

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

D. Zhang

**Dalin Zhang**

National Research Center of Railway Safety Assessment  
Beijing Jiaotong University  
Beijing, 100044, 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 high-speed train control system. The experiment results show that the method proposed in this paper has good efficiency and scalability for the analysis of big data with different structures, types and context sensitive from high-speed train control system.

**Keywords:** high-speed train control system, fuzzy RDF, 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 [3], CISCO Smart + Connected Railway [11], SIMENS Intelligent Train [20], 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) in 2008, which have 4 level totally [17]. 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 big 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 train 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 range to use DMS. How to extract the useful knowledge from DMS's massive data effectively 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 handle it well. 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 this demand. The challenge of semantic technology is the flow reasoning, that is in the process of continuous arrival of the data and 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 [2]. There are 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 Figure 1. The real-time data of CTCS-3 can be summarized as follows:

- (1) RBC. Generating the traffic permit according to the track circuit, interlocking and other information; Accepting vehicle equipment data through GSM-R.
- (2) GSM-R network. Two-way communication between vehicle equipment and ground equipment.
- (3) Transponder. Sending positioning, level conversion information, line parameters and temporary speed-limited to the vehicle equipment.
- (4) Secure computer. Generating the dynamic speed curve and monitoring the safe operation of the train.
- (5) Track circuit. Checking the train occupancy and sending traffic permit information.
- (6) CTC system. Realization of centralized control for the station signal equipment.

### 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 inter-operable, 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.

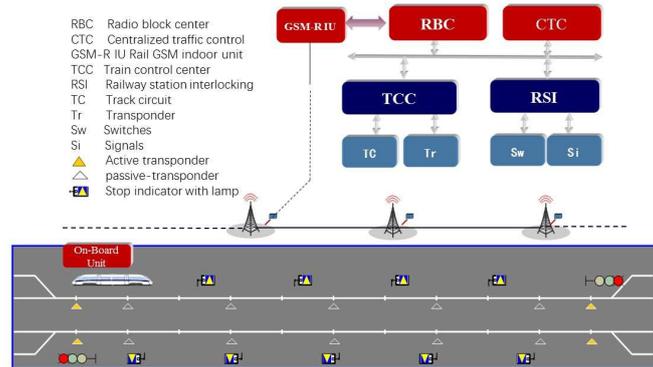


Figure 1: CTCS-3 system structure

**Resource description framework(RDF).** RDF [13] is a triplet  $\langle subject, predicate, object \rangle$ , 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 name space, such as DOI, ISBN etc.; *object* could 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 meta-data 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 [7] 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 it as a 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. The linked data is based on the existing web technology (HTTP, URL and HTML), and the web specification is further standardized and defined with 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 can be described as a linked data, so the linked data to big 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 [12] to convert CTCS-3 data into linked data. The 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 FRDF model  $FR$  is expressed as  $FR = \langle fs, fp, fo \rangle$ , where the  $fs$  is the fuzzy concept, and the  $fp$  is the set of fuzzy attributes for the subject, the  $fo$  is a set of fuzzy concept instances. The attributes  $fp$  is defined as  $fp = \{p_1, p_2, \dots, p_n\}$ , the  $p_i$  is an attribute that represents the relationship between the subject and the object;  $fo = \{o_1, o_2, \dots, o_n\}$ , the  $o_i$  represents a fuzzy instance. The function of the fuzzy degree in the FRDF 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. The "No G115", "train", "speed", "very", "fast" can total be described by the FRDF model as an instance. The "train" is a fuzzy concept, and it has the Fuzzy attribute "driving speed". The value of speed can be expressed as a group of instances  $\{S(\text{slow}), M(\text{middle}), F(\text{fast})\}$ .

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

## 2.4 FRDF model based on D-S theory

The D-S theory [10] 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)  $\sum_{A \subseteq \Theta} = 1$ ; Then,  $m : 2^\Theta \xrightarrow{\text{yields}} [0, 1]$  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.** 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  $Bel : 2^\Theta \xrightarrow{\text{yields}} [0, 1]$ ,  $Bel(A) = \sum_{B \subseteq A} m(B) = 1(\forall A \subseteq \Theta)$ .

