High-speed Train Control System Big Data Analysis Based on the Fuzzy RDF model and Uncertain Reasoning

D. Zhang

Dalin Zhang
National Research Center of Railway Safety Assessment,
Beijing Jiaotong University, Beijing, China
dalin@bjtu.edu.cn

Abstract: China High-speed train control system is a combination of computer, communication and control. Its events are diverse, including sensor data stream, GPS signal, GSM-R transmission data, real-time video monitoring data, train control software data, etc. These data have the typical characteristics of big data. If these data are well applied, this will be of great help to operations, maintenance, safety, passenger services, etc. This paper presents an efficient analysis method based on the fuzzy RDF model and uncertain reasoning for high-speed train control system big data. We have used the method proposed in this paper to analyze the data of the Beijing-Tianjin high-speed train control system. The experimental results show that the method proposed in this paper has good efficiency and scalability for the analysis of large data with different structures, types and context sensitive from train control system.

Key Words: train control system; fuzzy RDF; linked data; D-S theory; uncertain reasoning

1. Introduction

Railway transportation system is a complex system with many factors. It has been unable to meet the “high transport capacity, high efficiency, high safety, high quality of service” requirements relying on traditional theory and technical methods. The intelligent rail transport system has become a trend, such as the European rail traffic management system, the Japanese Cyber Rail, the US IRS (Intelligent Railway System), etc. In recent years, the IBM Smarter Railroad[1], CISCO Smart + Connected Railway[2], SIMENS Intelligent Train[3], they have promoted the development of the intelligent railway.

High-speed train control system makes use of the information provided by the ground, the distance between the target and the route, and generated the control curve automatically on the train control equipment. China has developed the standard of Chinese train control system(CTCS)[4] in 2008, which have 4 level totally. The CTCS-3 uses the GSM-R wireless communication for bidirectional transmission of information between the train and the ground. It uses the wireless blocking center (RBC) to generate traffic permits, uses the track circuit to achieve the train occupancy check, and uses a transponder to achieve train positioning. At the same time, it has the function of CTCS-2.

High-speed train control system is a combination of computer, communication and control. Its events are diverse, including sensor generated data stream, GPS signal, GSM-R transmission data, real-time video monitoring data, train control software data, etc. These data have the typical characteristics of large data, including massive data, distributed, complex, context-sensitive, etc. If
these data are well applied, this will be of great help to high-speed rail operations, maintenance, safety, passenger services, etc.

The dynamic monitoring system (DMS) is a comprehensive system for dynamic monitoring of the use status of the relevant equipment of the railway control system. It can analyze the data, guide the maintenance, and provide the basis for the decision. DMS consists of vehicle information collection, ground data center and query terminal. A variety of complex data analysis beyond our current ability to use DMS range. How to effectively extract the useful knowledge from DMS's massive data is an urgent need for the current situation. Because these excavated knowledges can help the maintenance department to achieve rapid and accurate fault diagnosis, and provide reasonable preventive maintenance decision support.

For big data, the current distributed database technology NoSQL and Hadoop can be well handled. If the semantic technology is used to analyze, the main challenge lies in the distributed query and reasoning. Such as streaming media, sensors, the current distributed computing technology can also meet the demand; the challenge of semantic technology is the flow reasoning - that is, in the process of continuous arrival of the data, the incomplete information should be used for reasoning. Such as distributed reasoning can use Hadoop / Storm and the flow reasoning can use continuous (dynamic) queries\(^5\). There are very many types of data in the DMS system, and the structure is very complex and the speed of change is very fast. It not only needs high performance but also needs management and real-time analysis. If it should be needed to meet the above three, there is no good solution.

In view of the above problems, this paper proposes a distributed real-time context-aware data processing method based on the fuzzy linked data and uncertain reasoning. This method extends the RDF model to a fuzzy RDF model, and propose real-time distributed context reasoning based D-S theory.

