD} > Dist_{SD}$ **then**
    The node is added in the forwarding Region, FN; Area of MPFn is measured; The nodes in the MPFn are identified;
else
    The node is in the neighbor list only;
end if
}
Procedure 2 ()
{
if $\Delta V > 0$ **then**
    The node is added in the forwarding Region, FN; CP_j and $ETX_{i,j}$ is measured; Link quality is measured using the Equation (9);
else
    The node is removed from the MPFn;
end if
}
Procedure 3 ()
{
for all $i = 0$; $i++$; $i < |MPFn|$ **do**
    {
    if Link Quality (i) > MAX **then**
        MAX = Link Quality (i)
        }
        The node in the MAX is selected as a router;
    end if
end for
}
}

each of the video frames in the compressed file into several segments. It was transmitted using MyUDP and received using MyUDPSink on the receiver side. The sender trace file retains the timestamp, the packet id, and the packet payload size for every UDP packet transmission, and the video trace file holds information about each frame. By utilizing the Tcl script, the network objects were configured for video streams. The video clip used in the simulation was 5.7 MB in size, 30 fps frame rate, with a pixel resolution of 720 X 576 in MPEG format. The simulation parameters are listed in Table 4.

Table 2: Simulation parameters

SIMULATOR	**Network Simulator 2**
Average total vehicle	50-100
density (Veh/km)	130
Vehicle max speed (km/h)	Randomly
Area	2000 X 100 m2
Communication Range	250 m
Interface Type	PHY/Wirelessphy
MAC Type	IEEE 802.11p
Queue Type	Droptail/Priority Queue
Queue Length	50 Packets
Antenna Type	Omni Antenna
Routing Protocol	RESP, LBRP, LSGR, and GPSR
Transport Agent	MyUDP, And MyUDPSINK
Application Agent	QMytraffictrace
Frame Rate	30 Fps
Input Video Format	YUV QCIF
Simulation Time	150s

6 Performance metric

The following performance metrics were used to evaluate the performance of scalable and reliable VANET routing under different density scenarios.

- Packet Delivery Ratio (PDR): It is the ratio between the total received packets and the generated
- Delay: It is the time taken to deliver the data from the source to the destination vehicle.
- Overhead : It refers to the total number of control packets involved in the scalable and reliable video transmission process.
- Frame Setback Ratio: The frame setback ratio is the ratio of the number of delayed frames and the number of forwarded frames per second. These were analyzed to show the impact of the routing schemes on the user perception in scenarios with different end-to-end distances and node densities.

7 Results and discussion

Two key performance metrics are shown in Figure 4, where the performance of the proposed protocol was compared to the performance of three other protocols - LBRP, LSGR, and GPSR. In this figure, the validated influence of vehicular traffic density on performance metrics was on the path lifetime and the overall transmission delay for the target content. In Figure 3a, the performance of the constructed path lifetime was compared. Recall that the path lifetime is determined by the link with the minimum lifetime. As can be seen from the figure, the proposed RESP produced the longest path lifetime since it preferred the neighboring vehicle that can maintain a high connectivity performance. Next in importance is that the LSGR showed the second-best path lifetime as a result of the consideration of both link lifetime and one-hop transmission distance information on the protocol design. On the contrary, without the consideration of the link connectivity property, the path tends to be fragile as illustrated in the GPSR and LBRP. Especially for GPSR, the constructed path is the most easily broken path with the largest inter-relay distance.

Figure 4(b) showed the observed delay of the RESP, LBRP, LSGR, and GPSR when the number of nodes was varied from 50 to 100. RESP achieved a very low packet delivery delay for varying nodes. There is a minimal variation in the packet delay between RESP and LBRP. Most of the delivered packets by RESP reached their destination on time, and this leads to a considerable delay in LSGR and GPSR. However, for the RESP, efficient packet transmissions were performed with a highly stable and less congested communication links. The influencing parameters in the RESP were the node stability and network collision. It minimized the com-

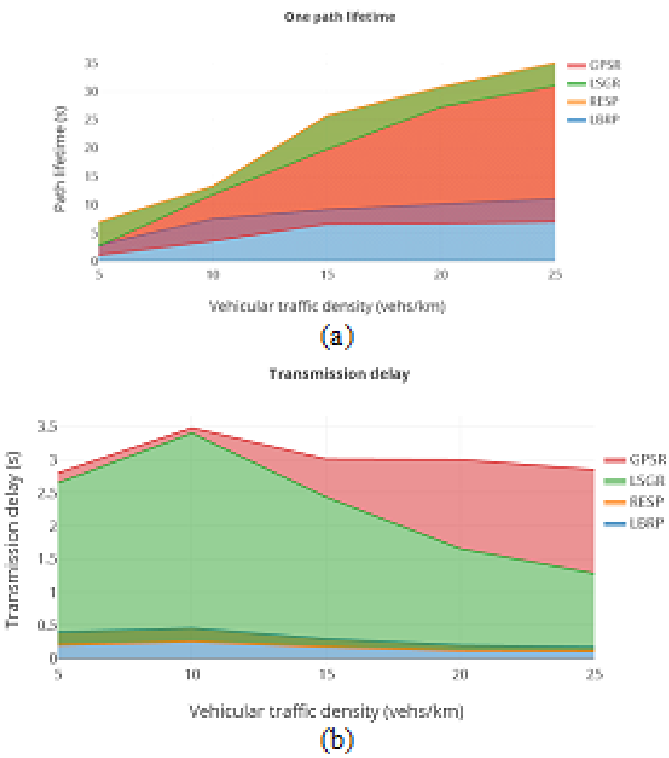


Figure 4: The three-performance metrics (a) Path lifetime; (b) Transmission delay. Link Lifetime-aware Beacon-less Routing Protocol (LBRP); Link State-aware Geographic Routing (LSGR); Greedy Perimeter Stateless Routing (GPSR)

munication delay without degrading the packet delivery ratio. Both RESP and GPSR limits the packet delay in the range of 0.01 to 0.03 seconds. However, LBRP delivers the packets in 0.04 seconds in the low-density network. When the number of nodes increased to 100, the data packets of LBRP were delivered with a delay of 0.02 seconds which was closer to the delay of RESP and GPSR at the same scenario. Even though RESP showed a better performance in terms of delay, the packet delivery ratio was less compared to those of LBRB, LSGR, and GPSR. The path lifetime metrics increases with vehicle speed when the energy factor must be considered in the holding time computation to increase network lifetime.

In Figure 5, the other three performance metrics were shown, where the performance of the proposed RESP was compared to the performance of LBRP, LSGR, and GPSR. In this figure, the validated influence of vehicular traffic density on the performance metrics was in the packet delivery ratio, overhead, and the overall frame setback ratio.

Figure 5 (a) displayed the packet delivery ratio of RESP, LBRP, LSGR, and GPSR when the number of nodes was varied from 50 to 100. The performance of these solutions was compared for scalability by increasing the vehicle density with the same data traffic. The number of vehicles in the network showed a huge impact on the packet delivery ratio of the 4 protocols. For instance, RESP improved the packet delivery ratio from 81.5% to 97.8% when the number of nodes increased from 50 to 100 compared to LBRP, LSGR, and GPSR. RESP achieved the highest delivery ratios because it does not have to select the intermediary vehicles in advance or wait to forward the data packets. Initially, LBRP attained a packet delivery ratio closer to

the RESP. With 70 to 80 nodes, LBRP improved the packet delivery ratio by 33% more than LSGR. The minimum node density reduced the probability of LBRP failing the communication links between the nodes. However, LSGR required more time to establish the entire path to the destination under a low-density environment and this increased its link breakage probability which resulted in a weaker packet delivery ratio. Beyond 60 nodes, LSGR drastically increased the packet delivery ratio compared to GPSR. The high dense scenario made the construction of the LSGR path faster. With a sufficient number of nodes, RESP improved the packet delivery ratio by 8.4% and 17.5% compared to LBRP, LSGR, and GPSR, respectively.

Figure 5(b) showed the overhead of the four protocols under various dense scenarios. As expected, increasing the number of nodes leads to a high overhead due to the increment of control packets. Due to the data transmission interruption in GPSR, the overhead fluctuates with various node density. For instance, LSGR spent 10980 packets to deliver data packets at a low dense scenario while decreasing the overhead to 30205 packets at a high dense scenario. RESP achieved a better result when using the hello packets to generate its neighbor's list and share the velocity information. Compared to the other protocols, RESP attained low overhead at a lower node number, but an increased overhead (from 8960 to 21600 packets) when the node number was increased from 50 to 100. LBRP conducted a receiver-based beaconless routing, however, it transmitted the content request and response packets. The number of these packets are increased compared to the number of packets carrying video contents under a highly dynamic network topology. For instance, LBRP increased the overhead from 15983 to 39875 packets when the number of nodes was increased from 50 to 100 in a high-density scenario but GPSR had the highest overhead rate (starting from 39008 and ending at 79502).

Figure 5(c) showed the result of the frame setback ratio. The delayed data packet is a primary reason behind the FSR value decrement. As expected, the FSR value was increased by all the protocols when the number of nodes was increased from 50 to 100. The RESP considers the relative velocity in link stability measurement rather than the node velocity. It improves the accuracy of node stability measurement and the performance of the RESP. For instance, RESP achieved 0.0014 of FSR when the number of nodes was increased from 50 to 100. Moreover, the number of neighboring nodes having the node identity in their corresponding MPFN list decides

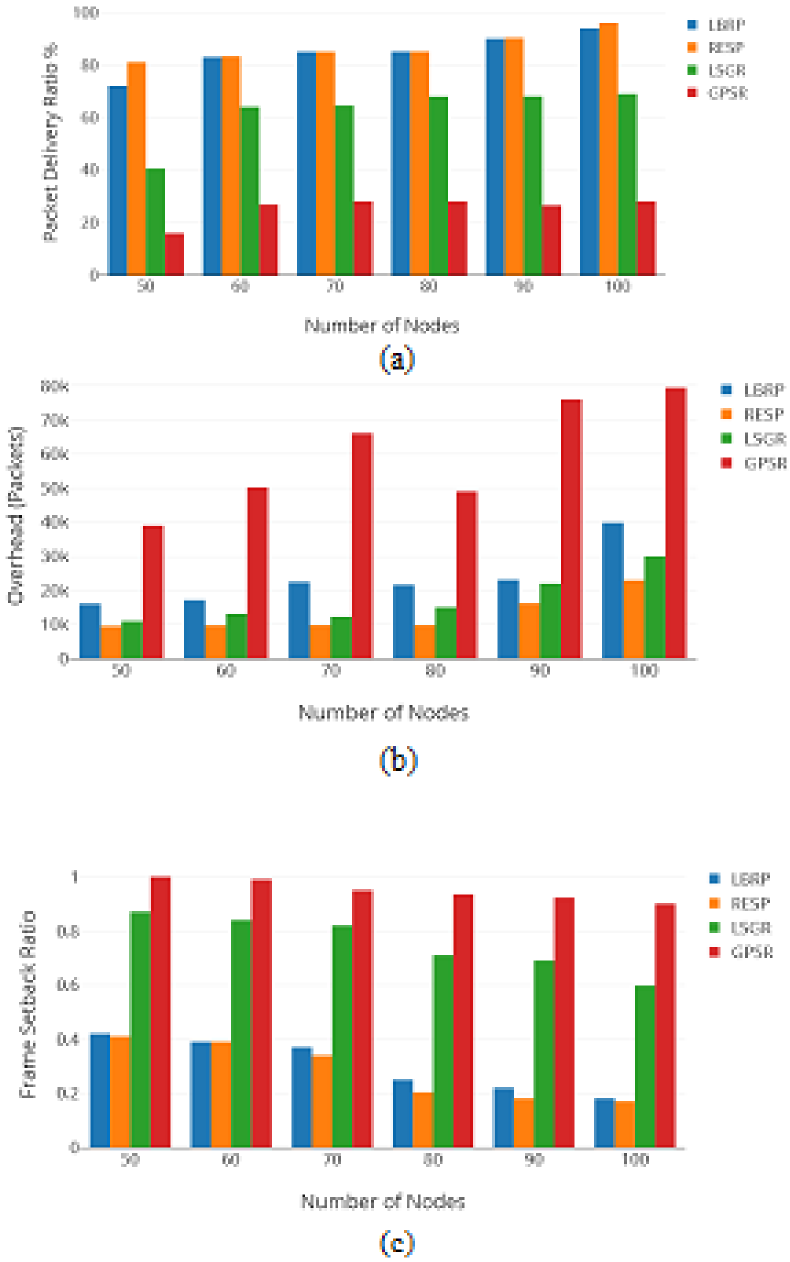


Figure 5: The three-performance metrics (a) Packet delivery ratio; (b) Overhead; (c) Frame Setback Ratio. Link Lifetime-aware Beacon-less Routing Protocol (LBRP); Link State-aware Geographic Routing (LSGR); Greedy Perimeter Stateless Routing (GPSR)

the packet collision at a given node. Considering this scenario, RESP made routing decisions and delivered data packets on time, thereby significantly reducing the frame setback ratio. Compared to LBRP, RESP reduced the frame setback ratio by 0.358 under a high dense scenario.

8 Conclusion

This work presented a scalable unicast geographic routing protocol for video streaming in VANETs. The building of a small region at the boundary of the communication range using an offset angle and applying the relay suitability measurement only on those nodes can reduce computational complexities without degrading the performance. In VANET scenarios, vehicle density, vehicle mobility, and destination mobility also reduce the scalability of protocols. The relay suitability measurement includes the forwarding probability and the delay measurement, and this ensures the trade-off between the node geographic advancement towards the destination and the link stability. This work conducted extensive experiments over VANET scenarios where the performance of the proposed and existing unicast routing protocols was compared in terms of scalability by increasing the vehicle density. The simulation results showed improvements in the RESP compared to the existing protocols. RESP was validated via simulation studies and showed to have a greater performance in terms of providing a high PDR and less packet delay with various network density. Moreover, it offered the lowest delay in delivering video frames. It is, therefore, concluded that RESP is a better scalable protocol for delivering video frames on time under a highly dynamic vehicular network topology.

Funding

The research work has been supported by Universiti Malaysia Pahang (UMP). Grant Id: RDU1703212.

Bibliography

[1] Ahuja, K.; Khanna, R.; Singh, B. (2011); Real time available-bandwidth estimation (ABE) algorithm based selection in heterogeneous network for wimax and 3G, *In Next Generation Mobile Applications, Services and Technologies (NGMAST), IEEE, 2011 5th International Conference on* , 169-174. 2011.

[2] Ayoob, A. A.; Su, G.; Wang, D.; Mohammed, M. N.; Hammood, O. A. (2017); Hybrid LTE-VANETs based optimal radio access selection, *In International Conference of Reliable Information and Communication Technology* , Springer, Cham, 189-200, 2017.

[3] Cabrera, G.; Niklander, S.; Cabrera, E.; Johnson, F. (2016); Solving a distribution network design problem by means of evolutionary algorithms, *Studies in Informatics and Control*, 25(1), 21-28, 2016.

[4] Chen, Y.; Weng, S.; Guo, W.; Xiong, N. (2016); A Game Theory Algorithm for Intra-Cluster Data Aggregation in a Vehicular Ad Hoc Network, *Sensors*, 16(2), 245, 2016.

[5] Cunha, F.; Villas, L.; Boukerche, A.; Maia, G.; Viana, A.; Mini, R. A.; Loureiro, A. A. (2016); Data communication in VANETs: Protocols, applications and challenges *Ad Hoc Networks*, 44, 90-103, 2016.

[6] De Felice, M.; Cerqueira, E.; Melo, A.; Gerla, M.; Cuomo, F.; Baiocchi, A. (2015); A distributed beaconless routing protocol for real-time video dissemination in multimedia VANETs, *Computer Communications*, 58, 40-52, 2015.

[7] Eze, E.C.; Zhang, S. J.; Liu, E. J.; Eze, J. C. (2016); Advances in vehicular ad-hoc networks (VANETs): Challenges and road-map for future development, *International Journal of Automation and Computing*, 13(1), 1-18, 2016.

[8] Hammood, O. A.; Kahar, M. N. M., Mohammed, M. N., Hammood, W. A., Sulaiman, J. (2018) ; The VANET-Solution Approach for Data Packet Forwarding Improvement, *Advanced Science Letters*, 24(10), 7423-7427, 2018.

[9] Hasan, R. A.; Mohammed, M. A.; Tapus, N.; Hammood, O. A. (2017); A comprehensive study: Ant Colony Optimization (ACO) for facility layout problem, *In Networking in Education and Research (RoEduNet), 16th RoEduNet Conference* , 1-8, IEEE, 2017.

[10] Hammood, O. A.; Kahar, M. N. M.; Mohammed, M. N. (2017); Enhancement the video quality forwarding Using Receiver-Based Approach (URBA), *Vehicular Ad-Hoc Network, In Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET), 2017 International Conference on* , IEEE, 64-67, 2017.

[11] Hu, M.; Zhong, Z.; Ni, M.; Baiocchi, A. (2016); Design and analysis of a beacon-less routing protocol for large volume content dissemination in vehicular ad hoc networks, *Sensors*, 16(11), 1834, 2016.

[12] Husain, A.; Sharma, S. C. (2015); Simulated analysis of location and distance based routing in VANET with IEEE802, *Procedia Computer Science*, 57, 323-331, 2015.

[13] Li, C.; Wang, L.; He, Y.; Zhao, C.; Lin, H.; Zhu, L. (2014); A link state aware geographic routing protocol for vehicular ad hoc networks, *EURASIP Journal on Wireless Communications and Networking*, 2014(1), 176, 2014.

[14] Lochert, C.; Mauve, M.; Fubler, H.; Hartenstein, H. (2005); Geographic routing in city scenarios, *ACM SIGMOBILE mobile computing and communications review*, 9(1), 69-72, 2005.

[15] Margaria, D.; Falletti, E. (2016); The local integrity approach for urban contexts: Definition and vehicular experimental assessment, *Sensors*, 16(2), 154, 2016.

[16] Menouar, H.; Lenardi, M.; Filali, F. (2007); Movement prediction-based routing (MOPR) concept for position-based routing in vehicular networks, *In Vehicular Technology Conference, 2007. VTC-2007 Fall. 2007 IEEE 66th* , IEEE, 2101-2105, 2017.

[17] Mezher, A. M.; Igartua, M. A. (2017); Multimedia Multimetric Map-aware Routing protocol to send video-reporting messages over VANETs in smart cities, *IEEE Transactions on Vehicular Technology*, 66(12), 10611-10625, 2017.

[18] Mohammed, M. N.; Nahar, A. K.; Abdalla, A. N.; Hammood, O. A. (2017); Peak-to-average power ratio reduction based on optimized phase shift technique, *In Communications and Information Technologies (ISCIT), 2017 17th International Symposium on* , IEEE, 1-6, 2017.

[19] Orchard, M.; Tang, L.; Saha, B.; Goebel, K.; Vachtsevanos, G. (2010); Risk-sensitive particle-filtering-based prognosis framework for estimation of remaining useful life in energy storage devices, *Studies in Informatics and Control*, 19(3), 209-218, 2010.

[20] Quadros, C.; Cerqueira, E.; Santos, A.; Lim, J.; Gerla, M. (2015); Beacon-less video streaming management for VANETs based on QoS and link-quality, *In Integrated Network Management (IM), 2015 IFIP/IEEE International Symposium on* , EEE, 191-198, 2015.

[21] Quadros, C.; Santos, A.; Gerla, M.; Cerqueira, E. (2016); QoE-driven dissemination of real-time videos over vehicular networks, *Computer Communications*, 91, 133-147, 2016.

[22] Reyes-Munoz, A.; Domingo, M.; Lopez-Trinidad, M.; Delgado, J. (2016); Integration of body sensor networks and vehicular ad-hoc networks for traffic safety, *Sensors*, 16(1), 107, 2016.

[23] Saif. A; Moath. M. AL; Ali. H. Al; Hussien. Z. (2014); A comprehensive survey on vehicular ad hoc network, *Journal of network and computer applications*, 37, 380-392, 2014.

[24] Sharef, B. T.; Alsaqour, R. A.; Ismail, M. (2013); Comparative study of variant position-based VANET routing protocols *Procedia Technology*, 11, 532-539, 2013.

[25] Song, C.; Tan, G.; Yu, C. (2017); An Efficient and QoE Supported Multichannel MAC Protocol for Vehicular Ad Hoc Networks *Sensors*, 17(10), 2293, 2017.

[26] Song, C. (2017); Performance Analysis of the IEEE 802.11 p Multichannel MAC Protocol in Vehicular Ad Hoc Networks, *Sensors*, 17(12), 2890, 2017.

[27] Spaho, E.; Ikeda, M.; Barolli, L.; Xhafa, F. (2013); Performance comparison of OLSR and AODV protocols in a VANET crossroad scenario, *In Information Technology Convergence* , Springer, Dordrecht, 37-45, 2013.

[28] Tripp-Barba, C.; Urquiza-Aguiar, L.; Igartua, M. A.; Rebollo-Monedero, D.; de la Cruz Llopis, L. J.; Mezher, A. M.; Aguilar-Calderon, J. A. (2014); A multimetric, map-aware routing protocol for VANETs in urban areas, *Sensors*, 14(2), 2199-2224, 2014.

[29] Xie, F.; Hua, K. A.; Wang, W.; Ho, Y. H. (2007); Performance study of live video streaming over highway vehicular ad hoc networks, *In Vehicular Technology Conference 2007. VTC-2007 Fall. 2007 IEEE 66th* , 8(1), 132, 2013.

[30] Xu, S.; Guo, P.; Xu, B.; Zhou, H. (2012); Study on qos of video communication over VANET, *In International Conference on Information Computing and Applications* , Springer, Berlin, Heidelberg, 730-738, 2012.

[31] Wan, J.; Liu, J.; Shao, Z.; Vasilakos, A. V.; Imran, M.; Zhou, K. (2016); Mobile crowd sensing for traffic prediction in internet of vehicles, *Sensors*, 16(1), 88, 2016.

[32] Zaimi, I.; Houssaini, Z. S.; Boushaba, A.; Oumsis, M.; Aboutajdine, D. (2016); Vehicular ad-hoc network: Evaluation of QoS and QoE for multimedia application, *International Conference on Networked Systems* , Springer, Cham, 367-371, 2016.

[33] Zhou, Z.; Dong, M.; Ota, K.; Chang, Z. (2015); Energy-efficient context-aware matching for resource allocation in ultra-dense small cells, *IEEE Access*, 3(9), 1849-1860, 2015.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 14(1), 39-55, February 2019.

A Prediction Algorithm based on Markov Chains for finding the Minimum Cost Path in a Mobile WSNs

G.A. Montoya, Y. Donoso

Germán A. Montoya*, Yezid Donoso
Systems and Computing Engineering Department
Universidad de Los Andes, Bogotá, Colombia
*Corresponding author:ga.montoya44@uniandes.edu.co
ydonoso@uniandes.edu.co

Abstract: In this paper we propose the usage of a prediction technique based on Markov Chains to predict nodes positions with the aim of obtain short paths at minimum energy consumption. Specifically, the valuable information from the mobility prediction method is provided to our distributed routing algorithm in order to take the best network decisions considering future states of network resources. In this sense, in each network node, the mobility method employed is based on a Markov model to forecast future RSSI states of neighboring nodes for determining if they will be farther or closer within the next steps. The approach is evaluated considering different algorithms such as: Distance algorithm, Distance Away algorithm and Random algorithm. In addition, with the aim of performing comparisons against optimal values, we present a mathematical optimization model for finding the minimum cost path between a source and a destination node considering all network nodes are mobile. This paper is an extended variant of [8][a].
Keywords: Markov model, RSSI, MWSN.

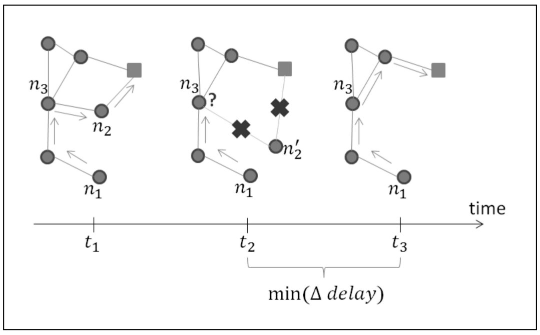
[a]Partially reprinted and extended, with permission based on Licence Number 4484761270350 © IEEE, from "2018 7th International Conference on Computers Communications and Control (ICCCC)."

1 Introduction

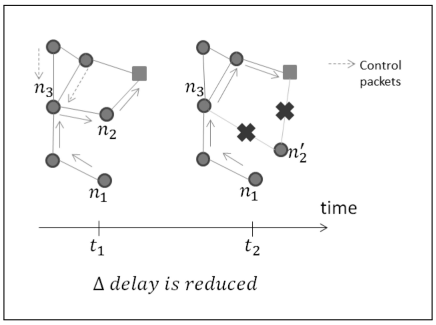
The advances of WSN have allowed attaching the sensors to an entity such as objects, animals or humans, to monitor a physical variable in its environment. However, the sensors are equipped with limited batteries whereby it is required to implement energy efficient routing techniques to extend the lifetime of the sensors as much as possible [3, 10]. In addition, communication disruptions caused by mobility in wireless sensor networks introduce undesired delays that affect the network performance in delay sensitive applications, such as health monitoring applications. Due to these kind of applications must provide a timeliness performance since they are dealing with health issues such as illness and continuous medical supervision, these applications must operate with minimum possible delay. In other words, a base station should no experiment delays from the information collected by the sensors [1, 2].

Given the scenario described above, a possible solution would consist to implement energy efficient routing techniques considering the sensor position to know the mobility level of the network. Based on this information, it is possible to determine the nodes that considerably affect the communication performance of the network. Some of these solutions propose the usage of sensors equipped with GPS devices, called GPS non-free approaches. However, these GPS non-free solutions have in most of cases drawbacks such as high implementation costs, delays for acquiring position information and non-accurate position information [4]. In addition, these types of solutions require an extra chip for the GPS [7], whereby more energy consumption is experimented. For these reasons, our work will not take into account sensors equipped with

GPS devices. Thus, in order to be aware of the network mobility we are going to use RSSI measurements, which indicates an approximated distance between two nodes.



(a) Problem



(b) Solution

Figure 1: Problem definition

The Figure 1.a) presents the problem we want to solve. Suppose we have a MWSN where at time t_1 there is a communication path between the source sensor node n_1 and the base station. However, at time t_2, the node n_2 moves away from the node n_3, causing a communication disruption for carrying the information from n_1 to the base station. Once n_3 has realized of this problem, at time t_3, n_3 has to perform routing corrections in order to reestablish the communication path between n_1 and the base station. The communication reestablishment between n_1 and the base station can be perfectly performed using routing techniques, but at the expense of introduce an undesired delay in this communication path. In some applications these delays can be omitted because do not affect the purpose itself of the application, but in other ones, such as delay sensitive applications like health monitoring, this disadvantage might mean a very low network performance.

Given the problem above, our proposal consists to use a predicting technique which is described in the figure 1.b) [5, 6]. It consists of the same situation showed in the Figure 1.a), but in this case, at time t_1, the node n_3 receive information that indicates the node n_2 will rapidly be away from its communication range, at time t_2. Given this information, n_3, at time t_1, is also analyzing a possible candidate which could replace n_2, in the case n_2 fails in a future time. If, indeed, at time t_2 the node n_2 fails because it has moved away from n_3, this node at time t_2 can promptly reestablish the communication path between n_1 and the base station, reducing the delay described in the Figure 1.a).

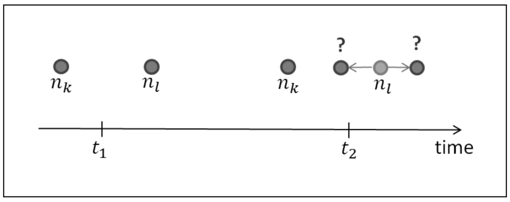
In order to solve the problem presented above, we propose to use a predicting method based on a Markov Model for estimating future RSSI states for a node with the aim of minimizing the delay experimented in the network. In this sense, our approach will be evaluated considering a Gauss-Markov mobility mode [9] where the mobility nodes can be considered predictable in order to test our prediction algorithm. Our work pretends to show an increasing network performance in terms of end-to-end delay and energy consumption against different algorithms such as: Distance algorithm, Distance Away algorithm and Random algorithm, which will be described in detail in the next sections. This paper is an extension of the work presented in ICCCC2018 [8]. In this sense, the extension consists to present a mathematical optimization model for finding the minimum cost path between a source and a destination node considering a mobile network with the aim of performing comparisons between optimal values and the algorithms proposed.

The remainder of the paper is organized as follows: Section 2 presents the prediction algorithm based on Markov Chains, that is, how the states are defined and the probability from going to one state to another. Section 3 shows the mathematical optimization model proposed, the objective function and the constraints. Finally, section 5 and 6 show the results and conclusions respectively.

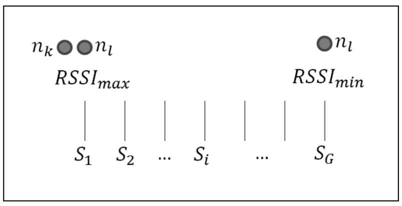
2 Prediction algorithm formulation

In order to solve the problem presented above, we propose to use a predicting method based on a Markov Model for estimating future RSSI states of a node with the aim of minimizing the delay experimented in the network . For this purpose, a detailed explanation, supported with the following figures, will be presented.

In relation to the Figure 2.a), suppose we have a network compound of two nodes: n_k and n_l, where n_l is a neighboring node of n_k. There are two times, t_1 and t_2, at which our little network is evolving in time. At time t_1 the node n_l is located at certain distance from n_k. However, at time t_2 we want to predict if n_l will be farther or closer (or at the same distance in t_1) from n_k.



(a) Possible movement of n_l



(b) RSSI States

Figure 2: Defining Markov states - 1

Respect to the Figure 2.b), there is a minimum and maximum distance at which n_l can be located in order to establish a communication link with n_k. At the minimum distance, n_l will have a maximum RSSI, $RSSI_{max}$, and, at the maximum distance, n_l will have a minimum RSSI, $RSSI_{min}$. At t_2, n_l could be located at any distance between $RSSI_{min}$ and $RSSI_{max}$.

Our goal consists to estimate the location between $RSSI_{min}$ and $RSSI_{max}$ at which n_l will be in a future time (in this case, t_2). Theoretically, there are infinite locations between $RSSI_{min}$ and $RSSI_{max}$, but for our model we assume discrete locations equitably spaced. These possible locations, at which n_l could be, are called states. In this sense, at a future time t_2, n_l could be at S_1, S_2, S_r or S_G, where G is the maximum number of states. The initial probability of n_l for being at any state S_i is $1/G$, which is called *Initial Probability Distribution of set S (π)*, can be expressed as follows:

$$\pi = \{P_{s_1}, P_{s_2}, ..., P_{s_G}\} \tag{1}$$

According to the Figure 2.a), suppose we want to know the probability to go from the the state S_2 to the state S_4, which is calculated with the following expression:

$$P_{24} = \frac{N(S_2, S_4)}{\sum_{j=1}^{G} N(S_2, S_j)} \tag{2}$$

Where $N(S_i, S_j)$ is the number of times that the state S_i follows state S_i.

This expression can be extensible for the rest of probabilities, as it is indicated in the following expression:

$$P_{ij} = \frac{N(S_i, S_j)}{\sum_{j=1}^{G} N(S_i, S_j)} \tag{3}$$

In this sense, we have the probability to go from any state S_i to any state S_j. These probabilities can be expressed in a matrix, which is called *Transition Matrix*:

$$T = \begin{bmatrix} P_{11} & P_{12} & ... & P_{1G} \\ P_{21} & P_{22} & ... & P_{2G} \\ . & . & ... & . \\ . & . & ... & . \\ P_{G1} & P_{G2} & ... & P_{GG} \end{bmatrix} \tag{4}$$

In relation to the figure 3.b), suppose that in a current time t_1, n_l is at state S_3 and we want to estimate the future state of n_l at a future time t_p. For this purpose, we can apply the following expressions:

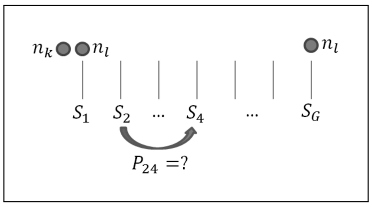
$$\pi_p = \pi * T^p \tag{5}$$

$$S_p = max\{\pi_p\} \tag{6}$$

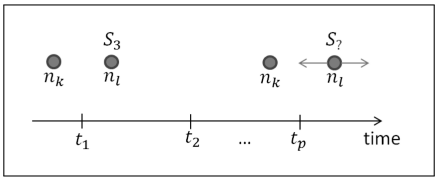
$$S_p = max\{P_{s_1}, P_{s_2}, ..., P_{s_G}\} \tag{7}$$

According to the expression 7, n_k can finally obtain the most probable future state at which n_l will be at a time t_p, and use this information for routing decisions in order to reduce the delay caused for probable communication disruptions in the future.

The present approach will be evaluated considering a Gauss-Markov mobility mode [9] where the mobility nodes can be considered predictable in order to test our prediction algorithm. Our work pretend to show an increasing network performance in terms of end-to-end delay and energy consumption against an approach without using a mobility prediction method and other approaches existent. Additionally, we will compare our algorithm results against a mathematical model optimization which minimizes energy consumption considering delay and network resources constraints.



(a) Probability to go from state S_2 to S_4.



(b) Predicting the future state of n_l.

Figure 3: Defining Markov states - 2

3 Mathematical model formulation

In this section our problem is enunciated and described in detail from a mathematical optimization point of view, as well as some assumptions are shown in order to simplify our proposed mathematical model. Remember that, in addition to the prediction algorithm approach, we present a mathematical optimization model for finding optimal values with purpose of performing valuable comparisons against the others algorithms described in this work. Now, we present the details and assumptions of our mathematical model proposal.

Based on the Figure 4, we will describe our problem:

- Mobile Network: Assume we have a mobile network, at which the nodes position changes across time periods. For this reason, the links cost between the network nodes also changes across time periods. This means that at each time period the network has particular links cost, different from the links cost at other time period. Given that at each time period the network have different links cost, we could say this reflects the network state in a given time period. For this reason, each network at a given time period will be called *Network State*. For instance, the *Network State* at time period *1* is called *Network State 1*, the *Network State* at time period *2* is called *Network State 2*, and so on. In other words, according to the Figure 4.a we have an initial network (Network State 1) compound by 4 nodes. As these nodes conform a network, there are interrelations between them that we will call *Links*. These links have a cost, which can be represented, for example, by the distance, the signal to noise ratio or the RSSI measurement between the nodes. In the next time period, the network costs at the Network State 1 change and then these new interrelations between the nodes are now the *Network State 2*. As the next time period occurs, the network at the Network State 2 becomes the network at the Network State 3, and this network will be the network at the network State 4, and so on.

- Nodes: Each node is denoted as n_{it} where i is number of the node and t is the network state of the node. Depending on the communication range, a node can communicate with another node in the direction described by the Figure 4. For example, n_{11} can communicate with n_{21} and n_{31}.

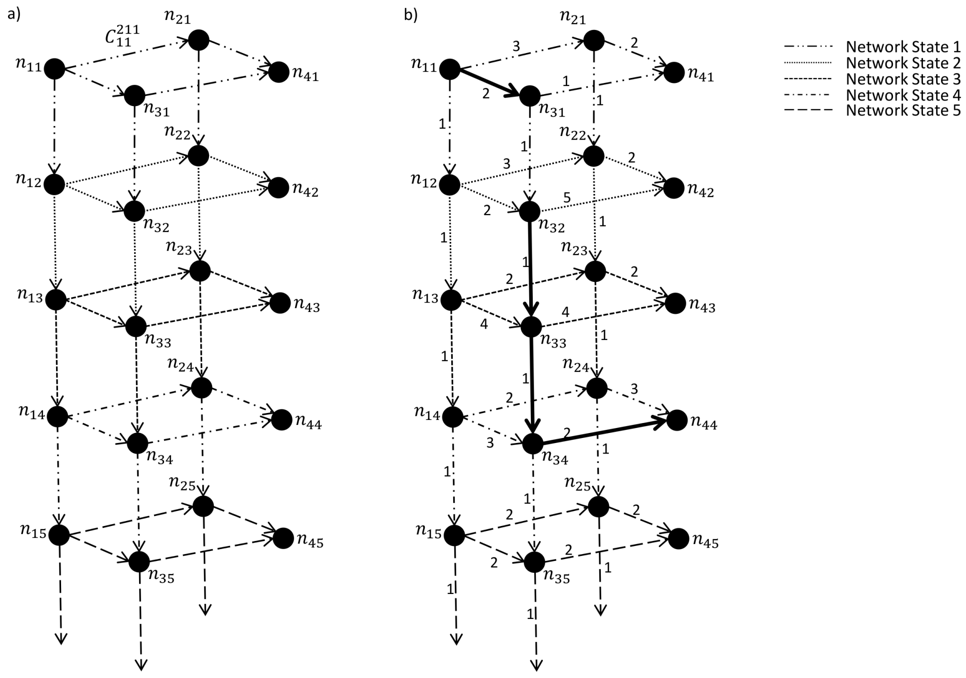


Figure 4: Problem scenario

- Buffer in each node: In telecommunication networks, a router or a sensor (a node) can decide not-sending its message, storing it in a buffer until it would be appropriated to send it. In our model, this situation is represented as a link between n_{11} and n_{12}, meaning that n_{11} can store its message in its buffer, that is, the node n_{12}.