**D-S rule.** For  $n$  limited mass functions  $m_1, m_2, \dots, m_n$  on the one recognition frame  $\Theta$ ,  $\forall A \subseteq \Theta$ ,  $A = A_1 \cap A_2 \cap \dots \cap A_n$ , then the synthesis of  $n$  belief functions is as follows:

$$m(A) = (m_1 \oplus m_2 \oplus \dots \oplus m_n)(A) = \frac{1}{K} \sum_A m_1(A_1) \cdot m_2(A_2) \cdot \dots \cdot m_n(A_n) \quad (1)$$

$$K = \sum_{A \neq \emptyset} m_1(A_1) \cdot m_2(A_2) \cdot \dots \cdot m_n(A_n) \quad (2)$$

The  $K$  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; When  $K = 1$ , it is 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 FRDF model has a very good scalability. The FRDF model can be expanded based on the D-S theory easily.

In the FRDF model based on the D-S theory(FRDFDS), the  $fs$  is still the fuzzy concept. The  $fp$  is the set of attributes for the subject which been expanded and contains some special evidence attributes and conclusion attributes, such as *hasConclusions*, *hasEvidences*, *hasBPAAvalue* and so on. The  $fo$  is a set of instances of concept which also been expanded and contains some special evidence and conclusion instances.

At present, there are five attributes which are added to the basic FRDF model: the attribute *hasConclusions*, the attribute *hasEvidences*, the BPA numeric attribute *hasBPAAvalue*, the *Bal* value attribute *hasBal*, and the time attribute *hasTimetag*. 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. In order to adapt to the context of the dynamic environment, diversity, uncertainty, we add data attribute *hasBel* and *hasBPAAvalue* for the conclusion and evidence, which are used to describe the probability of the values. 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  $\Theta$ , the evidence space  $E$ , the mapping  $F : E \rightarrow \Theta$  and the basic belief distribution value  $m$ , where  $\Theta$  and  $E$  are finite sets, the quaternion  $D = \langle \Theta, E, F, m \rangle$  is called a belief structure. Based on the basic FRDFDS model given in this paper, we can construct the belief structure  $D$ , and then create the instance of the uncertainty context information, and finally, use the evidence combination rule step by step reasoning process. The  $m$  is the distribution of the basic belief, which is a quantitative evaluation based on evidence. We can make a synthesis based on different sources of evidence through visiting their value of  $m$  correspondingly.

### 3 Multi-layer real-time uncertain context reasoning method

Context reasoning is the key to solve 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 [1], 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 symmetrical 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. 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  $(i + 1)^{th}$  layer can be inferred from the context of the  $i^{th}$  layer. The reasoning process combines the traditional rule-based reasoning and the uncertainty reasoning based on the D-S theory.

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  $\mu$ . 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.

---

**Algorithm 1** Multi-layer real-time uncertain context reasoning framework

---

**Require:** Input Context[layers][instances] = Context[leveal0][instances0]; //layers represent the level of context; instance is a FRDFDS model instance set;

**Ensure:** Output Context[layers][instances]; // the context at time t;

```

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

```

---

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:

$$m'_i(A_i) = \begin{cases} Bel(E)m_i(A_i) & \text{when } A_i \neq \Theta \\ 1 - \sum m'_i(A_i) & \text{when } A_i = \Theta \end{cases} \quad (3)$$

For different context applications, the *Bel* function is different, but the choice of *Bel* generally follows the following rules: (1) the more important evidence, the higher belief, the corresponding correction factor is great, that is, we assigned a big weight for the important evidence distribution. (2) the higher evidence conflict, the lower belief, the smaller corresponding correction factor, that is, there will be a larger amendment in the future.

