



**HAL**  
open science

# A Method of Ontology Evolution and Concept Evaluation Based on Knowledge Discovery in the Heavy Haul Railway Risk System

Tiancheng Cao, Wenxin Mu, Aurelie Montarnal, Anne-Marie Barthe-Delanoë

► **To cite this version:**

Tiancheng Cao, Wenxin Mu, Aurelie Montarnal, Anne-Marie Barthe-Delanoë. A Method of Ontology Evolution and Concept Evaluation Based on Knowledge Discovery in the Heavy Haul Railway Risk System. 20th Working Conference on Virtual Enterprises (PRO-VE), Sep 2019, Turin, Italy. pp.220-233, 10.1007/978-3-030-28464-0\_20 . hal-02406442

**HAL Id: hal-02406442**

**<https://imt-mines-albi.hal.science/hal-02406442v1>**

Submitted on 10 Feb 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# A Method of Ontology Evolution and Concept Evaluation Based on Knowledge Discovery in the Heavy Haul Railway Risk System

Tiancheng Cao<sup>1(✉)</sup>, Wenxin Mu<sup>1</sup>, Aurélie Montarnal<sup>2</sup>,  
and Anne-Marie Barthe-Delanoë<sup>3</sup>

<sup>1</sup> School of Economics and Management,  
Beijing Jiaotong University, Beijing, China  
{18120606, wxmu}@bjtu.edu.cn

<sup>2</sup> École des mines d'Albi-Carmaux, Albi, France  
aurelie.montarnal@mines-albi.fr

<sup>3</sup> Ecole Nationale Supérieure des Ingénieurs en Arts Chimiques  
et Technologiques, Toulouse, France  
annemarie.barthe@ensiacet.fr

**Abstract.** The risk pre-control of heavy haul railways is a collaborative scenario with multi-department linkage and the risk analysis model relies on multiple data sources. As a tool for knowledge formal modeling, Ontology and knowledge graph can achieve knowledge discovery, reasoning and decision support based on multi-dimensional heterogeneous data. This paper restores unusual context with participant behavior data as the core, establishes a basic Scenario-Risk-Accident Chain (SRAC) ontology framework. Under collaborative relationships formed by reasoning rules between context and risk, this paper establishes evolution mechanism of SRAC to introduce new knowledge, such as knowledge extracted from device detection data. New entities are added to the risk concept tree through semantic similarity algorithms. In addition, researchers added weight attribute to the risk ontology. With quantitative representation of risk concepts, this paper uses risk relevance mining to establish associated-subgraphs, establishes a new method for potential accident level assessment through maximum flow search mechanism.

**Keywords:** Ontology evolution · Risk knowledge reasoning · Semantic similarity · Maximum flow · Collaboration

## 1 Introduction

The train operation safety of heavy haul railways, as a systematic project, comprehensively consider the transportation organization, vehicle operating characteristics, signal system, personnel behavior and other factors to analyze the cause mechanism of the accident. Under the collaboration system of railway safety impact factors, both Analysis of Heterogeneous Knowledge and Knowledge Reasoning Ability should be involved in the risk analysis model. For the analysis of heterogeneous knowledge, for instance, some major accidents are caused by human factors, e.g. personal skills. If the

risk analysis model comprehensively analyzes the personnel behavior data and equipment operation data such as line orbit, the analysis conclusions about the accident cause mechanism are possible to mine implicit information: the combined effect of personal skill negligence and line aging makes the accident escalate, resulting in more serious consequences. The multi-dimensional analysis helps to advise on the later risk control and prevention in different aspects such as personnel management and equipment maintenance. That is, the decision support under multiple data sources can extract more implicit information than the model with single data source output. For knowledge reasoning, the risk analysis model should establish conceptual mapping and knowledge graphs within heterogeneous knowledge. For example, the aging of line on equipment layer can often infer that the security supervision mechanism on the management layer is not perfect, so the knowledge in management domain can be mapped with the knowledge in equipment domain. Based on concept mapping, this knowledge network structure makes the multi-data source not only a simple combination, but also it makes that the internal logic inside the knowledge is found by reasoning rules. The analysis results will be more intelligent, more accurately restore real situations of railway production environment (Fig. 1).