- Costs: As it was mentioned previously, a link has a cost. Then, there is a cost for sending a message from n_{11} to n_{21} called C_{11}^{21l}, and denoted as C_{it}^{jul}, that is, this is the cost to carry a message form the node i at the state t to the node j at the state u at the Network State l.

- Directed graph: In this example our goal consists to carry a message from the node 1 to the node 4. Then, our *Source* node is the node 1, and our *Destination* node is the node 4. In this sense, a directed graph is constructed from the Source to the Destination. For this reason, the links direction points to the Destination.

- Goal: Our goal consists to carry a message from the Source node to the Destination node using the neighbours nodes as forwarding nodes for passing a message, and even using the buffers, if it is necessary, for waiting an appropriated situation for sending the message. In this sense, we have to find the minimum cost path between a Source node and a Destination node considering the network is changing across time, that is, through the Network States. Additionally, for simplicity we assume only one link can be selected for sending the message per each Network State. This means that if a message is at the node n_{11}, this node at this Network State 1 can send a message to only one neighbour, n_{21} or n_{11}, or storing it in its buffer, that is, n_{12}.

- Example Result: According to the example shown in the Figure 4.b and based on the links cost, the minimum cost path from the Source node, n_{11}, to the Destination node, node 4, is the path compounded by the highlighted links: n_{11} to n_{31}, n_{32} to n_{33}, n_{33} to n_{34} and n_{34} to n_{44}. As we will describe later in the mathematical formulation, this result can be also expressed in terms of X: $X_{11}^{311} = 1$, $X_{32}^{332} = 1$, $X_{33}^{343} = 1$ and $X_{34}^{444} = 1$. The rest of X_{it}^{jul} values are zero.

Following is presented our mathematical model proposed for constructing a minimal cost path from a source node to a destination node considering a mobile network. The sets, parameters and variables required for the mathematical model are described in the Table 1.

Table 1: Sets, parameters and variables description for the mathematical model

Sets	**Description**
N	Set of network nodes.
S	Set of network states.
Parameters	**Description**
C_{it}^{juld}	Link cost at the state l from the node i at the state t to the node j at the state u.
Variables	**Description**
X_{it}^{jul}	Determines if the link at the state l from the node i at the state t to the node j at the state u is selected for building the path towards the *Destination* (Binary variable).
$Y_{i,l}$	Determines if the node i at the state l is selected as a forwarding node for building the path towards the *Destination* (Binary variable).
D_{jl}	Determines if the node j is selected at the destination state l (Binary variable).
DS_l	Determines if the state l is selected as a destination state (Binary variable).

Next, our mathematical model is described.

$$min \sum_{itjul} C_{it}^{jul} X_{it}^{jul} \tag{8}$$

Subject to:

$$\sum_{l>1} D_{jl} = 1 \qquad \forall j \in N \tag{9}$$

$$\sum_{l} D_{jl} = 1 \qquad \forall j \in N \mid j \neq Destination \tag{10}$$

$$D_{jl} * DS_l = D_{jl} \qquad \forall j \in N; \quad \forall l \in S \tag{11}$$

$$\sum_{l} DS_l = 1 \tag{12}$$

$$DS_l = 0 \qquad \forall l \in S \mid l = 1 \tag{13}$$

$$DS_l \sum_{i \in N} \sum_{t \in S} \sum_{j \in N} \sum_{u \in S} X_{it}^{jul} Y_{im} D_{jl} = DS_l \qquad \forall l, m \in S \mid m = l - 1 \tag{14}$$

$$DS_l \sum_i Y_{im} = DS_l \qquad \forall l, m \in S \mid m \leqslant l \tag{15}$$

$$DS_l \sum_i Y_{im} = 0 \qquad \forall l, m \in S \mid m > l \tag{16}$$

$$\sum_{i \in N} \sum_{t \in S} \sum_{u \in S} \sum_{l \in S} X_{it}^{jul} = 1 \qquad \forall j \in N \mid j = Destination \tag{17}$$

$$DS_l \sum_{i \in N} \sum_{t \in S} \sum_{j \in N} \sum_{u \in S} X_{it}^{jum} Y_{in} Y_{jm} = DS_l \qquad \forall l, m, n \in S \mid m > 1 \wedge m = l \wedge n = m - 1 \tag{18}$$

$$DS_l \sum_{i \in N} \sum_{t \in S} \sum_{j \in N} \sum_{u \in S} X_{it}^{jum} = DS_l \qquad \forall l, m \in S \mid m \leqslant l \tag{19}$$

$$\sum_{i \in N} \sum_{t \in S} \sum_{j \in N} \sum_{u \in S} X_{it}^{jul} Y_{jl} = 1 \qquad \forall i \in N \mid i = Source \quad \forall l \in S \mid l = 1 \tag{20}$$

The equation 8 corresponds to the objective function, which will try to find the X_{it}^{jul} variables with the less possible cost C_{it}^{jul}. The previous expressions are explained in the following items:

- Destination State Constraints (from 9 to 17): The following expressions are referred to the Destination State, that is, the state at which the Destination node is found at the minimum possible cost.

  - Defining D_{jl}: D_{jl} allows to obtain the Destination State l at which the Destination node j is found at the minimum possible cost. The expression 9 avoids that D_{jl} will be one at the first state. The equation 10 avoids D_{jl} will be one for a node different from the destination node.

  - Defining DS_l: DS_l allows to extract only the Destination State l at which the Destination node is found at the minimum possible cost. The expression 11 allows to know the state l at which D_{jl} was selected. The equation 12 indicates that only one destination state is possible. In the expression 13 we assume it is not possible that the destination state will be the first state.

  - Selecting the forwarding node: The forwarding node indicates the node selected at each state for constructing the minimum cost path. The expressions 14 and 15 restricts to one the number of Y_{jl} for each State less than the Destination State. The equation 16 restricts to zero the number of Y_{jl} for each State higher than the Destination State. The expression 17 indicates that it is possible only one link to the Destination node for all states, that is, only one state is selected, and for the rest of the states, the link must be zero.

- Intermediate State Constraints: These constraints allow selecting the predecessor node Y_{im} based on the current forwarding node Y_{jl}. In order to understand what these two types of nodes means, let's see an example. If we have a link between the nodes 1 and 2 in the direction from 1 to 2, the current forwarding node is 2 and the predecessor node is 1. The expression 18 allows to select the predecessor node at the intermediate states, where intermediate states refers to the states between the Destination and the Source States. The equation 19 restricts to one the number of X_{it}^{jul} for each state equal or less than the Destination State.

- Source State Constraint: The Source State indicates the State at which the Source node starts to construct the minimum cost path. The expression 18 restricts to one the number of X_{it}^{jul} for the Source state.

- Defining the First State Solution Constraint: All the constraints described above allow to find the minimum cost path between a Source node and a Destination node through several Network States. However, up to now our model does not consider the Destination State can be the first network state. For this reason it is necessary to apply the following post-processing pseudocode:

Algorithm 1 Post-processing pseudocode

$parameters\ Source,\ Destination$
$minSolution = MathModel(Source,\ Destination)$
$costFirstState = C_{it}^{jul} \quad | \quad i = Source, j = Destination, t = u = l = 1$
if $costFirstState < minSolution$ **then**
    $minSolution = costFirstState$
end if

This pseudocode basically indicates that if the cost between the Source and the Destination node is less than the solution found by the mathematical model, then the solution is at the first state, otherwise the solution is given by the mathematical model.

4 Implementation

We have designed a Mobile Wireless Sensor Network Simulator in MATLAB, which has the following basic network components:

- *Destination node:* it is the final node that will receive a data message. In our simulations this node will always be the last network node.

- *Source node:* This node will have a data message, which must arrive to the destination node. In our simulations this node will always be the first network node.

- *Connected node:* If a message arrive to this node, this node knows the path to achieve the destination node.

In order to test the Prediction technique above, our simulator is compound of the following main processes:

- *Forwarding node selection:* When a node has a data message, this process consists to select properly a neighbour node as a forwarding node, which is selected according to the following priorities:

  – If among the neighbour nodes there is the destination node, then, the forwarding node is the destination node.
  – If among the neighbour nodes there is not the destination node, but there is a connected node, then, the forwarding node is the connected node.
  – If among the neighbour nodes there is not a destination node neither a connected node, then, the forwarding node is a node given by the Predicion method.

- *Sink refreshing:* This process consists to determine which nodes will be connected nodes at each certain period. This refreshing process is required due to network mobility, since it causes that connected nodes established in a previous state period, they will not possibly be connected nodes in the next period.
- *Loop detection:* It is important that a message can achieve the destination node, whereby it is necessary avoiding the message fall into a loop.
- *Prediction at each k-state:* At each network state the *Transition Matrix* (T) is calculated for all network nodes, except the destination node. Remember that this Transition Matrix stores the probability of each node to be at certain distance level respect to their neighbour nodes.
- *Prediction for selecting a forwarding node:* As we say before, if among the neighbour nodes there is not a destination node neither a connected node, then, the forwarding node is a node given by the Predicion method. This forwarding node is selected based on the information given by the Transition Matrix.

In order to test the Prediction algorithm performance, we have designed more algorithms with the aim of doing comparisons and obtain valuable information. Next, there is a description of each algorithm respect to its forwarding node process selection:

- *Distance Algorithm:* Considering there is not a destination or a connected node among the neighbour nodes of a current node, the forwarding node is the node with the shortest distance to the current node. The current node is the one that currently has a message that must arrive to the destination node.
- *Distance Away Algorithm:* Considering there is not a destination or a connected node among the neighbour nodes, the forwarding node is the node with the longest distance to the current node.
- *Prediction Algorithm:* Considering there is not a destination or a connected node among the neighbour nodes, the forwarding node is the node with the best probability to be near to the current node.
- *Prediction Away Algorithm:* Considering there is not a destination or a connected node among the neighbour nodes, the forwarding node is the node with the best probability to be far away to the current node.
- *Random Algorithm:* Considering there is not a destination or a connected node among the neighbour nodes, the forwarding node is a random node.

5 Results

In this section we will present the main results for the different algorithms showed in the previous sections. The metrics used for showing these results are: *Energy Consumption* and *Hops*. The *Energy Consumption* metric indicates the energy wasted by all the network nodes until the destination node is found. The *Hops* metric indicates the amount of hops needed to find the destination node. These two metrics are showed versus the number of network nodes. In addition, in order to obtain valuable statistical results, the performance evaluation of each algorithm and the mathematical model was made for 10000 tests for each network size. The next figures show the performance of the different algorithms and the mathematical model proposed for finding the minimum cost path in the network.

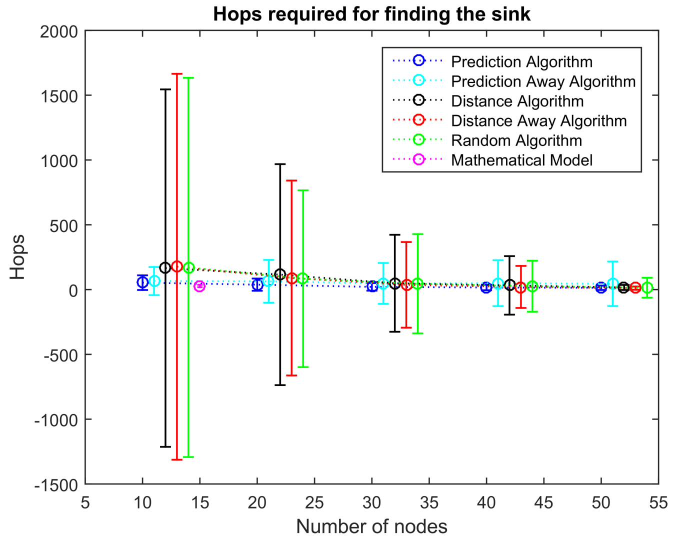


Figure 5: Hops performance along the network size

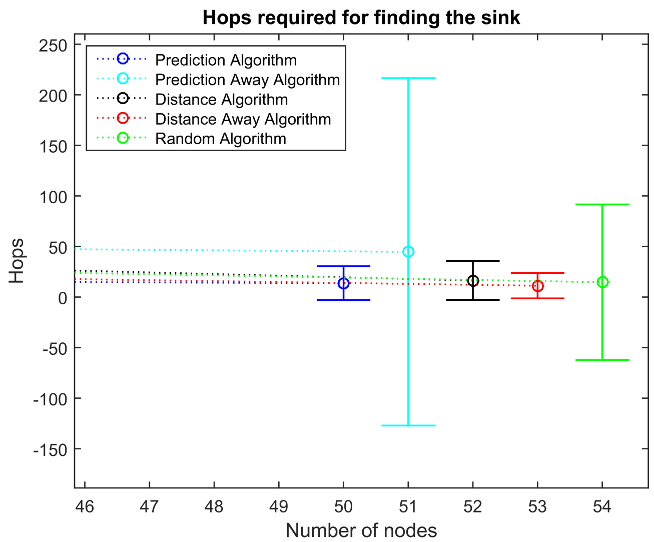


Figure 6: Hops for 50 nodes

The figure 5 shows the hops performance by the different algorithms along the network size. From this figure, as the size decreases the performance of Distance, Distance Away and Random algorithms gets worse because they require more hops to find the destination node. By contrast, the performance of Prediction and Prediction Away algorithms is better than the other ones because it requires less hops to find the destination node. This can be explained by the usage of prediction techniques, which offer more reliable paths. The following figures are focused in each network size.

The figure 6 shows the algorithms hops performance for 50 nodes. The number of hops for each algorithm is presented in the table 2. The best performance is obtained by the Distance algorithm, while the Prediction algorithm is second best because the network size is big (50 nodes), allowing more path alternatives for the Distance Away algorithm to find faster the destination node.

Table 2: Hops for 50 nodes

	Hops	**Ranking**
Prediction Algorithm	13.72	2
Prediction Away Algorithm	44.71	5
Distance Algorithm	16.30	4
Distance Away Algorithm	11.20	1
Random Algorithm	14.66	3

The figure 7 shows the algorithms hops performance for 40 nodes. The number of hops for each algorithm is presented in the table 3. The best performance is obtained by the Prediction algorithm because the number of nodes begins to decrease compared with the 50 nodes scenario, generating less path alternatives for the others algorithms and, then, thanks to the reliable feature given by the prediction technique, this algorithm can achieve faster the destination node.

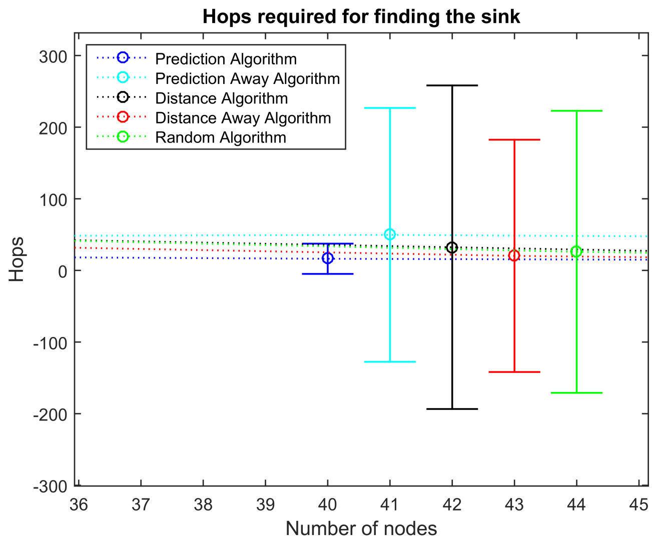


Figure 7: Hops for 40 nodes

Table 3: Hops for 40 nodes.

	Hops	**Ranking**
Prediction Algorithm	16.24	1
Prediction Away Algorithm	49.61	5
Distance Algorithm	32.38	4
Distance Away Algorithm	20.31	2
Random Algorithm	25.95	3

The figure 8 shows the algorithms hops performance for 30 nodes. The number of hops for each algorithm is presented in the table 4. The best performance is obtained again by the Prediction algorithm for the same reason as the previous figure. The less size of the network, the less path alternatives will have the rest of algorithms.

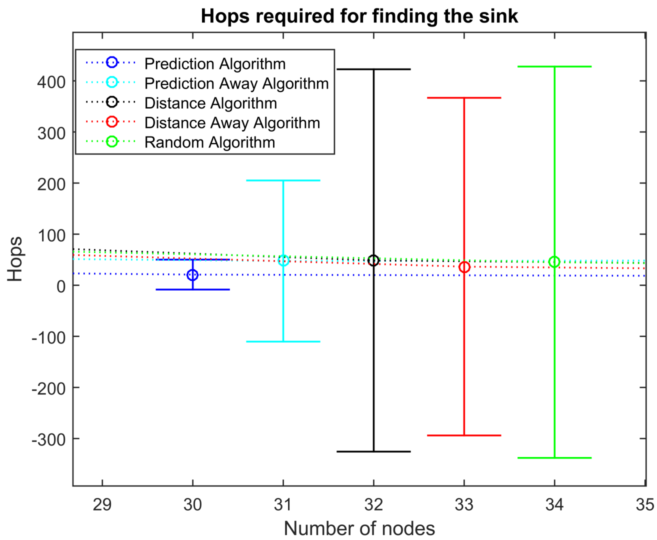


Figure 8: Hops for 30 nodes

Table 4: Hops for 30 nodes

	Hops	Ranking
Prediction Algorithm	20.86	1
Prediction Away Algorithm	47.38	4
Distance Algorithm	48.53	5
Distance Away Algorithm	36.42	2
Random Algorithm	45.05	3

The figure 9 shows the algorithms hops performance for 20 nodes. The number of hops for each algorithm is presented in the table 5.

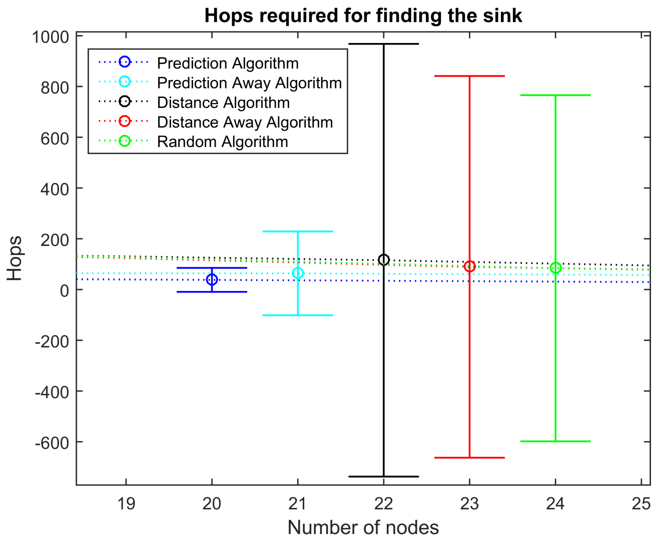


Figure 9: Hops for 20 nodes

Table 5: Hops for 20 nodes

	Hops	**Ranking**
Prediction Algorithm	37.99	1
Prediction Away Algorithm	63.75	2
Distance Algorithm	115.38	5
Distance Away Algorithm	89.30	4
Random Algorithm	83.85	3

The figure 10 shows the algorithms hops performance for 10 nodes. The number of hops for each algorithm is presented in the table 6. Here we can notice the large difference in terms of hops of using prediction techniques compared with not-using prediction techniques. This means that if our network has few nodes and, as a consequence, it is more difficult to find a path to the destination node, our prediction algorithm is capable of obtain a large advantage respect the others algorithms for finding the destination node. This advantage is represented in the hop different respect to the second algorithm in the ranking, which is 25.75 hops of difference. This comparison is among the Prediction and Prediction Away algorithms. However, if the comparison is done between the Prediction algorithm and the best algorithm that does not use prediction techniques (the Random Algorithm), the advantage of using the Prediction algorithm is even higher (45.85 hops). This indicates using prediction techniques are suitable when finding paths is a critical task, that is, when the network is compound of few nodes. Notice that in addition there are presented the results for the mathematical model, which obviously presents the best performance, showing a hops performance difference of 25.81 respect to the Prediction algorithm. Notice that the mathematical model tests where suitable in terms of time execution and memory usage for a maximum of 15 nodes. For instance, a test of 30 nodes or even 20 nodes, the mathematical model solution unfortunately never ended. For this reason, only solutions for 10 nodes is provided to be compared with the prediction routing algorithm and no-prediction routing algorithms.

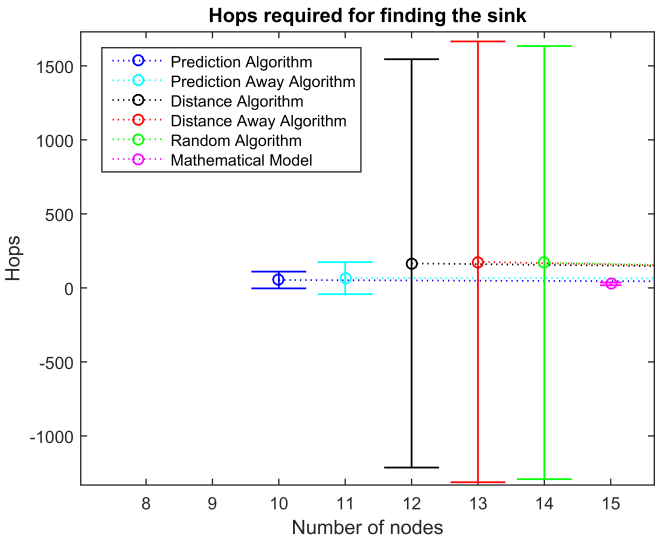


Figure 10: Hops for 10 nodes

Table 6: Hops for 10 nodes

	Hops	**Ranking**
Prediction Algorithm	53.38	1
Prediction Away Algorithm	65.75	2
Distance Algorithm	165.62	3
Distance Away Algorithm	176.07	5
Random Algorithm	170.92	4
Mathematical Model	27.57	

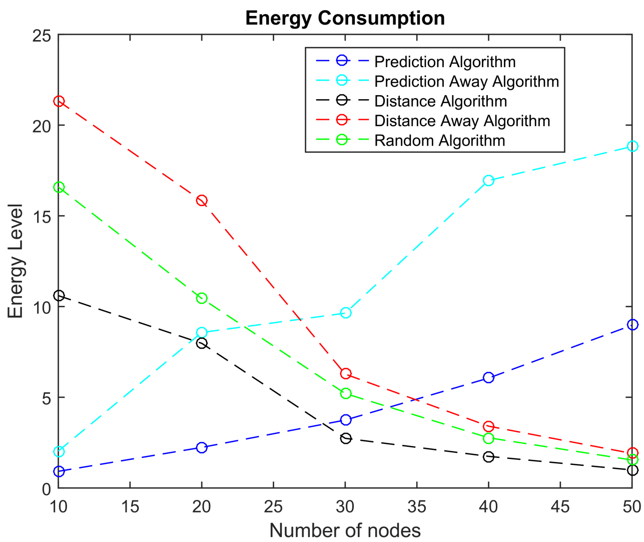


Figure 11: Energy consumption of the network

The figure 11 shows the algorithms energy consumption performance for different network size. This figure indicates that the Prediction Algorithm is suitable for low size network, showing that it starts to be efficient in terms of energy consumption from almost 30 nodes to 10 nodes.

6 Conclusions

We proposed the usage of a prediction technique in the context of a mobile wireless sensor network with the aim of the shortest path possible from a source node to a destination node. Employing this technique allowed to building the most reliable path for finding the destination node and at the same time it allowed to obtain the shortest path to the destination node. In other words, the reliability offered by the prediction technique allowed to select the most stable forwarding nodes in terms of their network connectivity. In this sense, through the prediction technique it was less likely that a data message would be in isolated network zones, and then, there was a higher probability for reaching the destination node by the data message. For this reason, when the number of network nodes was scarce, 10 or 20 nodes, the prediction algorithm performance was too high in comparison with the rest of the algorithms, obtaining 45.85 and 112.24 hops of difference with the second best no-prediction algorithm for the 10 and 20 nodes of network size. The impact of this finding is very interesting. Suppose a cattle application where the network nodes (20 nodes) changes each 100 miliseconds. This means that if we use the prediction algorithm, a data message will reach the destination node 11.22 seconds faster than the second best no-prediction algorithm. This time, 11.22 seconds, could be a significant advantage in delay sensitive applications where the timeliness is an imperative factor.

In terms of energy consumption, a prediction technique is suitable for scarce networks (10, 20 or 30 nodes) because the energy consumption was the less than the rest of algorithms. This energy performance besides to the hops performance make the prediction algorithm totally suitable for scarce networks, that is, MWSN applications where the number of nodes is not too high and it is required data messages arrive to the destination nodes as soon as possible.

Bibliography

[1] Ahmed A. A. (2013); An enhanced real-time routing protocol with load distribution for mobile wireless sensor networks, *Computer Networks,* 57, 2013.

[2] Ahmed A. A. (2007); *Real-Time Wireless Sensor Networks*, University of Virginia, 2007.

[3] Akyildiz, I. F.; Vuran, M. C. (2010); *Wireless Sensor Networks*, Vol. 4, John Wiley & Sons, Hoboken, 2010.

[4] Buchli, B.; Sutton, F.; Beutel, J. (2012); GPS-Equipped Wireless Sensor Network Node for High-Accuracy Positioning Applications, *Wireless Sensor Networks Lecture Notes in Computer Science,* Springer, 7158, 179-195, 2012.

[5] De Araujo, G. M.; J. Kaiser, J.; Becker, L. B. (2012); An Optimized Markov Model to Predict Link Quality in Mobile Wireless Sensor Networks, *Computers and Communications*, 307-312, 2012.

[6] De Araujo, G. M.; J. Kaiser, J.; Becker, L. B. (2014), Genetic Machine Learning Approach for Link Quality Prediction in Mobile Wireless Sensor Networks, *Cooperative Robots and Sensor Networks*, 1-14, 2014.

[7] Li, S.; Ma, X.; Wang, X.; Tan, M. (2011); Energy-efficient multipath routing in wireless sensor network considering wireless interference, *Journal of Control Theory and Applications,* 9(1), 127-132, 2011.

[8] Montoya, G. A.; Donoso, Y. (2018); A Prediction Algorithm based on Markov Chains for finding the Minimum Cost Path in a Mobile Wireless Sensor Network, *Proceedings of the 7th International Conference on Computers Communications and Control*, IEEE, 169 - 175, 2018.

[9] Torkestani, J. A. (2012); Mobility prediction in mobile wireless networks. *Journal of Network and Computer Applications,* 35(5), 1633-1645, 2012.

[10] Zheng, J.; Jamalipour, A. (2009); *Wireless Sensor Networks: A Networking Perspective*, Wiley, 2009.

[11] [Online] A Community Resource for Archiving Wireless Data At Dartmouth. http://crawdad.org/.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 14(1), 56-62, February 2019.

IoT Devices Signals Processing based on Multi-dimensional Shepard Local Approximation Operators in Riesz MV-algebras

D. Noje, R. Tarca, I. Dzitac, N. Pop

Dan Noje
1. University of Oradea
410087 Oradea, Universitatii, 1, Romania
dan.noje@gmail.com
2. Primus Technologies SRL
410248 Oradea, Onestilor, 3, Romania
dan.noje@primus-tech.ro

Radu Tarca*
University of Oradea
410087 Oradea, Universitatii, 1, Romania
*Corresponding author: rtarca@uoradea.ro

Ioan Dzitac
1. Aurel Vlaicu University of Arad
310330 Arad, Elena Dragoi, 2, Romania
ioan.dzitac@uav.ro
2. Agora University of Oradea
410526 Oradea, P-ta Tineretului 8, Romania,
idzitac@univagora.ro

Nicolae Pop
Institute of Solid Mechanics of the Romanian Academy
10141 Bucharest, Sector 1, Constantin Mille, 15, Romania
nicpop@gmail.com

Abstract: In this article we continue the study started in [8] to use Riesz MV-algebras for IoT devices signals processing. The Shepard local approximation operators introduced in [8] were defines such that to approximate single variable functions. In real industrial usage, the signals coming from IoT devices may be influenced by mode than a parameter, and thus we introduce generalized Shepard local approximation operators that can approximate multi-dimensional functions and some numerical experiments are considered.
Keywords: IoT devices; signal processing; Shepard local approximation operators; local approximation operators; approximation algorithms; Riesz MV-algebras, vectorial MV-algebras.

1 Introduction

With the aim to provide new mathematical tools that are useful to develop algorithms that are suitable for IoT devices signals processing, in [1] were introduced Shepard local approximation operators that can approximate one-dimension functions. Since in real industrial applications, the IoT devices signals are not depending on one single parameter, there is a real need to introduce generalized Shepard local approximation operators that can approximate multi-dimensional functions. This new approximation operators can be later used to develop software algorithms that act as input validators for industrial automated control systems [10, 11] based on the Riesz MV-algebra structure of the IoT devices signals [8].

In 1958, multivalued algebras, shortly named MV-algebras, were introduced by Chang [2, 3] as the algebraic structures corresponding to the ∞ - valued Lukasiewicz logic.

Definition 1. An MV-algebra is a structure $\mathcal{A} = (A, \oplus, \neg, 0_A)$ if and only if the following axioms are fulfilled:

$$(A, \oplus, \neg, 0_A) \text{ is an abelian monoid},$$

$$\neg\neg x = x,$$

$$x \oplus \neg 0_A = \neg 0_A,$$

$$\neg(\neg x \oplus y) \oplus y = \neg(\neg y \oplus x) \oplus x.$$

In an MV-algebra $\mathcal{A}$, the constant 1_A and the binary operations $\odot$ and $\ominus$ can be defined as follows:

$$1_A = \neg 0A,$$

$$x \odot y = \neg(\neg x \oplus \neg y),$$

$$x \ominus y = x \odot \neg y.$$

Also we can define a distance function $d : A \times A \to A$ as follows:

$$d(x, y) = (x \ominus y) \oplus (y \ominus x).$$

This distance as it is defined is a metric and plays a very important role in image and signal processing.

By introducing an additional external operation, in 2003, was defined the concept of Vectorial MV-algebras [6], shortly named VMV-algebras. Let consider an MV-algebra $\mathcal{A}$ and an external operation defined as follows:

$$\bullet : \mathbb{R}_+ \times A \to A.$$

Definition 2. The MV algebra $\mathcal{A}$ is an VMV-algebra if and only if the following axioms are fulfilled:

$$1 \bullet x = x, \forall x \in A,$$

$$(a + b) \bullet x = a \bullet x \oplus b \bullet x, \forall x \in A \text{ and } \forall a, b \in \mathbb{R}_+,$$

$$a \bullet (b \bullet x) \leq (a \cdot b) \bullet x, \forall x \in A \text{ and } \forall a, b \in \mathbb{R}_+,$$

$$d(a \bullet x, a \bullet y) \leq a \bullet d(x, y), \forall x, y \in A \text{ and } \forall a \in \mathbb{R}_+.$$

VMV-algebras inspired new algebraic structures, MV-modules [4] and Riesz MV-algebras [1].

Definition 3. An MV-algebra $\mathcal{A}$ is a truncated MV-module over the unital latticeal ring (R, v) if there is defined an external operation $\bullet : \mathbb{R}_+ \times A \to A$, such that the following properties are fulfilled for $\forall \alpha, \beta \in \mathbb{R}_+$ and $\forall x, y \in A$.

$$(\alpha + \beta) \bullet \ x = \alpha \bullet \ x \oplus \beta \bullet \ x,$$

$$\alpha \bullet (x \oplus y) = \alpha \bullet x \oplus \alpha \bullet y, \text{ if } x \leq \neg\, y,$$

$$\alpha \bullet (\beta \bullet x) = (\alpha \cdot \beta) \bullet \ x, \text{ if } \alpha, \beta \in [0, v].$$

If property

$$v \bullet x = x,$$

is also fulfilled, then $\mathcal{A}$ is an unital MV-module over unital ring $(\mathbb{R}, v)$.

Definition 4. If an MV-algebra is a truncated unital module over $(\mathbb{R}, 1)$, it is a Riesz MV-algebra.

Theorem 5. *In any Riesz MV-algebra, the following properties are fulfilled [4]:*

$$\alpha \bullet (\beta \bullet x) \leq (\alpha \cdot \beta) \bullet x,$$

$$d(\alpha \bullet x, \alpha \bullet y) \leq \alpha \bullet d(x, y),$$

$$0 \bullet x = 0,$$

$$\alpha \bullet 0 = 0,$$

$$x \leq y \Rightarrow \alpha \bullet x \leq \alpha \bullet y,$$

$$\alpha \leq \beta \Rightarrow \alpha \bullet x \leq \beta \bullet x,$$

$$\alpha \bullet (x \oplus y) \leq \alpha \bullet x \oplus \alpha \bullet y,$$

for any $x, y \in A$ and $\alpha, \beta \in \mathbb{R}_+$.

It was also proved that any Riesz MV-algebra is an VMV-algebra, but the reciprocal statement isn't true, but the set of values of IoT devices signals can be organized as Riesz MV-algebras [8].

In [1] was proved that Riesz MV-algebras are algebraic and topological structures for data processing, because:

Theorem 6. *Any method developed in the classical numerical analysis is applicable in Riesz MV-algebras if the Riesz MV-algebras operations are used.*

Based on this statement, in [8] were introduced Shepard local approximation operators on Riesz MV-algebras, to approximate one-dimension functions. Were considered a Riesz MV-algebra $\mathcal{A}$ and a function $f : [0, n] \to A$ and a Shepard kernel [9], which is a strictly decreasing function $K : [0, 1] \to \mathbb{R}_+$. Also, was considered the set

$$B(x, r) = \{y \in [0, n] \mid |x - y| \leqslant r\} \tag{1}$$

Definition 7. A Shepard local approximation operator, is a function $S : [0, n] \to A$ defined as follows:

$$S(f, x) = \oplus_{x_i \in B(x,r)} \frac{K\left(\frac{|x - x_i|}{r}\right)}{\sum_{x_i \in B(x,r)} K\left(\frac{|x - x_i|}{r}\right)} \bullet f(x_i)$$

where $\oplus$ and $\bullet$ are the Riesz MV-algebra operations.

2 Generalized Shepard local approximation operators for IoT devices signal processing

We us consider a Riesz MV-algbra $\mathcal{A}$ and a function $f : [0, n_1] \times [0, n_2] \times ... \times [0, n_m] \to A$. In this generalized situation, in the definition 1 of the set B the distance has to be replaced by a norm. The norms we consider in the numerical experiments are:

$$\|x\| = \sqrt{c_1^2 + c_2^2 + ... + c_m^2} \qquad \text{(euclidean norm)}$$

$$\|x\| = \max\left(|c_1|, |c_2|, ..., |c_m|\right) \qquad \text{(supremum norm)}$$

$$\|x\| = |c_1| + |c_2| + ... + |c_m| \qquad (l^1 \text{ norm})$$

where $x = (c_1, c_2, ..., c_m) \in [0, n_1] \times [0, n_2] \times ... \times [0, n_m]$.

Considering this, we can define the generalized Shepard local approximation operators as follows:

Definition 8. A generalized Shepard local approximation operator, is a function $Sg : [0, n_1] \times [0, n_2] \times ... \times [0, n_m] \to A$ defined as follows:

$$Sg\left(f, x\right) = \oplus_{x_i \in Bg(x,r)} \frac{K\left(\frac{\|x - x_i\|}{r}\right)}{\sum_{x_i \in Bg(x,r)} K\left(\frac{\|x - x_i\|}{r}\right)} \bullet f\left(x_i\right)$$

where $\oplus$ and $\bullet$ are the Riesz MV-algebra operations and

$$Bg\left(x, r\right) = \{y \in [0, n_1] \times [0, n_2] \times ... \times [0, n_m] \mid \|x - y\| \leqslant r\}. \qquad (2)$$

In [1] was considered that the set of possible values for IoT devices signals is the interval $[0, 2^t - 1]$, where t is the number of bits used to store these values. In [10] was proved that the structure $([0, 2^t - 1], \oplus, \neg, 0)$ is a MV-algebra, if the following definitions are used:

$$x \oplus y =_{def} \min(2^t - 1, x + y),$$

$$\neg\, x =_{def} 2^t - 1 - x,$$

$\forall x, y \in [0, 2^t - 1]$.

In [11] was proved that if we consider the external operation $\bullet : \mathbb{R}_+ \times [0, 2^t - 1] \to [0, 2^t - 1]$, defined as follows:

$$a \bullet x =_{def} \min(2^t - 1, a \bullet x),$$

$\forall a \in \mathbb{R}_+$ and $\forall x \in [0, 2^t - 1]$, the structure $([0, 2^t - 1], \oplus, \neg, 0, \bullet)$ is a vectorial MV-algabra and is easy to see that this is also a Riesz MV-algebra, as mentioned in [8].

If we use the above definition of $\oplus$ and $\bullet$ operations and the formula of the general Shepard local approximation operator from Definition 8, we can define an algorithm that can be used to fill in the missing data of signals received from IoT devices.