---

**Algorithm 2** Uncertainty reasoning based on the D-S theory

---

**Require:** Input a context instance in instance set; //an instance in one layer context

**Ensure:** 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
4:     // conclusion is also a FRDFDS model instance.
5:     get its evidence set and
6:     calculate the BPA numeric and the Bal value;
7:     calculate the conflict factor;
8:     if conflict_factor ≤ k then
9:       calculate the Bal value for this conclusion;
10:    else
11:      calculate the Modified Bal value for evidence set;
12:      calculate the Bal value for this conclusion;
13:    end if
14:  end for
15: }
```

---

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 following modified belief formula:

$$m'(A) = (m'_1 \oplus m'_2 \oplus \dots \oplus m'_n)(A) = \frac{\sum_A \omega_1 m'_1(A_1) \cdot \omega_2 m'_2(A_2) \cdots \omega_n m'_n(A_n)}{\sum_{A \neq \emptyset} \omega_1 m'_1(A_1) \cdot \omega_2 m'_2(A_2) \cdots \omega_n m'_n(A_n)} \quad (4)$$

$$\omega_i = \frac{\sum_{j=1, j \neq i}^n (1 - d_{BPA}(m_i, m_j))}{\sum_{i=1}^q \sum_{j=1, j \neq i}^n (1 - d_{BPA}(m_i, m_j))} \quad (5)$$

The  $d_{BPA}(m_i, m_j)$  is the distance of  $m_i$ , and  $m_j$ . The formula (5) reflects the different importance and belief of the evidence which resolves the one-vote veto problem and the robustness problem in the conflict evidence combination. It makes full use of the conflict evidence information and avoids the valid information loss.

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. As shown in Algorithm 3, this paper uses fuzzy similarity to update the fuzzy data on the same data node. 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 is as follows.

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

$$\psi(M, N) = 1 - \frac{1}{|\Theta|} \sum_i^n |\mu_M(A_i) - \mu_N(A_i)| \quad (6)$$

---

**Algorithm 3** Real-time context update based on similarity

---

**Require:** Input Context[layers][instances] at time t and t-1;// the context at time t; layers represent the level of context; instance is a FRDFDS model instance set;

**Ensure:** Output the updated Context[layers][instances];