2. CTCS-3 Data Context Modeling

2.1 CTCS-3 Data

The system structure of CTCS-3 is shown in the following Fig. 1. The real-time data of CTCS-3 can be summarized as follows: (1) RBC data. According to the track circuit, interlocking road and other information to generate traffic permit; through the GSM-R wireless communication system will drive permission, line parameters, temporary speed limit transmission to the CTCS-3 level vehicle equipment; through GSM-R wireless communication system to accept vehicle equipment sent location and train data and other information. (2) GSM-R network. Used to realize two-way communication between vehicle equipment and ground equipment. (3) Transponder. To the vehicle equipment transmission positioning and level conversion information; to the vehicle equipment transmission line parameters and temporary speed limit and other information to meet the needs of the backup system. (4) Secure computer. According to the ground equipment provided by the traffic permit, line parameters, temporary speed limit and other information and train parameters, according to the target distance continuous speed control mode to generate dynamic speed curve to monitor the safe operation of the train. (5) Track circuit. To achieve the train occupancy check, send traffic license information to meet the needs of the backup system. (6) CTC system. Such as train operation monitoring, automatic tracking number, to the point of automatic acquisition, the actual operation of the map automatically generated, dispatch command network issued, the station traffic logs automatically generated, since further realization of the station signal equipment centralized control, The road through the road and shunt control.
2.2 Linked Data

At present, more and more big data applications began to introduce semantic technology, which makes the description of the data more standardized and rich in machine comprehensible semantics. Rich semantic links make the system more open and interoperable, and make big data analysis deep into the “knowledge” level, which requires big data technology can provide a wealth of related functions and simple reasoning ability. Big data technology can effectively solve the distributed environment of the web scale of unstructured information management and utilization issues. And the associated data brings the rich formalized semantics, which is a tool for cross-domain integration and intelligent analysis.

Resource description framework (RDF): RDF is a triplet \( \langle \text{subject}, \text{predicate}, \text{object} \rangle \), which describes the generic model of the resource description, where the subject is a resource with a uniform identifier (URI), or a blank node with no namespace, such as DOI, ISBN Etc.; object can be a resource with a URI, a blank node, or a string value; and a predicate that represents the relationship between the subject and the object.

More and more big data applications are described and encoded using the RDF model. The metadata and ontology make the data semantically recognizable by the machine, enrich the semantics of big data, and make the data have better interoperability. Since the linked data is semantically described, it is no longer a piece of information. When we combine linked data with big data mining, which provides a powerful tool for big data analysis, which implements semantic-based integration in the data. When using the linked data technology in big data applications, we can call the linked big data application.

Linked data. Linked data is a data application form, which uses the URI as a data identifier, and RDF triplet structure as a data model, and based on HTTP, which is a simplified realization of the semantic web, and the intention is to build “data web “.

Linked data (linked data) of the distribution process. The linked data is based on the existing web technology (HTTP, URL and HTML), and the web specification is further standardized and defined with four basic principles.

Depending on the principle of the linked data, the elements of the triad of the linked data should be encoded as RDF as much as possible. Especially the “subject”, must be able to access in an open HTTP way, and the RDF description should contain other more data link. The data in the linked data is not an independent, context-free abstract data, but a clear knowledge unit with a URI.
identifier and an RDF description (including cross-domain links), which is the basic semantic unit managed in the web. Any one thing, people, institutions, places, events, concepts, etc. can be described as a linked data, so the linked data to large data development is inevitable.

2.3 Fuzzy RDF Model

The purpose of context modeling is to describe the data and its environment, which plays an extremely important role in building a context-aware system. At present, the context modeling methods are key-value model, tag configuration model, graph model, object-oriented model, based on the logical model, ontology model. Ontology model has the advantages of knowledge sharing, logical reasoning, easy knowledge reuse, and so on. In this paper, the use of extended RDF modeling method is also a kind of ontology.

This study used Drupal to convert CTCS-3 data into linked data. RDF data model is the basis of linked data, but in specific applications, will inevitably use some domain ontology, such as FOAF, Dublin Core, SKOS, OWL, SIOC and so on. Drupal also supports the import of external ontologies, which can define the mapping between these external and local content models. Drupal provides a way for mapping content types, fields, and nodes to “classes”, “attributes”, and “objects” in the RDF triplet model, namely, subject, predicate, object. In the field of CTCS-3 data analysis, some concepts are not deterministic, such as speed, security, etc., so the RDF model needs to be extended. The context representation of this paper is based on the fuzzy RDF model, and it is optimized for CTCS-3 data processing.

**Fuzzy RDF model (FRDF).** The fuzzy RDF FR is expressed as FR = \( \{ fs, fp, fo \} \), where fs is the fuzzy concept, fp is the set of fuzzy attributes for the subject, fo is a set of fuzzy concept instances. The attributes fp is defined as \( fp = \{ p_1, p_2, ... p_n \} \), \( p_i \) represents an attribute that represents the relationship between the subject and the object; \( fo = \{ o_1, o_2, ... o_n \} \), \( o_i \) represents a fuzzy instance. The function of the fuzzy degree in the fuzzy RDF model is \( \mu: fo \rightarrow [0,1] \). If \( \mu_{o_i} = 1 \), we call \( o_i \) concrete instance.