In [1] were considered the following types of kernels:

$$K(u) = \frac{1}{u^{2\lambda}}, \qquad \text{(Shepard kernel)}$$

$$K(u) = e^{-\lambda u^2}, \qquad \text{(Exponential kernel)}$$

$$K(u) = \left(\frac{\sin\left(q\pi u\right)}{\sin\left(\pi u\right)}\right)^{2\lambda}, \qquad \text{(Shepard-Jackson kernel)}$$

where λ is a parameter that can influence the performance of obtained results, and q is the degree of the Shepard-Jackson kernel. Also in [1] was determined by numerical experiments that the best results are given by Shepard and exponential kernels, thus only these will be considered in the following numerical experiments.

Let now consider an industrial rectangle shape grid of IoT sensors that collect temperature of an environment. The temperatures collected at a certain moment can be represented by a two-dimension function

$$f : [0, n_1] \times [0, n_2] \to [0, 2^t - 1].$$

In the formula of Definition 8, $x_i = (c_{i_r}, c_{i_c})$ are the sensors located on row c_{i_r} and column c_{i_c} have transmitted data and $x = (c_r, c_c)$ is the sensor located in row c_r and column c_c that was not transmitted data.

The fill in algorithm has the following steps:

1. A kernel has to be selected;
2. The parameter λ is set;
3. The radius r, that influence how many received values are considered in the approximation of missing values, is set;
4. A grid traversing method is selected and each missing value is approximated.

3 Numerical results

In the approximation process, there is also the possibility to use the previously new approximated values or to ignore them. Also to reduce the computational complexity, we replace the ball defined in 2 with a square having the side length $2r + 1$ and we used the supremum norm.

A grid of 31×31 sensors was considered, and we assumed that the collected values should be $f(x) = \frac{(c_r+c_c)}{2} \cdot \sin(\frac{\pi(c_r+c_c)}{40})$. It was considered a continuous function, because one of the purposes of this algorithm is to be used to approximate missing values collected by the new fiber optic Bragg grating sensors system designed to monitor the ethanol fermentation during the bioethanol and wine production. This new fiber optic Bragg grating sensor system was developed using financing through grant PN-III-P2-2.1-PED-2016-1955.

The approximation error was determined using the formula

$$P = \sum_{x \in [0,n_1] \times [0,n_2]} \left| f(x) - Sg(f, x) \right|,$$

because we are interested in the overall error cumulated error.

3.1 Experiment 1

In this experiment we assume that we received only data transmitted by sensors that are located on rows and columns that are both even numbers. Several parameterizations are considered. After running the tests, we get the approximation errors listed as follows:

Parametrization	*Shepard kernel*	*Exponential kernel*
$r = 2, \lambda = 2$	48.1403	74.1296
$r = 2, \lambda = 10$	36.3588	36.3465
$r = 2, \lambda = 20$	36.3587	36.3587
$r = 3, \lambda = 2$	58.0658	183.302
$r = 3, \lambda = 10$	36.3588	42.9482
$r = 3, \lambda = 20$	36.3587	36.5376
$r = 5, \lambda = 2$	68.9431	476.092
$r = 5, \lambda = 10$	36.3588	110.7
$r = 5, \lambda = 20$	36.3587	54.642
$r = 10, \lambda = 2$	80.0381	1431.53
$r = 10, \lambda = 10$	36.3588	463.169
$r = 10, \lambda = 20$	36.3587	237.138

3.2 Experiment 2

In this experiment we assume that we couldn't receive data transmitted by sensors that are located on rows and columns that are both odd numbers. Same parameterizations like in the previous experiment are considered. After running the tests, we get the approximation errors listed as follows:

Parametrization	*Shepard kernel*	*Exponential kernel*
$r = 2, \lambda = 2$	15.1187	20.5379
$r = 2, \lambda = 10$	13.1902	13.2046
$r = 2, \lambda = 20$	13.1902	13.1902
$r = 3, \lambda = 2$	18.6535	58.1864
$r = 3, \lambda = 10$	13.1902	14.2934
$r = 3, \lambda = 20$	13.1902	13.2233
$r = 5, \lambda = 2$	21.7479	152.12
$r = 5, \lambda = 10$	13.1902	33.5822
$r = 5, \lambda = 20$	13.1902	16.5552
$r = 10, \lambda = 2$	24.8594	455.862
$r = 10, \lambda = 10$	13.1902	146.969
$r = 10, \lambda = 20$	13.1902	74.6371

4 Conclusion

Nowadays, industrial information systems are depending on signals received form IoT devices. There can be several problems in acquiring data from these IoT devices, problems that can led to missing values. Without a complete set of data, the automation of processes isn't possible or is not satisfying enough. The algorithm proposed in this paper has the role to fill in the missing values of signals sent by IoT devices. As mentioned in [8], for industrial usage of the algorithm, this methods should be further developed to determine the proper set of parameters for each of the kernels, based on the particularities of the industrial processes handled and on the amount of missing values. Depending on the constrains of the real processes that have to be modeled, other error measures can be considered as well.

Funding

This research was funded by Unitatea Executiva pentru Finantarea Invatamantului Superior, a Cercetarii, Dezvoltarii si Inovarii (UEFISCDI) Romania, grant number 198PED/2017, grant code: PN-III-P2-2.1-PED-2016-1955.

Author contributions

The authors contributed equally to this work.

Conflict of interest

The authors declare no conflict of interest.

Bibliography

[1] Bede, B.; Di Nola, A. (2004), Elementary calculus in Riesz MV-algebras, *International Journal of Approximate Reasoning*, 36, 129-149, 2004.

[2] Chang, C.C. (1958), Algebraic analysis of many valued logics, *Trans. Amer. Math. Soc.*, 88, 467-490, 1958.

[3] Chang, C.C. (1959), A new proof of the completeness of the Lukasiewicz axioms, *Trans. Amer. Math. Soc.*, 93, 74-80, 1959.

[4] Di Nola, A.; Flondor, P.; Leustean, I. (2003), MV-modules, *Journal of Algebra*, 2003, 261, 21-40, 2003.

[5] Noje, D. (2002), Using Bernstein Polynomials for image zooming, *Proceedings of the Symposium Zilele Academice Clujene, Computer Science Section*, 99-102, 2002.

[6] Noje, D.; Bede, B. (2003), Vectorial MV-algebras, *Soft Computing*, 7(4), 258-262, 2003.

[7] Noje, D.; Bede, B. (2001), The MV-algebra structure of RGB model, *Studia Universitatis Babes-Bolyai, Informatica*, XLVI, 1, 77-86, 2001.

[8] Noje, D.; Dzitac I.; Pop N.; Tarca, R. (2019), IoT devices signals processing based on Shepard local approximation operators defined in Riesz MV-algebras, *Informatica*, submitted for publication 2018.

[9] Shepard, D. D. (1968), A two dimensional interpolation function for irregularly spaced data, *Proceedings of 23rd Nat. Conf. ACM*, 517—524, 1968.

[10] [Online]. *Sisteme informatice*, Universitatea Stefan cel Mare Suceava. Available online: http://www.seap.usv.ro/~sorinv/PSI.pdf, Accessed on 5 October 2018.

[11] [Online]. *Sisteme Informatice Industriale*, Universitatea Politehnică din Bucuresti. Available online: http://shiva.pub.ro/?page_id=345, Accessed on 2 October 2018.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 14(1), 63-77, February 2019.

Using an Adaptive Network-based Fuzzy Inference System to Estimate the Vertical Force in Single Point Incremental Forming

S.G. Racz, R.E. Breaz, O. Bologa, M. Tera, V.S. Oleksik

Sever-Gabriel Racz, Radu-Eugen Breaz*, Octavian Bologa,
Melania Tera, Valentin-Stefan Oleksik
Department of Industrial Machines and Equipment
Lucian Blaga University of Sibiu, Romania, 550024 Sibiu, Victoriei, 10, Romania
gabriel.racz@ulbsibiu.ro
*Corresponding author: radu.breaz@ulbsibiu.ro
octavian.bologa@ulbsibiu.ro,
melania.tera@ulbsibiu.ro, valentin.oleksik@ulbsibiu.ro

Abstract: Manufacturing processes are usually complex ones, involving a significant number of parameters. Unconventional manufacturing processes, such as incremental forming is even more complex, and the establishment of some analytical relationships between parameters is difficult, largely due to the nonlinearities in the process. To overcome this drawback, artificial intelligence techniques were used to build empirical models from experimental data sets acquired from the manufacturing processes. The approach proposed in this work used an adaptive network-based fuzzy inference system to extract the value of technological force on Z-axis, which appears during incremental forming, considering a set of technological parameters (diameter of the tool, feed and incremental step) as inputs. Sets of experimental data were generated and processed by means of the proposed system, to make use of the learning ability of it to extract the empirical values of the technological force from rough data.
Keywords: adaptive network-based fuzzy inference system, CNC milling machines, incremental forming, technological force.

1 Introduction

Manufacturing processes are usually involving a significant number of parameters. Complex analytical models were developed to express the dependencies between these parameters, for every manufacturing processes, but the complexity of the factors involved in these processes often affect the utility and accuracy of such models.

To compensate this drawback, methods involving artificial intelligence techniques, such as fuzzy logic, artificial neural networks, adaptive network-based fuzzy inference systems and adaptive neuro fuzzy inference systems were used to extract significant information, such as empirical relationships between parameters, form large amount of manufacturing processes rough data, where analytical dependencies between inputs and outputs were difficult to be derived. These methods make use of the learning ability of the artificial intelligence systems to overcome the complexity of the phenomenon involved in machining processes.

A synthesis regarding the main ideas and activities or the renowned scientist Lofti Zadeh, the father of fuzzy logic was presented in [8]. Among others, fuzzy data processing was pointed as a very useful application of fuzzy logic. A comprehensive presentation regarding the fuzzification of classical structures was introduced in the survey work [9], which can be used as starting theoretical support for every research using fuzzy logic methods and tools. Fuzzy linguistic modeling was the subject of the review presented in [12], which stated the fact that fuzzy linguistic modeling has proven as a useful tool for information accessing systems.

As presented above, fuzzy logic has numerous useful applications, industrial control being one of the most important ones. A survey paper synthesizing the industrial applications of fuzzy control was introduced in [21]. Another work regarding modeling and simulation of a combined PID-fuzzy control structure for a telerobot was introduced in [11].

The work presented in [25] used fuzzy logic to predict the surface roughness in milling processes. A fuzzy inference system using cutting regime parameters (such as cutting speed, and feed per tooth, cutting forces values), parameters linked with the tool geometry (tool nose radius) and parameters related to the cutting conditions (the use of cutting fluids) as inputs was designed. The system was able to calculate the surface roughness as an output, in an empirical way. Experimental work unfolded during this research had shown that the data predicted by the fuzzy system were close to the experimental ones.

The development of a system based upon fuzzy adaptive networks to overcome the complexity introduced by high-speed machining in turning operations was introduced in [15]. A model based upon fuzzy adaptive networks was built and used to predict surface roughness in turning. Experimental data was used to validate the proposed model.

The researches presented in [16] proposed the use of adaptive neuro-fuzzy inference systems to predict the surface quality obtained by means of end milling. Four parameters (feed, speed, vibration and depth of cut) were used as inputs, while the considered output parameter was roughness. The experimental data set was used for both training and validation for the proposed technique. The approach was found to be superior to the use of artificial neural networks from the point of view of prediction accuracy. The authors considered adaptive neuro-fuzzy inference systems as hybrid intelligent techniques, while the use of artificial neural networks was considered a simple intelligent one.

In [4] the research team developed an adaptive neuro-fuzzy inference systems model to predict the white layer thickness and surface roughness for an electro discharge machining process. Some process parameters such as duration of the pulse, feed rate of the wire, open circuit voltage and dielectric pressure were used as inputs for the adaptive neuro fuzzy inference systems system. Experimental data was used for validating the approach, which was found to provide feasible results.

The work presented in [25] compared the use of adaptive neuro fuzzy inference systems and backpropagation artificial neural networks to predict the values of some specific process parameters (roughness, tool wear and metal removal rate) for an electro discharge machining process. The above-mentioned parameters were used as outputs, while capacitance, feed rate, gap voltage and threshold were considered as inputs. The results had also shown that adaptive neuro-fuzzy inference systems outperformed artificial neural networks for the prediction task.

The fusion of a support vector machine with adaptive neuro-fuzzy inference systems for fault detection of an industrial system was introduced in [22]. A simulation study had proven that the fusion-based approach performed better than individual support vector machine and adaptive neuro fuzzy inference systems.

Artificial intelligence techniques, mainly their ability to cope with experimental sets of data and extract and/or predict crisp empirical values for significant output parameters were also used in [19], were an air pollution forecasting system based upon artificial neural networks was designed and tested. Robot control is another filed of application of such computational intelligence methods. The work from [6] used reduced training examples for artificial neural networks to control a robot with six degrees of freedom.

2 Single point incremental forming - technological forces

Single point incremental forming (SPIF) is an unconventional forming process which allows the user to manufacture prototypes or small batches of sheet metal parts by combining three or more technological movements in a flexible manner [14] [18] [10] [2].

A brief description of the SPIF process principle is depicted in (Fig. 1). The blank (2) is fixed by means of the blank holder (3). In order to realize the shape of the sheet metal part, one of the active elements, usually the punch (1) has an axial feed vertical movement direction, continuous or incremental (with the step s), while the other element, the active plate (4) carries out a plane horizontal movement.

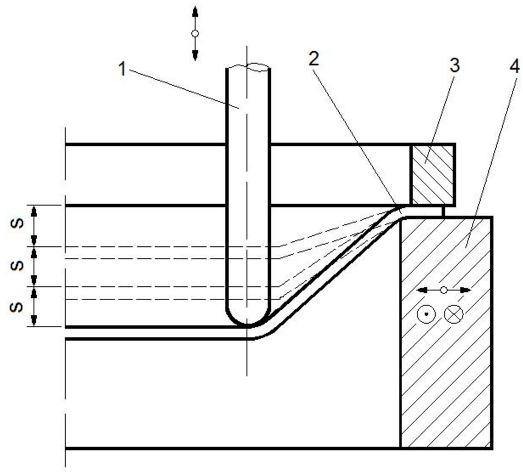


Figure 1: Single point incremental forming - principle

The SPIF process still has many drawbacks, which hinder its industrial implementation. One of them is related with the lack of specific technological equipment. Prototype machines are used in research facilities for unfolding the process. A literature survey has reported also CNC milling machines and serial industrial robots [5], [23] as technological equipment used for the process, with a special emphasis, by the usage point of view, on CNC milling machines. Consequently, it may be stated that the process is unfolded most frequently on CNC milling machines, which are sensitive and expensive equipment.

Another drawback of the process is related to the difficulty of estimating the technological forces, which became very important when CNC milling machines are involved. Forces exceeding the allowable values on each kinematic chain could appear and damage the CNC machine. Some modern CNC working centers are equipped with electronic safety systems, which can stop the movements on all axes when the resistant forces and torques are surpassing the allowable values, but this is not a general situation. There are CNC machines which are not fitted with such systems, or these systems could be too slow to stop the process before damages occur. The user must take into consideration both scenarios (best and worst case) and, of course try to avoid the risk of damaging the machine. It is here noticeable that even if the safety systems intervene and stops the process, productivity will be affected (Fig. 2).

A lot of researches were conducted recently to develop some methods of calculating and assessing the technological forces within the SPIF process, not only for avoiding machine damage, but also because accurate information about the values of the technological forces are also important to predict the plastic behavior of the sheet metal part.

In [7] the influence of the tool diameter, vertical depth increment and steepness of the part's wall or wall angle is studied, by manufacturing a simple geometry part, a cone. The result shows

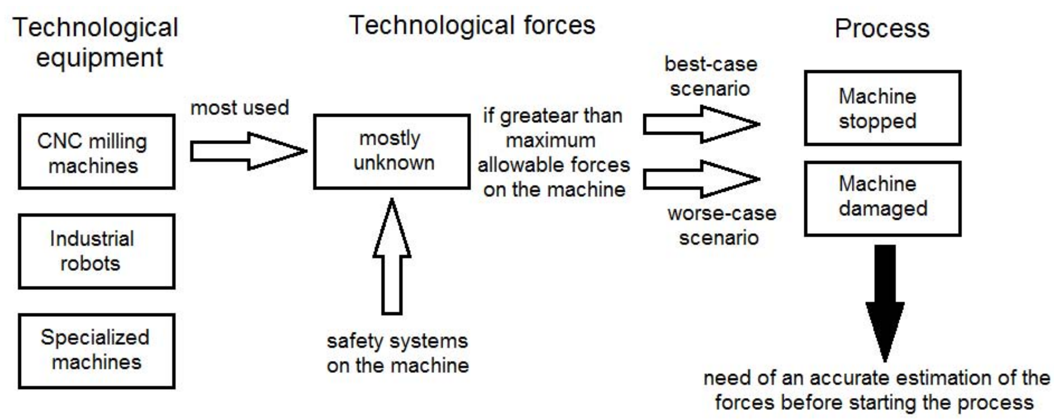


Figure 2: Flowchart of the influence of unknown technological forces upon the CNC milling machines

that if the vertical step size, tool diameter or wall angle are increased the forces will increase as well.

The research presented in [1] was focused on establishing practical formulae enabling the calculation of forces occurring during the single point incremental forming process, based on experimental results as well as analytical relations derived from finite element analysis (FEA) results. The experiments were unfolded for manufacturing a conical part. The formula for the vertical force takes into consideration the characteristics of the material (tensile strength), the material thickness, the diameter of the tool, the vertical step (expressed by the influence of it on the scallop height) and the wall angle of the conical part.

Another approach, presented in [17], was focused on experimental measurement on forming forces in the conical frustum-forming process to analyze the trend of forming force. Finally, analytical models for tangential force prediction were proposed and validated through experimental data. An empirical combined model has been constructed which could provide the prediction of tangential force with an average error less than 11 % for the experimental conditions employed in the experimental research.

The work presented in [20] compared the formulae developed in [1] with experimental results. It was verified that the proposed analytical formula has been adequate for constant wall angle slopes but limited to predict the forming force of variable wall angle geometries.

In [26], a discussion regarding the process parameters upon the technological forces is presented, based upon a synthesis of the results presented in the literature. According to that, the main parameters which influence the behavior of the forces are:

- Tool diameter (proportional influence, an increase in the tool diameter leads to an increase in the force value);
- Vertical step size (proportional influence);
- Sheet thickness (proportional influence);
- Wall angle slope (proportional influence).

It is noticeable the fact that the influence of these parameters (mainly the wall angle slope was synthesized and reported for different shapes of the parts (cone, conical frustum and pyramid) and for different materials (steel and aluminum alloys).

Most of the researches reported that the experimental measurements of the technological forces employed specialized devices, such as 3D dynamometers or custom designed data acquisition systems, which are very expensive and nod widely available. The latest mentioned systems also take a long time for setting-up and calibration.

Some previous work of the authors of this paper was presented in [24]. A reduced experimental data set was used in that approach and only two parameters were considered as input for the adaptive network-based fuzzy inference system (feed in the horizontal plane and depth). It was considered that only two parameters are not enough to describe the evolution of the vertical technological force. The approach presented here significantly developed the preliminary adaptive network-based fuzzy inference system presented in [24], by using more parameters within the SPIF process as inputs and significantly extending the amount of data processed by the system (268 data sets instead of 90 data sets).

In this paper, the authors set forth a method for measuring the technological forces which can be used even at workshops level and involves no supplementary costs, out of the initial costs of the CNC milling machine. Also, using a neuro-fuzzy approach, the experimentally gathered database was organized as a technological knowledge base, which allow the user to assess the value of the vertical technological force outside the process (table 4).

Table 1: Process parameters

Process parameters considered in the literature	**Process parameters considered in the proposed approach**	**Observations**
Tool diameter	Tool diameter	
Vertical step size	Vertical step size	
Sheet thickness	-	The approach was to develop knowledge databases for every sheet thickness of a given material, so sheet thickness was not considered as an input variable
Wall angle slope	-	The influence of the wall angle slope could not be predicted by analytical approaches [20]. Moreover, industrial implementation of the process involves manufacturing parts with highly variable wall angle slope
-	Feed	
-	Spindle speed	Considered only during initial approach

The proposed approach indicates as influence factors for the value of the vertical technological force, in the initial phase, the tool diameter, the vertical step size, the feed and the spindle speed. In order to preserve the degree of generality, no particular geometry of the processed part was taken into consideration (cone, conical frustum or pyramids), so no wall angles were considered as influence factors in this stage of the researches.

3 Method for calculating the vertical technological force

The technological movements required by the SPIF process are performed on CNC milling machine tools by means of the feed drives. The typical structure of a linear feed drive, either horizontal or vertical consists of a rotary servomotor axially coupled with a ball screw system. The schematic diagrams of a horizontal and a vertical feed drives are presented in (Fig. 3 a and b).

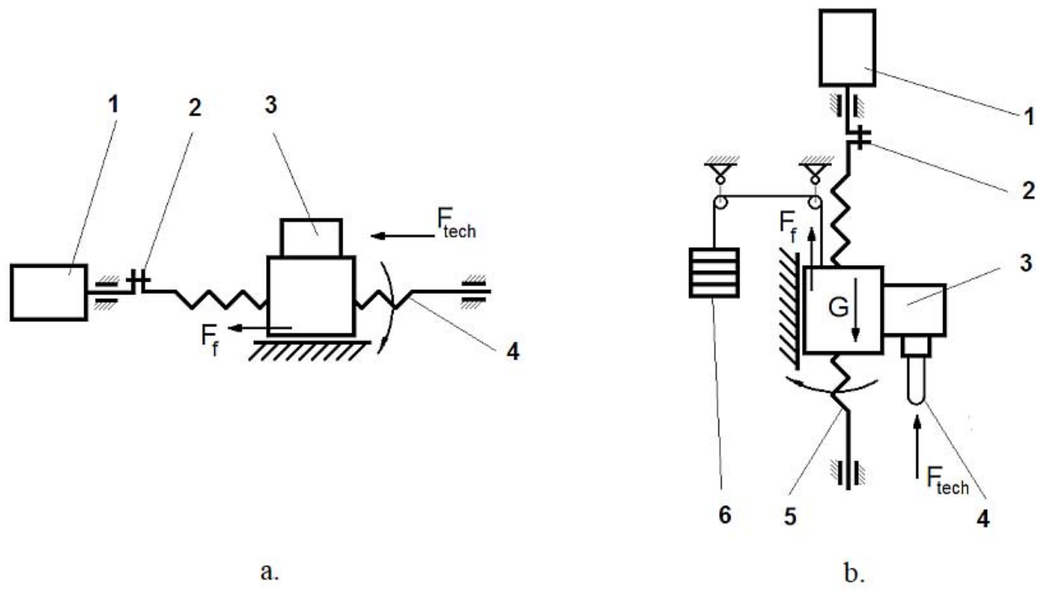


Figure 3: Schematic loads diagram for a horizontal (a) and a vertical (b) feed drive

The notations from (Fig. 3 a) have the following significations: 1 - drive motor, 2 - axial coupling, 3 - horizontal slide and part, 4 - lead ball-screw.

The notations from (Fig. 3 b) have the following significations: 1 - drive motor, 2 - axial coupling, 3 - vertical slide and working unit (with the main spindle), 4 - tool, 5 - lead ball-screw, 6 - counterweight for balancing the weight of the working unit.

The drive motor of a CNC feed drive must overcome both the static and the dynamic torques which appear during the machining process:

$$T_{mot} = T_{st} + T_{din} \tag{1}$$

where:
T_{mot} - motor torque;
T_{st} - resistant static torque;
T_d - resistant dynamic torque.
The total static torque for a horizontal feed drive may be calculated with the following relation:

$$T_{st} = \frac{(F_{tech} + F_f)p}{2\pi\eta_{tot}} \tag{2}$$

where:
F_{tech} - resistant technological force [N];
F_f - friction force;
p - ball-screw step [mm];
i - transfer ratio between motor and ball screw;
η_{tot} - total mechanical efficiency.

For the considered CNC milling machine-tool the value of step is p = 4.233 mm When vertical feed drives are considered, the friction force opposes the weight force of the machine slide (G), consequently, relation (2) must be rewritten as:

$$T_{st} = \frac{(F_{tech} - G + F_f)p}{2\pi\eta_{tot}} \tag{3}$$

where only the down movement has been taken into consideration (the technological force appears only when the tool is driven down, towards the workpiece).

A 3-axis (X, Y, Z) Haas CNC Mini Mill was used as technological equipment. On each axis, the actuating motor is directly coupled with the ball screw, so the transfer ratio i=1. According to this, the resistant static torque on Z-axis (vertical direction) may be expressed as:

$$T_{st} = \frac{(F_{techZ} - G + F_{fZ})p}{2\pi\eta_{tot}} \tag{4}$$

The CNC equipment displays at the level of operator panel the current value of the motor torque, for each axis, as a percent value. The percent value is linked with a certain crisp value of the motor torque, which is different from both nominal and maximal values. If the readings are done during the constant velocity phase, the dynamic torque (4) may be considered zero and consequently neglected.

To calculate the amount of the motor torque which is considered as 100%, the values of the static torque in the no-load conditions were subtracted from the values in load conditions, according to the following relations:

$$\Delta T_{st} = \frac{(F_{techZ} - G + F_{fZ})p}{2\pi\eta_{tot}} - \frac{(-G + F_{fZ})p}{2\pi\eta_{tot}} = \frac{F_{techZ} \cdot p}{2\pi\eta_{tot}} \tag{5}$$

where:
ΔT_{stZ} - difference between static torque in load and no-load conditions on Z-axis.

Expressing the left-hand side of equation (5) based on the maximum feed force allowed on Z-axis (8896 N, taken from the machine documentation) leads to:

$$\Delta T_{max\_stZ} = \frac{F_{techZ} \cdot p}{2\pi\eta_{tot}} = \frac{8896 \cdot 4.233 \cdot 10^{-3}}{2 \cdot 3.14 \cdot 0.9} = 6.65[Nm] \tag{6}$$

where:
ΔT_{max_stZ} - the maximum difference between static torque in load and no-load conditions on Z-axis.

The efficiency factor for the screw nut assembly was considered $\eta = 0.9$. It can be observed that the values of ΔT_{max_stZ} is in-between the values of the nominal torque ($T_n = 5.39$ Nm) and maximum torque ($T_{max} = 13.8$ Nm) of motors.

Considering Δt_Z the difference between the load and no-load static torques read from the operator panel of the machine (in percent), the following relation may be written:

$$\Delta T_{stZ} = \frac{\Delta T_{max\_stZ} \cdot \Delta t_Z}{100} \tag{7}$$

Equation (7) allows now the user to calculate of the vertical technological force, by using the difference between the values of the torques, displayed by the CNC equipment. The percent displayed values of torques can be replaced by crisp numerical values, knowing the 100% reported motor torque calculated by means of equations (6) and (7). The technological forces can now be calculated using the relation:

$$F_{techZ} = \frac{2\pi\eta_{tot}}{p}\Delta T_{stZ} = \frac{2\pi\eta_{tot}}{p} \cdot \frac{\Delta T_{max\_stZ} \cdot \Delta t_Z}{100} \tag{8}$$

For a given movement cycle, it is only necessary to program the same cycle in no-load conditions, prior to unfolding it in load conditions, to be able to read and subtract the no-load and load-conditions torques.

To validate the proposed calculation method, the technological force on Z-axis was measured using a data acquisition system. The differences between measured and calculated values were found to be in the range of 1% resolution of the CNC operator panel display, leading to a maximum error of 88.6 N. Considering that for incremental forming of steel sheets the vertical technological force could reach values greater than 2000 N, this error was considered acceptable.

4 Generating the experimental database

An experimental program was unfolded to gather the data which will be processed by the adaptive network-based fuzzy inference system. The DC04 steel, with a thickness of 0.7 millimeters was used as workpieces material. Simple toolpaths were programmed (Fig. 4), and the punch was moved along them. On each component of the toolpath (segments from 1 to 10, each one with an overall length of 10 millimeters) the manufacturing inputs were modified. The inputs taken into consideration were tool diameter, spindle speed, feed on horizontal plane, incremental vertical step.

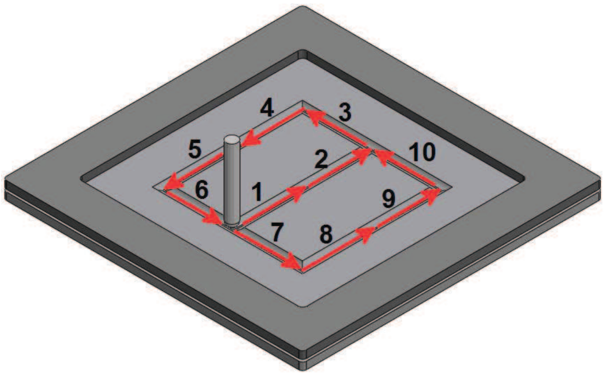


Figure 4: Toolpaths used for generating the experimental database

Some examples of the processed parts are presented in (Fig. 5).

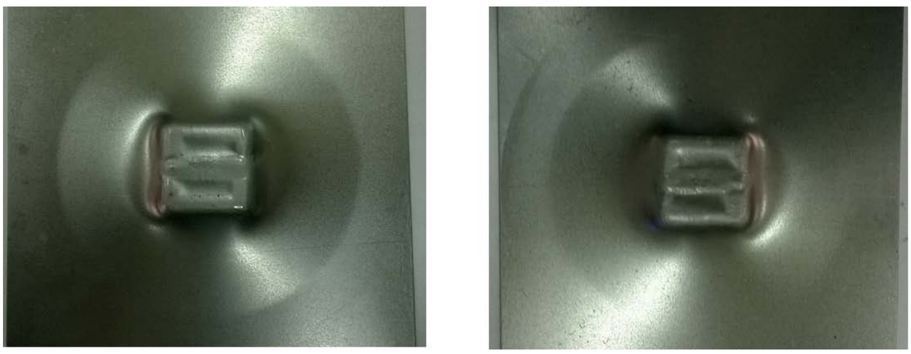


Figure 5: Examples of processed parts

As presented in equation (1), the dynamic resistant dynamic torque also appears during the acceleration and deceleration phases of the movement. To avoid its influence, the data form the operator panel was read in the constant velocity phase of the movements, when due to zero value of the linear and angular acceleration, the dynamic torque is also zero (Fig. 6). Every toolpath was generated in no-load and load conditions and the no-load values were subtracted form the first ones.

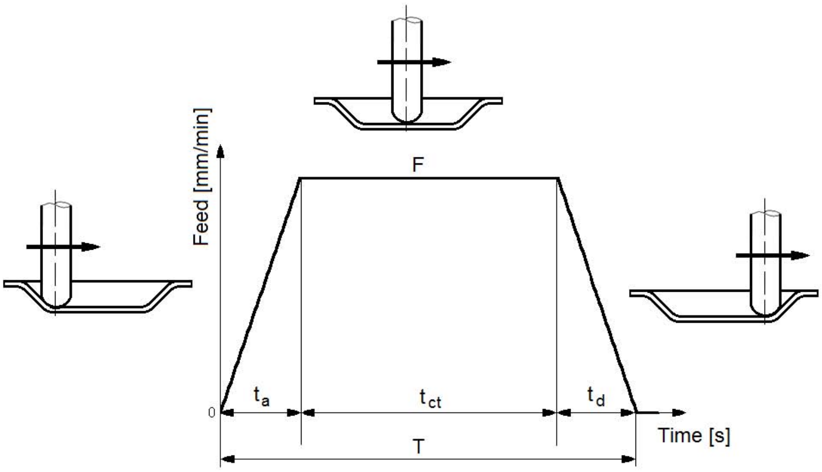


Figure 6: The readings from the operator panel were made after reaching the constant velocity on the axis

The data form the operator panel of the CNC equipment were captured by a digital camera and afterwards the needed information was extracted and processed. The image processing method used in this approach was described in detail in [3].

After this process, the Δt_Z values were obtained and using equation (8) the vertical technological forces could be calculated.

For generating the rough data to be processed by the proposed adaptive network-based fuzzy inference system, a total of 268 experimental data sets were obtained by experimental tests. A reduced selection of these data is presented in (table 1).

Table 2: Example of experimental data set (selection)

Tool diameter [mm]	**Spindle speed [rev/min]**	**Feed on horizontal plane [mm/min]**	**Vertical step [mm]**	**Technological force [N]**
8	400	200	-1	1790.161
8	360	130	-2	1716.338
8	450	175	-2	1492.168
10	300	50	-1	977.211
10	480	200	-2	2043.60
10	450	90	-2	1066.049
12	110	290	-3	1425.840
12	120	350	-2.5	2186.288
12	400	350	-2.5	1901.120

For each group of parameters from table two variation intervals were stated. The values for the tool diameter were set to only three values: 8, 10 and 12 mm. For the spindle speed, the limits were set between 0 and 500 mm/min while for the feed on horizontal plane the limits were set between 50 and 350 mm/min. For the vertical step increments of 1, 2, 2.5 and 3 mm were used. The vertical step was increased using the above-mentioned increments up to a total depth of 7.5 mm.

5 Adaptive network-based fuzzy inference system

The first stage of building the adaptive network-based fuzzy inference system model involved a selection process in order to find the most influential combination of input parameters upon the output parameter. As stated before, the data sets have taken into consideration four parameters as inputs (tool diameter, spindle speed, feed on horizontal plane, vertical step) and one parameter as output (technological force).

It was considered that a system with four inputs could be to complex and consequently a method of selecting only three inputs (the optimal ones) was considered. To fulfill this goal, the Matlab *exhsrch* command (from Fuzzy Logic Toolbox) was used. The function *exhsrch* performs an exhaustive search within the available inputs to select the set of inputs that most influence the output. Essentially, *exhsrch* builds an adaptive network-based fuzzy inference model for each combination and trains it for one epoch and reports the performance achieved. The results from *exhsrch* (Fig. 7) indicate that *"Diameter"*, *"Feed"* and *"Step"* form the best combination of three input attributes. The training (continuous line with circle markers) and checking errors (dotted line with asterisks markers) are getting distinguished, indicating the outset of overfitting. The left-most input set of variables in (Fig. 7) has the least error or in other words the most relevance with respect to the output.

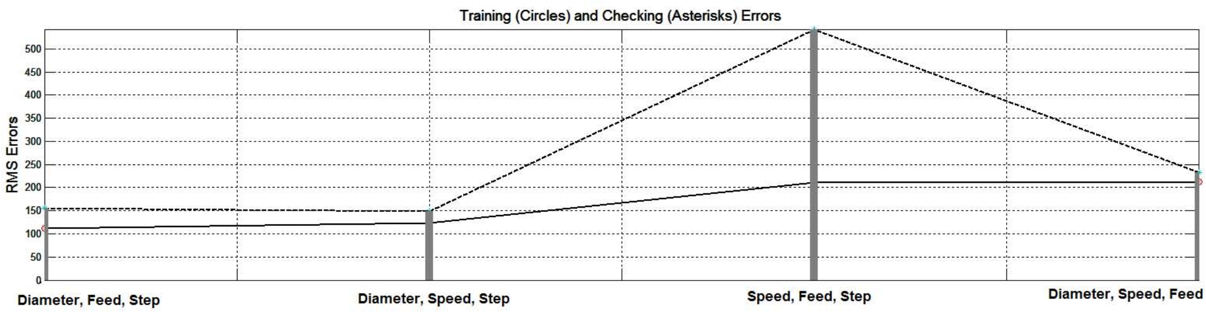


Figure 7: Tool diameter, vertical feed and incremental step - the optimal combination of three inputs

These results are in accordance with the literature survey, which in most of the cases considered speed as the least influential factor within the SPIF process. Most of the authors consider that a nonzero speed of the punch is only required as a measure to reduce friction and it only influences the overall roughness of the surface.

With the selection of input parameters done, the adaptive network-based fuzzy inference system was designed to calculate the vertical technological force, considered as an output.

As for any adaptive network-based fuzzy inference system, a training and checking process had to be unfolded. The experimental data sets were divided into training and checking sets. The division was made by selecting the odd values from the 268 experimental data set as training set while the even values were selected as checking set. At the beginning of the training process, a preliminary fuzzy inference system was generated. At this initial stage the fuzzy system does not reflect entirely the behavior of the considered system. During the training process, artificial neural networks were used to improve the fuzzy system, by means of a learning process. The learning process was completed after 3 stages (epochs), needed for the modelling errors to reach acceptable values. At the end of the learning process, the fuzzy inference system can accurately reflect the real relationships between inputs and output, for the entire range of variations.

The adaptive network-based fuzzy inference system was built and trained using the interactive graphical user interfaces provided by Matlab software packages and its modules, Fuzzy Logic and Neural Network toolboxes. The structure of the fuzzy inference system is presented in (Fig. 8). Triangular membership functions, presented in (Fig. 9) for the *"feed"* input, were selected. Figure 9 also presents the intervals of variation for the selected input.