```

1: Procedure update_context(Context[layers][instances]) {
2:   for (all all layers  $\in$  N) do
3:     for (all FRDFDS model instance in Context[layers][instances] ) do
4:       initialize similarity threshold value = s;
5:       get the previous t-1 instance p_instance;
6:       calculate the index of similarity between p_instance and instance;
7:       if index  $\leq$  s then
8:         make an additional with previous
9:       end if
10:    end for
11:  end for
12: }
```

---

## 4 Experiment and evaluation

### 4.1 DMS

As shown in Figure 2, 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. 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 framework

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

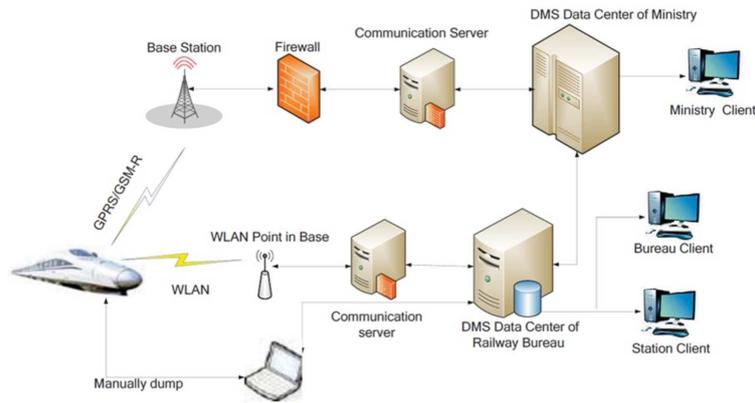


Figure 2: DMS network structure

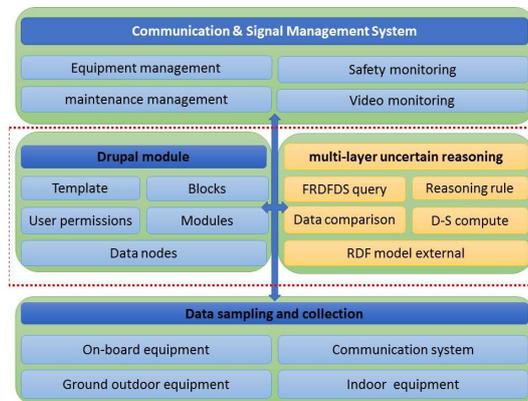


Figure 3: Data analysis integration framework

in Figure 3. In the Figure 3, 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, which is a set of functions that can be used in all aspects of railway operations; The base part is the DMS data pool.

It is easy to see from the left of the middle of the Figure 3 that our analysis method is based on Drupal module [12]. The RDF is the standard framework for semantic web and also is a recommended framework for linking data. The linked data technology enables data to be interoperability, reusable, and easier to use. Drupal 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 to a FRDFDS model based on the D-S theory firstly. All data in the DMS 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 plug-in also was expanded and programmatically filters and displays content. We implemented the algorithms presented in this paper as two functional plug-ins deployed to Drupal modules.

### 4.3 Experiment analysis

As shown in Figure 2, 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. The experiments were divided into 3 groups simultaneously on 3 IBM x3650 M4 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 [10]. We have implemented the method in paper [10] and compared it with our reasoning method.

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 Figure 4(Bar graph). According to Figure 4, 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.

The second set of experiments analyzes the data of two work areas, which doubled the amount of experimental data and focus on the efficiency of the algorithm in a distributed data environment. As shown in the Figure 4(Line graph), we compare the results of the first set of experiments with the experimental 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 Figure 5, 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.

We have implemented Jusselme's method [10] 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 Jusselme's in 24 point-in-time, and the results are shown in Figure 6. In the initial period,

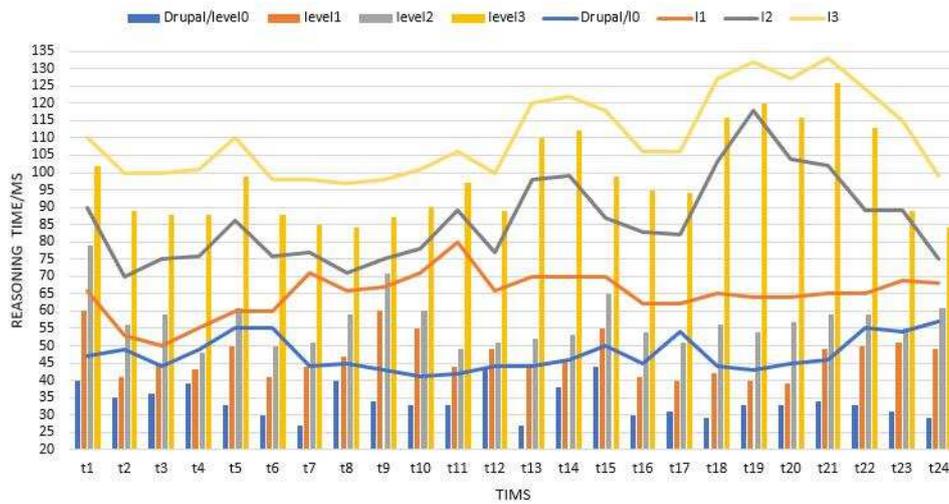


Figure 4: Performance comparison on different data sets

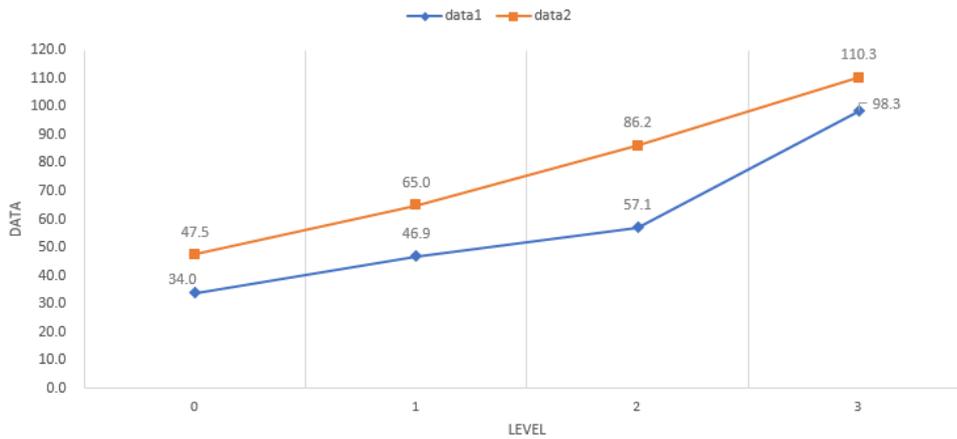


Figure 5: Performance comparison on different data density

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. The paper [10] 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

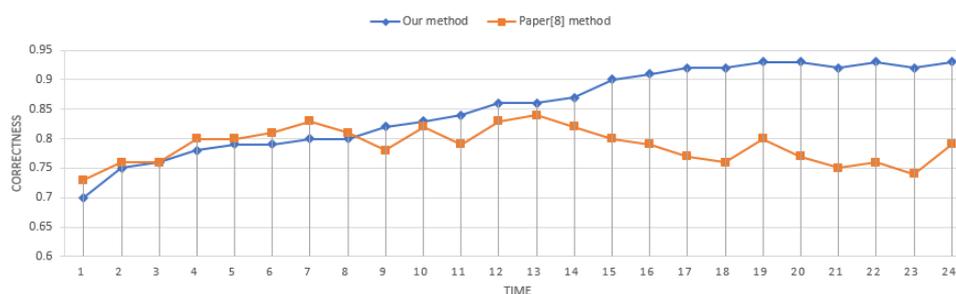


Figure 6: Evaluate the correctness of the proposed method and method in paper [10]

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.