For example, the “the No G115 train” is a fuzzy concept, meanwhile, it is an instance of another concept “train”, “the speed of No G115 train is very fast” is an example of CTCS-3 data, which can be described by the FRDF model. Firstly, the “No G115”, “train” “speed”, “very”, “fast” can total be described by the fuzzy RDF model as an instance; And the “train” is a fuzzy concept, and it has the Fuzzy attribute “driving speed”, and the value of speed can be expressed as a group of instances (“S (slow)”, “M (middle)”, “F (fast)”).

**Fuzzy contexts.** According to the FRDF model, Fuzzy context FC is defined as \( FC = \langle FS, FP, FO \rangle \), FS is a set of fuzzy concepts, FP is a fuzzy attribute set that represents the relationship between the subject and the object, and FO is a set of fuzzy objects.

2.4 FRDF Model Based On D-S Theory

The D-S theory\(^8\) is a generalization of the Bayesian reasoning method, which mainly takes advantage of the Bayesian conditional probability in probability theory, and requires a priori probability. The D-S theory does not require a priori probability, can describe “uncertainty”, is widely used to deal with uncertain data. The D-S theory classifies the information that people are interested in some identification frameworks.

**Basic probability assignment (BPA).** For any of the recognition frames \( \Theta \) and its subset \( A \), If the following conditions are met: (1) \( m(\phi) = 0 \); (2) \( \sum_{\Theta \in \Theta} m(A) = 1 \); Then, \( m: 2^\Theta \rightarrow [0,1] \) is called the BPA function. \( m(A) = 0 \) is the basic probability assignment of \( A \),
which indicates the degree of belief in A. The BPA is also called the mass function.

**Belief function (Bel).** To express the degree of trust in the proposition, D-S theory introduces the concept of belief function Bel, and its relationship with BPA satisfies \( \text{Bel}: 2^\Theta \rightarrow [0,1], \text{Bel}(A) = \sum_{B \subseteq A} m(B) = 1(\forall A \subseteq \Theta) \).

**D-S rule.** For \( n \) limited mass functions \( m_1, m_2, \ldots, m_n \) on the one recognition frame \( \Theta \), \( \forall A \subseteq \Theta \), then the synthesis of \( n \) belief function is

\[
m(A) = (m_1 \oplus m_2 \oplus \cdots \oplus m_n)(A) = \frac{1}{K} \sum_{A=A_1 \cap A_2 \cap \cdots \cap A_n} m_1(A_1) \cdot m_2(A_2) \cdots m_n(A_n)
\]

\[
K = 1 - \sum_{A_1 \cap A_2 \cap \cdots \cap A_n = \emptyset} m_1(A_1) \cdot m_2(A_2) \cdots m_n(A_n)
\]

K is called the normalization factor, which reflects the degree of conflict among the evidence. When \( k = 0 \), it is called incomplete conflict; when \( 0 < k < 1 \), it is called incomplete conflict; \( k = 1 \), called full conflict.

D-S theory does not well reflect the structure of contextual information and the relationship between them, so its application is limited. To play its superiority, it must use a valid context information representation. The D-S theory uses the evidence recognition framework \( E \) and the conclusion recognition framework \( \Theta \) to divide the information. The FRDF model divides the context into two layers: the context of the low-level acquisition and the context of the high-level reasoning, which can naturally be combined to establish the FRDF model based on the D-S theory. The RDF model has very good scalability. FRDF model can be expanded based on the D-S theory easily.

The FRDF model based on the D-S theory is expressed as \( \text{FRDFDS} = (f_s, f_{p_{D-S}}, f_{o_{D-S}}) \), where the \( f_s \) is still the fuzzy concept. The \( f_{p_{D-S}} \) is the set of fuzzy attributes for the subject which been expanded and contains some special evidence attributes and conclusion attributes, such as “hasConclusions”, “hasEvidence”, “hasBPAvalue” and so on. The \( f_{o_{D-S}} \) is a set of instances of fuzzy concept which also been expanded and contains some special evidence and conclusion instances. The uncertainty context model based on FRDFDS is shown in Fig. 2.