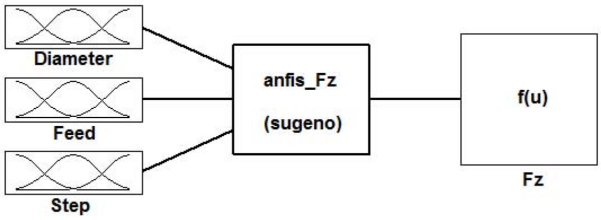


Figure 8: Structure of the fuzzy system

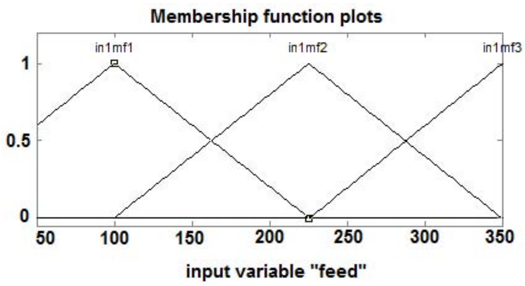


Figure 9: Membership function for the input variable *"feed"*

The final structure of the proposed system, presented in (Fig. 10) has three inputs, an output, 5 levels of ANN, 193 neurons and 81 fuzzy rules.

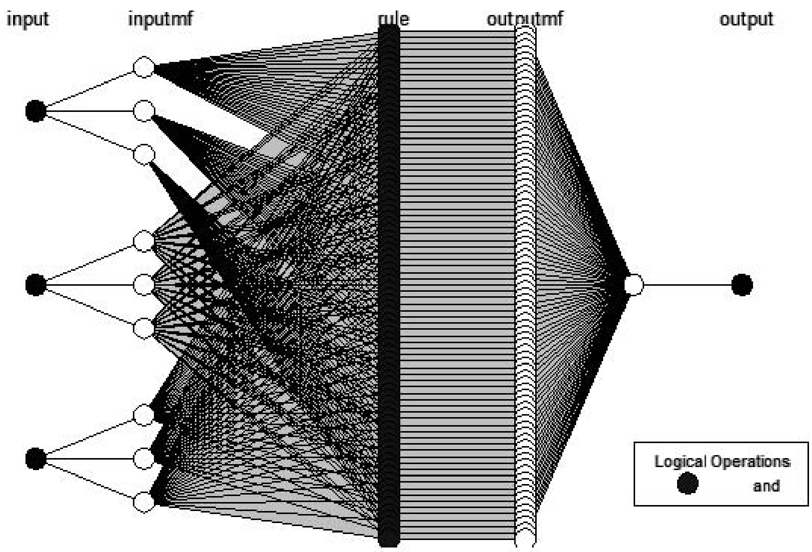


Figure 10: The structure of the neural networks for the proposed system

The adaptive network-based fuzzy inference system has an interactive graphical user interface, presented in (Fig. 1), which allows the user to calculate the output (vertical force) for any intermediary value from the inputs variation ranges. Thus, the system can predict the result for data which were not experimentally available. The user can either drag the vertical bars through any values of the two input variables, or manually introduce input data for them and the model will automatically calculate the value of the output (technological vertical force).

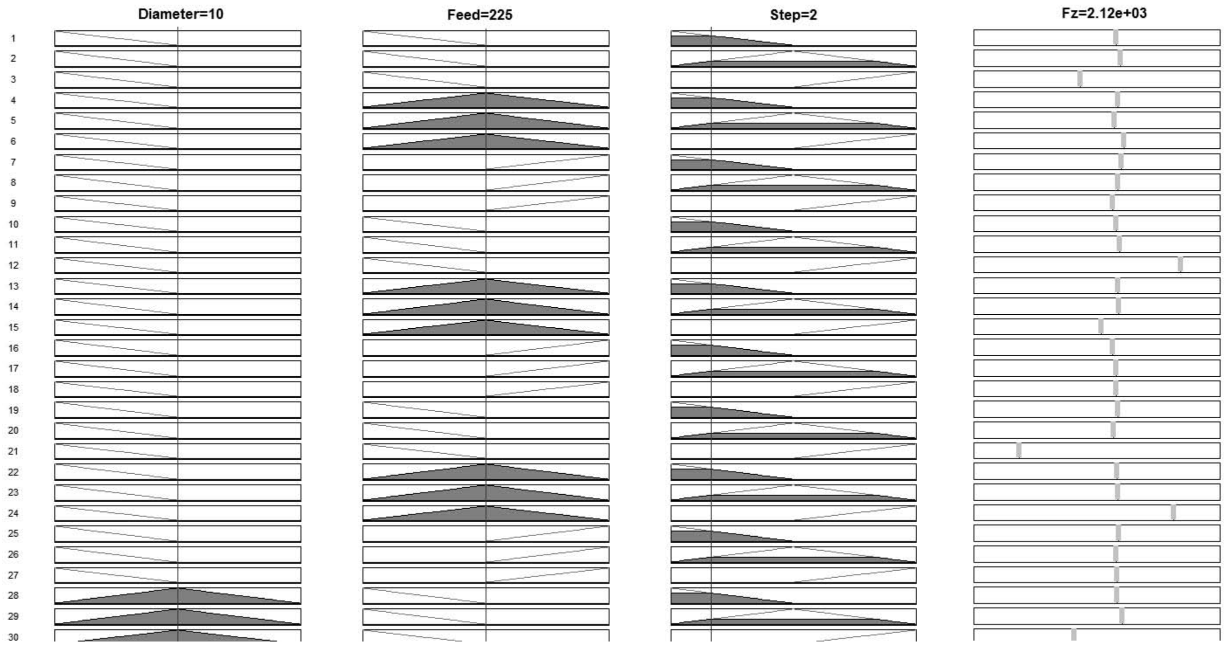


Figure 11: Graphical user interface of the adaptive network-based fuzzy inference system (only first 31 rules are visible in the figure, out of 81)

No related fuzzy models were found in the literature for calculating the vertical technological force during SPIF. To evaluate the efficiency of the proposed model, a comparison between it and a linear regression model was unfolded. The comparison was made by comparing the RMSE (Root mean square) values against checking data for both models. For the proposed model the RMSE was 153.353, while for the linear regression model the RMSE error was 159.290. This result proves to a certain extent the usefulness of the proposed model. However, it should be stressed the fact that further experimental tests must be performed to validate the proposed model for different sheet thicknesses and different materials.

The experimental data (both training and checking data sets) and the adaptive network-based fuzzy inference system (.fis file to be run using the graphical user interface ANFIS from Matlab) can be downloaded from the following link:

https://www.dropbox.com/s/s0gdpjap2istku5/anfis.zip?dl=0

6 Conclusion

The approach presented in this paper did not aim to develop an analytical expression for calculating the vertical force values, but to develop an adaptive network-based fuzzy inference system which allows the user to estimate in an empirical way the force for a given set of process parameters and their variation intervals.

Numerical simulation, using finite elements analysis is today able to calculate the forces within the SPIF process and this allows to evaluate if the process is safe or not for the technological equipment. However, numerical simulation requires both very expensive software packages and highly qualified personnel, able to develop and work with complex and accurate FEA models of the incremental forming process and run the above-mentioned simulations.

The approach presented in this paper offers a cost-effective alternative to that, the method requiring some simple experiments and a cost-effective software package. Moreover, the building of the network-based fuzzy system is a straightforward process, based only upon Matlab integrated graphical user interfaces and does not require a highly trained user.

The proposed method is dedicated to speed the adoption of the SPIF process by the industry, by offering a simple, yet (considered by authors) effective method for estimating the forces, which can be applied at workshops level, rather than in research laboratories.

Future researches will be oriented of taking into consideration different sheet thicknesses and different materials, besides the already considered parameters, to build a more comprehensive knowledge base.

Funding

This research was partially funded by the Romanian Ministry of Research and Innovation CCCDI-UEFISCDI, project number PN-III-P1-1.2-PCCDI-2017-0446/nr. 82PCCDI/2018, within PNCDI III, project title: "Smart Manufacturing Technologies for Advanced Production of Parts from Automotive and Aeronautics Industries".

Author contributions

The authors contributed equally to this work.

Conflict of interest

The authors declare no conflict of interest.

Bibliography

[1] Aerens, R.; Eyckens, P.; Van Bael, A.; Duflou, J. R. (2010); Force prediction for single point incremental forming deduced from experimental and FEM observations, *International Journal of Advanced Manufacturing Technology*, 46(9-12), 969–982, 2010.

[2] Behera, A.K.; De Sousa R.A.; Ingarao, G.; Oleksik, V. (2017); Single point incremental forming: An assessment of the progress and technology trends from 2005 to 2015, *Journal of Manufacturing Processes*, 27, 37–62, 2017.

[3] Breaz, R.; Bologa, O.; Tera, M.; Racz, G. (2013); Determination of Technological Forces in the Incremental Forming Process, *Applied Mechanics and Materials*, 371, 133–137, 2013.

[4] Caydas, U.; Hascalik, A.; Ekici, S. (2009); An adaptive neuro-fuzzy inference system (ANFIS) model for wire-EDM, *Expert Systems with Applications*, 36, 6135-6139, 2009.

[5] Ceretti, E.; Giardini, C.; Attanasio, A. (2004); Experimental and simulative results in sheet incremental forming on CNC machines, *Journal of Materials Processing Technology*, 152(2), 176–184, 2004.

[6] Ciupan, E.; Lungu, F.; Ciupan, C. (2015); ANN Training Method with a Small Number of Examples Used for Robots Control, *International Journal of Computers Communications & Control*, 10(5), 643–653, 2015.

[7] Duflou, J. R.; Szekeres, A.; Vanherck, P. (2005); Force Measurements for Single Point Incremental Forming: An Experimental Study, *Advanced Materials Research*, 6-8, 441–448, 2005.

[8] Dzitac, I.; Filip, F.G.; Manolescu, M.J. (2017); Fuzzy logic is not fuzzy: World-renowned computer scientist Lotfi A. Zadeh, *International Journal of Computers Communications* & *Control*, 12(6), 748–789, 2017.

[9] Dzitac, I. (2015); The fuzzification of classical structures: A general view, *International Journal of Computers Communications* & *Control*, 10(6), 772–788, 2015.

[10] Gatea, S.; Ou, H.; Mccartney, G. (2016); Review on the influence of process parameters in incremental sheet forming, *International Journal of Advanced Manufacturing Technology*, 87, 479–499, 2016.

[11] Haidegger, T.; Kovacs, L.; Precup, R.-E.; Benyo, B.; Benyo, Z.; Preitl, S. (2012); Simulation and control for telerobots in space medicine, *Acta Astronautica*, 181(1), 390–402, 2012.

[12] Herrera-Viedma, E.; Lopez-Herrera, A.G. (2010); A review on information accessing systems based on fuzzy linguistic modelling, *International Journal of Computational Intelligence Systems*, 3(4), 420–437, 2010.

[13] Hyacinth Suganthi, X.; Natarajan, U.; Sathiyamurthy, S.; Chidambaram, K. (2013); Prediction of quality responses in micro-EDM process using an adaptive neuro-fuzzy inference system (ANFIS) model, *International Journal of Advanced Manufacturing Technology*, 68, 339-347, 2013.

[14] Jeswiet, J.; Micari, F.; Hirt, G.; Bramley A.; Duflou, J.; Allwood, J.(2005); Asymmetric Single Point Incremental Forming of Sheet Metal, *Annals of CIRP*, 54, 623–650, 2005.

[15] Jiao, Y.; Lei, S.; Pei, Z.J.; Lee, E.S. (2004); Fuzzy adaptive networks in machining process modeling: surface roughness prediction for turning operations, *International Journal of Machine Tools and Manufacture*, 44(15), 1643–1651, 2004.

[16] Kumanan, S.; Jesuthanam, C. P. (2008); Ashok Kumar, R.; Application of multiple regression and adaptive neuro fuzzy inference system for the prediction of surface roughness, *International Journal of Advanced Manufacturing Technology*, 35, 778-788, 2008.

[17] Li, Y.; Liu, Z.; Lu, H.; Daniel, W. J. T.; Liu, S.; Meehan, P. A. (2014); Efficient force prediction for incremental sheet forming and experimental validation, *International Journal of Advanced Manufacturing Technology*, 73, 571-587, 2014.

[18] Micari, F.; Ambrogio, G.; Filice, L. (2007); Shape and dimensional accuracy in single point incremental forming: State of the art and future trends, *Journal of Materials Processing Technology*, 191(1-3), 390-395, 2007.

[19] Oprea, M.; Mihalache, S. F.; Popescu, M. (2017); Computational Intelligence-based PM2.5 Air Pollution Forecasting, *International Journal of Computers Communications* & *Control*, 12(3), 365–380, 2017.

[20] Perez-Santiago, R.; Bagudanch FrigolĂŠand I.; Garcia-Romeu de Luna, M.L. (2011); Force Modeling in Single Point Incremental Forming of Variable Wall Angle Components, *Key Engineering Materials*, 473, 833-840, 2011.

[21] Precup, R.-E.; Hellendoorn, H. (2011); A survey on industrial applications of fuzzy control, *Computers in Industry*, 62, 213-226, 2011.

[22] Salahshoor, K.; Kordestani, M.; Khoshro, M.S. (2010); XFault detection and diagnosis of an industrial steam turbine using fusion of SVM (support vector machine) and ANFIS (adaptive neuro-fuzzy inference system) classifiers, *Energy*, 35, 5472–5482, 2010.

[23] Schafer, T.; Schraft, R.D. (2005); Incremental sheet metal forming by industrial robots, *Rapid Prototyping Journal*, 11(5), 278–286, 2005.

[24] Tera, M.; Breaz, R.E.; Bologa, O.; Racz, S.G.(2015); Developing a Knowledge Base about the Technological Forces within the Asymmetric Incremental Forming Process, *Key Engineering Materials*, 651, 1115-1121, 2015.

[25] Tseng, T.-L.; Konada, U.; Kwon, Y. (2016); A novel approach to predict surface roughness in machining operations using fuzzy set theory, *Journal of Computational Design and Engineering*, 3, 1-13, 2016.

[26] Velosa De Sena, J.I. (2015); *Advanced numerical framework to simulate Incremental Forming Processes*, Ph.D. Thesis, University of Aveiro, Portugal, 2015.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 14(1), 78-89, February 2019.

A Chi-square Distance-based Similarity Measure of Single-valued Neutrosophic Set and Applications

H.P. Ren, S.X. Xiao, H. Zhou

Haiping Ren
Teaching Department of Basic Subjects
Jiangxi University of Science and Technology,330013 Nanchang, China
9520060004@jxust.edu.cn

Shixiao Xiao*
Chengyi University College
Jimei University,361021 Xiamen, China
*Corresponding author:xiaoshixiao@jmu.edu.cn

Hui Zhou
School of Mathematics and Computer Science
Yichun University,336000 Yichun, China
huihui7978@126.com

Abstract: The aim of this paper is to propose a new similarity measure of single-valued neutrosophic sets (SVNSs). The idea of the construction of the new similarity measure comes from Chi-square distance measure, which is an important measure in the applications of image analysis and statistical inference. Numerical examples are provided to show the superiority of the proposed similarity measure comparing with the existing similarity measures of SVNSs. A weighted similarity is also put forward based on the proposed similarity. Some examples are given to show the effectiveness and practicality of the proposed similarity in pattern recognition, medical diagnosis and multi-attribute decision making problems under single-valued neutrosophic environment.
Keywords: Chi-square distance measure, similarity measure, multi-attribute decision making, single-valued neutrosophic set.

1 Introduction

Since fuzzy set was first proposed by Zadeh and has achieved a great success in various fields due to its capability of handling uncertainty [4, 6, 9, 14, 21, 29]. Over the last decades, some extended fuzzy sets have been introduced by researchers, such as intuitionistic fuzzy set [1], vague set [10], interval-valued intuitionistic fuzzy set [2] and hesitant fuzzy set [28]. As an extension of Zadeh's fuzzy set, intuitionistic fuzzy set can better describe the situation when decision making process exists decision makers' hesitation than Zadeh's fuzzy set by adding a non-membership degree parameter [32]. In recent years, intuitionistic fuzzy set has received a lot attention and been applied to many fields, such as management decision, pattern recognition and medical diagnosis [5, 7, 18, 20, 27]. In practice, indeterminate and inconsistent information may occur, then intuitionstic fuzzy set cannot deal with these situations well because it only contains the true membership degree and the false membership degree (non-membership degree). For example, when an authority wants to choose the best candidate, ten experts are invited to take part in the decision. For one candidate he gained 10 votes from the experts. There are 3 votes yes", 2 votes "no", 2 "gave up"" and 3 "undecided". In this case, intuitionistic fuzzy set cannot describe it well. To overcome this shortcoming, Smarandache [26] introduced a concept of neutrosophic set, which is an extension of intuitionistic fuzzy set from philosophical point of view. Neutrosophic set is defined as a set containing the degree of truth, indeterminacy, and falsity. For afore-mentioned example, the vote result can be expressed by a neutrosophic set. However, the original neutrosophic set is difficult to apply in practical problems. To applied it easily in science and engineering fields, Wang et al. [30] introduced the concept of single-valued neutrosophic set (SVNS), which is a subclass of Smarandache's neutrosophic set. Because SVNS is easy to express, it has been a useful mathematical tool for handling various practical problems involving imprecise, indeterminacy, and inconsistent data [11, 15, 19, 23, 25].

In recent years, the study and applications of information measures of fuzzy sets have received a lot attention. Similarity measure is one of the most important measurement tools for comparing the degree of similarity between two objects. Since Li and Chen [17] introduced the definition of the similarity measure between two intuitionistic fuzzy sets. Since then intuitionistic fuzzy similarity measures have received great attention. From a different point of view, many similarity measures are proposed and applied to solve various practical problems of MADM, pattern recognition and medical diagnosis, etc. [12,13,16,22,31,34]. As an extension of intuitionistic fuzzy set, some similarity measures of neutrosophic set is developed from those of intuitionistic fuzzy sets, and some new similarity measures are also proposed, but the references are still rare ([35]). Based on vector similarity functions, some similarity measures between simplified neutrosophic sets are put forwards, such as similarity measures based on Jaccard, Dice, and cosine functions [33–35].

We find that the existing similarity measures have shortcomings, and the detail analysis can be found in Example 1. Then the aim of this paper is to develop a new similarity measure of SVNSs based on Chi-square distance measure, which is an important measure in statistical theory. We will show the advantage of the proposed similarity measure with existing similarity measures of SVNSs through comparison with some numerical examples. Three examples are provided to demonstrate the effectiveness and practicality of the proposed similarity in the application of pattern recognition, medical diagnosis and multi-attribute decision making.

The remains of this article are organized as follows: Section 2 will recall some basic concepts and properties of SVNSs and similarity measure. Section 3 introduces a new similarity measure between SVNSs based on ordinary Chi-square distance measure, and put forward a weighted similarity for further applications. Section 4 develops the applications of the proposed similarity measure with some examples. Finally, conclusions are provided in Section 5.

2 Preliminary knowledge

In this section, some basic concepts and properties of SVNSs and similarity measure are presented. Smarandache [26] originally introduced a concept of neutrosophic set from philosophical point of view.

Definition 1 Let X be a universal set. A set is called a neutrosophic set, if it is characterized by three parameters: truth-membership function $T_A(x)$, indeterminacy-membership function $I_A(x)$ and falsity-membership function $F_A(x)$ .That is A has the following form:

$$A = \{ < x, T_A(x), I_A(x), F_A(x) > |x \in X\}.$$

Here$T_A(x), I_A(x), F_A(x) : X \to]^-0, 1^+[$, $]^-0, 1^+[$ is non-standard interval, and they satisfy

$$^-0 \leq \sup T_A(x) + \sup I_A(x) + \sup F_A(x) \leq 3^+$$

A neutrosophic is defined as a set containing the degree of truth, indeterminacy, and falsity. However, the original neutrosophic set is difficult to apply in practical problems. To apply it easily in science and engineering fields, Wang et al. [30] introduced single-valued neutrosophic set (SVNS), which is a subclass of SmarandacheĄŻs neutrosophic set.

Definition 2 Let X be a universal set. A set A is called a SVNS, if it is characterized by three parameters: truth-membership function $T_A(x)$, indeterminacy-membership function $I_A(x)$ and falsity-membership function $F_A(x)$. That is, A has the following form:

$$A = \{< x, T_A(x), I_A(x), F_A(x) > |x \in X\}.$$

Here$T_A(x), I_A(x), F_A(x) : X \to [0,1]$, and they satisfy $0 \leq T_A(x) + I_A(x) + F_A(x) \leq 3$. For convenience, when $X = \{x\}$, we briefly denote the element $< x, T_A(x), I_A(x), F_A(x) >$ of A by $< T_A, I_A, F_A >$.Element $< T_A, I_A, F_A >$ is often named as a single-valued neutrosophic value (SVNV).

Definition 3 [30] Let X be a universal set, and $A = \{< x, T_A(x), I_A(x), F_A(x) > |x \in X\}$and $B = \{< x, T_B(x), I_B(x), F_B(x) > |x \in X\}$ are two SVNSs in X, then

(i) The complement of a SVNS A is

$$A^c = \{< x, F_A(x), 1 - I_A(x), T_A(x) > |x \in X\}$$

(ii)$A \subseteq B$ if and only if

$$T_A(x) \leq T_B(x), I_A(x) \geq I_B(x), F_A(x) \geq F_B(x),$$

for all x in X. (iii) $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

In the following discussion, we always use $SVNSs(X)$ to denote the set of all SVNSs in X. Definition 4 will introduce the definition of a similarity measure between two SVNSs A and B.

Definition 4 [35] Let A and B be two SVNSs, and S is a mapping $S : SVNSs(X) \times SVNSs(X) \to [0,1]$. We call $S(A,B)$ the similarity measure between A and B if it satisfies the following properties:

(i) $0 \leq S(A,B) \leq 1$;

(ii)$S(A,B) = 1$ if and only if $A = B$;

(iii)$S(A,B) = S(B,A)$;

(iv) If $A \subseteq B \subseteq C$, then $S(A,C) \leq \min\{S(A,B), S(B,C)\}$.

3 A new Chi-square distance-based similarity

This section contains two subsections. The first subsection will propose a new similarity measure between two SVNSs based on Chi-square distance measure. The second subsection will compare the proposed similarity measure with existing similarity measures of SVNSs.

3.1 A new proposed similarity based on Chi-square distance

This section will propose a new similarity measure between two SVNSs based on Chi-square distance measure. The name of the Chi-square distance measure is derived from Pearson's Chi-squared test statistic $\chi^2(\mathbf{x}, \mathbf{y}) = \sum\limits_{i=1}^{n} \frac{(x_i - y_i)^2}{x_i}$, which is used to compare two discrete probability distributions. However, as a distance measure, the function $d(x, y)$ should be symmetric for two objects x and y .Then Chi-square distance measure of two real vectors $\mathbf{x} = (x_1, x_2, ..., x_n)$ and $\mathbf{y} = (y_1, y_2, ..., y_n)$ is proposed as the following formula [24]:

$$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \frac{(x_i - y_i)^2}{x_i + y_i} \tag{1}$$

Chi-square distance measure is one of most important distance measure used in face recognition [24]. To avoid the fact that the denominator is zero, the numerical value is meaningless. For later use, we proposed a revised version of Chi-square distance measure formula as follows:

$$d(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n} \frac{(x_i - y_i)^2}{2 + x_i + y_i} \tag{2}$$

Note that the constant 2 can be changed as any other positive number. Let $X = \{x_1, x_2, ..., x_n\}$ be a universe set. Then for two given SVNSs $A = \{< x_i, T_A(x_i), I_A(x_i), F_A(x_i) > |x_i \in X\}$ and $B = \{< x_i, T_B(x_i), I_B(x_i), F_B(x_i) > |x_i \in X\}$, the new neutrosophic fuzzy information measure based on Chi-square distance $S \triangleq S\,(A,B)$ is constructed as follows:

$$\begin{aligned} S = 1 - \tfrac{1}{2n} \sum_{i=1}^{n} \Big[& \tfrac{(T_A(x_i) - T_B(x_i))^2}{2 + T_A(x_i) + T_B(x_i)} + \tfrac{(I_A(x_i) - I_B(x_i))^2}{2 + I_A(x_i) + I_B(x_i)} \\ & + \tfrac{(F_A(x_i) - F_B(x_i))^2}{2 + F_A(x_i) + F_B(x_i)} + |m_A(x_i) - m_B(x_i)| \Big] \end{aligned} \tag{3}$$

Where $m_j(x_i) = \frac{1 + T_j(x_i) - F_j(x_i)}{2}$, $j = 1, n$.

To prove the information measure (3) is a valid similarity measure, we need the following lemma which can be easily proved by straightforward calculation.

Lemma 1 Let a, b, c be three non-negative real numbers, and $0 \le a \le b \le c$. Then

(i) $\frac{(a-c)^2}{2+a+c} \ge \frac{(a-b)^2}{2+a+b}$

(ii) $\frac{(a-c)^2}{2+a+c} \ge \frac{(b-c)^2}{2+b+c}$

Theorem 1 Let $X = \{x_1, x_2, ..., x_n\}$ be a universe set. $A = \{< x_i, T_A(x_i), I_A(x_i), F_A(x_i) > |x_i \in X\}$ and $B = \{< x_i, T_B(x_i), I_B(x_i), F_B(x_i) > |x_i \in X\}$ are two SVNSs. Then information measure $S\,(A,B)$ given by (2) is a valid similarity measure between SVNSs A and B.That is, $S\,(A,B)$ satisfies the properties (i)-(iv) of Definition 4.

Proof (i) It is obvious that $0 \le S\,(A,B) \le 1$.

(ii) When $A = B$, i.e. $T_A(x_i) = T_B(x), I_A(x_i) = I_B(x_i), F_A(x_i) = F_B(x_i)$, for all x_i in X. Then

$$m_A(x_i) = \frac{1 + T_A(x_i) - F_A(x_i)}{2} = \frac{1 + T_B(x_i) - F_B(x_i)}{2} = m_B(x_i)$$

Hence we have $S\,(A,B) = 1$.

(iii) The result is obvious.

(iv)If $A \subseteq B \subseteq C$, i.e.

$$T_A(x_i) \le T_B(x) \le T_C(x), I_A(x_i) \ge I_B(x_i) \ge I_C(x_i), F_A(x_i) \ge F_B(x_i) \ge F_C(x_i),$$

Then $0 \le m_A(x_i) \le m_B(x_i) \le m_C(x_i)$,for all x_i in X.

Consequently, we can get

$$|m_A(x_i) - m_C(x_i)| \ge |m_A(x_i) - m_B(x_i)|, |m_A(x_i) - m_C(x_i)| \ge |m_B(x_i) - m_B(x_i)|$$

Thus by (i) of Lemma 1, we have

$$\frac{(T_A(x_i) - T_C(x_i))^2}{2 + T_A(x_i) + T_C(x_i)} \ge \frac{(T_A(x_i) - T_B(x_i))^2}{2 + T_A(x_i) + T_B(x_i)},$$

$$\frac{(I_A(x_i) - I_C(x_i))^2}{2 + I_A(x_i) + I_C(x_i)} \ge \frac{(I_A(x_i) - I_B(x_i))^2}{2 + I_A(x_i) + I_B(x_i)},$$

$$\frac{(F_A(x_i)-F_C(x_i))^2}{2+F_A(x_i)+F_C(x_i)} \geq \frac{(F_A(x_i)-F_B(x_i))^2}{2+F_A(x_i)+F_B(x_i)}.$$

Then we can easily conclude that $S(A,C) \leq S(A,B)$. By (ii) of Lemma 2, we have

$$\frac{(I_A(x_i)-I_C(x_i))^2}{2+I_A(x_i)+I_C(x_i)} \geq \frac{(I_B(x_i)-I_C(x_i))^2}{2+I_B(x_i)+I_C(x_i)},$$

$$\frac{(F_A(x_i)-F_C(x_i))^2}{2+F_A(x_i)+F_C(x_i)} \geq \frac{(F_B(x_i)-F_C(x_i))^2}{2+F_B(x_i)+F_C(x_i)}.$$

$$\frac{(T_A(x_i)-T_C(x_i))^2}{2+T_A(x_i)+T_C(x_i)} \geq \frac{(T_B(x_i)-T_C(x_i))^2}{2+T_B(x_i)+T_C(x_i)},$$

Then we can easily conclude that $S(A,C) \leq S(B,C)$. Hence $S(A,C) \leq \min\{S(A,B), S(B,C)\}$. This completes the proof of Theorem 1. If we consider the important degree of $x_i \in X = \{x_1, x_2, ..., x_n\}$, then we can establish a weighted similarity measure $S_W \triangleq S_W(A,B)$ between SNNSs A and B as follows:

$$S_W = 1 - \frac{1}{2}\sum_{i=1}^{n} w_i \Big[\frac{(T_A(x_i)-T_B(x_i))^2}{2+T_A(x_i)+T_B(x_i)} + \frac{(I_A(x_i)-I_B(x_i))^2}{2+I_A(x_i)+I_B(x_i)} + \frac{(F_A(x_i)-F_B(x_i))^2}{2+F_A(x_i)+F_B(x_i)} + |m_A(x_i)-m_B(x_i)|\Big] \tag{4}$$

where w_i $(i = 1, 2, ..., n)$ is the important degree of the element x_i, they satisfy $w_i \in [0, 1]$ and $\sum_{i=1}^{n} w_i = 1$. If we set $w_i = \frac{1}{n}$ $(i = 1, 2, ..., n)$, then $S_W(A,B) = S(A,B)$. Similar to the proof process of $S_R(A,B)$in Theorem 1, we can easily prove that the weighted similarity measure $S_W(A,B)$ is also a valid similarity between two SNVSs A and B . That is $S_W(A,B)$ satisfies the properties (i)-(iv) of Definition 4.

3.2 Comparison of various similarity measures

To demonstrate the validness and performance of the new proposed similarity measure, some numerical examples are used to compare it with existing similarity measures: Jaccard similarity $S_J(A,B)$, Dice similarity $S_D(A,B)$, Cosine Similarity $S_C(A,B)$, Improved cosine similarity $C_1(A,B)$and $C_2(A,B)$, Tangent function-based similarity $T_1(A,B)$, $T_2(A,B)$, and Cotangent function-based similarity $CoT_1(A,B)$, $CoT_2(A,B)$. These similarity measures are given as follows([33], [34], [35]):

$$S_J(A,B) = \frac{1}{n}\sum_{i=1}^{n}\frac{S_{J1}}{S_{J2}}, \tag{5}$$

where $S_{J1} = T_A(x_i)T_B(x_i)+I_A(x_i)I_B(x_i)+F_A(x_i)F_B(x_i)$ and $S_{J2} = (T_A^2(x_i)+I_A^2(x_i)+F_A^2(x_i))+(T_B^2(x_i)+I_B^2(x_i)+F_B^2(x_i)) - S_{J1}$

$$S_D(A,B) = \frac{1}{n}\sum_{i=1}^{n}\frac{2(T_A(x_i)T_B(x_i)+I_A(x_i)I_B(x_i)+F_A(x_i)F_B(x_i))}{(T_A^2(x_i)+I_A^2(x_i)+F_A^2(x_i))+(T_B^2(x_i)+I_B^2(x_i)+F_B^2(x_i))}, \tag{6}$$

$$S_C(A,B) = \frac{1}{n}\sum_{i=1}^{n}\frac{T_A(x_i)T_B(x_i)+I_A(x_i)I_B(x_i)+F_A(x_i)F_B(x_i)}{\sqrt{T_A^2(x_i)+I_A^2(x_i)+F_A^2(x_i)}\sqrt{T_B^2(x_i)+I_B^2(x_i)+F_B^2(x_i)}}, \tag{7}$$

$$C_1(A,B)=\frac{1}{n}\sum_{i=1}^{n}\cos\left[\frac{\pi}{2}\max\left(|T_A(x_i)-T_B(x_i)|,|I_A(x_i)-I_B(x_i)|,|F_A(x_i)-F_B(x_i)|\right)\right], \quad (8)$$

$$C_2(A,B)=\frac{1}{n}\sum_{i=1}^{n}\cos\left[\frac{\pi}{6}\left(|T_A(x_i)-T_B(x_i)|+|I_A(x_i)-I_B(x_i)|+|F_A(x_i)-F_B(x_i)|\right)\right], \quad (9)$$

$$T_1(A,B)=1-\frac{1}{n}\sum_{i=1}^{n}\tan\left[\frac{\pi}{4}\max\left(|T_A(x_i)-T_B(x_i)|,|I_A(x_i)-I_B(x_i)|,|F_A(x_i)-F_B(x_i)|\right)\right], \quad (10)$$

$$T_2(A,B)=1-\frac{1}{n}\sum_{i=1}^{n}\tan\left[\frac{\pi}{12}\left(|T_A(x_i)-T_B(x_i)|+|I_A(x_i)-I_B(x_i)|+|F_A(x_i)-F_B(x_i)|\right)\right], \quad (11)$$

$$CoT_1(A,B)=\frac{1}{n}\sum_{i=1}^{n}\cot\left[\frac{\pi}{4}\max\left(|T_A(x_i)-T_B(x_i)|,|I_A(x_i)-I_B(x_i)|,|F_A(x_i)-F_B(x_i)|\right)\right], \quad (12)$$

$$CoT_2(A,B)=\frac{1}{n}\sum_{i=1}^{n}\cot\left[\frac{\pi}{4}+\frac{\pi}{12}\left(|T_A(x_i)-T_B(x_i)|+|I_A(x_i)-I_B(x_i)|+|F_A(x_i)-F_B(x_i)|\right)\right], \quad (13)$$

Example 1 Suppose that $X=\{x\}$, we consider pattern recognition problems with six pairs of SVNSs shown in Table 1. The calculated numerical values of these 9 existing similarity measures and proposed similarity measure are shown in Table 1.

Table 1: Values of the different similarity measures under different pairs of (A,B)

	Case 1	**Case 2**	**Case 3**	**Case 4**	**Case 5**	**Case 6**
A	<0.3,0.3,0.4>	<0.3,0.3,0.4>	<1,0,0>	<0.4,0.2,0.6>	<0.4,0.4,0.2>	<0.4,0.4,0.2>
B	<0.4,0.3,0.4>	<0.4,0.3,0.3>	<0,1,1>	<0.2,0.2,0.3>	<0.5,0.2,0.3>	<0.5,0.3,0.2>
$S_J(A,B)$	0.9737	0.9429	0	0	0.8500	0.9474
$S_D(A,B)$	0.9867	0.9706	0	0	0.9189	0.9730
$S_C(A,B)$	0.9910	0.9706	0	Null	0.9193	0.9733
$C_1(A,B)$	**0.9877**	**0.9877**	0	0	0.9511	0.9877
$C_2(A,B)$	0.9986	0.9945	0	**0.8660**	0.9781	0.9945
$T_1(A,B)$	**0.9213**	**0.9213**	0	0	0.8416	0.9213
$T_2(A,B)$	0.9738	0.9476	0	**0.7321**	0.8949	0.9476
$CoT_1(A,B)$	0.8541	0.8541	0	0	0.7265	0.8541
$CoT_2(A,B)$	0.9490	0.9004	0	**0.5774**	0.8098	0.9004
$S(A,B)$	0.9713	0.9463	0	**0.5833**	**0.9886**	**0.9714**

From Table 1, we can see that the similarity measures $C_1(A,B)$ and $T_1(A,B)$ cannot carry out the recognition between Case 1 and Case 2. For Case 4, there are only four reasonable similarity measures $C_2(A,B)$, $T_2(A,B)$, $CoT_2(A,B)$ and $S(A,B)$ intuitively consistent. An interesting counter-intuitive case occurs when three SNVSs $A =< 0.4, 0.4, 0.2 >$, $B =< 0.5, 0.2, 0.3 >$ and $C =< 0.5, 0.3, 0.2 >$. They can be written in forms of intuitionistic fuzzy values as: $A =< 0.4, 0.2 >$, $B =< 0.5, 0.3 >$ and $C =< 0.5, 0.2 >$, respectively. In this case, Boran and Akay [3] pointed out that it is expected that the similarity degree between A and B should not be less than the similarity degree between A and C since they are ordered as $A \leq B \leq C$ according to score function and accuracy function. However, the similarity degree between A and C is greater than the similarity degree between A and B when the existing similarity measures are used (The results can be found in Table 1), which does not seem to be reasonable. Table 1 shows that our proposed similarity measure is in agreement with this analysis. According to the above analysis, the proposed similarity measure is the most reasonable similarity measure.

4 Applications

In the following discussion, we will give two examples in pattern recognition and medical diagnosis to demonstrate the effectiveness and practicability of the proposed similarity measure.

Example 2 Assume that there are two patterns in $X = \{x_1, x_2\}$. The two patterns are expressed by SVNSs, which are shown as follows:

$$A_1 = \{< x_1, 0.2, 0.0, 0.2 >, < x_2, 0.2, 0.0, 0.2 >, < x_3, 0.2, 0.0, 0.2 >\},$$

$$A_2 = \{< x_1, 0.4, 0.0, 0.4 >, < x_2, 0.4, 0.0, 0.4 >, < x_3, 0.4, 0.0, 0.4 >\}.$$

Assume that there is an object

$$B = \{< x_1, 0.3, 0.0, 0.3 >, < x_2, 0.3, 0.0, 0.3 >, < x_3, 0.2, 0.0, 0.3 >\}$$

Our task is to classify the object B in A_1 or A_2 . According to the recognition principle of maximum similarity measure between SVNSs, the process of assigning the object B to A_1 or A_2 is described by

$$k = arg \max_{1 \leq i \leq 2} \{S_R(A_i, B)\} \tag{14}$$

By Eq. (3), we can get the similarity measures between A_1, A_2 with B : $S(A_1, B) = 0.9700, S(A_2, B) = 0.9415$. Then the pattern B is classified in A_1 according to the recognition rule given by Eq. (14). This result is consistent with our intuition.