## 5 Related research

### 5.1 Railway intelligent transportation system

The railway management department has been using the "failure - repair" work mode. However, efficient railway operations require maintenance and repair equipment timely, so as to avoid the occurrence of a failure, by analyzing the evolution of equipment status [25]. 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 [18] 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 one railway intelligent transportation system(RITS), which is proposed according to the actual situation of China's railway and makes each information system work together. In order to realize the RITS, many scholars have carried out many researches, such as meta-data sharing based on meta-data [5], general data mode based on XML [15], IoT [24], cloud computing technology [8], knowledge reasoning [4] and so on.

In summary, the scholars have begun to study the use of cloud computing and distributed systems to collect and analysis related information for rail transport, infrastructure management, maintenance departments [21] . 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 data analysis system in heterogeneous environments. There are various context representation models were presented [19], including key-value model, object oriented model and ontology-based model. The 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 [6] attempts to integrate fuzzy logic into ontology design structure [16]; The distributed fuzzy reasoning Petri net and fuzzy ontology [22] were used for distributed fuzzy reasoning extensively [14].

The challenge of the context sensitive system is mainly to make the right decision for the user's context in real-time. The processing context data in an intelligent way is called context reasoning. In recent years, there has been some research work on the processing of context sensitive events. Zhou et al. proposed a similarity based on context reasoning, which defines the similarity between the context models [26]. This method does not require the initial context information, thus reduces the complexity of the reasoning process. Helmer et al. described the framework of context event processing, and summarized the current context support the commonly used event processing systems [9]. Ashish et al. proposed a context sensitive and complex event processing method based on ontology [23]. However, the most of current articles discuss the idea and framework of context sensitive complex event processing, which lack the details of 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 high-speed train control system. The experiment 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.

## Acknowledgment

This work is sponsored by the National Natural Science Foundation of China under Grant No. 61502029.

## Bibliography

- [1] Anicic D. et al. (2011), Etalis: Reasoning in Event-Based Distributed Systems, Volume 347 of the series *Studies in Computational Intelligence*, Springer, 99-124, 2011.
- [2] Aniello L., Baldoni R., Querzoni L. (2013), Adaptive online scheduling in storm, *Proceedings of the 7th ACM international conference on Distributed event-based systems*, ACM, 207-218, 2013.
- [3] Dierkx K. (2009), *The Smarter Railroad: An Opportunity for the Railroad Industry*, IBM Institute for Business Value, 2009.
- [4] Feljan A.V. et al. (2017), Framework for Knowledge Management and Automated Reasoning Applied on Intelligent Transport Systems, arXiv preprint arXiv:1701.03000, 2017.
- [5] Gregor D. et al. (2016), A methodology for structured ontology construction applied to intelligent transportation systems, *Computer Standards & Interfaces*, 47, 108-119, 2016.
- [6] Gu L. et al. (2014), Trust Model in Cloud Computing Environment Based on Fuzzy Theory, *International Journal of Computers Communications & Control*, 9(5), 570-583, 2014.