![Fig. 2 Uncertainty context model based on FRDFDS](image-url)
The five attributes are added to the basic FRDF model: the attribute “hasConclusions”, the attribute “hasEvidences”, the BPA numeric attribute “hasBPAValue”, the Bal value attribute “hasBal”, and the time attribute “hasTimetag”. As shown in the figure above, in the context of the FRDFDS model, all data is represented as a unified FRDFDS model, which is the advantage of the linked data, and each data can be considered as a node of the FRDFDS model instance. The “hasTimetag” attribute is used to get the context information for the time. The “hasConclusions” attribute may map a single instance or multiple instance instances. Likewise, the attribute “hasEvidences” corresponds to all Evidence instance collections. In the pervasive computing environment, a conclusion context information may be another conclusion of the evidence context information, so in order to adapt to the context of the dynamic environment, diversity, uncertainty, for the conclusion and Evidence add data attribute “hasBel” And “hasBPAValue”, used to describe the probability of the value. The above FRDFDS model is universal, just as the basic RDF model, the relevant attributes and instances can be added to express the uncertainty information of the context.

**Belief structure.** Given a recognition frame Θ, the evidence space E, the mapping $F: E \rightarrow \Theta$ and the basic belief distribution value m, where Θ and E are finite sets, the quaternion $D = \langle \Theta, E, F, m \rangle$ is called a trust structure. Based on the basic RDFDS model, we can construct the trust structure D based on the uncertain RDF concept model given in this paper, and then creates the ontology instance of the uncertainty context information, and finally uses the evidence combination rule Step by step reasoning process. The mapping between the ontology model and the trust structure is shown in Fig. 2. In the definition of the basic belief of the distribution of m is based on evidence to make the proposition, access to m value, you can on different sources of evidence on the synthesis of the evidence.

### 3. Multi-Layer Real-Time Uncertain Context Reasoning Method

Context reasoning is the key to solving the problem of uncertainty reasoning, and which often depends on the context model. According to the FRDFDS model proposed in this paper, three kinds of reasoning mechanisms can be used in context reasoning\textsuperscript{[9]}, namely, ontology reasoning, rule reasoning and evidence theory reasoning. The rule-based reasoning is to generate new knowledge by matching existing facts with predefined rules. For example, when the sensor information is collected: “John is currently in the bedroom, the indoor curtains closed, the light intensity is dark”, you can infer that “John is sleeping”. The reasoning process of the ontology-based is like the rule-based reasoning process, except that the rules used are defined by the OWL language itself to obtain knowledge that is implicit in explicit definitions and declarations, such as the symmetric property “SymmetricProperty”, the transitive property “TransitiveProperty”, etc. For the reasoning based on D-S theory, the evidence set and the conclusion set are expressed by multiple discriminant frameworks respectively. The synthetic method provides the hierarchical reasoning model of evidence. Uncertainty is transferred layer by layer based on the synthesis rules of D-S Theory. The hierarchical model can add the evidence based on the original synthesis, greatly improving the efficiency of evidence synthesis.

CTCS-3 is the complex application of the Internet of things. the logic-based reasoning method will become more complex. To solve the problem of large data, this paper proposes a distributed multi-layer real-time uncertain context reasoning framework based on FRDFDS model, as shown in Fig. 3.
In this framework, the first layer generates the context state according to the FRDFDS model instance, in which the FRDFDS instance is generated by the complex event processing engine. Since the event processing agent may be multiple, the FRDFDS instance is distributed. CTCS-3 data is collected in real-time, so the context state is continually updated with updated data in real-time. The context state of the i+1 layer can be inferred from the context of the lower layer. The reasoning process combines the traditional rule-based reasoning and the uncertainty reasoning based on the D-S theory.

```
Input: Context[levels][instances] = Context[0][instances0];
//layers represent the level of context; instance is a FRDFDS model instance set;

Output: Context[levels][instances]; // the context at time t;

1 Procedure Main (){
2     initialize Context[levels][instances];
3     initialize context update = false;
4     Procedure reasoning_framework () {
5         for all layers ∈ N and layers > 0 do
6             for all FRDFDS model instance in Context[levels][instances] do
7                 if instance is an uncertain then
8                     Call D-S theory reasoning procedure ds_reasoning ();
9                 else
10                    Call rule-based reasoning Interface;
11             end for
12         end for
13     }  
14     if update = false then
15         Call update_context(Context[levels][instances]);
16         update = true;
17 }
```

Algorithm1: multi-layer real-time uncertain context reasoning framework.