Example 3 We consider the following pattern recognition problem: There are three patterns A_1, A_2 and A_3 , which are represented by SVNSs in universe set $X = \{x_1, x_2, x_3\}$, as follows:

$$A_1 = \{< x_1, 1.0, 0.2, 0.0 >, < x_2, 0.8, 0.3, 0.0 >, < x_3, 0.7, 0.1, 0.1 >\}$$
$$A_2 = \{< x_1, 0.8, 0.1, 0.1 >, < x_2, 1.0, 0.1, 0.2 >, < x_3, 0.9, 0.2, 0.1 >\}$$
$$A_3 = \{< x_1, 0.6, 0.3, 0.2 >, < x_2, 0.8, 0.2, 0.3 >, < x_3, 0.6, 0.3, 0.2 >\}$$

Given an unknown pattern B, which is represented by the SVNS:

$$B = \{< x_1, 0.5, 0.3, 0.2 >, < x_2, 0.6, 0.3, 0.2 >, < x_3, 0.8, 0.2, 0.1 >\}$$

Our task is to classify the pattern B in one of the classes A_1, A_2 and A_3 . According to the recognition principle of maximum similarity measure between SVNSs, the process of assigning the pattern B to A_k $(k = 1, 2, 3)$ is described by

$$k = arg \max_{1 \leq i \leq 3} \{S_R(A_i, B)\} \tag{15}$$

By Eq.(3), we can get the similarity measures between B with A_i $(i = 1, 2, 3)$:
$S(A_1, B) = 0.9115, S(A_2, B) = 0.8813, S(A_3, B) = 0.9345.$

Then the pattern B is classified in A_3 according to the recognition rule given by Eq. (15). Some medical diagnosis problems are very complex. Physicians need to use modern medical technologies to obtain a lot of information available to physicians for the help of decision, but the information is often incomplete, indeterminate and inconsistent. The SVNSs proposed by Wang et al. [30] can be better choice to express this kind of information than ZadehĄŻs fuzzy sets and intuitionistic fuzzy sets. Now in Example 4 we will utilize the proposed similarity measure to solve a class of medical diagnosis problems.

Example 4 The medical diagnosis problem is adapted from De et al. [8]. Let $Q =$ Q_1 (Viral fever), Q_2 (Malaria), Q_3 (Typhoid), Q_4 (Stomach problem), Q_5 (Chest problem) be a set of diagnoses (diseases) and $S=$ s_1(Temperature), s_2(Headache), s_3 (Stomach pain), s_4 (Cough), s_5(Chest pain) be a set of symptoms. Each diagnosis Q_i $(i = 1, 2, 3, 4, 5)$ can be represented by SVNSs as follows:

$$Q_1 = \{< s_1, 0.4, 0.6, 0.0 >, < s_2, 0.3, 0.2, 0.5 >, < s_3, 0.1, 0.2, 0.7 >, < s_4, 0.4, 0.3, 0.3 >, < s_5, 0.1, 0.2, 0.7 >\}$$

$$Q_2 = \{< s_1, 0.7, 0.3, 0.0 >, < s_2, 0.2, 0.2, 0.6 >, < s_3, 0.0, 0.1, 0.9 >, < s_4, 0.7, 0.3, 0.0 >, < s_5, 0.1, 0.1, 0.8 >\}$$
$$Q_3 = \{< s_1, 0.3, 0.4, 0.3 >, < s_2, 0.6, 0.3, 0.1 >, < s_3, 0.2, 0.1, 0.7 >, < s_4, 0.2, 0.2, 0.6 >, < s_5, 0.1, 0.0, 0.9 >\}$$
$$Q_4 = \{< s_1, 0.1, 0.2, 0.7 >, < s_2, 0.2, 0.4, 0.4 >, < s_3, 0.8, 0.2, 0.0 >, < s_4, 0.2, 0.1, 0.7 >, < s_5, 0.2, 0.1, 0.7 >\}$$
$$Q_5 = \{< s_1, 0.1, 0.1, 0.8 >, < s_2, 0.0, 0.2, 0.8 >, < s_3, 0.2, 0.0, 0.8 >, < s_4, 0.2, 0.0, 0.8 >, < s_5, 0.8, 0.1, 0.1 >\}$$

Suppose there are two patients P_1 and P_2 , with respect to all the symptoms, can be represented by the following SVNSs:

$$P_1 = \{< s_1, 0.8, 0.1, 0.1 >, < s_2, 0.6, 0.3, 0.1 >, < s_3, 0.2, 0.0, 0.8 >, < s_4, 0.6, 0.3, 0.1 >, < s_5, 0.1, 0.3, 0.6 >\}$$

$$P_2 = \{< s_1, 0.0, 0.2, 0.8 >, < s_2, 0.4, 0.4, 0.2 >, < s_3, 0.6, 0.3, 0.1 >, < s_4, 0.1, 0.7, 0.2 >, < s_5, 0.1, 0.8, 0.1 >\}$$

Our aim is to determine the patients P_1 and P_2 belong to which diagnosis of Q_j $(j = 1, 2, 3, 4, 5)$, respectively. Because the medical diagnosis problem is actually a pattern recognition problem, then we can use the diagnosis rule as follows: If $k = arg \max_{1 \leq j \leq 5} \{S_R(Q_j, P_i)\}$, then we assign the patient P_1 and P_2 to the diagnosis Q_k.

$$S(Q_1, P_1) = 0.9160, S(Q_2, P_1) = 0.9360,$$
$$S(Q_3, P_1) = 0.9000, S(Q_4, P_1) = 0.7220$$

and $S(Q_5, P_1) = 0.6640,$

$$S(Q_1, P_2) = 0.8103, S(Q_2, P_2) = 0.7259,$$
$$S(Q_3, P_2) = 0.8043, S(Q_4, P_2) = 0.8435$$

and $S(Q_4, P_2) = 0.7553$ Then, By the above diagnosis rule, we can assign the patient P_1 to the diagnosis Q_2 (Malaria), and P_2 to the diagnosis Q_4 (Stomach problem). This result is in agreement with the one obtained in De et al. [8].

Example 5 (Multi-attribute decision making) We consider a MADM problem adopted from Ye [33]. A manufacturing company wants to select the best global supplier form a set of four suppliers $A = \{A_1, A_2, A_3, A_4\}$ whose core competencies are evaluated according to the four attributes $O = \{O_1, O_2, O_3, O_4\}$: o_1 (the level of technology innovation) , o_2 (the control ability of flow), o_3 (the ability of management) , o_4 (the level of service). The attributes are all benefit attributes. The weight vector for the four attributes determined by decision maker is

$$\mathbf{W} = (w_1, w_2, w_3, w_4)^T = (0.30, 0.25, 0.25, 0.20)^T$$

Suppose that the evaluation value of the alternative $A_i (i = 1, 2, 3, 4)$ with respect to $o_j (j = 1, 2, 3, 4)$ is a SNVN $a_{ij} =< T_{ij}, I_{ij}, F_{ij} >$, which is obtained from a questionnaire of a domain expert. For example, when we ask the opinion of an expert about an alternative A_1 with respect to an attribute o_1 , he/she may say that the possibility in which the good statement is 0.5 and the poor statement is 0.3 and the degree in which he/she is not sure is 0.1. For the neutrosophic notation, it can be expressed as $a_{11} =< T_{11}, I_{11}, F_{11} >$. The evaluation values are listed in Table 2.

Table 2: Evaluation values of each alternative with respect to each attribute1

Alternatives	o_1	o_2	o_3	o_4
A_1	<0.75,0.2,0.3>	<0.7,0.2,0.3>	<0.65,0.2,0.25>	<0.75,0.2,0.1>
A_2	<0.8,0.1,0.2>	<0.75,0.2,0.1>	<0.75,0.2,0.1>	<0.85,0.1,0.2>
A_3	<0.7,0.2,0.2>	<0.78,0.2,0.1>	<0.85,0.15,0.1>	<0.76,0.2,0.2>
A_4	<0.8,0.2,0.1>	<0.85,0.2,0.2>	<0.7,0.2,0.2>	<0.86,0.1,0.2>

Now, we will propose a decision making method based on the proposed similarity measure to solve this problem and the detail steps is given as follows:

Step 1 Determine the ideal solution A^* as follows:

$$A^* = (< T_j^*, I_j^*, F_j^* >)_{1\times 4} = (< \max_{1\le i\le 4}(T_{ij}), \min_{1\le i\le 4}(I_{ij}), \max_{1\le i\le 4}(F_{ij}) >)_{1\times 4}$$

Step 2 According to Eq. (4), calculate similarity measures between each alternative $A_i (i = 1, 2, 3, 4)$ and the ideal solution A^* as follows:

$$S(A_1, A^*) = 0.9859, S(A_2, A^*) = 0.9955, S(A_3, A^*) = 0.9919$$

and $S(A_4, A^*) = 0.9942$.

Step 3 According to the similarity measure values, the ranking order of the four suppliers is $A_2 \succ A_4 \succ A_3 \succ A_1$. Hence, the best supplier is A_2 , which is in agreement with the result obtained by using weighted projection similarity and weighted Dice similarity methods(Ye [33]).

5 Conclusion

Neutrosophic sets are suitable to model the indeterminate and inconsistent information occurred in many practical problems. In this paper, we have proposed a new Chi-square distance-based similarity measure of SVNSs. The new proposed similarity measure is a valid similarity measure and it can also overcome the counter-intuitive cases of the existing similarity measures by using some numerical examples. We have given the applications of the proposed similarity measure in pattern recognition and medical diagnosis. Furthermore, a multi-attribute decision making method is proposed through an example in which attribute values are expressed with SVNVs.

As a prospect, the MADM method proposed in this paper could be applied to other MADM problems, such as the risk evaluation, credit evaluation. In the future work, we shall extend the proposed similarity to clustering analysis and image processing.

Funding

This paper was supported by the National Natural Science Foundation of China (No.71661012) and Foundation of Jiangxi Educational Committee (No. GJJ170496).

Author contributions

The authors contributed equally to this work.

Conflict of interest

The authors declare no conflict of interest.

Bibliography

[1] Atanassov, K. T.(1986); Intuitionistic fuzzy sets, *Fuzzy Sets and Systems*, 20(1), 87–96, 1986.

[2] Atanassov, K.; Gargov, G. (1989); Interval-valued intuitionistic fuzzy sets, *Fuzzy Sets and Systems*, 31(3), 343–349, 1989.

[3] Boran, F. E; Akay, D. (2014); A biparametric similarity measure on intuitionistic fuzzy sets with applications to pattern recognition, *Information Sciences*, 255, 45–57, 2014.

[4] Bozic, M.; Ducic, N.; Djordjevic, G.; Slavkovic, R. (2017); Optimization of Wheg Robot Running with Simulation of Neuro-Fuzzy Control, *International Journal of Simulation Modelling*, 16(1), 19–30, 2017.

[5] Chaira, T.; Panwar, A. (2014); An Atanassov's intuitionistic fuzzy kernel clustering for medical image segmentation, *International Journal of Computational Intelligence Systems*, 7(2), 360–370, 2014.

[6] Dalman H.; Gazel N.; Sivri M. (2016); A fuzzy set-based approach to multi-objective multi-item solid transportation problem under uncertainty, *International Journal of Fuzzy Systems*, 18(4), 716–729, 2016.

[7] Das, S.; Guha, D.; Dutta, B. (2016); Medical diagnosis with the aid of using fuzzy logic and intuitionistic fuzzy logic, *Applied Intelligence*, 45(3), 850–867, 2016.

[8] De S. K.; Biswas, R.; Roy, A. R. (2001); An application of intuitionistic fuzzy sets in medical diagnosis, *Fuzzy Sets & Systems*, 117(2), 209–213, 2001.

[9] Dubois, D.; Prade, H.; Esteva, F. (2015); Fuzzy set modelling in case-based reasoning, *International Journal of Intelligent Systems*, 13(4), 345–373, 2015.

[10] Gau, W. L.; Buehrer, D. J.(1993); Vague sets, *IEEE Transactions on Systems Man & Cybernetics*, 23(2), 610–614, 1993.

[11] Huang, H. (2016); New distance measure of single-valued neutrosophic sets and its application, *International Journal of Intelligent Systems*, 31(10), 1021–1032, 2016.

[12] Hung, W. L.; Yang, M. S. (2007); Similarity measures of intuitionistic fuzzy sets based on Lp metric, *International Journal of Approximate Reasoning*, 46(1), 120–136, 2007.

[13] Hwang, C. M.; Yang, M. S. (2013); New construction for similarity measures between intuitionistic fuzzy sets based on lower, upper and middle fuzzy sets, *International Journal of Fuzzy Systems*, 15(3), 371–378, 2013.

[14] Kar, A. K. (2015); A hybrid group decision support system for supplier selection using analytic hierarchy process, fuzzy set theory and neural network, *Journal of Computational Science*, 6,23–33, 2015

[15] Karaaslan, F. (2017); Correlation coefficients of single valued neutrosophic refined soft sets and their applications in clustering analysis, *Neural Computing & Applications*, 28(9), 2781–2793,2017.

[16] Le, H. S.; Phong, P. H. (2016); On the performance evaluation of intuitionistic vector similarity measures for medical diagnosis, *Journal of Intelligent & Fuzzy Systems*, 31(3), 1–12, 2016.

[17] Li D. F.; Cheng C. T. (2012); New similarity measures of intuitionistic fuzzy sets and application to pattern recognitions, *Pattern Recognition Letters*, 23(1), 221–225, 2002.

[18] Li D. F.; Ren H. P. (2015); Multi-attribute decision making method considering the amount and reliability of intuitionistic fuzzy information, *Journal of Intelligent & Fuzzy Systems*, 28(4), 1877–1883, 2015.

[19] Liu, P. D. (2016); The aggregation operators based on archimedean t-conorm and t-norm for single-valued neutrosophic numbers and their application to decision making, *International Journal of Fuzzy Systems*, 18(5), 1–15, 2016.

[20] Meng, F.; Chen, X. (2016); Entropy and similarity measure of AtanassovĄŻs intuitionistic fuzzy sets and their application to pattern recognition based on fuzzy measures, *Pattern Analysis and Applications*, 19(1), 11–20, 2016.

[21] Mousavi, M.; Yap, H. J.; Musa, S. N.; Dawal, S. Z. M. (2017); A Fuzzy Hybrid GA-PSO Algorithm for Multi-Objective AGV Scheduling in FMS, *International Journal of Simulation Modelling*, 16(1), 58–71, 2017.

[22] Nguyen, H. (2016); A novel similarity/dissimilarity measure for intuitionistic fuzzy sets and its application in pattern recognition, *Expert Systems with Applications*, 45, 97–107, 2016

[23] Peng, J. J.; Wang, J. Q.; Wang, J.; Zhang, H. Y; Chen, X. H. (2016); Simplified neutrosophic sets and their applications in multi-criteria group decision-making problems, *International Journal of Systems Science*, 47(10), 2342–2358, 2016.

[24] Perlibakas, V. (2004); Distance measures for PCA-based face recognition, *Pattern Recognition Letters*, 25(6), 711–724, 2004.

[25] Pramanik, S.; Pramanik, S.; Giri, B. C. (2015); TOPSIS method for multi-attribute group decision-making under single-valued neutrosophic environment, *Neural Computing & Applications*, 27(3), 727–737, 2015.

[26] Smarandache, F. (1999); A Unifying Field in Logics. Neutrosophy: Neutrosophic Probability, Set and Logic, *American Research Press*, 1999.

[27] Tavana, M.; Zareinejad, M.; Caprio, D. D.; Kavianic M. A. (2016); An integrated intuitionistic fuzzy AHP and SWOT method for outsourcing reverse logistics, *Applied Soft Computing*, 40, 544–557, 2016.

[28] Torra V. (2010); Hesitant fuzzy sets, *International Journal of Intelligent Systems*, 25(6), 529–539, 2010.

[29] Vasavi, C.; Kumar G. S.; Murty, M. S. N.(2016); Generalized differentiability and integrability for fuzzy set-valued functions on time scales, *Soft Computing*, 20(3), 1093–1104, 2016.

[30] Wang, H. B.; Smarandache, F.; Zhang Y. Q.; Sunderraman R. (2010); Single valued neutrosophic sets, *Multispace and Multistructure*, 4, 410–413, 2010.

[31] Xu, Z. S.; Chen, J. (2008); An overview of distance and similarity measures of intuitionistic fuzzy sets, *International Journal of Uncertainty, Fuzziness and Knowlege-Based Systems*, 16(4), 529–555, 2008.

[32] Xu Z. S.; Zhao N. (2016); Information fusion for intuitionistic fuzzy decision making: an overview, *Information Fusion*, 28, 10–23, 2016.

[33] Ye, J.(2015); The generalized Dice measures for multiple attribute decision making under simplified neutrosophic environments, *Intelligent* & *Fuzzy Systems*, 31(1), 663–671, 2015.

[34] Ye J.; Fu, J. (2016); Multi-period medical diagnosis method using a single valued neutrosophic similarity measure based on tangent function, *Comput Methods Programs Biomed*, 123, 142–149, 2016.

[35] Ye, J. (2017); Single-valued neutrosophic similarity measures based on cotangent function and their application in the fault diagnosis of steam turbine, *Soft Computing*, 21(3), 1–9, 2017.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 14(1), 90-106, February 2019.

A Fuzzy Group Decision-making Model for Determining the Most Influential Persons in the Sustainable Prevention of Accidents in the Construction SMEs

Z. Turskis, S. Dzitac, A. Stankiuviene, R. Šukys

Zenonas Turskis
Faculty of Civil Engineering
Vilnius Gediminas Technical University
Sauletekio al. 11, Vilnius, Lithuania, 10223
zenonas.turskis@vgtu.lt

Simona Dzitac
University of Oradea
Universitatii, 1, Oradea, Romania, 410087
simona@dzitac.ro

Aušra Stankiuviene*, Ritoldas Šukys
Faculty of Civil Engineering
Vilnius Gediminas Technical University
Sauletekio al. 11, Vilnius, Lithuania, 10223
*Corresponding author: ausra.stankiuviene@vgtu.lt
ritoldas.sukys@vgtu.lt

Abstract: Safety in the sustainable construction is a game with uncertain reality, with human's life. Managers in power, involved in the construction project life cycle, create rules for the game and are the main players. The project's success can be perceived differently by stakeholders. A company's competitiveness, successful implementation of the projects and effective safety management depends on the strategic allocation of its human resources in order to select a proper project team in alignment with employee capabilities. This requires detailed factors reflecting their role (in line with their interests and attitudes) and knowing which of them has a decisive influence on the successful implementation of the project. The real data describing key factors in this concept can be provided as uncertain conditions. This work presents a novel integrated modified fuzzy group decision-making approach to select and rank the most influential persons ensuring the sustainable prevention of accidents at work in the small and medium-sized construction enterprises. This model includes the Delphi method and fuzzy extensions of Eckenrode's criteria rating method. The proposed model could be expanded in order to select the most suitable individuals for sustainable management of safety and, moreover, for the effective implementation of safety and health measures.
Keywords: construction site, accidents prevention, occupational safety, risk management, fuzzy, rating, group, MCDM.

1 Introduction

Sustainability is defined as a condition of equilibrium, stability, and interoperability, enabling people to meet their needs and support ecosystems [42]. The application of sustainability principles in any activity, including the construction, is becoming increasingly important for the development in terms of economic, social effectiveness, safety, and sustainability [1]. The uncertain challenges affect planning and practices [17]. A growth in the construction industry creates a demand for suitable materials and techniques [70]. The construction integrates the achievements of the various fields of science, and proper safety management is inseparable [27]. In the modern world buildings and infrastructures are becoming increasingly complex [47]. The construction sustainability, safety culture and zero accidents at work aspiration are challenges of crucial importance for modern society. Knowledge and education, to enhance understanding and adaptation of these paradigms, are fundamental to solving these dilemmas [6, 19] widespread in organizations even nowadays when many organizations work on several complex projects at the same time [47]. Project leaders control many risks and disadvantages during of construction works and play different roles in safety management in the construction process. A successful implementation of sustainable projects and effective implementation of safety management are the characteristics of a company's competitiveness in the market [74] and reduce environmental, social and economic losses [31]. Consequently, effectiveness of safety management has become a core topic for researches.

The construction industry is one of the most dangerous sectors of the economy. During the last decade, the number of accidents at work has increased. The small and medium-sized enterprises (SMEs) have a higher accidents risk compared to large construction enterprises [25, 53]. The increased number of accidents at work in the construction impacts on a sustainable working environment. A reduction of accidents at work, as a priority, must be included in the action plans of construction projects [41]. Hasle et al. [23] has investigated SMEs accidents causation criterion and prevention actions, and concluded that the owners of SMEs may deal with safety and health issues quite differently compared to those in the large construction enterprises.

Many factors influencing accidents at work in the construction are still largely unexplained. In general, accidents at the construction sites could be qualified as defects of the safety management system which occur due to the number of aspects, including technical, technological, organisational and other types of factors which affect human safety performance and working environment [33, 34, 52]. Indrees et al. [26] presented that psychological factors such as an attitude of workers, a workload, organizational relationships, a clarity of roles, a mental stress, job insecurity, and job satisfaction strongly affect workers' safety. Scientists and practitioners mainly concentrated on a risk assessment process [4, 56, 59].

Dozens of researchers concentrate on reactive (accident frequency and the seriousness index) [53] and proactive indicators (organisational, technical and behavioural), and each of them has their strengths and weaknesses for the safety performance [24,36,39,43,45,60]. Yau [66] and Zhou et al. [76] presented that the development of a safety culture significantly reduced the number of accidents. After reviewing the literature, it can be concluded that most of the accidents at work in SMEs construction sites are essentially related to negligent management or insufficient employee safety awareness [67]. Management problems are mainly due to inappropriate management skills, lack of opportunities to implement safety and health strategies, non-compliance with labour safety laws and regulations, and the inappropriate use of security measures [33, 58].

According to Pasman et al. [46], a prediction what could happen under specified conditions is the most elementary step to the better safety.

Meanwhile, it is evident that accidents at work in the construction is a result of behaviour of stakeholders (owners, developer, project leaders, contractors, subcontractors, occupational safety and health (OSH) specialists, OSH coordinators). Therefore, the stakeholders must understand principles of the sustainable construction and the impact of safety. The selection, preparation, and implementation of the right strategies for effective prevention of accidents is the biggest challenge for the most enterprises [18]. Therefore, it is very important to know which is the most influential person and which person can continuously improve the level of safety, ensure a sustainable and proper prevention of accident at work in construction.

Until now, there is a lack of research on who the main stakeholders on onsite construction are and how they influence the effective safety on SME's construction site. This research examines the understanding of the role of stakeholders and their importance in ensuring the management of accident prevention in projects. Therefore, all stakeholders need to be identified, categorized and their concerns, interests and influence need to be analysed and prioritized using group fuzzy model.

The authors consider that this work presents a useful approach to the accidents prevention in the construction, and suggest a novel integrated group fuzzy modified approach to determine the most influential person in the prevention of accidents at work in the SME's. This model integrates the Delphi technique and a modified fuzzy extension of Eckenrode's rating method. The presented model will be expanded and applied for the sustainable construction project development and the ranking of the main persons for preparation and implementation of a proper safety prevention projects.

2 Fuzzy MCDM safety management problems' solution models in construction site

A successful project is considered when all risks are properly managed in all project stages. The selection of proper techniques and strategies depends on all players involved in construction project implementation. The project's owner, project manager, project OSH coordinator, contractor, subcontractor, OSH coordinator, OSH specialist are examples of stakeholders that are influential persons in accidents prevention during the design and organization of the project. Stakeholders must constantly seek and promote their mutual interests in reducing any risks and improving their management skills. It should be emphasized that all stakeholders involved in construction project need to be aware of the risks they are facing and how they apply risk management in order to create better working conditions.

Safety in the sustainable construction is a stakeholders' game against nature, environment, and with human's life. The stakeholders and project participants play significant role in this game dealing with problems. The owner or client of the project is the person concerned and has overall responsibility for project management. It is often related to the financial part; while other organizations or individuals interested in the project have contractual relations with the project owner. However, the project owner can transfer management responsibility to other stakeholders in the project. But it is not exempted from liability for the safety and health of workers in accordance with established legislation. The project owner (client) or project manager designates one or more OSH coordinators at the site who coordinate the preparation of a safety and health plan and ensures that site workers apply all the preventive measures they need to take on their safety and health site. In addition, employers or their designated OSH specialist must comply with the minimum safety and health requirements applicable at construction sites and must take into account the directions of the OSH coordinator.

It is obvious that stakeholders, participating in project management and accident prevention, must constantly seek and promote their mutual interests in reducing any risks and improving their management skills. It is important that the project clearly defined the role and influence of stakeholders in preventing accidents. In addition, stakeholders need to be equally guided by the need to manage the threats, opportunities and uncertainties associated with developmental achievements and commit to reaching the desired goal in the relationship. There are direct and indirect reciprocal relations between the participants and the organization or project; because everyone has the ability to influence and can affect other activities. Due to the variety of different stakeholders involved in the project, the risk of conflicts increases, and this can increase the risk associated with the construction project and may affect the successful completion of the project. In the implementation of the project, the co-operation plan of stakeholders can be considered

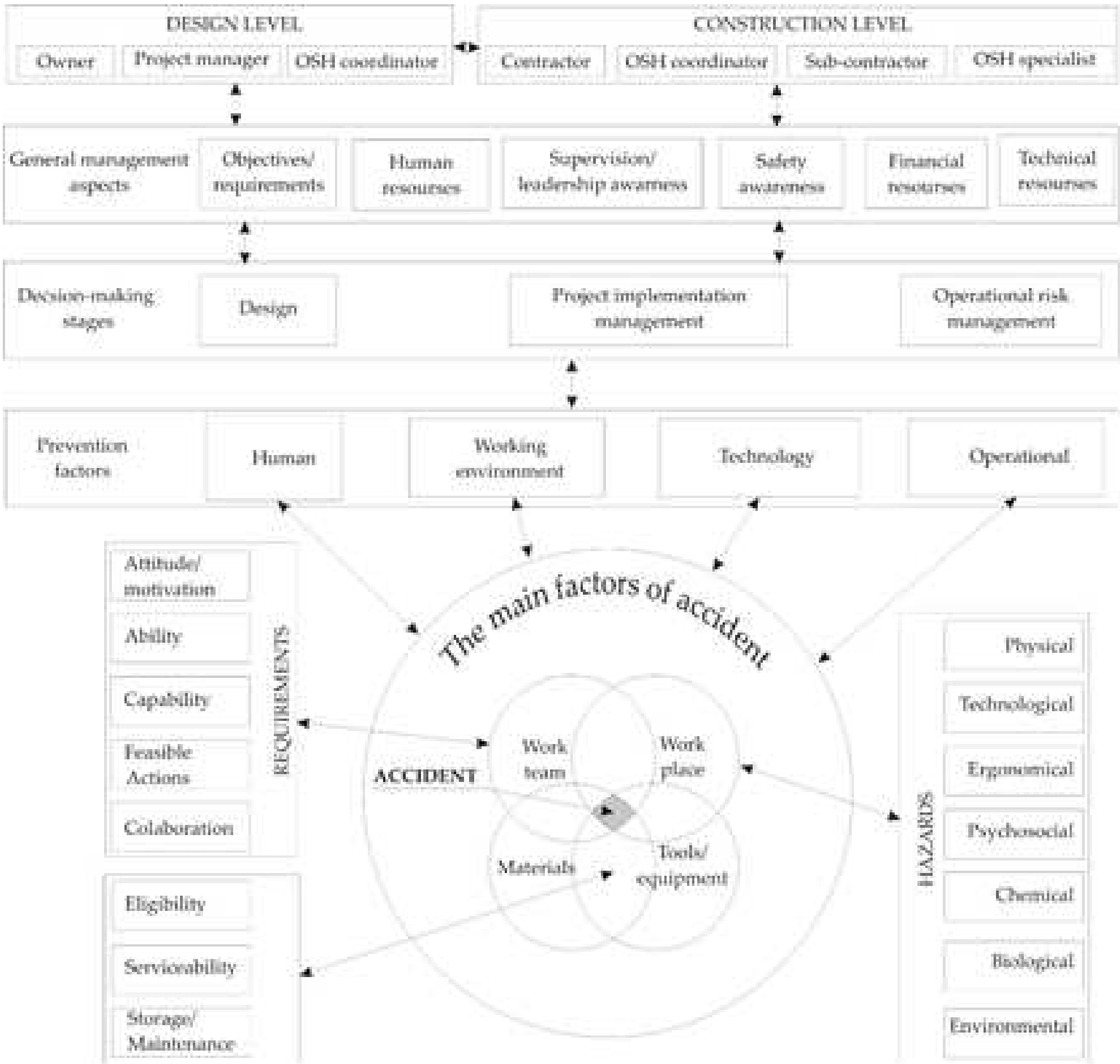


Figure 1: The main influencing factors of accident at work

an important part of overall risk management. The stakeholder input may be beneficial to the project, or, conversely, may jeopardize the success of the project. An assessment of the role of project stakeholders at the initial design stage and the ability to manage unpredictable responses throughout the construction phase are crucial to the success of the project.

The solutions are not completely known, more than half of them are uncertain and insecure. The uncertainty and insecurity arise from many rules which has permanently ordained, and it depends on how many approaches are caused by mental factors, while the solution's implementation in practice without any limitations becomes possible. While the accidents at work cannot be exactly predicted, in general, the solution becomes possible only when risk management is carried out under the current real-life conditions. The problem of measuring according to some criteria in construction is known as multi-criteria decision-making (MCDM) [75]. Whereas there is no method which is the most suitable for all decision-making situations, many MCDM models and their extensions have been developed [40]. A number of authors promote their own MCDM methods and models. However, there is a lack of methods and models of evaluating the efficiency of personnel. MCDM methods in the construction can be used on the national, organizational and project levels. Even so, the most assessment methods are trying to find out how to make the most economic construction decisions, and mostly all these decisions are only intended for economic objectives [55]. Therefore, MCDM approaches are very useful to solve different management problems.

A selection of safe and effective management systems in the construction is a complex MCDM task. The Analytic Hierarchy Process (AHP) [49] is the most widely used MCDM method. A lot of studies have been made on MCDM methods and applications. Usually the AHP is applied to calculate significance of criteria. Šaparauskas et al. [51] assessed and prioritised the relative importance of various criteria based on the Saaty's Analytic Hierarchy Process (AHP). A multiplicative exponential weighting (MEW) was applied in order to calculate the best performed project [67]. Keshavarz et al. [29] introduced a new method of Evaluation based on Distance from Average Solution (EDAS) for multi-criteria inventory classification problems. Zavadskas et al. [71] merged two different multi-criteria decision-making (MCDM) methods and developed an original Weighted Aggregated Sum Product assessment method (WASPAS), while the original combination of three different MCDM methods was introduced as a new Multiple Objective Optimisation on the basis of Ratio Analysis Plus Full Multiplicative Form (MULTIMOORA) method.

The similar problems could be solved using different methods. Zavadskas et al. [72] applied three different hybridised well-known MCDM techniques to rank and assess different technological alternatives in construction: Step-Wise Ratio Assessment - Technique for Order of Preference by Similarity to Ideal Solution (SWARA-TOPSIS), ELimination and Choice Expressing the REality (Et Choix Traduisant la REalité (SWARA-ELECTRE III Elimination)), Multicriteria Optimization and Compromise Solution (VIsekriterijumska optimizacija i KOmpromisno Resenje in Serbian language (SWARA-VIKOR)). Recently Turskis and Juodagalvienė [61] presented a novel peculiar model, which is based on ten different multi-criteria decision-making methods: Game Theory, AHP, Simple Additive Weighting (SAW), Multiplicative Exponential Weighting (MEW), TOPSIS, EDAS, Additive Ratio Assessment (ARAS), Full Multiplicative form, Laplace Rule, and Bayes Rule.

A philosophy and logic are the basis of mathematical models of real-life issues. In 1657 Christopher Wren in his chair of astronomy at Gresham College said that mathematical models could be used when dealing with uncertainty [65]. Most of the construction safety management problems includes vague and uncertain values of criteria. The fuzzy set theory provides a decision framework to incorporate imprecise judgments inherent in the personnel selection process. In 1965 Zadeh [69] introduced the fuzzy set theory to deal with inaccurate and uncertain data, even in situations when information is based on subjective evaluations and defined in the lexical conditions. Zadeh used fuzzy set theory to a gradual transition from one class to another in the development of decision support models. It was later applied in MCDM applications [8, 32]. Risk management decision determined by various fuzzy criteria and used by practitioners and supervisors in the construction sector [10]. The fuzzy set theory and fuzzy extensions of multi-criteria decision-making models play principal roles in the risks' weighting, responses and choices of means to manage them [15, 20].

The group decision-making processes are necessary to design and evaluate a set of different alternatives. One of the most important task is to reject those alternatives that do not meet lower bounds of the significant criteria values. For a long time, a rigorous agreement was seen as a final group's opinion. In the most cases, a group of experts who make real-life decisions have no strict and steady opinion about the same criteria and alternatives. An agreement of the group reached when the most dominant players agree with the criteria ratings and performances of the considered alternatives. Real-life problems' modelling and solution lead the group of decision-makers to situations when models are based on vague logic. In addition, most often the models based on the criteria rating in words. Such type of ratings cannot be replaced by the strict (crisp) numerical values. Fuzzy set theory allows decision makers to use incomplete or partially obtained information into the problem solving model [62]. A fuzzy set is characterized by a membership (characteristic) function which assigns to each object a grade of membership ranging [69]. Various types of membership functions are available. In this research the most commonly used triangular membership function is used [14]. A fuzzy triangular number will be denoted as *(α, β, γ) (α* - lower value of fuzzy number*, β* - modal value of fuzzy number*, γ* - upper value of fuzzy number). Van Laarhoven and Pedrycz [63] introduced the basic operations of fuzzy triangular number.

Many researchers investigated such problems and developed different fuzzy models to solve problems of safety and risk management in the construction. Debnath et al. [12] developed a fuzzy inference model for assessing occupational risks in the construction sites. Grassi et al. [21] presented a fuzzy multi-criteria model for risk evaluation in the workplaces. Liu and Tsai [38] used the fuzzy ANP analysis to identify and assess the underlying hazards and causes.

3 An integrated Delphic-Eckenrode's Likert-type scale-based fuzzy rating

It is very important to identify the importance of the activities of the different process managers before starting to assess the key challenges of workplace safety's management, efficiency level of safety solutions and quality improvement. To order to achieve this, experts can use weighting methods for criteria. There are a lot of different subjective approaches for assessing weights [44]: AHP [49], ANP [50] expert method [73], SWARA [28, 30], FARE (FActor RElationship) [21], etc.

Nominal group technique Delphi [13, 37] is a useful tool solving complicated problems which needs expert data. This is a group decision-making process and includes idea generation, problem description, data assessment, and generation of feasible alternatives.

Five types of measurement scales (nominal, ratio, ordinal, interval, or fuzzy) commonly could be used to measure criteria values and importance [64]. Lot of scales evaluating importance of criteria are proposed in scientific papers. For example, rating of perceived exertion Borg scale [7] is a frequently used quantitative measure of perceived exertion (rated exertion on a scale of 6-20). A data is expected to fit into one or another of the absolute categories. In any case, decision-makers must apply methods appropriate to their data and to the questions they wish to answer. Likert scales were introduced by Likert in 1930s [35] as a tool for the measurement and assessment of attitudes. Since then Likert-type scales have become more popular in many fields of decision-making, including management [57]. It is the most widely used approach to scale responses in survey research. The reason for this is that the Likert scale is a very simple tool to use and can be analysed effectively as interval or fuzzy scales [1]. The Likert-type scale is often used interchangeably with rating scale, even though these two ones are not synonymous.

In 1965 Eckenrode [16] presented seminal work on criteria weights elicitation. The six methods were compared for their reliability and efficiency in collecting the judgment data: Ranking [2, 3]; Rating [35]; The Three Paired Comparisons Methods: Partial Paired Comparisons I Buel [9]; Partial Paired Comparisons II; Complete Paired Comparisons [5]; Successive Comparisons Churchman et al. [11, 50].

Eckenrode [16] found no significant differences among the techniques. The values calculated by all of methods correlates. Rating is effective for personal assessment, and it's especially effective for group decision making. It works well because it forces expert to get clarity on his own criteria and create a shared set of criteria. Eckenrode's Rating method is selected and modified by applying basics of fuzzy sets theory in this study.

The problem solution flow chart is presented in Figure 2.