- [7] Hartig O., Bizer C., Freytag J.C. (2009), Executing SPARQL queries over the web of linked data, *The Semantic Web-ISWC*, 293-309, 2009.
- [8] He S., Song R., Chaudhry S.S. (2014), Service-oriented intelligent group decision support system: application in transportation management, *Information systems frontiers*, 16(5), 939-951, 2014.
- [9] Helmer S., Poulouvassilis A., Xhafa F. (2011), Introduction to Reasoning in Event-Based Distributed Systems, *Reasoning in Event-Based Distributed Systems*, Springer Berlin Heidelberg, Vol 347, 1-10, 2011.
- [10] Jusselme A.L., Grenier D., Bosse E. (2001)O, A new distance between two bodies of evidence, *Information fusion*, 2(2): 91-101, 2001.
- [11] Kondepudi S., Baekelmans J. (2012), *Service Delivery Platform: The Foundation of Smart+ Connected Communities*, Cisco Smart+ Connected Communities Institute, 2012.
- [12] Kumar N. et al. (2016), *Drupal 8 Development: Beginner's Guide*, Packt Publishing Ltd, 2016.
- [13] Lassila O., Swick R R. (1999), Resource description framework (RDF) model and syntax specification, *W3C Recommendation*, 22 February 1999.
- [14] Liu H.C. et al. (2017), Fuzzy Petri nets for knowledge representation and reasoning: A literature review, *Engineering Applications of Artificial Intelligence*, 60, 45-56, 2017.
- [15] Medjoudj M., Yim P. (2007), Extraction of critical scenarios in a railway level crossing control system, *International Journal of Computers Communications & Control*, 2(3): 252-268, 2007.
- [16] Nadaban S. (2015), Fuzzy continuous mappings in fuzzy normed linear spaces, *International Journal of Computers Communications & Control*, 10(6), 74-82, 2015.

- [17] Ning B., Tang T., Qiu K., Gao C., Wang Q. (2004), CTCS-Chinese Train Control System, *Computers in Railways*, WIT Press, 393-399, 2004.
- [18] Ning B. et al. (2006), Intelligent railway systems in China, *IEEE Intelligent Systems*, 21(5), 80-83, 2016.
- [19] Perera C. et al. (2014), Context aware computing for the internet of things: A survey, *IEEE Communications Surveys & Tutorials*, 16(1): 414-454, 2014.
- [20] Roop S.S., Ruback L.G. (2001), Intelligent rail crossing control system and train tracking system, *U.S. Patent*, 6, 179-252, 2001.
- [21] Tan X., Ai B. (2011), The issues of cloud computing security in high-speed railway, *Electronic and Mechanical Engineering and Information Technology (EMEIT), 2011 International Conference on. IEEE*, 8, 4358-4363, 2011.
- [22] Rehman Z., Kifor C.V. (2016), An Ontology to Support Semantic Management of FMEA Knowledge, *International Journal of Computers Communications & Control*, 11(4), 507-521, 2016.
- [23] Taylor K., Leidinger L.(2011), Ontology-driven complex event processing in heterogeneous sensor networks, *ESWC'11 Proceedings of the 8th extended semantic web conference on The semantic web: research and applications* , Part II, 285-299, 2011.
- [24] Zhang N. et al. Optimization scheme of forming linear WSN for safety monitoring in railway transportation, *International Journal of Computers Communications & Control*, 9(6), 800-810, 2014.
- [25] Zheng W., Hu N. (2015), Automated test sequence optimization based on the maze algorithm and ant colony algorithm, *International Journal of Computers Communications & Control*, 10(4): 593-606, 2015.
- [26] Zhou H., Wang Y., Cao K. (2013), Fuzzy DS theory based fuzzy ontology context modeling and similarity based reasoning, *Computational Intelligence and Security (CIS), 2013 9th International Conference on, IEEE*, 707-711, 2013.