The overall algorithm framework is proposed by this paper in algorithm 1. The algorithm 1 combines traditional rule-based reasoning and the uncertainty reasoning based on the D-S theory. From the algorithm 1, we can see that all context can be expressed by a FRDFDS model instance, and this framework is multi-layer and real-time and has a good adaptability. In algorithm 1, the fuzzy degree is described by the function µ. When all reasoning is finished, the framework will trigger an update_context procedure. the procedure ds_reasoning() and update_context() will be present in algorithm 2 and algorithm 3.
Input: a context instance in instance set; //an instance in one layer context
Output: a context instance in instance set; //the context instance after running a reasoning process;

1 Procedure ds_reasoning(instance) {
   2 initialize conflict factor threshold value = k;
   3 for all conclusions in instance do // conclusion is also a FRDFDS model instance.
      4 get its evidence set and calculate the BPA numeric and the Bal value;
      5 calculate the conflict factor conflict_factor;
      6 if conflict_factor <= k then
         7 calculate the Bal value for this conclusion;
      8 else
         9 calculate the Modified Bal value for evidence set;
         10 calculate the Bal value for this conclusion;
   11 end for
12 }

Algorithm2: uncertainty reasoning based on the D-S theory.

The algorithm 2 gives a complete uncertainty reasoning process based on D-S theory. The whole process is based on the belief structure which can map the FRDFDS model perfectly. The classical evidence combination formula requires that all evidence of participation in the synthesis have the same degree of importance. However, in the pervasive computing environment, the evidence obtained from different sources of evidence may differ in importance and belief, so that the basic credibility of the evidence needs to be corrected before the evidence is synthesized to reflect the different importance of the evidence and belief. The formula for correcting the belief is as follows:

\[
m'_i(A_i) = \begin{cases} 
  Bel(E)m_i(A_i), & A_i \neq \emptyset \\
  1 - \sum m'_i(A_i), & A_i = \emptyset 
\end{cases}
\]

For different context applications, the Bel method is different, but the choice of Bel generally follows the following rules: (1) the importance of the evidence, the higher the belief, the greater the corresponding correction factor, that is, to the important evidence distribution Big weight. (2) the higher the conflict of evidence, the lower the belief, the smaller the corresponding correction factor, that is, it is to be a larger amendment, the conflict part of the confidence to more unknown areas.

The weighted combination rule is proposed to fuse the evidence, which solves the shortcomings of the evidence theory in the case of high degree of conflict between the evidence. However, the reasoning model of the method is not adaptive. Therefore, this paper improves the combination rule with the modified belief formula, the formula is as follows:

\[
m'(A) = (m_1 \oplus m_2 \oplus \cdots \oplus m_n)(A) = \frac{\sum_{A=A_1 \cap A_2 \cap \cdots \cap A_n} \omega_1 m'_1(A_1) \omega_2 m'_2(A_2) \cdots \omega_n m'_n(A_n)}{1 - \sum_{A_1 \cap A_2 \cap \cdots \cap A_n=\emptyset} \omega_1 m'_1(A_1) \omega_2 m'_2(A_2) \cdots \omega_n m'_n(A_n)}
\]

\[
\omega_i = \frac{\sum_{j=1,j \neq i}^n (1 - d_{BPA}(m_i, m_j))}{\sum_{i=1}^n \sum_{j=1,j \neq i}^n (1 - d_{BPA}(m_i, m_j))}
\]

\[d_{BPA}(m_i, m_j)\] is the distance of \(m_i\) and \(m_j\). The formula (4) reflects the different importance and belief of the evidence; resolves the “one-vote veto” problem and the “robustness” problem in the
conflict evidence combination, makes full use of the conflict evidence information, avoids the valid information loss.

\[
\text{Input: } \text{Context}\{\text{layers}\}[\text{instances}] \text{ at time } t \text{ and } t-1;
\]

// the context at time t; layers represent the level of context; instance is a FRDFDS model instance set;

\[
\text{Output: } \text{the updated Context}\{\text{layers}\}[\text{instances}];
\]

\[
\begin{align*}
\text{Procedure update\_context}&(\text{Context}\{\text{layers}\}[\text{instances}]) \{ \\
\quad &\text{for all } \text{layers} \in \mathbb{N} \text{ do} \\
\quad &\quad \text{for all FRDFDS model instance in Context}\{\text{layers}\}[\text{instances}] \text{ do} \\
\quad &\quad \quad \text{initialize similarity threshold value } = s; \\
\quad &\quad \quad \text{get the previous } t-1 \text{ instance } \text{p\_instance}; \\
\quad &\quad \quad \text{calculate the index of similarity between } \text{p\_instance} \text{ and instance}; \\
\quad &\quad \quad \text{if index } \leq s \text{ then} \\
\quad &\quad \quad \quad \text{make an additional with } \text{previous}; \\
\quad &\quad \quad \text{end if} \\
\quad &\quad \text{end for} \\
\quad &\text{end for} \\
\} \\
\end{align*}
\]

Algorithm3: real-time context update based on similarity.