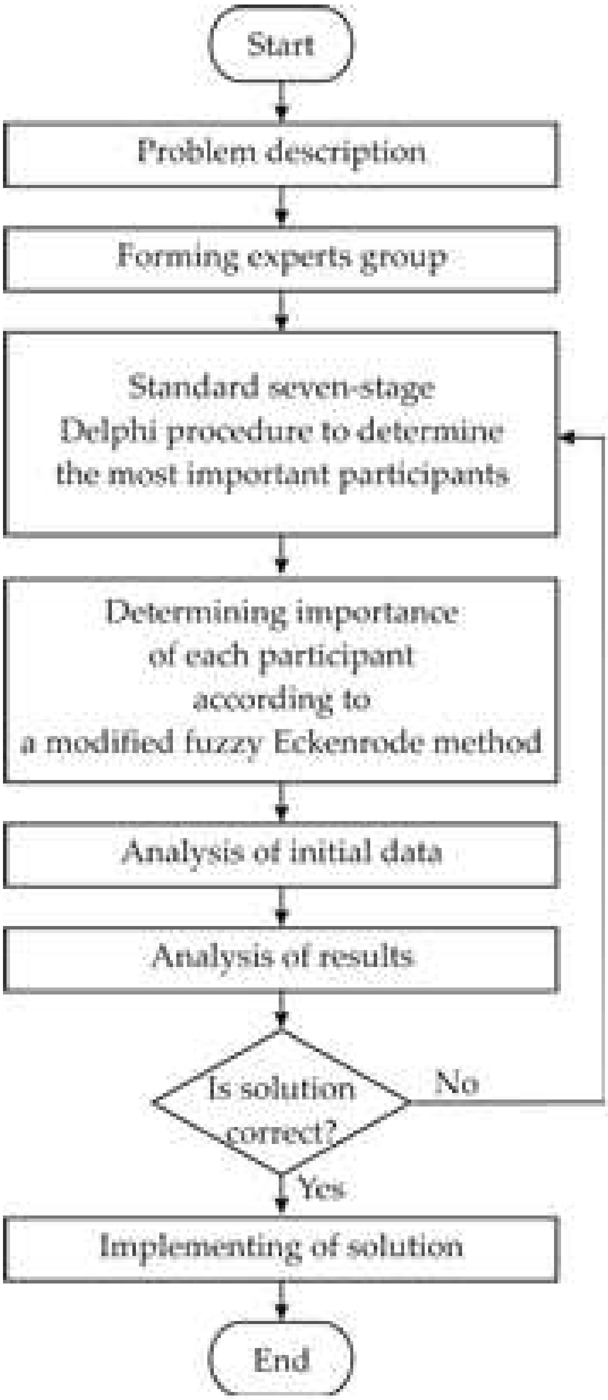


Figure 2: Proposed framework of the study

At the beginning of the problem solution experts determined the main participants (stakeholders) influencing accidents prevention in the construction. Each participant differently affects accidents prevention processes, and therefore has different importance in these problems solving. Delphi method selects the most important stakeholders of the problem under consideration. A

Table 1: Likert-type scale to determine participant's importance

(TFN = Triangular Fuzzy Number)

Importance abbreviation	Value	Importance level	TFN α	TFN γ	TFN β
L	1	Little	0.5	1	1.5
M	2	Medium	1	1.5	2
I	3	Important	1.5	2	2.5
V	4	Very	2	2.5	3
E	5	Extreme	2.5	3	3.5

team of experts determined a set of the most important participants (Table 1). The Likert-type scale (important, irrelevant, ...) made below for the significance of the participant's importance (Fig. 3).

Rating: The raw rating assigned by the judge to each criterion against the scale of 0 to 5 (5 most valuable or important) treated as follows (Table 2 and 3):

$$w_{cj} = p_{cj} / \sum_{c=1}^{m} p_{cj} \tag{1}$$

where,
w_{cj} - weight computed for criterion c from the rating given by judge j p_{cj} - rating given by judge j to criterion c, and w_c is calculated as follows:

$$w_c = \sum_{j=1}^{n} w_{cj} / \sum_{j=1}^{n} \sum_{c=1}^{m} w_{cj} \tag{2}$$

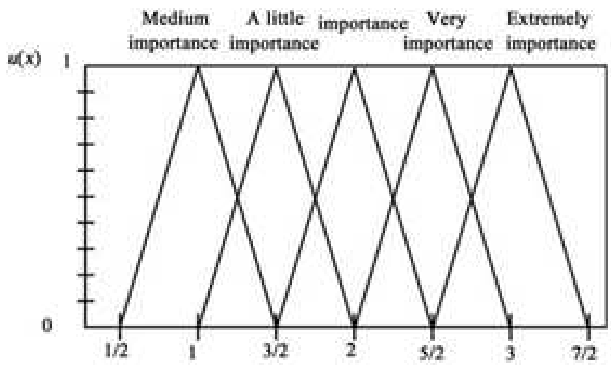


Figure 3: Likert-type scale to determine participant's importance

The equations (1) and (2) using rules of fuzzy arithmetic are modified as follows:

$$\tilde{w}_{cj} = \tilde{p}_{cj} / \sum_{c=1}^{m} \tilde{p}_{cj} = \left(p_{c\alpha j} / \sum_{c=1}^{m} p_{c\gamma j};\ p_{c\beta j} / \sum_{c=1}^{m} p_{c\beta j};\ p_{c\gamma j} / \sum_{c=1}^{m} p_{c\alpha j} \right); \tag{3}$$

and

$$\tilde{w}_c = (w_{c\alpha};\ w_{c\beta};\ w_{c\gamma}) = \sum_{j=1}^{n} \tilde{w}_{cj} / \sum_{j=1}^{n} \sum_{c=1}^{m} \tilde{w}_{cj} =$$
$$\left(\sum_{j=1}^{n} w_{c\alpha j} / \sum_{j=1}^{n} \sum_{c=1}^{m} w_{c\gamma j};\ \sum_{j=1}^{n} w_{c\beta j} / \sum_{j=1}^{n} \sum_{c=1}^{m} w_{c\beta j};\ \sum_{j=1}^{n} w_{c\gamma j} / \sum_{j=1}^{n} \sum_{c=1}^{m} w_{c\alpha j} \right) \tag{4}$$

where,

$w_{jl} - \min_k y_{jk},\ j = \overline{1,n},\ k = \overline{1,\ p}$ is minimum possible value, $w_{j\alpha} = \left(\prod_{k=1}^{p} y_{jk}\right)^{\frac{1}{p}},\ j = \overline{1,\ n},\ k = \overline{1,p}$ is the most possible value and $w_{j\beta} = \max_k y_{jk}, j = \overline{1,n},\ k = \overline{1,p}$ is the maximal possible value of j-th criterion $w_{j\gamma} = \min_k y_{jk},\ j = \overline{1,n},\ k = \overline{1,p}$.

A defuzzification should be applied before final decisions are made. The defuzzification is a process of producing a quantifiable result in crisp logic, given fuzzy logic, and corresponding membership degrees. A common and useful defuzzification technique is a centre of gravity. This method is selected in the case study.

$$w_c = \frac{1}{3} \left(w_{c\alpha} +\ w_{c\beta} +\ w_{c\gamma} \right). \tag{5}$$

3.1 Problem solution: Fuzzy group multi-criteria method in assessing the most influential person in sustainable prevention of accidents at work

In order to ensure sustainable prevention of accidents at work in the construction, stakeholders should use an expert advice which will allow to select the effective and rational alternatives. To achieve the above mentioned goals, an integrated determination method of criteria significance depending on their characteristics, is developed. The problem could be solved based on the survey of experts' data. All of experts have university degree in a civil engineering and actively work with accidents prevention. At first step there was formed a team of eighteen OSH specialists, OSH coordinators, project contractors and subcontractors from different construction companies which employs less 50 people. All of them were male. Respondents were 26 – 55 years old (26-30 – 6; 31-35 – 4; 36-40 – 4; 41-45 – 2; 46-50 – 1; 51-55 – 1) and their work experience in the construction was 2 – 25 years (2-5 – 2; 6-10 – 5; 11-15 – 6; 16-20 – 1; 21-25 – 4).

The standard seven-stage Delphi procedure is applied in the case study. Firstly, a facilitator explains to the participants the purpose and procedure of the problem. Secondly, members of the group silently explain their opinion about the solution (criteria), with a short explanation in written not consulting or discussing their ideas with other participants. This ensures that all participants get an opportunity to make an equal contribution. Thirdly, a facilitator encourages a sharing and a discussion of reasons for the choices made (criteria) by each group member to identify a common ground. Fourthly, participants verbally explain in details all presented ideas which are not clear for all participants of the groups or further details about any of the ideas that colleagues have produced and which may not be clear to them. Fifthly, a facilitator eliminates duplicate solutions (criteria) from the list of all solutions, and the members proceed to rank the solutions starting from the most important to the least important. Sixthly, a facilitator involves a prioritizing procedure of the recorded ideas in relation to the original problem. Following the voting and ranking process, a facilitator asks some questions to participants who have a different opinion about ranks from average alternatives (criteria) ranking. Seventhly, a final ranking and rating of alternatives (criteria) should be done.

The participants of experts' group, based on their experience of accidents at work prevention in SME's enterprises, generated as many ideas as they can. The ideas were evaluated against all consequences and sequels factors. Then experts ranked and rated the influential persons in

Table 2: Participants' importance (PI) lexical evaluation based on Likert-type scale

PI	Experts																	
	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}	E_{13}	E_{14}	E_{15}	E_{16}	E_{17}	E_{18}
P_1	E	M	V	I	M	M	L	M	M	M	E	I	E	I	V	M	M	L
P_2	M	L	L	I	M	I	M	V	L	M	M	V	I	V	L	I	L	M
P_3	E	I	V	E	E	E	V	E	I	V	E	E	E	E	V	V	E	I
P_4	V	V	V	E	E	E	I	E	E	I	E	E	E	E	I	V	I	E
P_5	I	E	E	E	E	E	E	E	V	E	E	E	E	E	M	I	V	V
P_6	I	M	L	L	E	I	I	V	V	V	V	I	E	E	V	E	E	E

Table 3: Participants' importance level expressed by triangular fuzzy numbers corresponding to the linguistic scale

	E_1			E_2				E_{15}				E_{18}		
	α	β	γ	α	β	γ	...	α	β	γ	...	α	β	γ
P_1	2.5	3	3.5	1	1.5	2	...	2	2.5	3	...	0.5	1	1.5
P_2	1	1.5	2	0.5	1	1.5	...	0.5	1	1.5	...	1	1.5	2
P_3	2.5	3	3.5	1.5	2	2.5	...	2	2.5	3	...	1.5	2	2.5
P_4	2	2.5	3	2	2.5	3	...	1.5	2	2.5	...	2.5	3	3.5
P_5	1.5	2	2.5	2.5	3	3.5	...	1	1.5	2	...	2	2.5	3
P_6	1.5	2	2.5	1	1.5	2	...	2	2.5	3	...	2.5	3	3.5

prevention of accidents at work. Later, the resulting data were analysed and the list of the main participants such as a developer (P_1), a project leader (P_2), a contractor (P_3), OSH specialist (P_4), OSH coordinator (P_5) and a subcontractor (P_6) was made. Then the experts were requested to rate the main participants according to linguistic significance scale. Finally, linguistic variables and converted to fuzzy numbers and ranks of influential persons determined.

Fuzzy participants' importance values defuzzified as follows (Figure 4):

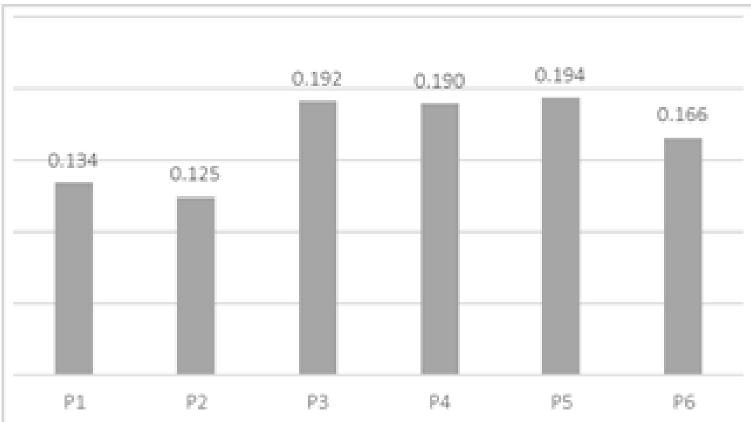


Figure 4: Final assessment of the influential person in sustainable prevention of accidents at work in construction

The last stage is a calculation of a relative importance index (RI) of each considered participant (Figure 5):

$$RI_c = \frac{w_c}{\max_c w_c} \tag{6}$$

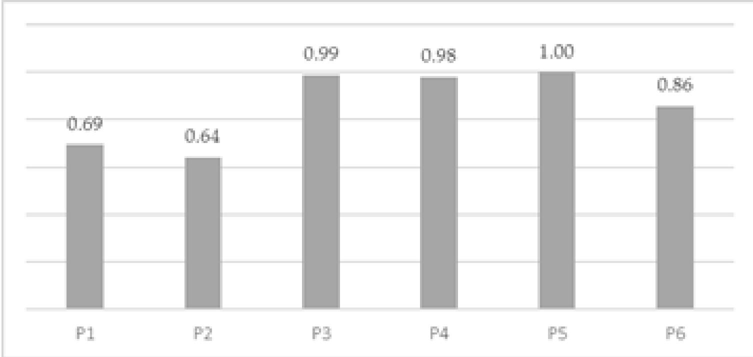


Figure 5: RI-relative importance of the influential person in sustainable prevention of accidents at work in the construction

4 Results

Key stakeholders, including the developer, the project leader, the OSH specialist, the OSH coordinator, the contractor, and the subcontractor, are the most influential persons who can prevent accidents at work in the SMEs construction sites.

The results of the investigation show that these stakeholders could have a positive impact on the sustainable prevention of accidents at work for SMEs. The importance of criteria has influence on their values in the multi-attribute utility function.

Besides, it is necessary to define values of quantitative and qualitative criteria (or people). The data identifying the most influential persons in the prevention of accidents at work are shown in Figure 5. According to the shown values of the relative importance RI, the main participants are divided into three groups. The most influential persons are the OSH coordinator (P_5), the contractor (P_3) and the OSH specialist (P_4) whose RI rates are the highest 1.00, 0.99 and 0.98 respectively. It means that the OSH coordinator, the OSH specialist and the contractor play a special role in the construction process, including n the areas of preventing an accident at work. Another influential person is the subcontractor (P_6) at level 0.86. He represents the second most influential group of persons. At the third importance level is the client (P_1) and the project leader (P_2) with the lowest RI levels – 0.69 and 0.64 respectively.

The above results show that leadership positions can have a positive impact on sustainable safety practices in the SME's construction enterprises. Managers such as the OSH coordinator, the OSH specialist and the contractor can also help to realise the targeted strategy of collectivism and consumer communication between a contractor and subcontractors.

5 Conclusions

The successful implementation of a sustainable modern construction project in a real environment is influenced by many complex procedural, social, economic, political and technological factors and affects various groups of people differently.

Therefore, the best solutions can be achieved by applying scientific methods involving a large amount of information and calculations. An important problem for each country is prevention of accidents at work in the SME's which dominate in the construction market.

Introducing necessary precautions on the construction sites prior to construction begins with an aim to ensure a safe and healthy environment is one of key tasks of building project leaders (stakeholders). Each participant of the construction process plays his own and specific role in the execution process of construction projects and stands against various risks.

Their different attitudes, rights, abilities, and responsibilities in the prevention of accidents at work depends on education, work experience in management of construction projects, skills and knowledge of specific approaches to prevent accidents. Besides education, an ability to cope with risk is very important factors.

The developer, the project leader, the OSH specialist, the OSH coordinator, the contractor, and the subcontractor are the most influential persons who have the ability to an effective sustainable prevention of accidents at work in the construction.

The initial data in similar group decision-making approaches are presented by experts in words. Each of experts has own opinion about criteria values. The significance of expert estimations was assessed with the help of the modified fuzzy group Eckenrode's rating method. The proposed method is superior to conventional techniques because the proposed method has the capability to make group decisions in fuzzy environment. Therefore, fuzzy set theory is a powerful tool to solve such problems.

It is very important to know which of the project life cycle 's participants, depending on their position in the decision hierarchy, is the most influential on direct decisions, or on the choice of project management groups which can most effectively implement projects and supervise these processes.

It is also necessary to determine the most effective strategy for a sustainable development.

When designing a project team, the size, complexity, duration, type of the project and specific requirements for the stakeholders should be evaluated.

The research shows that the most influential persons in a sustainable prevention of accidents at work in the SME's construction enterprises rank as follows: the OSH coordinator, the contractor, the OSH specialist (rates from 1 to 0.98), the subcontractor (rates as 0.86), the client, and the project leader (rates 0.69 and 0.64 respectively). The model presented in this study is suitable for identifying and selecting key players and their influence or for determining the importance of criteria in the multi-criteria utility function. The results and findings will be expanded and used by future researchers.

Bibliography

[1] Allen, D. E.; McAleer, M., Singh, A. K. (2017); Risk measurement and risk modelling using applications of Vine copulas, *Sustainability*, 9(10), 1762, 2017.

[2] Barrett, M. (1914; . A comparison of the Order of Merit method and the method of Paired Comparisons. *Psychological Review*, 21(4), 278–294, 1914.

[3] Bartlett, C. J.; Heermann, E.; Rettig, S. (1960); A comparison of six different scaling techniques. *The Journal of Social Psychology*, 51(2), 343–348, 1960.

[4] Bejinariu, C.; Darabont, D. C.; Baciu, E. R.; Georgescu, I. S.; Bernevig-Sava, M. A.; Baciu, C. (2017); Considerations on Applying the Method for Assessing the Level of Safety at Work. *Sustainability*, 9(7), 1263, 2017.

[5] Bock, R.D.; Jones, L.V. (1963); *The Measurement and Prediction of Judgemental Response: Statistical Methods. U. S. Army Test and Evaluation Command Project 7-S-0901-OIE*. Psychometric Laboratory, University of North Carolina, 1963.

[6] Bonenberg, W.; Kapliński, O. (2018); The Architect and the Paradigms of Sustainable Development: A Review of Dilemmas. *Sustainability*, 10(1), 100, 2018.

[7] Borg, G. (1970); Perceived exertion as an indicator of somatic stress. *Scandinavian Journal of Rehabilitation Medicine*, 2(2), 92–98, 1970.

[8] Buckley, J. J. (1984); The multiple judge, multiple criteria ranking problem: a fuzzy set approach. *Fuzzy Sets Syst*, 13, 25–37, 1984.

[9] Buel, W.D. (1960); A simplification of Hay's method of recording paired comparisons. *Journal of Applied Psychology*, 44(5), 347–348, 1960.

[10] Chatterjee, K.; Pamucar, D.; Zavadskas, E. K. (2018); Evaluating the performance of suppliers based on using the R'AMATEL-MAIRCA method for green supply chain implementation in electronics industry *Journal of Cleaner Production*, 184, 101–129, 2018.

[11] Churchman, C. W.; Ackoff, R.L.; Arnoff, E.L. (1957); *Introduction to Operations Research*. New York: Wiley. 1957.

[12] Debnath, J.; Biswas, A.; Sivan, P.; Sen, K.N.; Sahu, S. (2016); Fuzzy inference model for assessing occupational risks in construction sites, *International Journal of Industrial Ergonomics*, 55, 114–128, 2016.

[13] Delbecq, A.L.; VandeVen, A.H. (1971); A Group Process Model for Problem Identification and Program Planning. *Journal of Applied Behavioral Science*, 7, 466–491, 1971.

[14] Dubois, D.; Prade, H. (1978); Operations on fuzzy numbers. *International Journal of Systems Science*, 9(6), 613–626, 1978.

[15] Dzitac, I.; Filip, F.G.; Manolescu, M.J.(2017); Fuzzy Logic Is Not Fuzzy: World-renowned Computer Scientist Lotfi A. Zadeh, *International Journal of Computers Communications & Control*, 12(6), 748-789, 2017.

[16] Eckenrode, R.T. (1965); Weighting multiple criteria. *Management Science*, 12(3), 180–192, 1965.

[17] Eizenberg, E.; Jabareen, Y. (2017); Social sustainability: A new conceptual framework. *Sustainability*, 9(1), 68, 2017.

[18] Finkbeiner, M.; Schau, E. M.; Lehmann, A.; Traverso, M. (2010); Towards life cycle sustainability assessment. *Sustainability*, 2, 3309–3322, 2010.

[19] Gerbec, M. (2017); Safety change management–A new method for integrated management of organizational and technical changes, *Safety Science*, 100, 225–234, 2017.

[20] Ghassemi, A.; Darvishpour, A. (2018); A novel approach for risk evaluation and risk response planning in a geothermal drilling project using DEMATEL and fuzzy ANP. *Decision Science Letters*, 7(3), 225–242, 2018.

[21] Ginevicius, R. (2011); A new determining method for the criteria weights in multicriteria evaluation, International *Journal of Information Technology & Decision Making*, 10(6), 1067–1095, 2011.

[22] Grassi, A.; Gamberini, R.; Mora, C.; Rimini, B. (2009); A fuzzy multi-attribute model for risk evaluation in workplaces. *Safety Science*, 47(5), 707–716, 2009.

[23] Hasle, P.; Kines, P.; Andersen, L.P. (2009); Small enterprise owners' accident causation, attribution and prevention. *Safety Science*, 47(1), 9–19, 2009.

[24] Hinze, J.; Thurman, S.; Wehle, A. (2013); Leading indicators of construction safety performance. *Safety Science*, 51(1), 23–28, 2013.

[25] Hoła, B.; Szóstak, M. (2017); The methodology of analysing the accident rate in the construction industry. In *Procedia Engineering, Proceedings of Modern Building Materials, Structures and Techniques (MBMST)* 2016, Vilnius, Lithuania; Juozapaitis A., Daniūnas A., Zavadskas E. K., 172, 355–362, 2017.

[26] Idrees, M. D.; Hafeez, M.; Kim, J.Y. (2017); Workers' age and the impact of psychological factors on the perception of safety at construction sites, *Sustainability*, 9(5), 745, 2017.

[27] Jiang, W.; Zhao, X.; Ni, J. (2017); The impact of transformational leadership on employee sustainable performance: The mediating role of organizational citizenship behavior. *Sustainability*, 9(9), 1567, 2017.

[28] Keršuliene, V.; Zavadskas, E.K.; Turskis, Z. (2010); Selection of rational dispute resolution method by applying new step-wise weight assessment ratio analysis (SWARA). *Journal of business economics and management*, 11(2), 243–258, 2010.

[29] Keshavarz Ghorabaee, M.; Zavadskas, E. K.; Olfat, L.; Turskis, Z. (2015); Multi-criteria inventory classification using a new method of evaluation based on distance from average solution (EDAS). *Informatica*, 26(3), 435–451, 2015.

[30] Keshavarz-Ghorabaee, M.; Amiri, M.; Zavadskas, E.K.; Turskis, Z.; Antucheviciene, J. (2018); An Extended Step-Wise Weight Assessment Ratio Analysis with Symmetric Interval Type-2 Fuzzy Sets for Determining the Subjective Weights of Criteria in Multi-Criteria Decision-Making Problems. *Symmetry*, 10, 91, 2018.

[31] Khasreen, M. M.; Banfill, P.F.; Menzies, G.F. (2009); Life-cycle assessment and the environmental impact of buildings: a review. *Sustainability*, 1(3), 674–701, 2009.

[32] Leberling, H. (1981); On finding compromise solution in multi-criteria problems using the fuzzy min-operator. *Fuzzy Sets Syst*, 6, 105–110, 1981.

[33] Leung, M.; Chan, I.Y.S.; Yu, J. (2012); Preventing construction worker injury incidents through the management of personal stress and organizational stressors. *Accident Analysis and Prevention*, 48, 156–166, 2012.

[34] Liao, P.C.; Liu, B.; Wang, Y.; Wang, X.; Ganbat, T. (2017); Work paradigm as a moderator between cognitive factors and behaviors–A comparison of mechanical and rebar workers. *KSCE Journal of Civil Engineering*, 21(7), 272814–2525, 2017.

[35] Likert, R. (1932); A technique for the measurement of attitudes. *Archives of Psychology*, 140(22), 1–55, 1932.

[36] Lingard, H.; Wakefield, R.; Cashin, P. (2011); The development and testing of a hierarchical measure of project OHS performance, *Eng., Constr. Arch. Manage*, 18(1), 30–49, 2011.

[37] Linstone, H.A.; Turoff, M. (2002); *The Delphi Method: Techniques and Applications*, Addison-Wesley Publishing Company: Advanced Book Program, Vol. 18, 2002.

[38] Liu, H. T.; Tsai, Y. L. (2012); A fuzzy risk assessment approach for occupational hazards in the construction industry, *Safety Science*, 50(4), 1067–1078, 2012.

[39] Liu, Y.J.; Chen, J.L.; Cheng, S.Y.; Hsu, M.T.; Wang, C.H. (2014); Evaluation of safety performance in process industries. *Process Safety Progress*, 33(2), 166–171, 2014.

[40] Maliene, V.; Dixon-Gough, R.; Malys, N. (2018); Dispersion of relative importance values contributes to the ranking uncertainty: Sensitivity analysis of Multiple Criteria Decision-Making methods, *Applied Soft Computing Journal*, 67, 286–298, 2018.

[41] McSween, T.; Moran, D. J. (2017); Assessing and Preventing Serious Incidents with Behavioral Science: Enhancing Heinrich's Triangle for the 21st Century. *Journal of Organizational Behavior Management*, 37(3–4), 283–300, 2017.

[42] Morelli, J. (2011); Environmental Sustainability: A Definition for Environmental Professionals. *Journal of Environmental Sustainability*, 1(1), 2, 2011.

[43] Moura, R.; Beer, M.; Patelli, E.; Lewis, J.; Knoll, F. (2017); Learning from accidents: Interactions between human factors, technology and organisations as a central element to validate risk studies, *Safety Science*, 99, 196–214, 2017.

[44] Nakhaei, J.; Bitarafan, M.; Lale Arefi, S.; Kapliński, O. (2016); Model for rapid assessment of vulnerability of office buildings to blast using SWARA and SMART methods (a case study of Swiss Re Tower). *Journal of Civil Engineering and Management*, 22(6), 831–843, 2016.

[45] Papazoglou, I.A.; Aneziris, O.N.; Bellamy, L.J.; Ale, B.J.M.; Oh, J. (2017); Multi-hazard multi-person quantitative occupational risk model and risk management, *Reliability Engineering & System Safety*, 167, 310–326, 2017.

[46] Pasman, H.J.; Rogers, W.J.; Mannan, M.S. (2017); Risk assessment: What is it worth? Shall we just do away with it, or can it do a better job? *Safety Science*, 99, 140–155, 2017.

[47] Peldschus, F.; Zavadskas, E. K.; Turskis, Z.; Tamosaitiene, J. (2010); Sustainable assessment of construction site by applying game theory. *Inzinerine Ekonomika-Engineering Economics*, 21(3), 223–237, 2010.

[48] Ruzgys, A.; Volvaciovas, R.; Ignatavicius, C.; Turskis, Z. (2014); Integrated evaluation of external wall Insulation in residential buildings using SWARA-TODIM MCDM method. *Journal of Civil Engineering and Management*, 20(1), 103–110, 2014.

[49] Saaty, T. L. (1980); *The Analytic Hierarchy Process: Planning, Priority Setting, Resources Allocation.* New York: McGraw, 1980.

[50] Saaty, T. L. (1996); *Decision Making with Dependence and Feedback: The Analytic Network Process*, Pittsburgh, Pennsylvania: RWS Publications, 1996.

[51] Šaparauskas, J.; Zavadskas, E. K.; Turskis, Z. (2011); Selection of facade's alternatives of commercial and public buildings based on multiple criteria. International *Journal of Strategic Property Management,* 15(2), 189–203, 2011.

[52] Seker, S.; Zavadskas, E. K. (2017); Application of Fuzzy DEMATEL Method for analyzing occupational risks on construction sites, *Sustainability*, 9(11), 2083, 2017.

[53] Sinclair, R. C.; Cunningham, T.R.; Schulte, P. A. (2013); A model for occupational safety and health intervention diffusion to small businesses. *Am. J. Ind. Med.*, 56(12), 1442–1451, 2013.

[54] Sinelnikov, S.; Inouye, J.; Kerper, S. (2015); Using leading indicators to measure occupational health and safety performance. *Safety Science*, 72, 240–248, 2015.

[55] Sivilevičius, H.; Zavadskas, E.K.; Turskis, Z. (2008); Quality attributes and complex assessment methodology of the asphalt mixing plant, *Baltic Journal of Road & Bridge Engineering*, 3(3), 161–166, 2008.

[56] Stankiuviene, A.; Cyras, P.; Vakriniene, S. (2008); Risk identification in technical regulation. *The 7th International Conference Environmental Engineering*, Selected Papers, ISBN, 978-9955-28-256-3, Vilnius Gediminas Technical University Publishing House: Technika, 341–349, 2008.

[57] Stoklasa, J., Talášek, T., Kubatova, J., Seitlova, K. (2017). Likert Scales in Group Multiple-criteria Evaluation. *Journal of Multiple-Valued Logic & Soft Computing*, 29(5), 425–440, 2017.

[58] Šukys, R.; Čyras, P.; Šakėnaitė, J. (2011); Economical loss due to non-compliance with requirements for personnel safety and health in Lithuanian construction sector, *Journal of Civil Engineering and Management*, 17(2), 168–176, 2011.

[59] Sun, M.; Zheng, Z.; Gang, L. (2018); Uncertainty Analysis of the Estimated Risk in Formal Safety Assessment, *Sustainability*, 10, 321, 2018.

[60] Tremblaya, A.; Badri, A. (2018); Assessment of occupational health and safety performance evaluation tools: State of the art and challenges for small and medium-sized enterprises, *Safety Science*, 101, 260–267, 2018.

[61] Turskis, Z.; Juodagalvienė, B. (2016); A novel hybrid multi-criteria decision-making model to assess a stairs shape for dwelling houses, *Journal of Civil Engineering and Management*, 22:8, 1078–1087, 2016.

[62] Turskis, Z.; Lazauskas, M.; Zavadskas, E. K. (2012); Fuzzy multiple criteria assessment of construction site alternatives for non-hazardous waste incineration plant in Vilnius city, applying ARAS-F and AHP methods, *Journal of Environmental Engineering and Landscape Management*, 20(2), 110–120, 2012.

[63] Van Laarhoven, P.J.M.; Pedrycz, W. (1983); A fuzzy extension of Saaty's priority theory, *Fuzzy Sets and Systems*, 11(1–3), 229–241, 1983.

[64] Velleman, P. F.; Wilkinson, L. (1993); Nominal, ordinal, interval, and ratio typologies are misleading, *The American Statistician*, 47(1), 65–72, 1993.

[65] Whinney, M. D. (1971); *Christopher Wren*, Praeger Publishers: New York, 1971.

[66] Yau, B. (2014); Occupational safety culture index e Measuring the community and employees awareness, attitude and knowledge towards workplace safety and health in Hong Kong [power point slides], *XX World Congress on Safety and Health at Work 2014 e Global Forum for Prevention.* Frankfurt (Germany), 2014.

[67] Yilmaz, F.; Alp, S. (2016); Underlying factors of occupational accidents: the case of Turkey, *Open Journal of Safety Science and Technology*, 6, 1–10, 2016.

[68] Yoon, K.; Hwang, C. (1995); *Multiple Attribute Decision Making: An introduction;* Sage Publications, London, 1995.

[69] Zadeh, L. A. (1965); Fuzzy sets. *Inf Control*, 8, 338–353, 1965.

[70] Zagorskas, J.; Zavadskas, E. K.; Turskis, Z.; Burinskienė, M.; Blumberga, A.; Blumberga, D. (2014); Thermal insulation alternatives of historic brick buildings in Baltic Sea Region. *Energy and buildings* , 78, 35–42, 2014.

[71] Zavadskas, E. K.; Antuchevičienė, J.; Šaparauskas, J.; Turskis, Z. (2013); MCDM methods WASPAS and MULTIMOORA: verification of robustness of methods when assessing alternative solutions, *Economic Computation and Economic Cybernetics Studies and Research*, 47(2), 5–20, 2013.

[72] Zavadskas, E. K.; Turskis, Z.; Volvačiovas, R.; Kildienė, S. (2013); Multi-criteria assessment model of technologies, *Studies in Informatics and Control*, 22(4), 249–258, 2013.

[73] Zavadskas, E. K.; Vilutienė, T. (2006); A multiple criteria evaluation of multi-family apartment block's maintenance contractors: I—Model for maintenance contractor evaluation and the determination of its selection criteria, *Building and environment*, 41(5), 621–632, 2006.

[74] Zavadskas, E. K.; Vilutienė, T.; Turskis, Z.; Šaparauskas, J. (2014); Multi-criteria analysis of Projects' performance in construction, *Archives of Civil and Mechanical Engineering*, 14(1), 114–121, 2014.

[75] Zavadskas, E.K.; Kaklauskas, A.; Turskis, Z.; Kalibatas, D. (2009); An approach to multi-attribute assessment of indoor environment before and after refurbishment of dwellings, *Journal of Environmental Engineering and Landscape Management*, 17(1), 5–11, 2009.

[76] Zhou, Z.; Goh, Y.M.; Li, Q. (2015); Overview and analysis of safety management studies in the construction industry, *Safety Science*, 72, 337–350, 2015.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 14(1), 107-123, February 2019.

A Latent-Dirichlet-Allocation Based Extension for Domain Ontology of Enterprise's Technological Innovation

Q. Zhang, S. Liu, D. Gong, Q. Tu

Qianqian Zhang, Shifeng Liu,
Daqing Gong*, Qun Tu
School of Economics and Management
Beijing Jiaotong University, China
100044 No.3 Shangyuancun, Haidian, Beijing, China
15113121@bjtu.edu.cn, shfliu@bjtu.edu.cn
*Corresponding author:gongdq@bjtu.edu.cn
17113133@bjtu.edu.cn

Abstract: This paper proposed a method for building enterprise's technological innovation domain ontology automatically from plain text corpus based on Latent Dirichlet Allocation (LDA). The proposed method consisted of four modules: 1) introducing the seed ontology for domain of enterprise's technological innovation, 2) using Natural Language Processing (NLP) technique to preprocess the collected textual data, 3) mining domain specific terms from document collections based on LDA, 4) obtaining the relationship between the terms through the defined relevant rules. The experiments have been carried out to demonstrate the effectiveness of this method and the results indicated that many terms in domain of enterprise's technological innovation and the semantic relations between terms are discovered. The proposed method is a process of continuously cycles and iterations, that is the obtained objective ontology can be re-iterated as initial seed ontology. The constant knowledge acquisition in the domain of enterprise's technological innovation to update and perfect the initial seed ontology.
Keywords: Latent Dirichlet Allocation (LDA), ontology extension, enterprise's technological innovation, semantic web, text mining.

1 Introduction

With the pace of globalization of economy accelerated significantly, the market has stepped into the information age from the era of industrialization. As the market demand changes at a faster pace, the competition of the market has become extremely fierce. In this context, the technological innovation is increasingly becoming the inner motivation and main source of enterprise development. There is important significance to evaluate the capability of enterprise technological innovation scientifically and efficiently to set the technological innovation policy for government, revise technological innovation strategy reasonably for enterprises and improve the technological innovation ability. The evaluation of enterprise's technological innovation ability has drawn extensive attention of much scholars. Although much progress has been made on the theoretical research of enterprise's technological innovation [13, 33]. There still exist many problems such as evaluation mechanism and evaluation methodology, namely, the biggish subjectivity of evaluating indexes, strong dependence on declared data of evaluated enterprise, low evaluation accuracy, poor coincidence of evaluate results, etc. Enterprises produce large amounts of textual information in technological innovation process, including technological innovation activity report, meeting minutes, annual report and patent file. Hence, enterprises need to not only make use of these documents but to mine and discover valuable and hidden knowledge from large collections of data. It is also a pressing problem to transform massive textual data into knowledge that can serve and utilize for technological innovation of enterprise and provide decision-making for technological innovation of enterprise. Therefore, it is the field which has not been involved by using techniques of text mining and machine learning to analyze massive textual information that generated by enterprise's technological innovation and further determined the enterprise's technological innovation ability from objective data. In this paper, we deal with three major problems as follows:

- Is it possible to discover the concepts from large amount of textual corpus of domain of enterprise's technological innovation?
- Is it possible to build rules for semantic relationship recognition to make the enterprise's technological innovation ontology subsumption hierarchy?
- Is it possible to make the enterprise's technological innovation domain ontology extension automatically?

To improve this situation, this paper presents an approach to extract core concepts from large textual data and proposes a new method of building rules for semantic relationship recognition based on LDA algorithm. The rest of paper is organized as follows, section 2 provides some background knowledge concerning concept and relative literature reviews. In section 3 explains the proposed methods, while section 4 presents the experimental results. Section 5 concludes the paper.

2 Background knowledge and related works

2.1 Technological innovation capability

Technological Innovation Capability (TIC) has become the key to improve productivity and maintain competitiveness in the constantly fluctuating environments for enterprises. However, the definition of TIC is hard to agree upon since the technological innovation involves numerous organizational functions and resources integration among various department [26]. The concept of innovation originally from the innovation theory proposed by Schumpeter. On the base of it, Burgelaman *et al.* [5] put forward that all TIC can be defined as a series of characteristics in an organization facilitating and supporting an innovation strategy. Based on differing perspectives, there are many scholars proposed various components of TICs of a firm [22, 30].Therefore, the measurement of TIC is difficult and complicated since the perceive objectives and criteria for TIC is different. Tsai *et al.* [24] established an evaluation model for the TIC of high-tech industries based on the AHP method. Wang and Chang [25] proposed a model for diagnose the value of TIC in enterprise and established an evaluation system by AHP method. Wang *et al.* [26] evaluated and analyzed TIC combined with fuzzy evaluation and non-additional fuzzy evaluation. Deng *et al.* [12] established a TIC evaluation system by factor analysis and the fuzzy synthetic assessment method is used to evaluate TIC. Guan *et al.* [13] developed an innovation measurement framework based on the traditional DEA method. By looking at literatures of the measurements of TIC [8, 32], few studies can avoid to involve the subjective judgement, previous experience and uncertain assessment by experts.

2.2 Ontologies construction and extension

In the last decade, many scholars have done a lot of researches on ontology definition, construction, extension and application aspects. Ontologies were defined as "an explicit specification of shared conceptualization" [14] provide the key to machine-processable data on Semantic Web, being fundamental components for sharing, reusing as well as reasoning over knowledge domains [1]. Although there is a great progress in knowledge acquisition and ontology construction, the current ontology construction methods still rely heavily on manual parsing and existing knowledge bases. The process of ontology learning and extending is a costly, time-consuming and error-prone task when done manually. With the constant emergence of new domain knowledge, the domain ontology automatic updates are facing new challenge.

Many researchers have engaged into ontology construction and enriching automatically in recent years. In previous work, the machine learning and statistical analysis method has great advantages in accuracy and recall rate and has been proposed to solve this problem [9]. For instance, Jeroen *et al.* [10] proposed the subsumption method and a hierarchical clustering algorithm to arrange the domain terms hierarchically and compared the two methods performances. Researches [18] and [23] using the fuzzy mechanism to extract domain concept and generate the domain ontology through the fuzzy conceptual clustering. Khan and Luo [17] presented a modified self-organizing tree algorithm (SOTA) which is performs better than the hierarchical agglomerative clustering (HAZ) on ontology construction automatically. Gilles *et al.* [1] put forward a Mo'k workbench which is a framework using the agglomerative clustering techniques to generate concept hierarchies from parsed corpora. Cimiano and Völker [6] presented a Text2Onto which implementing variety algorithms and techniques for ontology learning. However, most of the existing methods require a certain scale of supervised training corpus as the learning object, and the result seldom consider semantic-aware which is difficult to recognize the relationship between terms in domain ontology.

Although the ontology construction and extending automatically has been achieved some progress, there are still some problems in this field. For example, the non-taxonomic relationship among terms were often omit in the ontology hierarchical relations construction. Besides, the parameter setting in the model and complex computing in the process cause the heavy computing burden and make the model overfitting which limits their application.

2.3 LDA topic model

The latent topic discovery researches have gained much attention to hierarchical relation learning in recent years. Latent topic discovery is invented to overcome the bottleneck of bag-of words processing model in information retrieval area, trying to advance the text processing technology from pattern to semantic calculation [23]. For the research in latent topic discovery, an earlier work in literatures is Latent semantic indexing (LSI), which is a retrieval technique to learn latent topic by performing a matrix decomposition (SVD) on the term-document matrix [31]. Through this technique, latent topics are revealed which are actually distributions over the words of the term space of the corpus [10]. For example, the work in [3] uses the technique of LSI to identify relationships among entities in large collections of text. The author in [4] also using the LSI for discovering new information relevant to a given topic in large textual databases. Although the LSI based on SVD having some early success on latent topic discovery and relationship identification, it lacks rigorous mathematical and statistical basis and the SVD decomposition is time-consuming. Probabilistic Latent Semantic Indexing (PLSI) was proposed to extend the LSI assuming which associates a latent context variable with each word occurrence and can deal with synonymy and polysemous words. The author in [16] proposed that PLSI has been considered as an unsupervised learning method used in the task of text learning. The work in [15] also using the PLSI to represent sentences and queries as probability distributions over latent topic to solve the multi-document summarization problem. Other than LSI and PLSI, the algorithm of Latent Dirichlet Allocation (LDA) is more advantageous since LDA model can avoid overfitting and large sets of parameters.

LDA model, proposed by David Blei *et al.* [2], is a statistical topic model and can analyzes hidden topics in large-scale data. Ontology learning using LDA model is a relative new research approach. Elias *et al.* [34, 35] used the LDA model for discovery of topics that represent ontology concepts and comparing the high-probability terms in topics to arrange concepts in a subsumption hierarchy. However, it cannot infer subsumption relations in the case where a topic subsumes only one other topic. Yeh and Yang [29] developed an automatic domain ontology construction for historical documents. LDA model was used to extract latent topic from raw textual Chinese Recorder data and the basic cosine similarity with hierarchical agglomerative clustering is used to clustering the topic, but the relationship between the topic cannot be defined since the clustered latent topic is a hierarchical tree structure. Francesco *et al.* [7] present an automatic terminological ontological learning system which the common hypernyms between the aggregate root node and aggregate words are determined through the LDA model and then added the semantically similar root node to the ontology. however, the measurement in large set of data may cause heavy computing burden. Ni *et al.* [20] also used the LDA model to select the domain terms and through the word association analysis to discover the hierarchical relations among domain terms. Raghuveer [21] using the LDA model to obtain the topics from legal documents and clustering legal judgments by cosine similarity.

3 The proposed method

The paper has combined the ontology technique and LDA topic model, used the initial seed ontology guiding the LDA model to obtain the concept in the field of enterprise technological innovation. Adding the new concept to the initial domain ontology by defined rules to realize the iteratively updating and perfection of ontology. The framework of enterprise's technological innovation domain concept acquisition contains the following four modules:

- The module of seed ontology introducing. The paper needs to construct a seed ontology to guide the concept acquisition for enterprise's technological innovation domain. The basic concept and relationship of seed ontology in domain enterprise technological innovation mainly extracted from Chinese Classified Thesaurus. The protege 4.3 was used to visualize the construction of seed ontology. More details will be introduced in the next chapter.

- The module of text preprocessing. This is the process of converting a text into individual words or sequences of words which using the Natural Language Processing (NLP) technique including of word segmentation, Part-of-Speech (POS)tagging, stop-word filtering preprocessed the collected Chinese textual documents. Two words merging needs to satisfy adjacency and frequent co-occurrence both, the calculation method as follows. In order to guarantee the semantic accuracy after word segmentation, the method of entropy was adopted to merge the words [27, 28].

  $$E(w_{m-1}, w_m) = \frac{p(w_{m-1}w_m)}{minp(w_{m-1}), p(w_m)} \quad (1)$$

  where $p(w_m)$ denotes the frequency of word w_m in documents and $p(w_{m-1}w_m)$ denotes the continuous frequency of word w_{m-1} and w_m in documents.

- The module of mining domain specific terms. LDA (Latent Dirichlet Allocation) is a three-level hierarchical Bayesian model which proposed by Blei [34]. It assumes that each document in corpus is represented as random mixtures over latent topic, where topic is characterized by a distribution over all the words. LDA is constructed for documents with "bag-of-words" which uses the statistical information of words to represent text in vector space and explores the probabilistic relationships between words and text. In this paper, we use the LDA model was described below.

LDA taking the corpus D which after the preprocessing by module B as input and output the topic distributions and the distribution of words for each topic by training. The LDA generates the words in a two-stage process: words are generated from topics and topics are generated by documents. The graphical model of LDA is shown in Fig. 1. The terms of LDA was defined as follows:

A document is a sequence of N words denoted by $w = (w_1, w_2, \cdots, w_n)$ where w_n is the nth word in the sequence, and a corpus is a collection of M documents denoted by $D = d_1, d_2, \cdots, d_M$;

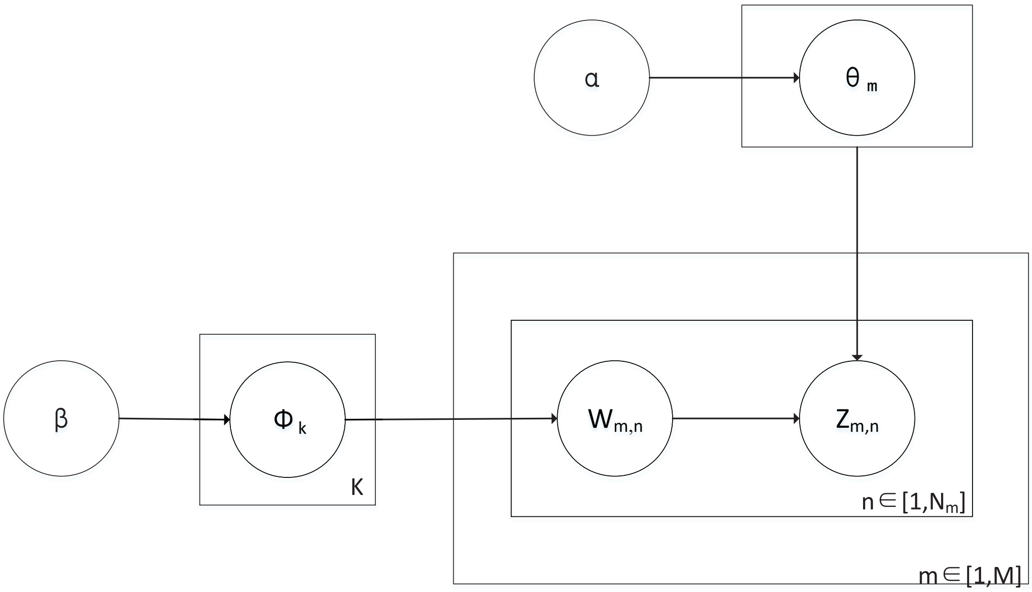


Figure 1: Graphical model representation of LDA

α and β are Dirichlet prior hyperparameters; All the words in document M will be clustered into Z topics, for each topic $Z \in 1, 2, \cdots, k$, sample a word distribution $\phi_k \sim Direchlet(\beta)$;

- Choose $N \sim Possion(\xi)$
- Choose a topic distribution $\theta_m \sim Direchlet(\alpha)$
- For each of word $w_{m,n}$ in mth document:
  * Choose a topic of the word $Z_{m,n} \sim Multinomial(\theta_m)$
  * Choose a word $w_{m,n} \sim Multinomial(\phi_{Z_{m,n}})$

Since the process to generate the topic for M documents are independent of one another, we can have M conjugated structures and the generative process of probabilistic of topics in corpus is as follows:

$$\begin{aligned} p(\vec{z}|\vec{\alpha}) &= \prod_{m=1}^{M} p(\vec{z_m}|\vec{\alpha}) \\ &= \prod_{m=1}^{M} \frac{\Delta(\vec{n_m} + \vec{\alpha})}{\Delta(\vec{\alpha})} \end{aligned} \tag{2}$$

The process to generate words for K topics are independent of one another, we can have K conjugated structures and the probabilistic of words in corpus is as follows:

$$\begin{aligned} p(\vec{w}|\vec{z},\vec{\beta}) &= \prod_{k=1}^{k} p(\vec{w_k}|\vec{z_k},\vec{\beta}) \\ &= \prod_{k=1}^{k} \frac{\Delta(\vec{n_k}+\vec{\beta})}{\Delta(\vec{\beta})} \end{aligned} \tag{3}$$

Thus, within a document, the probability distribution over words specified by the LDA model is given as follows:

$$\begin{aligned} p(\vec{w},\vec{z}|\vec{\alpha},\vec{\beta}) &= p(\vec{w},\vec{z}|\vec{\beta}) * p(\vec{z}|\vec{\alpha}) \\ &= \prod_{k=1}^{k} \frac{\Delta(\vec{n_k}+\vec{\beta})}{\Delta(\vec{\beta})} * \prod_{m=1}^{M} \frac{\Delta(\vec{n_m}+\vec{\alpha})}{\Delta(\vec{\alpha})} \end{aligned} \tag{4}$$

Thus, in this paper, the LDA topic model was used to train the term candidate set which obtained by the module of text preprocessing and to obtain the word probabilistic of domain concepts (topics) as shown in Fig.2.

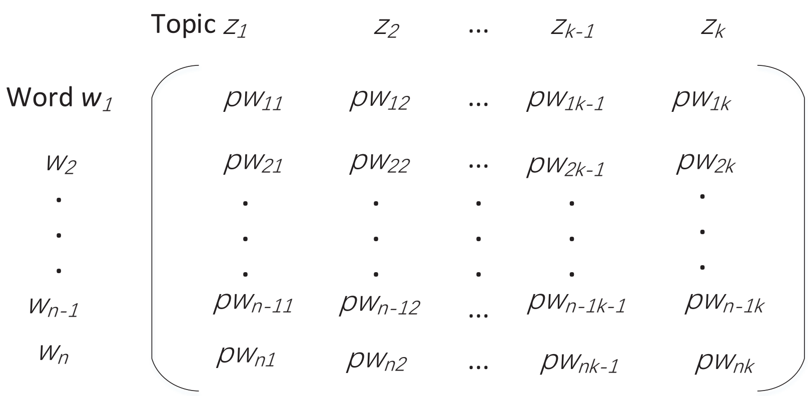


Figure 2: Words distribution probabilistic of topics

where pw_{nk} represents the probability of the word n in the topic k.

- The module of domain ontology updating. The module is the key point and difficulty of this paper. Take each concept in the initial enterprise technological innovation ontology as a document into the module (3) trained LDA model. We can get the topics probabilistic of documents as shown in Fig.3. Where, a corpus is a collection of M ontology concepts denoted by $C = (c_1, c_2, \cdots, c_{m-1}, c_m)$;

Where pz_{km} represents the probability of the topic k in concept(document) m.

According to the LDA algorithm, we can get the term probabilistic of documents, namely, the probabilistic of words in documents and concepts in initial ontology denoted as $p(w_n|c_m)$. Then by using the relevant rules to judge the relationship between topics generated by LDA model and concepts in initial domain ontology.

$$p(w_n|c_m) = \sum_{j=1}^{K} p(w_n|z=j) * p(z=j|c_m) \tag{5}$$

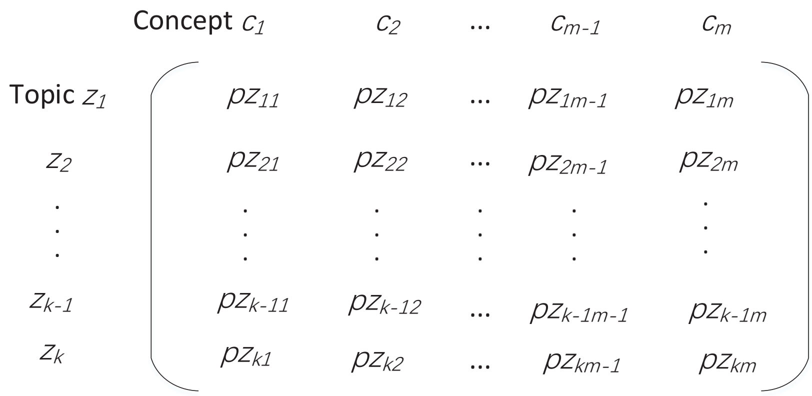


Figure 3: Topics distribution probabilistic of documents

When the $p(w_n|c_m)$ greater than the threshold value TH and the word n is not in the list of $C = (c_1, c_2, ..., c_{m-1}, c_m)$, therefore, the term w_n is an associated term of c_m.

$$p(w_n|c_m) > TH \tag{6}$$

Algorithm: Rules of the semantic relationship recognition were defined as follows:

$$\begin{aligned} W(W_n, C_m) &= \frac{p(W_n|C_m)}{p(z=j|C_m) + p(W_n|C_m)} \\ &= \frac{\sum_{j=1}^{K} p(w_n|z=j) * p(z=j|c_m)}{p(z=j|C_m) + \sum_{j=1}^{K} p(w_n|z=j) * p(z=j|c_m)} \end{aligned} \tag{7}$$

- Rule 1: Rules for synonymy relations recognition. If the $W(Wn, Cm) \geq 0.01$, the semantic relationship between word and concept is equivalent, namely, the related terms extracted by LDA is equal to the existed concept.

- Rule 2: Rules for hyponymy relations recognition. When the Rule 1 cannot be satisfied, if the $W(Wn, Cm) \geq 0.004$, the word includes the concept, namely, the related terms extracted by LDA is superclass of the existed concept, the relationship as "is-a" or "subclass".

- Rule 3: Rules for correlation recognition. When the Rule 1 and Rule 2 are cannot be satisfied, the relationship between existed concept and related terms can be recognized as related or using people to identify the specific semantic relationship by external knowledge base.

Based on the above rules, the semantic relations between the existing concepts and their related terms are identified, add the obtained related terms and semantic relations to the original ontology O, the original ontology O was updated to Oi.

4 Experiments and result

4.1 Ontology acquisition

Enterprise ontology and TOVE (Toronto Virtual Enterprise Ontology) are the most popular ontology-based enterprise modeling methodologies. The two projects all point out the common key influencing factors in the process of enterprise ontology construction including of resources, organization, strategy, market and activity. In this paper, the five factors also considered as the first class of the enterprise's technological innovation ontology. The Chinese Classified Thesaurus has clear semantic structure which is more suitable for the extraction of concepts and relationship between concepts. Transforming thesaurus into ontology through further concepts analysis and semantic relationship adjustment of the words in F27 category of Enterprise Economy in Chinese Classified Thesaurus. There are 5 concepts extracted from the thesaurus including of Innovation resources, Marketing innovation, Strategic innovation, Organizational innovation and Innovation activities. The nested composite view provides a representation of the interrelation between the first classes in the entire ontology structure. It is convenient for considering whether the constructed domain ontology meets actual needs. The nested composite view of enterprise's technological innovation domain is shown as Fig.4. The relationship between domain ontology concepts includes the hyponymy relations and complex non-hierarchical relationship for specific application. The Fig.5 shows that the relationship between domain ontology concepts which takes the Strategic innovation as the center and reflects the complex relationship between concepts. The ontology of enterprise's technological innovation is a prototype, in which many concepts and relationships are still insufficient and need to continuously improved.

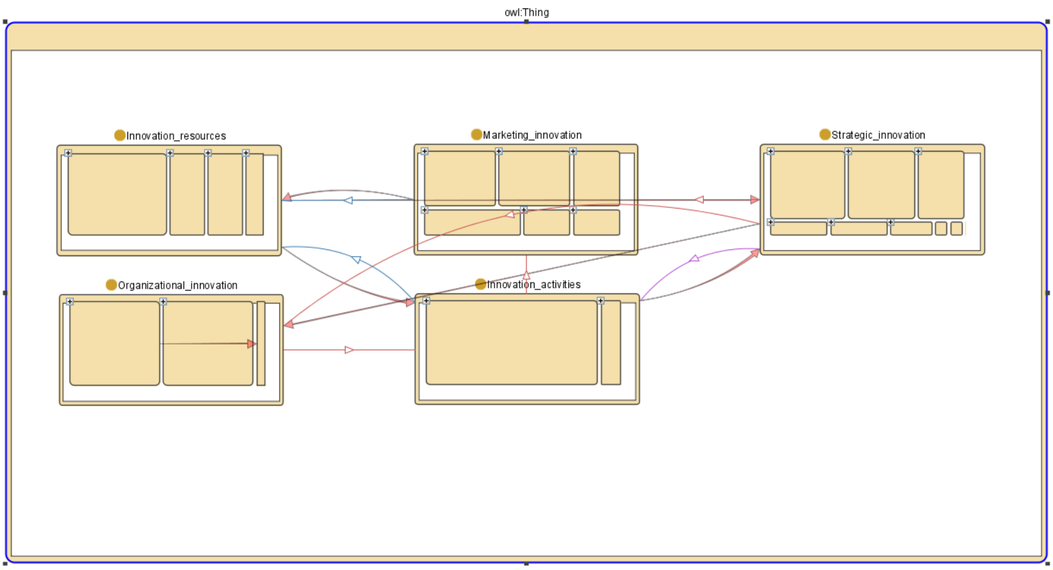


Figure 4: Nested composite view of enterprise technological innovation domain

4.2 Textual data collection

There are two aspects to collect the textual data of enterprise technological innovation, one is the internal information generated from daily production activities such as internal R&D, innovation activities, etc. The other type of collected data is generated when enterprise interacting with external customers and partners by social networks, mobile applications, etc. 863 sets of valid data are obtained which includes of 413 enterprise technology centers.

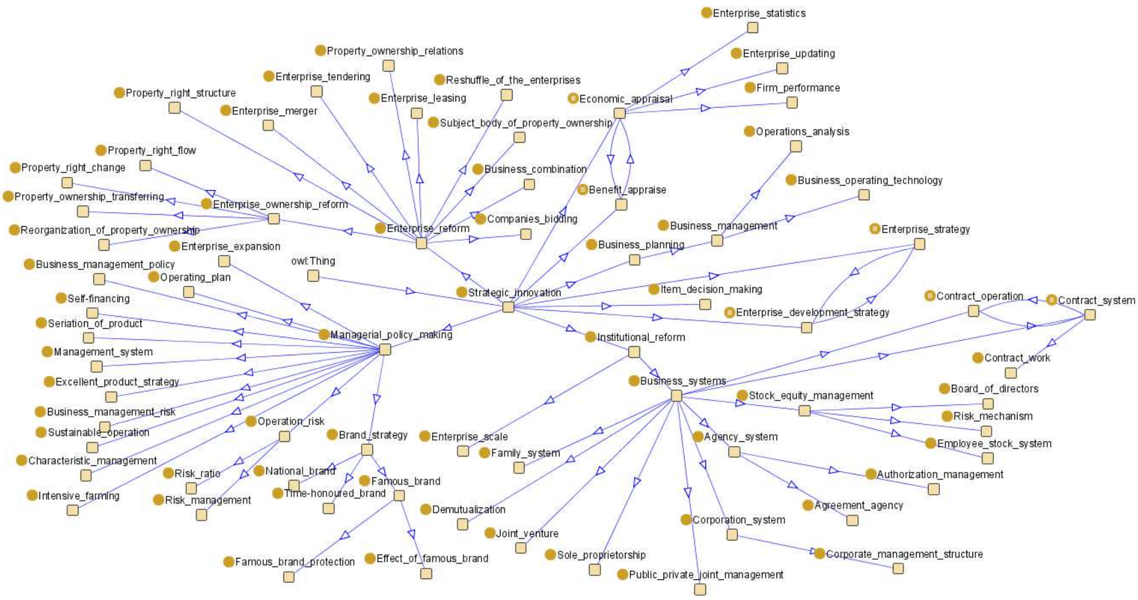


Figure 5: Visualization of relationship between concepts

4.3 Text preprocessing

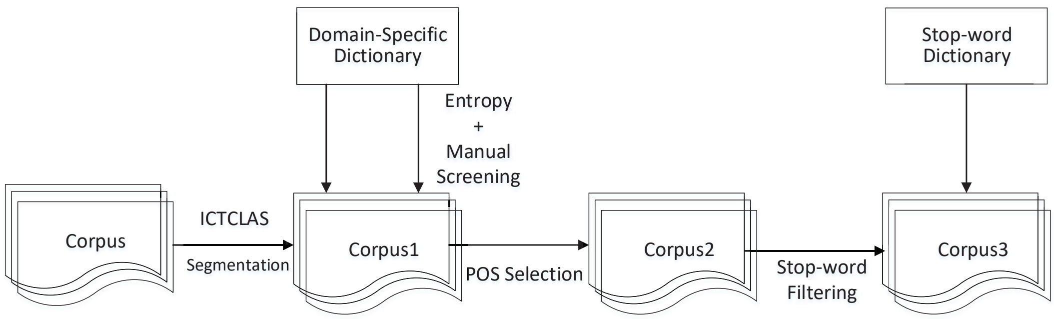


Figure 6: Process of text preprocessing

Chinese segmentation

Firstly, constructing the domain-specific dictionary for the field of enterprise's technological innovation by widely collected materials such as the cell thesaurus and imported the dictionary into the ICTCLAS segmentation system [32] which developed by the Chinese Academy of Sciences. Secondly, the result of segmentation will appear the problem due to a Chinese phrase was wrongly divided into many words. For example, the "enterprise's technological innovation" was divided into three small-grained words such as "enterprise", "technological" and "innovation". The method of entropy was adopted to merge the words which shown as equation (1). Combining two words that satisfy the conditions into a new phrase and adding to the domain-specific dictionary by manual screening. Then, segment the source document and iterate repeatedly.

POS selection

The documents of enterprise's technological innovation are the synthetic texts, in which, nouns are more representative important for semantic information in source documents. Hence, selecting the nouns and the word similar to nouns as the research object such as the verb with noun function, the adjective with noun function, etc.

Elimination of stop-words

Useless words selected from the domain of enterprise's technological innovation is used to build stop words dictionary. Filtering the stop words in documents which processed by the above two steps. It can reduce the size of the indexing structure considerably by elimination of stop words.

4.4 Mining domain terms from text corpus based on LDA

- Terms selection. According to the word frequency of terms in all corpora, the word frequency of [50, 1000] were selected as terms to represent each document in vector space model.

- Optimal number of topics. The perplexity index is adopted in optimal topic selection. Perplexity is an effective measurement to verify the model generalization ability. A lower perplexity indicates the better generalization performance. The perplexity is defined as follows:

$$perplexity(W_n|C_m) = e^{\frac{-\sum \log(p(W_n|C_m))}{N}} \tag{8}$$

  Where $p(W_n|C_m)$ is the probability of each word in candidate term set, N is the number of words.

  The perplexity of all documents generated under different topic numbers is shown as Fig.7 It looks like the 160-topic model has the lowest perplexity score. Hence, the optimal number of topic 160 (k=160) is selected for all corpus by perplexity analysis. The smoothing parameters α and β were fixed at 0.1 and 0.3. The threshold TH was set to 0.001.

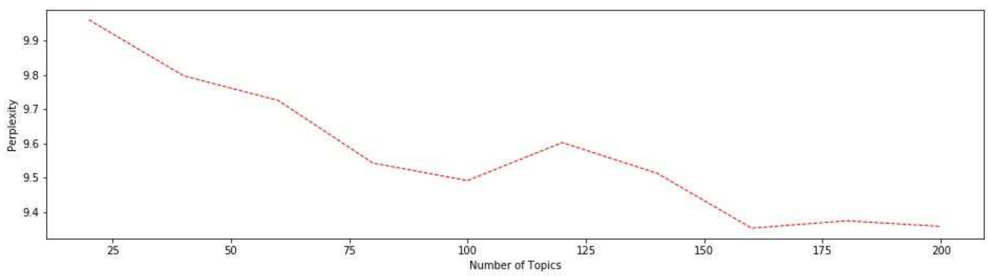


Figure 7: Perplexity result on enterprise technological innovation corpora for LDA model

- When the number of topics is 160, the LDA topic modelling is carried out to obtain the distribution of terms, that is each topic comprises of a series of related words. The order of the top terms of each topic is arranged by the probability and presented in the Table.1, in which only the first 10 topics with high probability of topic distribution were shown. The Fig.8 shows the resulting graph visualization LDA model for top terms of topics.

Table 1: The distribution probability of topics and words when topic K=160

Topic 5	$P(w_n\|k=160)$	**Topic 13**	$P(w_n\|k=160)$	**Topic 23**	$P(w_n\|k=160)$
Technology	0.001665396	Expert	0.0021303003	Patent	0.0095653171
Material	0.0014465271	Doctor	0.0009977445	Name	0.0047526103
Technique	0.0012552338	Senior engineer	0.0007695822	Number	0.0042560613
New Material	0.0007991531	Counselor	0.0006120452	Technology	0.0031178757
Product	0.0005905334	Master	0.0005237754	Information	0.0028379832
Technological innovation	0.0005886399	Bachelor	0.0005183982	Type	0.0027637654
High-performance	0.0005476442	Post-doctoral	0.0003545205	Invent	0.0038754337
Precision	0.0005436358	Degree	0.0002742724	Copyright	0.0016382793
New-technology	0.0004137853	Professor	0.0002726482	Authorized-patent	0.0007759799
New-product	0.0003711188	College	0.0002557352	Conservation	0.0006379388
Stability	0.0003021399	Associate-professor	0.0001320247	Authorization	0.0002794288
Practical	0.0002943092	Academic	0.0001297645	Intellectual-property	0.0002544165
Topic 27	$P(w_n\|k=160)$	**Topic 30**	$P(w_n\|k=160)$	**Topic 39**	$P(w_n\|k=160)$
Project	0.0008273301	Enterprise	0.0010061373	New-product	0.0008340621
Types	0.0008237544	Name	0.0009097208	New-techniques	0.0008340621
Invisible-asset	0.0004237544	Development Organization	0.0008218773	Name	0.0004272026
Fix asset	0.0004223029	Contact-telephone	0.0006368324	Market Occupancy	0.0004272025
Equipment	0.0004208453	Organization	0.0004940051	Profit	0.0004272025
Facility	0.0003637534	Company	0.0004912549	Period	0.0004272025
Total-amount	0.0003230094	Department	0.0004209615	Sale-quota	0.0004263008
Quantity	0.0003034324	Laboratory	0.0004209615	Sales-volume	0.0004262023
Cost	0.0002784593	Contact person	0.0004209615	Competitive	0.0004262010
Fund	0.0002764534	Research-institute	0.0004209615	Economic-benefit	0.0004260232
Amount	0.0002230895	Contact details	0.0004209615	Popularization	0.0003037646
Instrument	0.0002234943	Information	0.0003026468	Technical management	0.0003037564

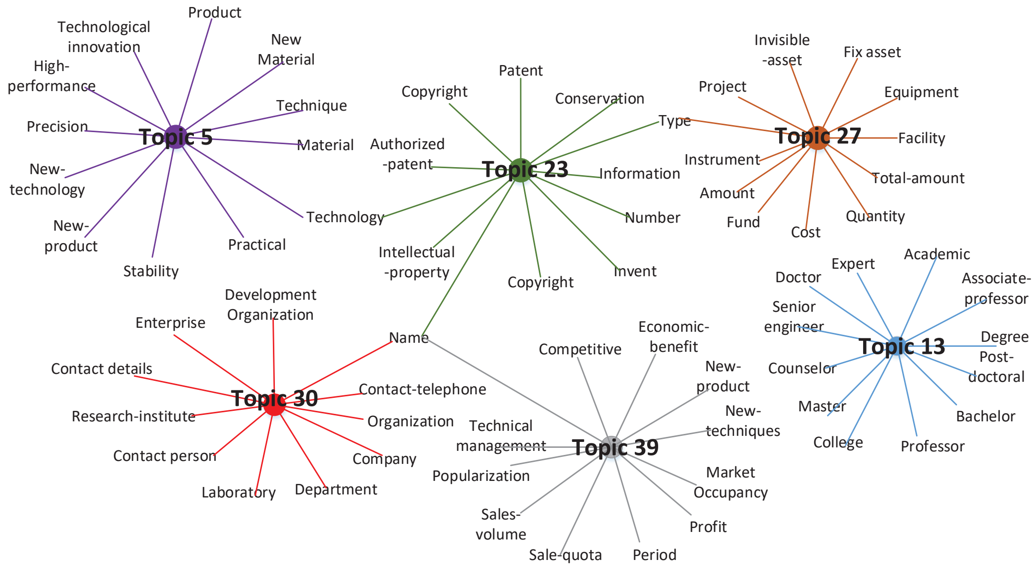


Figure 8: Graph of LDA for top terms of topics

4.5 Learning hierarchical relations among terms

Using the trained LDA model to infer each concept in the initial ontology and taking each concept (or word) as a document to calculate the topic probability of the document. Identify the semantic relations between existing concepts and its related terms, and add the related terms as the domain ontology concept to the appropriate position of the existing ontology to complete an update process of the domain ontology. Table.2 shows the results of the conceptual related terms extraction and relations recognition.

Table 2: Related terms extraction and relations recognition

Existing Concepts	Topic	$P(c_m\|z=j)$	Related terms	Weights	Appli-cable rules	Semantic relations
Profit Management	79	0.438621	Innovation Resources	0.00413586	(2)	Subclass
			Total Amount	0.00624005	(2)	Subclass
Technical Information	139	0.388462	Material	0.00651796	(2)	Subclass
			Painting	0.00331878	(3)	Related
			Alloy	0.00243573	(3)	Related
Visible Asset	27	0.236543	Fixed asset	0.01342533	(1)	Equivalent
			Equipment	0.00523433	(2)	Subclass
			Instrument	0.00243234	(2)	Subclass
Visible Asset	27	0.236543	Fixed asset	0.01342533	(1)	Equivalent
			Equipment	0.00523433	(2)	Subclass
			Instrument	0.00243234	(2)	Subclass
Technical-Quality	5	0.388462	Precision	0.00257653	(3)	Related
			New Technology	0.00323643	(3)	Related

Existing Concepts	Topic	$P(c_m \mid z = j)$	Related terms	Weights	Applicable rules	Semantic relations
High-tech Product	136	0.446243	Strategy Innovation	0.00276406	(3)	Related
High-tech Product	136	0.446243	Technological Innovation	0.00143524	(3)	Related
Product Innovation	23	0.237643	Patent	0.00332763	(3)	Related
			Brand	0.00236232	(3)	Related
			Copyright	0.00323422	(3)	Related
			New Product	0.00332542	(3)	Related
Staff Management	13	0.376432	Expert	0.00335476	(3)	Related
			Doctor	0.003276543	(3)	Related
			Degree	0.002387432	(3)	Related
			Senior engineer	0.003723423	(3)	Related
Wage management	63	0.3412663	Wage	0.01472652	(1)	Equivalent
			Subsidy	0.00234653	(3)	Related
			Bonus	0.00334523	(3)	Related
			Insurence	0.00343263	(3)	Related

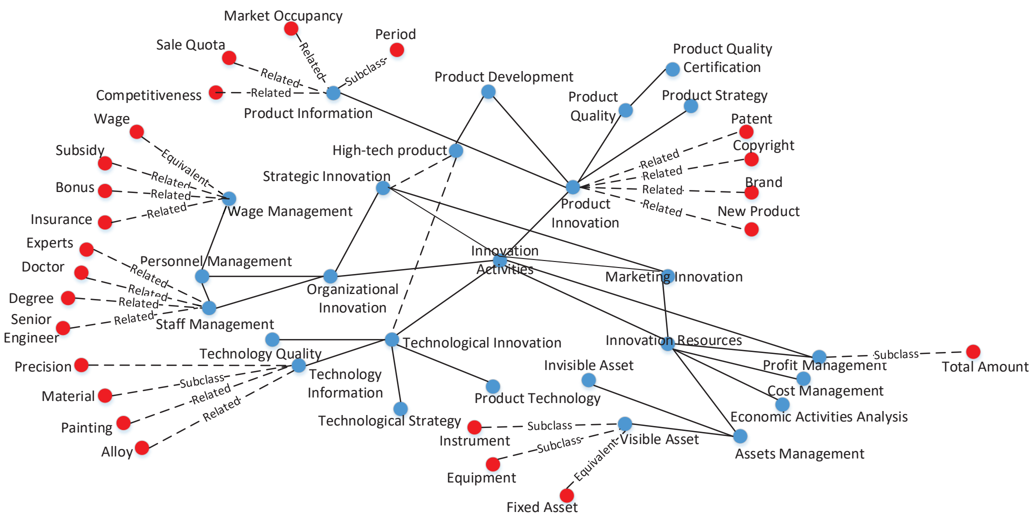


Figure 9: Parts of produced enterprise technological innovation domain ontology

The Fig.9 shows part of produced enterprise's technological innovation domain ontology. The blue dots represent the original terms of initial domain ontology and the red dots stand for the produced new terms. The original relations among entities in ontology are shown with solid lines, the dashed lines represent the new relations. The total amount of new terms in enterprise technological innovation domain ontology has updated about 163, the figure only shows parts of the result due to space limitation.

By looking at the literature of ontology evaluation, there are two approaches for measuring the ontology including of manual evaluation by human experts and gold standard-based approaches [11]. The first evaluation approach presents the learned ontology to one or more human experts and judge how far the extracted information is correct. The second method compare the learned ontology with a previously created gold ontology which example for this kind of evaluation can be found in papers like [34]. The degree of matching between learned ontology and gold ontology determines the precision of learning ontology. The evaluation of ontologies when these ontologies are produced by an automated learning procedure is an open field of research. Since the enterprise's technological innovation is a new developing academic field which has not formed a generally acknowledged ontology yet. Therefore, the manual evaluation by human experts was the best way so far. The research chosen 5 groups and 20 terms and relations for each group in the updated enterprise's technological innovation domain ontology randomly. The assisted algorithm like following equation was defined as the ration between the right terms and relationships which evaluated by human experts and the total terms and relationships in ontology. According to the validation about the correct terms and relations with domain experts, the result of accuracy test is shown as Table 3.

$$precision = \frac{righttermsandrelationships}{totaltermsandrelationshipsontology} \tag{9}$$

Table 3: The accurate rate of the concepts in enterprise technological innovation domain

No.	**Number of groups**	**Accurate number of groups**	**Precision**
Group 1	20	19	95%
Group 2	20	19	95%
Group 3	20	18	90%
Group 4	20	17	85%
Group 5	20	19	95%

Compared with the traditional ontology construction methods such as OntoLearn and Text2-Onto, the proposed method has same precision which the average accuracy rate is 92%. The semantic content and relationship in the produced ontology is basically correct. The proposed automatic ontology extension method reduces the manual labor for ontology updating and solved the problem of automatic domain ontology acquisition and dynamic maintenance.