In order to improve the speed of analysis, we do not look back at historical data, but only analyze and update the current node data based on rules, D-S theory. This paper uses the transitive property “TransitiveProperty” to update the state attribute information in Drupal module instead of using based a finite state machine method. For fuzzy data on the same data node, this paper uses fuzzy similarity to update it. In the distributed, multi-layer, real-time uncertain context updating based on similarity, the former context in the different node is fused by the D-S evidence theory. Based on the similarity of fuzzy sets, the new context is fused to the existing context in the context update processing. The fuzzy similarity calculation rules are as follows.

**Fuzzy set similarity.** The recognition framework is \( \Theta = \{A_1, A_2, \ldots, A_n\} \). Fuzzy set \( M \) and \( N \) are two random fuzzy subsets, and the similarity between them is

\[
\psi(M, N) = 1 - \frac{1}{|\Theta|} \sum_{i} |\mu_M(A_i) - \mu_N(A_i)|.
\]

4. **Experiment**

4.1 **DMS**

To monitor real-timely the operation of CTCS-3 equipment, China high-speed train equipped with DMS which consists of on-board information detecting equipment, ground data center and inquiry terminal. The on-board information detecting equipment collect data from ATP, transponder, track circuit and RBC, and then transmits them to the ground center through GPRS/GSM-R/WLAN network. Through this remote monitoring method, the ground center can monitor and deal with the working states and faults of on-board signal equipment. The business logic of DMS includes data acquisition, storage and analysis, and command. The main function of data acquisition part is data sampling and aggregation. The storage and analysis part mines and analyses the running status data
of all related systems comprehensively. And it can realize the early warning analysis and fault diagnosis. Based on the storage and analysis, the command part provides data sharing service for daily production and emergency dispatching of the electrical department, and it can improve the efficiency of the electrical department daily work.

4.2 Data Analysis System Design

At present, the capability of analysis needs to be further improved in DMS. The work of this paper can be seen as an extension of DMS which is integrated with DMS in the way as shown below. As shown in the Fig. 4, the middle part of the figure is the work of this paper and which is marked with a dotted line; the top part is the DMS UI (the “skin”), which is a set of functions that can be used in all aspects of railway operations; the base part is the collection of nodes—the DMS’s data pool.

![Fig.4 The network topology of DMS](image)

Fig.4 The network topology of DMS

It is easy to see from the left of the middle of the Fig. 5 that our analysis method is based on Drupal module\(^{[10]}\). The resource description framework (RDF) is the standard framework for semantic Web and is a recommended framework for linking data movement. Linked Data technology enables network data to be interoperable, reusable, and easier to use. Drupal 7 is the first mainstream content management system to support semantic web technology in its core, which can publish linked data by exposing content with RDF. Our main work in this paper is layout on the right of the middle of the figure. We extend the RDF model to a FRDFDS model based on the D-S theory firstly. All data in the DMS’s data pool will be expressed as a FRDFDS data. We also
customize the data reasoning rule on the nodes by our reasoning rule module. Accordingly, the
query plugin also was expanded and programmatically filter and display content. We implemented
the algorithm presented in this paper as two functional plugins deployed to Drupal Modules.

4.3 Experimental Analysis

As shown in Fig. 4, there are two types of DMS, one is deployed in the railway bureau and the
other in the Ministry of Railways. Our experiment was deployed DMS data center in Beijing
communication and signaling section of Beijing railway bureau. We selected Beijing-Tianjin
inter-city train control system data as the research object. Specifically, we selected two work area’s
real-time data from 6 am to 6 pm in one day as experimental data. We collect the performance of
the algorithms every half hour and analyze it. Our experiment is detailed, we try to find the various
problems of our algorithm. Experiments were divided into 3 groups simultaneously on 3 servers
respectively. The first group analyzes the data of a work area. Its purpose is to analyze the influence
of the number of layers on the performance of the reasoning algorithms. The second group analyzes
the data of two work areas. Its purpose is to analyze the effect of the increase of the number of
nodes on the performance of the algorithm. The third group also analyzes one work area data, but
its analysis algorithms were replaced by proposed in paper [8]. We have implemented the method in
paper [8] and compared it with the reasoning method in this paper.