5 Conclusion and future work

This paper presented an automatic ontology extension method for the domain of enterprise's technological innovation. The main contributions of this paper present as follows: Firstly, this paper proposes an ontology-based LDA topic model for concept extraction and applies it to the realm of enterprise technological innovation, which not only discover the concepts from large amount of textual corpus, but also can provides data support for ontology construction. Secondly, this article takes a huge amount of enterprise technological innovation information in unstructured texts as the data source and proposes a method of building rules for semantic relationship recognition based on LDA topic probability distribution, and the process of automated domain ontology updating based on the LDA topic model is realized. Finally, the experiment results demonstrate the efficiency and validation of proposed method. The method focuses on discovering the domain terms via latent topics found by LDA algorithm from plain text corpus and recognizing the semantic relations among domain terms based on word association analysis. The proposed method is a process of continuously cycles and iterations, the domain ontology of enterprise's technological innovation will be updated and perfected automatically with the constant knowledge acquisition in the domain. The paper introduces the ontology on the basis of the LDA topic model and the ontology is extended by the obtained related topics. The proposed method is an improvement for the single LDA algorithm.

The future work needs to solve several problems, firstly, improving the proposed method to achieve a better performance and continuing exploring automatic evaluation approaches on thesaurus constructing methods. Secondly, using the constructed enterprise technological innovation ontology and combined with the text mining methods to construct the mechanism of evaluation for enterprise's technological innovation.

Funding

This paper is supported by the Fundamental Research Funds for the Central Universities (2018YJS051,B18RC00070) and Beijing Social Science Funds (18JDGLA018).

Bibliography

[1] Bisson, G.; Nédellec, C. Canamero, D.(2000); Designing Clustering Methods for Ontology Building-The Mo'K Workbench, *ECAI workshop on ontology learning*, 31, 2000.

[2] Blei, D.M.; Ng, A.Y.; Jordan, M.I. (2003); Latent dirichlet allocation, *Journal of machine Learning research*, 3(Jan), 993–1022, 2003.

[3] Bradford, R.B. (2006); Relationship discovery in large text collections using latent semantic indexing, *Proceedings of the Fourth Workshop on Link Analysis, Counterterrorism, and Security*, 2006.

[4] Bradford, R.B. (2005); Efficient discovery of new information in large text databases, *International Conference on Intelligence and Security Informatics*, 374–380, 2005.

[5] Burgelman, R.A.; Maidique, M.A.; Wheelwright, S.C. (1996); *Strategic Management of Technology and Innovation*, Chicago,IL:lrwin, 1996.

[6] Cimiano, P.; and Völker, J. (2005); text2onto, *International conference on application of natural language to information systems*, 227–238, 2005.

[7] Colace, F.; De Santo, M.; Greco, L.; Amato, F.; Moscato, V.; Picariello, A. (2014); Terminological ontology learning and population using latent dirichlet allocation, *Journal of Visual Languages & Computing*, 25(6), 818-826, 2014.

[8] Dai, Y.; Wu, W.; Zhou, H.B.; Zhang, J.; Ma, F.Y. (2018); Numerical simulation and optimization of oil jet lubrication for rotorcraft meshing gears, *International Journal of Simulation Modelling*, 17(2), 318–326, 2018.

[9] Dai, Y.; Zhu, X.; Zhou, H.; Mao, Z.; Wu, W.(2018); Trajectory tracking control for seafloor tracked vehicle by adaptive neural-fuzzy inference system algorithm, *International Journal of Computers, Communications & Control* 13(4), 465–476, 2018.

[10] De Knijff, J.; Frasincar, F.;Hogenboom, F. (2013); Domain taxonomy learning from text: The subsumption method versus hierarchical clustering *Data & Knowledge Engineering*, 83, 54-69, 2013.

[11] Dellschaft, K; Staab, S. (2008); Strategies for the evaluation of ontology learning, *Ontology Learning and Population*, 167, 253–272, 2008.

[12] Deng, L; Wang, X; Lin, Y; He, F.Z. (2005); Model of Multiple Fuzzy Synthetical Evaluation for Enterprise Technology Innovation, *Journal of Chongqing University (Natural Science Edition)*, 7, 004, 2005.

[13] Guan, J.C.; Yam, R.C.; Mok, C.K.; Ma, N. (2006); A study of the relationship between competitiveness and technological innovation capability based on DEA models, *European Journal of Operational Research*, 170(3), 971-986, 2006.

[14] Guarino, N.; Poli, R. (1993); Toward principles for the design of ontologies used for knowledge sharing, *In Formal Ontology in Conceptual Analysis and Knowledge Representation, Kluwer Academic Publishers, in press. Substantial revision of paper presented at the International Workshop on Formal Ontology*, 1993.

[15] Hennig, L. (2009); Topic-based multi-document summarization with probabilistic latent semantic analysis, *Proceedings of the International Conference RANLP-2009*, 144–149, 2009.

[16] Hofmann, T. (2001); Unsupervised learning by probabilistic latent semantic analysis, *Machine learning*, 42(1-2), 177–196, 2001.

[17] Khan, L.; Luo, F. (2002); Ontology construction for information selection, *Proceeding of Tools with Artificial Intelligence*, 122-127, 2002.

[18] Lee, C.S.; Kao, Y.F.; Kuo, Y.H.; Wang, M. H. (2007); Automated ontology construction for unstructured text documents, *Data & Knowledge Engineering*, 60(3), 547–566, 2007.

[19] Liu, Q.; Zhang, H.; Yu, H.; Cheng, X. (2004); Chinese lexical analysis using cascaded hidden markov model, *Journal of Computer Research and Development*, 41(8), 1421–1429, 2004.

[20] Ni, N.; Liu, K.; Li, Y. (2011); An automatic multi-domain thesauri construction method based on lda, *2011 10th International Conference on Machine Learning and Applications Workshops*, 235-240, 2011.

[21] Raghuveer, K. (2012); Legal documents clustering using latent dirichlet allocation, *International Journal of Applied Information Systems*, 2(1), 34-37, 2012.

[22] Saunila, M.; Ukko, J. (2012); A conceptual for the measurement of innovation capability and its effects, *Baltic Journal of Management*, 7(4), 355–375, 2012.

[23] Tho, Q.T.; Hui, S.C.; Fong, A.C.M.; Cao, T.H. (2006); Automatic fuzzy ontology generation for semantic web, *IEEE transactions on knowledge and data engineering*, 18(6), 842-856, 2006.

[24] Tsai, M.T; Chuang, S.S; Hsieh W.P. (2008); Using Analytic Hierarchy Process to Evaluate Organizational Innovativeness in High-Tech Industry, *Decision Sciences Institute 2008 Annual Meeting (DSI)*, 1231-1236, 2008.

[25] Wang, T. J; Chang, L. (2011); The development of the enterprise innovation value diagnosis system with the use of systems engineering, *System Science and Engineering (ICSSE), 2011 International Conference on IEEE*, 373–378, 2011.

[26] Wang, C; Lu, I; Chen, C. (2008); Evaluating firm technological innovation capability under uncertainty, *Technovation*, 28(6), 349–363, 2008.

[27] Wei, W.; Guo, C.; Chen, J.; Tang, L.; Sun, L. (2017); CCODM: conditional co-occurrence degree matrix document representation method, *Soft Computing*, 1-17, 2017.

[28] Wei, W.; Guo, C.; Chen, J.;Zhang, Z. (2017); Textual topic evolution analysis based on term co-occurrence: A case study on the government work report of the State Council (1954–2017), *Intelligent Systems and Knowledge Engineering*, 1-6, 2017.

[29] Yeh, J.H.; Yang, N. (2008); Ontology construction based on latent topic extraction in a digital library, *International Conference on Asian Digital Libraries*, 93–103, 2008.

[30] Yliherva, J. (2004); Management model of an organization's innovation capabilities; development of innovation capabilities as part of the management system, *dissertation, Department of Industrial Engineering and Management, University of Oulu.*

[31] Zhang, W.; Zhang, Z.; Chao, H.C.; Tseng, F.H. (2018); Kernel mixture model for probability density estimation in Bayesian classifiers. Data Mining and Knowledge Discovery, *Data Mining and Knowledge Discovery*, 32(3), 675–707, 2018.

[32] Zhang, W.; Zhang, Z.; Qi, D.; Liu, Y. (2014); Automatic crack detection and classification method for subway tunnel safety monitoring, *Sensors* , 14(10), 19307–19328, 2014.

[33] Zhao, W.; Zeng, Y. (2011); Construction and design of evaluation index system of innovative enterprises on innovative capacities, *Science and Technology Management Research*, 1, 005, 2011.

[34] Zavitsanos, E.; Paliouras, G.; Vouros, G.A.; Petridis, S. (2010); Learning subsumption hierarchies of ontology concepts from texts, *Web Intelligence and Agent Systems: An International Journal*, 8(1), 37-51, 2010.

[35] Zavitsanos, E.; Paliouras, G.; Vouros, G.A.; Petridis, S. (2010); Discovering subsumption hierarchies of ontology concepts from text corpora, *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, 402–408, 2007.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 14(1), 124-134, February 2019.

Text Classification of Public Feedbacks using Convolutional Neural Network Based on Differential Evolution Algorithm

S. Zhang, Y. Chen, X.L. Huang, Y.S. Cai

Shuai Zhang, Yong Chen
School of Information
Zhejiang University of Finance and Economics
No.18 Xueyuan Street, Xiasha, Hangzhou 310018, China
zhangshuai@zufe.edu.cn, chenyong@zufe.edu.cn

Xiaoling Huang*
School of International Education
Zhejiang University of Finance and Economics
No.18 Xueyuan Street, Xiasha, Hangzhou 310018, China
*Corresponding author: huangxl@zufe.edu.cn

Yishuai Cai
School of Information
Zhejiang University of Finance and Economics
No.18 Xueyuan Street, Xiasha, Hangzhou 310018, China
caiyishuai@zufe.edu.cn

Abstract: Online feedback is an effective way of communication between government departments and citizens. However, the daily high number of public feedbacks has increased the burden on government administrators. The deep learning method is good at automatically analyzing and extracting deep features of data, and then improving the accuracy of classification prediction. In this study, we aim to use the text classification model to achieve the automatic classification of public feedbacks to reduce the work pressure of administrator. In particular, a convolutional neural network model combined with word embedding and optimized by differential evolution algorithm is adopted. At the same time, we compared it with seven common text classification models, and the results show that the model we explored has good classification performance under different evaluation metrics, including accuracy, precision, recall, and F1-score.
Keywords: public feedback, deep learning, text classification, convolutional neural network, differential evolution algorithm.

1 Introduction

In recent years, online feedback has played an increasingly important role in linking government and citizens. People can directly express their problems and opinions to the government departments by submitting their feedbacks online. It will help the government to better carry out their work. However, how to deal with hundreds of public feedbacks a day efficiently and deliver them to the appropriate departments is a big challenge for government administrators. Therefore, it is very necessary to explore a method to realize the automatic classification of public feedbacks, and to help government administrators improve their working efficiency and reduce their workload.

With the development of information technology, the field of text mining has attracted more and more scholars' attention [14, 20]. Feedback classification belongs to the text classification, which is an important branch of text mining. Text classification includes text segmentation, stop-word removal, word vector representation, feature selection and classification [14]. In order to improve the accuracy of classification, several studies have improved the selection of text features and classifier [5,35]. Meanwhile, with the development of deep learning technology, convolutional neural network (CNN) [17] has shown excellent performance in solving many problems such as visual recognition [1], image recognition [10], and text classification [32].

In this study, we apply CNN to classify public feedbacks and assign them corresponding classes of labels, indicating which department should be responsible for the problems formulated in the public feedbacks. We obtained the public feedback data from a municipal government in China. The data was processed by distributed representation and word embedding method [15], and then using the CNN to automatically extract the deep text features and complete the classification of public feedbacks. To enhance the classification accuracy of the CNN, we adopt a heuristic optimization algorithm, differential evolution (DE) algorithm [25], for the selection and optimization of network parameters. In order to verify the effectiveness of the explored model, we compared the performance with other common text classification models, and the results show that the CNN has a good advantage in dealing with the classification of public feedbacks.

The remainder of this paper is organized as follow. Section 2 describes the related work on text classification and deep learning. Section 3 presents the content of the data, the way of text representation, the selected text classification model, the parameter optimization algorithm, and the corresponding performance evaluation metric. Section 4 analyzes and discusses the classification performance. In the last section of this paper, we made a summary and put forward some prospects for the future.

2 Related work

In recent years, data mining and analysis technology has been widely concerned and applied in various fields [6,7]. Among them, as an important part of data analytics, classification technology has been developed rapidly in text processing [8], image processing [34] and other fields. Specially, text classification is to classify text into predetermined categories according to the characteristic of the text, which has been applied in some fields. For example, Li et al. [19] proposed a hybrid classification model based on sentiment dictionary, support vector machine (SVM) and k-nearest neighbor (KNN) for sentiment classification of Chinese micro-blogs. Chau et al. [4] improved neural network algorithm to achieve multilingual text classification task. Sabbah et al. [27] proposed four modified frequency-based term weighting schemes, which is combined with common text classifiers such as SVM, KNN, naive Bayes (NB) and extreme learning machine, and tested in the text classification corpora. Liu and Peng [22] considered the statistics of positive and negative samples in the method of term frequency-inverse document frequency (TFIDF), and experiments were carried out in three real-world datasets. Nevertheless, these machine learning classification models only extract shallow text features, but cannot automatically and deeply extract text features hidden in context.

At the same time, deep learning technology has been applied to the field of text classification. Li et al. [21] used deep belief networks for the classification of web spam. Sun et al. [29] used a deep neural network model based on restricted boltzmann machine optimization to realize the sentiment classification of micro-blog text; its experimental results show that this model is more suitable to deal with the classification of short-length text than other traditional classification models, such as NB or SVM. In addition, CNN is also a popular deep learning algorithm. Unlike other deep learning algorithms, CNN has the characteristics of weight sharing and local perception [13]. These two characteristics make CNN show its excellent performance in the ability of spatial feature extraction and high-dimensional data processing. Because the distribution of text has some spatial correlation, CNN has been widely used to solve the problem of text classification. For instance, Gando et al. [12] used a deep CNN model to achieve the automatic classification of illustrations in photographs. To solve the problem of data sparseness in the process of text classification, Wang et al. [32] proposed a model of hybrid word embedding clustering and CNN, and its effectiveness is verified by two open benchmarks. In addition, in order to improve the performance of the neural network, Ijjina and Chalavadi [18] proposed using genetic algorithm to optimize the weights of network parameters in order to improve image classification accuracy. To improve the efficiency of model fitting, Trivedi et al. [31] proposed to optimize the parameters of CNN fully connected layer by using genetic algorithm.

As far as we know, there are few studies on improving the work efficiency of government administrators in handling the classification of public feedbacks. Accordingly, to ease the burden of government administrators, we applied a CNN model to automatically classify public feedbacks by combining word embedding and optimizing network parameters with DE algorithm.

3 Method

In this section, we will describe in detail the text classification model based on CNN used in this study. The flow chart of the model, as shown in Figure 1, including three modules: text representation, classifier training, and performance evaluation.

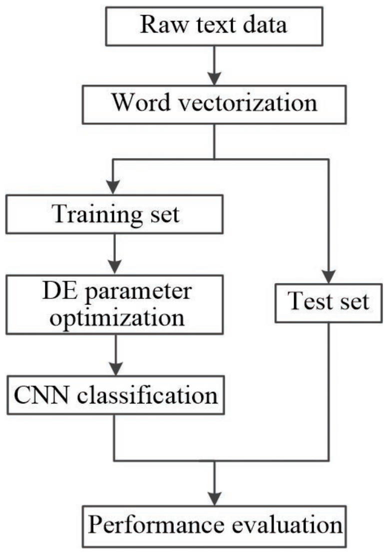


Figure 1: The flow chart of the text classification model

3.1 Data collection

In this paper, the dataset contains 4257 short-length texts, covering 22 categories, which is provided by a municipal government in China. The dataset is divided into three parts: training set, validation set, and test set, with the proportion of 80%, 10%, and 10%, respectively. The name of each category and the corresponding number of feedbacks are detailed in Table 1. Each feedback describes some suggestions or opinions written by the public to the relevant government departments. The meaning of each category is that the problem formulated in the feedback should be solved by the corresponding department in that category. In addition, it should be noted that each text has only one corresponding category.

Table 1: Names and feedback numbers of 22 categories

Category Name	**Number of Feedbacks**
Population and Family Planning Commission	42
Disabled Persons' Federation	29
Security Supervision Bureau	39
National Tax Bureau	43
City Management Enforcement Bureau	503
Electric Power Bureau	119
Housing Administration	382
Industrial and Commercial Bureau	152
Public Security Bureau	688
Sports Bureau	57
Bureau of Land and Resources	174
Environmental Protection Bureau	131
Commission for Discipline Inspection	98
Traffic Bureau	190
Education Bureau	401
Brigade Committee	66
Weather Bureau	56
Health Bureau	256
Pricing Bureau	507
Tobacco Bureau	112
Post Office	120
Court	92

3.2 Text representation and text classification model

In this section, we will describe the details of the text classification model. The structure of the model is shown in Figure 2.

The first part of the structure is the operation of text representation. In view of the fact that most words are meaningful to Chinese text classification as well as the ability of CNN to process high dimensional data, this study will not remove the stop words like general text classification. To reduce the dimension disaster of text representation, we will use word embedding method [15] to reduce the dimension of word vector space. Assuming that the vector dimension of each word is M, by using distributed representation method [15], map each word to a dictionary of common words with a capacity of 5000. The corresponding matrix size of a sentence will be 5000*M. However, we assume that a sentence has up to N words, and by word embedding we can reduce the matrix size to N*M, of which N is far less than 5000.

The second part of the structure is the operation of text classification. First, the above-mentioned word vector matrix is used as input of CNN model. By using 256 different convolution kernels and padding operation [23] for feature extraction, we can get 256 corresponding feature maps that have the same size as that before the convolutional operation. Second, in order to reduce the number of parameters of the network and prevent the overfitting issues, we took a max-pooling operation [23] on the mapping results. It should be mentioned that the max-pooling operation preserves the salient features extracted and discard the less important features, which can facilitate the fitting of the model. Then, by linking the results of the pooling with the neurons in the fully connected layer and using the rectified linear unit incentive function [24],

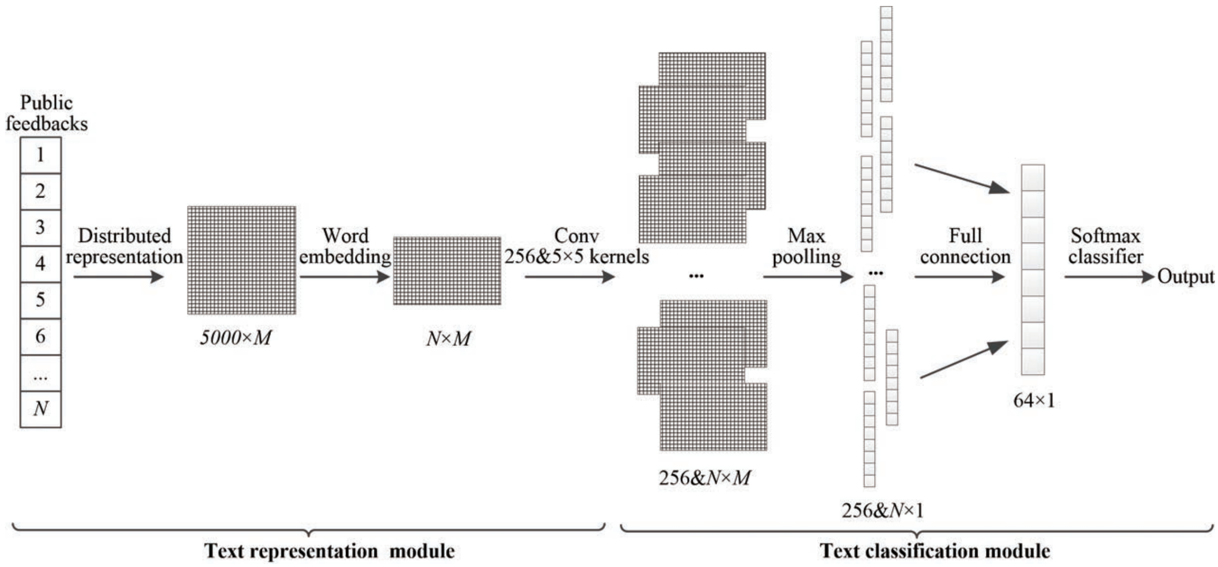


Figure 2: The structure of CNN classification model

the output vector is obtained. Finally, the output vector is put into the Softmax classifier [28] to get the classification results, which is the probability of classifying into each category. In this study, the size of the convolution kernels we use is 5*5, and the loss function is cross-entropy [2].

3.3 Parameter optimization

As we know, the structure of neural network and its training parameters have a great influence on the performance of network. The parameters include the size of convolution kernel, the number of neurons in the fully connected layers, the learning rate, etc. Therefore, in this paper, we will use the differential evolution algorithm [25] to optimize the parameters of convolution neural network. DE is a parallel direct search algorithm proposed by Price in 1996, which has strong global optimization capability. It includes four steps: initializing the population, mutation, crossover, and selection [25] . The implementation steps of the algorithm are shown in Table 2. A chromosome [25] consists of six genes, the number of neurons in the fully connected layer, the number of convolution filters, the size of batch, the values of dropout [23], and the learning rate, respectively. Among them, the first four values are integers and the last two values are decimal. In addition, the fitness function of DE algorithm is the error evaluation function of CNN during training.

3.4 Evaluation metric

In this study, we use four metrics to evaluate the effectiveness of the model used, including accuracy, precision, recall, and F1-score. The calculation of these metrics is shown in Equations (1-4).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

Table 2: The pseudocode of the DE algorithm

```
Initialize population size P, chromosome length L, crossover rate CR, and
chromosomes
While (termination condition is not satisfied)
    For i in range(0, P)
      Random selection of three chromosomes x_a ,x_b, x_c, then perform the mutation
      operation and get the new chromosome v_i
        For j in range(0, L)
            If rand(0,1) < CR or rand(0,L) == j
                Perform the crossover operation
            Else
                Do not perform crossover operation
            End if
            Get the new chromosome u_i
        End for
    End for
    For i in range(0, P)
        If fitness(u_i)> fitness(x_i)
            Update chromosome x_i
        Else
            Not update chromosome x_i
        End if
    End for
End while
The optimal fitness of the chromosome as the final output
```

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

Where TP represents the numbers of correctly classified as the target category, TN represents the number of correctly classified as another category, FP represents the numbers of misclassified as target category, and FN represents the numbers of misclassified as another category. We can draw from the formula that the value of all the metrics is between 0 and 1, and the closer the value is to 1, the better the classification performance of the model is.

4 Computational experiments

This section will evaluate the performance of the model. We compare the CNN model based on DE optimization (CNN-DE) with the common text classification models to verify the effectiveness of the model. All experiments were developed under the Python version 3.6. The core configuration of the computer includes Inter Core i7-8700k processors with a 3.7GHz basic frequency and 32G running memory.

In order to verify the effectiveness of the CNN-DE model, this paper compares it with the commonly used seven text classification models in the same dataset. Among them, six models are combined by the basic classifier and TFIDF method. The classifier includes NB [33] decision tree (DT) [26], random forest (RF) [3], SVM [30], KNN [9], and gradient boost decision tree (GBDT) [11]. TFIDF is a common word frequency based vector representation method. DT, RF, and GBDT are based on tree structure for classification and forecasting, which has the advantages including simple structure and fast training speed. SVM is a supervised learning algorithm by constructing the maximal margin hyperplane, which is often used to solve the multi-dimensional nonlinear classification problem. KNN is an unsupervised learning algorithm, which can make classification according to the preset number of classes. In addition, the long short-term memory network (LSTM) [16] model does not adopt the TFIDF method, and the features are extracted automatically directly. LSTM is a deep neural network algorithm with memory function, which has often been used to solve the problem of sequence forecasting and classification. For ease of description, the comparison models are recorded as TFIDF-NB, TFIDF-DT, TFIDF-RF, TFIDF-SVM, TFIDF-KNN, TFIDF-GBDT, and LSTM, respectively.

In addition, all of these models are developed in a Spyder environment, where the TFIDF-NB, TFIDF-DT, TFIDF-RF, TFIDF-SVM, TFIDF-KNN, and TFIDF-GBDT models are implemented using the Sklearn toolkit, the CNN-DE and LSTM are implemented using open source deep learning framework Tensorflow 1.8.0.

The performance evaluation results of CNN-DE and other comparison models are shown in Figure 3. Table 3 lists the detailed evaluation results. We can find that the CNN-DE model outperforms with the other seven models under the same text dataset according to the four performance evaluation metrics used. The accuracy, precision, recall, and F1-score of CNN-DE model are 0.829, 0.830, 0.829, and 0.825, respectively. More specifically, according to the values of the accuracy, precision and recall, the CNN-DE model has improved 15.1%, 12.3%, and 15.1%, respectively compared to the TFIDF-DT model, and the F1-score of the CNN-DE model is 14.7% better than the TFIDF-KNN model. It is worth noting that the values of precision and recall of CNN-DE model are not only relatively high, but also the difference between the two is small compared with other models. It shows that the CNN-DE model has a good balance in dealing with text classification problems. This is because the CNN-DE model not only makes use of the word frequency information of text, but also automatically extracts the abstract semantic information, and then excavates the deeper text features to achieve better classification performance. At the same time, the DE algorithm optimizes the network parameters of CNN and further improves the performance of the model. However, for other comparative models, the classifiers can only extract the shallow text feature information, and the deep high-dimensional feature cannot be obtained. On the other hand, the word vector representation method based on TF-IDF is prone to cause dimension disaster problem, thus reducing the classification efficiency. In addition, the parameters of the models have been adjusted by repeated experiments, and there is still room for further improvement.

Furthermore, the evaluation results of CNN-DE model for each category are shown in Table 4, where ‘Support’ represents the number of test samples for each category. We can observe that the CNN-DE model has good classification performance under sufficient sample size. For example, the Post Office category has a test sample of 14 feedbacks, with a F1-score of 0.93.

Table 3: Detailed information on the evaluation results of each model

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
CNN-DE	0.829	0.830	0.829	0.825
LSTM	0.716	0.730	0.716	0.715
TFIDF-NB	0.728	0.762	0.728	0.717
TFIDF-DT	0.678	0.707	0.678	0.682
TFIDF-RF	0.739	0.783	0.739	0.744
TFIDF-SVC	0.700	0.829	0.700	0.710
TFIDF-KNN	0.683	0.720	0.683	0.678
TFIDF-GBDT	0.711	0.779	0.711	0.726

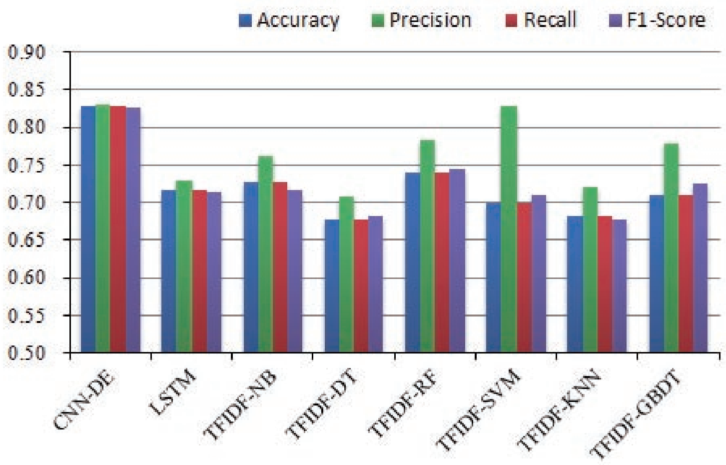


Figure 3: Performance comparison of the text classification models

Table 4: The evaluation results for each category of the CNN-DE model

Category Name	**Precision**	**Recall**	**F1-score**	**Support**
Population and Family Planning Commission	0.71	0.50	0.59	10
Disabled Persons' Federation	1.00	1.00	1.00	4
Security Supervision Bureau	1.00	1.00	1.00	1
National Tax Bureau	1.00	1.00	1.00	2
City Management Enforcement Bureau	0.85	0.89	0.87	53
Electric Power Bureau	1.00	0.81	0.90	16
Housing Administration	0.90	0.88	0.89	40
Industrial and Commercial Bureau	0.61	0.85	0.71	13
Public Security Bureau	0.72	0.81	0.76	59
Sports Bureau	0.00	0.00	0.00	3
Bureau of Land and Resources	0.80	0.84	0.82	19
Environmental Protection Bureau	0.86	0.86	0.86	22
Commission for Discipline Inspection	0.86	0.86	0.86	7
Traffic Bureau	0.77	0.48	0.59	21
Education Bureau	0.91	0.85	0.88	34
Tourism Committee	0.80	1.00	0.89	8
Weather Bureau	1.00	0.80	0.89	5
Health Bureau	0.81	0.84	0.82	25
Pricing Bureau	0.92	0.94	0.93	48
Tobacco Bureau	0.78	0.70	0.74	10
Post Office	0.93	0.93	0.93	14
Court	0.69	0.75	0.72	12

5 Conclusion and future work

In order to help government administrator improve the efficiency of handling public feedbacks and reduce the burden of work, we applied the CNN text classification model to automatically classify feedbacks. The model consists of text representation and text classification. Before training CNN classifier, to avoid the problem of dimension disaster caused by the excessive size of text input vectors, we used the word embedding method to reduce dimension. To improve the fitting ability of CNN, we used the DE algorithm to optimize the network parameters of CNN. To verify the effectiveness of the model, we compared the CNN-DE model with other seven common text classification models in terms of accuracy, precision, recall, and F1-score. The experimental results show that the CNN-DE model has better performance under each evaluation metric.

Meanwhile, this study still has some limitations that need to be addressed in future work. For example, the amount of data used in the model training is limited, and it has a certain influence on the fitting of the model. In the future, we will use a richer set of data to train model to improve its accuracy. In addition, we will improve the DE algorithm and explore more effective parameter optimization methods to optimize the CNN structure, such as genetic algorithm.

Acknowledgment

This work was supported by National Social Science Foundation of China (No. 17BGL047), National Natural Science Foundation of China (No. 51875503, No. 51475410).

Bibliography

[1] Bai, D.D.; Wang, C.Q.; Zhang, B.; et al. (2018); Sequence searching with CNN features for robust and fast visual place recognition, *Computers & Graphics*, 70, 270–280, 2018.

[2] Bishop, C.M. (1995); *Neural Networks for Pattern Recognition*, Oxford University Press, UK, 1995.

[3] Breiman, L. (2001); Random forests, *Machine Learning*, 45(1), 5–32, 2001.

[4] Chau, R.N.; Yeh, C.S.; Smith, K.A. (2005); A neural network model for hierarchical multilingual text categorization, *In Proceedings of the 2nd International Symposium on Neural Networks*, May 30–June 1, Chongqing, China, 238–245, 2005.

[5] Chen, J.N.; Huang, H.K.; Tian, S.F.; et al. (2009); Feature selection for text classification with Naive Bayes, *Expert Systems with Applications*, 36(3), 5432–5435, 2009.

[6] Dai, Y.; Wu, W.; Zhou, H.B.; et al. (2018); Numerical simulation and optimization of oil jet lubrication for rotorcraft meshing gears, *International Journal of Simulation Modelling*, 17(2), 318–326, 2018.

[7] Dai, Y.; Zhu, X.; Zhou, H.; et al. (2018); Trajectory tracking control for seafloor tracked vehicle by adaptive neural-fuzzy inference system algorithm, *International Journal of Computers Communications & Control*, 13(4), 465–476, 2018.

[8] Du, C.; Huang, L. (2018); Text classification research with attention-based recurrent neural networks, *International Journal of Computers Communications & Control*, 13(1), 50–61, 2018.

[9] Duda, R.O.; Hart, P.E. (1973); *Pattern Classification and Scene Analysis*, Wiley, USA, 1973.

[10] Ferreira, A.; Giraldi, G. (2017); Convolutional neural network approaches to granite tiles classification, *Expert Systems with Applications*, 84, 1–11, 2017.

[11] Friedman, J.H. (2001); Greedy function approximation: a gradient boosting machine, *Annals of Statistics*, 29(5), 1189–1232, 2001.

[12] Gando, G.; Yamada, T.; Sato, H.; et al. (2016); Fine-tuning deep convolutional neural networks for distinguishing illustrations from photographs, *Expert Systems with Applications*, 66, 295–301, 2016.

[13] Goodfellow, I.; Bengio, Y.; Courville, A. (2016); *Deep Learning*, The MIT Press, 2016.

[14] Gupta, V.; Lehal, G.S. (2009); A survey of text mining techniques and applications, *Journal of Emerging Technologies in Web Intelligence*, 1(1), 60–76, 2009.

[15] Hinton, G.E. (1986); Learning distributed representations of concepts, *In Proceedings of the 8th Annual Conference of the Cognitive Science Society*, August 15–17, Hillsdale, Canada, 1–12, 1986.

[16] Hochreiter, S.; Schmidhuber, J. (1997); Long short-term memory, *Neural Computation*, 9(8), 1735–1780, 1997.

[17] Hubel, D.H.; Wiesel, T.N. (1962); Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *The Journal of Physiology*, 160(1), 106–154, 1962.

[18] Ijjina, E. P.; Chalavadi, K.M. (2016); Human action recognition using genetic algorithms and convolutional neural networks, *Pattern Recognition*, 59, 199–212, 2016.

[19] Li, F.F.; Wang, H.T.; Zhao, R.C.; et al. (2017); Chinese micro-blog sentiment classification through a novel hybrid learning model, *Journal of Central South University*, 24(10), 2322–2330, 2017.

[20] Li, N.; Wu, D.D. (2010); Using text mining and sentiment analysis for online forums hotspot detection and forecast, *Decision Support Systems*, 48(2), 354–368, 2010.

[21] Li, Y.C.; Nie, X.Q.; Huang, R. (2018); Web spam classification method based on deep belief networks, *Expert Systems with Applications*, 96, 261–270, 2018.

[22] Liu, L.; Peng, T. (2014); Clustering-based method for positive and unlabeled text categorization enhanced by improved TFIDF, *Journal of Information Science and Engineering*, 30(5), 1463–1481, 2014.

[23] Mou, L.C.; Ghamisi, P.; Zhu, X.X. (2018); Unsupervised spectral-spatial feature learning via deep residual conv-deconv network for hyperspectral image classification, *IEEE Transactions on Geoscience and Remote Sensing*, 56(1), 391–406, 2018.

[24] Nair, V.; Hinton, G.E. (2010); Rectified linear units improve restricted boltzmann machines, *In Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, June 21–24, Haifa, Israel, 807–814, 2010.

[25] Price, K.V. (1996); Differential evolution: a fast and simple numerical optimizer, *In Proceedings of the North American Fuzzy Information Processing Society*, June 19–22, New York, USA, 524–527, 1996.

[26] Quinlan, J.R. (1987); Simplifying decision trees, *International Journal of Man-machine Studies*, 27(3), 221–234, 1987.

[27] Sabbah, T.; Selamat, A.; Selamat, M.H.; et al. (2017); Modified frequency-based term weighting schemes for text classification, *Applied Soft Computing*, 58, 193–206, 2017.

[28] Socher, R.; Perelygin, A.; Wu, J.; et al. (2013); Recursive deep models for semantic compositionality over a sentiment treebank, *In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, October 18–21, Seattle, USA, 1631–1642, 2013.

[29] Sun, X.; Li, C.C.; Ren, F.J. (2016); Sentiment analysis for Chinese microblog based on deep neural networks with convolutional extension features, *Neurocomputing*, 210, 227–236, 2016.

[30] Suykens, J.A.; Vandewalle, J. (1999); Least squares support vector machine classifiers, *Neural Processing Letters*, 9(3), 293–300, 1999.

[31] Trivedi, A.; Srivastava, S.; Mishra, A.; et al. (2018); Hybrid evolutionary approach for devanagari handwritten numeral recognition using convolutional neural network, *Procedia Computer Science*, 125, 525–532, 2018.

[32] Wang, P.; Xu, B.; Xu, J.M.; et al. (2016); Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification, *Neurocomputing*, 174, 806–814, 2016.

[33] Zhang, W.; Zhang, Z.; Chao, H.C.; et al. (2018); Kernel mixture model for probability density estimation in Bayesian classifiers, *Data Mining and Knowledge Discovery*, 32(3), 675–707, 2018.

[34] Zhang, W.; Zhang, Z.; Qi, D.; et al. (2014); Automatic crack detection and classification method for subway tunnel safety monitoring, *Sensors*, 14(10), 19307–19328, 2014.

[35] Zhou, Y.; Li, Y.W.; Xia, S.X. (2009); An improved KNN text classification algorithm based on clustering, *Journal of Computers*, 4(3), 230–237, 2009.

Author index