These three IBM x3650 M4 servers are connected with DMS data center through a high-speed
network, each with 2 Intel Xeon E5-2600 processors and 4GB memory, and the operating system of
them are ubuntu 14.04. The Drupal 8.0 with our algorithm was deployed in every server
respectively which can automatically generate a variety of FRDFDS nodes.

The first group experiment tested the distributed multi-level real-time uncertain context
reasoning method. The results are shown in Fig. 6. According to Fig. 6, we recorded the time of the
four-tier query respectively, where the first layer is Drupal based data query time, it reflects the
performance of Drupal; the second layer, the third layer, the fourth layer were a higher level of
knowledge reasoning inquire. We summed up the 24 knowledge which needs to reason through the
algorithm in the preliminaries. This knowledge includes certainty and uncertainty, which are
distributed in these three layers. As can be seen from the figure, the results showed that the
algorithm performance is not significantly decreased when the amount of data is increased to a
certain value from the level0 to level2; the level3 layer calculation has increased significantly. This
is because we deployed several uncertain reasoning, for example, safety, flexibility and so on.

As shown in the Fig. 7, we compare the results of the first set of experiments with the experimental

![Event partition by spatial context](image-url)
results of the second group (dense data and sparse data), when the number of data increases, the performance of the methods is decreased, because the increase of the number leads to more complexity of the state rule and uncertain calculation. The column chart in the figure represents the one work area sparse data, and the line chart represents the two areas dense data. But our algorithm does not increase much with the increase of experimental data, this is because the method is based on the current node, and we will update the current node based on the fuzzy similarity at the end of each calculation, as shown in Algorithm 3. This shows that our algorithm has good scalability.

To more intuitive statistics of our algorithm efficiency, we counted the average process time of each layer by our algorithms with the increase in the number of layers in different types of data. As shown in the following Fig. 8, when the algorithm handles the data (data1) of one work area, the average time of level 0 is 34.0 ms; the average time of level1 is 46.9 ms; the average time of level2 is 57.1 ms; the average time of level3 is 98.3 ms. When the algorithm handles the data of two work areas (data2), the average time of level 0 is 47.5 ms; the average time of level1 is 65.0 ms; the average time of level2 is 86.2 ms; the average time of level3 is 110.3 ms. With the increase in data, level3 processing time increased by 12 ms. This also shows that it has a better performance, because our method uses a similarity-based method, avoiding the calculation of many context nodes of the top level.

![Fig.7 Performance comparison on two data set](image1)

![Fig.8 Performance comparison of our method on two data set](image2)

We have implemented Jousselme’s method and compared it with the reasoning method in this paper. The purpose of experiment 3 is to evaluate the correctness of the proposed method and
Jousselme’s in 24 point-in-time, and the results are shown in Fig. 9. In the initial period, the correctness of this algorithm is slightly lower than the paper 12 method; at time point 9, the correctness of this method begins to be higher than the paper 12 method; then, the correctness of the method is improved, and at time point 15, the correctness of this method reaches a steady state.

Due to the existence of conflicting evidence, the use of classical evidence theory for reasoning may result in incorrect results, and improved combinatorial rules can improve this situation. Paper 8 method proposed a weighted combination of rules for evidence fusion, to solve the shortcomings of high conflict between the evidence, but the lack of self-adaptability of the method. On the basis of this idea, this paper first gives the belief distribution table by the expert and then adjusts it by human intervention in the concrete calculation. If there is no human intervention, the original credit distribution table is calculated. In the first eight time points, we have a corresponding fine-tuning of the credit rating table and achieved good results.

In general, as can be seen from the above experiments, the research method in this paper is effective in dealing with the distributed context sensitive complex event of the internet of things. What’s more, when in large-scale networking applications, it has better performance than the general method and scalability. In addition, this method has high correctness in combination with expert dynamic intervention, and the experimental results show that it has attractive usability in the field of dynamic control.

![Graph showing correctness over time](image)

Fig.9 Evaluate the correctness of the proposed method and method in [8]

5. Related Research

5.1 Railway Intelligent Transportation System

The railway management department has been using the “failure - repair” mode of work. However, efficient railway operations require timely maintenance and repair of equipment, so as to avoid the occurrence of a failure, by analyzing the evolution of equipment status\(^{[11]}\). The distributed inspection and monitoring system continues to collect infrastructure status data and train operating status data, which generate relevant information through conversion. The evolutionary trend of this information can be analyzed to gain knowledge of the evolution of the infrastructure state, which provides decision support for preventive maintenance of the infrastructure.

In recent years, with the development of railway information construction and intelligent transportation management, China's railway transport system\(^{[12]}\) has gradually built a number of application-oriented management information systems. At present, these information management systems deal with different aspects of vehicle, infrastructure operation and maintenance management, which have independent organizational structure and produce different forms of data. If we want to use the data of each information system to provide support for the upper application, you must first solve the problem of data sharing. Based on the above, we can realize the intelligent
processing of data, using data fusion analysis, expert system, data mining, knowledge reasoning and other technologies, in the field of warning, traffic control, integrated scheduling, resource management, operation management, fault analysis. This is the railway intelligent transportation system (RITS), which is proposed according to the actual situation of China's railway. Through the coordination of various information systems. RITS makes each information system work together. In order to realize the RITS, many scholars have carried out research on basic theoretical methods such as metadata sharing based on metadata\textsuperscript{13}, general data model research based on XML\textsuperscript{14}, research of Internet of things\textsuperscript{15}, cloud computing technology research\textsuperscript{16}, knowledge reasoning\textsuperscript{17} and so on.

In the aspect of infrastructure maintenance, the gray prediction system theory is applied to the orbital geometric irregularity detection data, which can reflect the stochastic volatility of the track quality with time. It is suitable for long-term forecasting and provides decision support for the rational development of maintenance and repair plan.

In summary, the scholars have begun to study the use of cloud computing and distributed systems to collect data and related information\textsuperscript{18}, the use of data mining technology for rail transport and infrastructure management, maintenance departments to provide useful information. However, the development of these work is mainly focused on transport planning and operation and management. In the communication, signal system, there is no relevant research and applications are reported.

The train control system and the GSM-R network used in the high-speed train are large and complex systems. The application of new technologies makes the workload and the difficulty of work greatly increased in the infrastructure maintenance department. A variety of complex data analysis process greatly exceeded our current maintenance capacity. In addition, our maintenance personnel do not have enough time and energy to analyze the data, which leads to a lot of data idle, and has not been fully utilized. How to effectively analyze the massive data produced by various detection and monitoring devices, and get useful rules and knowledge that is an urgent task for high-speed rail. It helps the maintenance department to achieve rapid and accurate fault diagnosis and provide reasonable support for preventive maintenance and maintenance decisions.

5.2 Context Sensitive Event Processing

The context model plays an important role in the development and application of the system in the context. In the context of the relevant research, various context representation models were presented\textsuperscript{19}, including key-value model, object oriented model and ontology-based model. Ontology is the best model of event context representation, but the traditional ontology cannot handle uncertain knowledge which limits its application in uncertain event processing, so in recent years there have been some fuzzy ontology model and reasoning research: The logic based fuzzy model\textsuperscript{20} attempts to integrate fuzzy logic into ontology design structure\textsuperscript{21}; the distributed fuzzy reasoning Petri net and fuzzy ontology\textsuperscript{22} were used for distributed fuzzy reasoning extensively\textsuperscript{23}.

Context awareness plays an important role in heterogeneous environments. The challenge of the context sensitive system is mainly to make the right decision for the user's context in real-time. Processing context data in an intelligent way is called context reasoning. Zhou et al proposed a similarity based on context reasoning, which defines the similarity between the context models\textsuperscript{24}. This method does not require the initial context information, thus reduces the complexity of the reasoning process.

In recent years, there has been some research work on the processing of context sensitive events: Helmer et al described the framework of context event processing, and summarized the
current context support the commonly used event processing systems\textsuperscript{[25]}; Ashish et al. proposed a context sensitive and complex event processing method based on ontology\textsuperscript{[26]}; However, most of the current articles discuss the idea and framework of context sensitive complex event processing, lacking of details of the context of the introduction of complex event processing algorithms.

6. Conclusion

China High-speed train control system is a combination of computer, communication and control. Its data is diverse, including sensor generated data stream, GPS signal, GSM-R transmission data, real-time video monitoring data, train control software data, etc. This paper presents an efficient analysis method based on the fuzzy linked data and uncertain reasoning for high-speed train control system big data. We have used the method proposed in this paper to analyze the data of the Beijing-Tianjin high-speed train control system. The experimental results show that the method proposed in this paper has good efficiency and scalability for the analysis of large data with different structures, types and context sensitive from train control system. The work of this paper is based on the real practical application. In the future, there is still a lot of work to be done, such as the adaptive distribution of belief value, RDF representation of expert knowledge, the architecture of reasoning system based on Drupal.

Acknowledgement

This work is sponsored by the National Natural Science Foundation of China under Grant No. 61502029. We are grateful to the anonymous reviewers for their comments on earlier drafts of this paper.

Bibliography