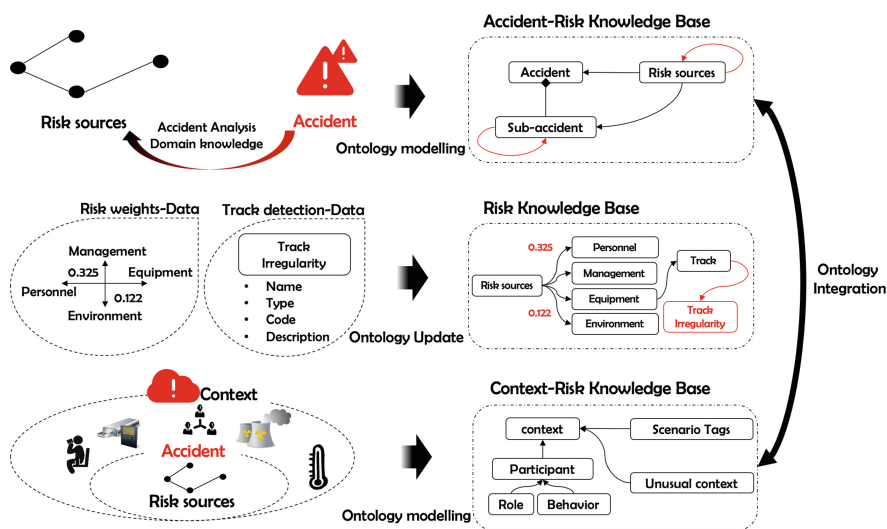


Fig. 1. Research framework

In order to make the risk analysis model constructed in this paper have the above two capabilities, it is necessary to conceptualize knowledge, establish inter-concept relationship and reasoning rules. In this paper, Sect. 3 introduces the SRAC model (Scenario-Risk-Accident Chain). This model uses the association rule mining and expert rules reasoning to construct a knowledge reasoning framework from the unusual context knowledge to the risk source knowledge, so that the ontology has an initiatory reasoning ability. It is the basis for the implementation of ontology evolution. Section 4

defines concept mapping rules and concept updating mechanism of risk ontology. This section introduces the weight attribute of safety indicators and other knowledge such as “track irregularity” into the original ontology. In the environment supporting ontology evolution and heterogeneous knowledge integration, this paper constructs a new risk evaluation method in Sect. 5. The implementation methods are based on knowledge graph and semantic similarity. Based on the SRAC reasoning framework and the ontology evolution algorithms, this paper tend to construct a knowledge reasoning evolution platform (KREP) in risk reasoning domain.

## 2 Related Works

This paper mainly studies the risk knowledge discovery and heterogeneous knowledge integration and interoperability under the multi-sector collaboration scenario in the heavy haul railway domain. With the knowledge discovery and ontology as keywords, 22 strong related articles are searched and screened in the web of science. The research of knowledge discovery is mainly divided into three categories: (1) Related techniques and algorithms for knowledge extraction from data. (2) Research on the construction of knowledge management and reasoning models. (3) Research on the data and knowledge interoperability in collaborative scenarios.

At the algorithm level of extracting knowledge from data, data features and semantic relationships are the focus of the research. In the scenario of big data, in multi-database integration environment, [1] give general algorithms and basic algorithms for different aspects of network paradox knowledge discovery. Based on rough set theory, [2] use parallel-reduction algorithm for knowledge extraction and it is suitable for large data sets with different roughness. In the process of constructing enterprise knowledge graphs, the inconsistency and knowledge conflicts are solved by [3] using the associated data paradigm algorithm. In the research of knowledge reasoning models, Ontology and Semantic Web [4] are widely used for organization the scattered knowledge extracted by factor analysis, cluster analysis methods and difference matrix [5], such as hotspot information. In addition to traditional classification, [6] also explored the fuzzy representation of knowledge and rules.

For the construction of knowledge management models and reasoning models, it involves generalized representation of knowledge and meta-model abstraction based on business activities. Based on ontology, Petri net [7], BPMN model [8], etc., the main research object is the definition of the rules and concept attributes in the model layer and the relationship between the established features in the business. The knowledge base and the feature library are continuously enriched based on existing models in the running environment. [9] recommended investment types for investors by mining the frequent characteristics of stock price changes, and the concept extraction is also transferred from indicators to contextual information such as the subject of sale and the scope of amount. The impact of context on decision-making is increasingly valued by researchers and the real-time access to information [10]. For the information overload problem of browsing rather than search process, [11] used social network analysis method to analyze the edge of important knowledge map, so as to guide the important knowledge of different user types in learning field to achieve good results. The

knowledge management model begins to provide contextual interfaces, and the flexibility of the interface is also a problem that needs to be solved. [12] in the study of process behavior prediction problems, established a decomposition machine model and active k-tuples, it is easy to add known features of the process model, rather than pre-defined hard rules. [13] also expressed the necessity to mine the generalized connection of multi-dimensional knowledge from the unified process of preprocessing, mining and post-processing of knowledge discovery.

For research on data and knowledge interoperability, focusing on applied research and architectural design, [14] construct a knowledge management framework for distributed health care systems consisting of data- and knowledge-bases, it combines patient data and the mined knowledge to enable decision making in a higher level. [15] have emphasized the importance of interaction and iteration of the knowledge discovery process through case studies. The complex synergy between the dynamics of business scenarios and business objects constitutes the heterogeneity and time-varying of perceived data. In an open collaborative scenario [16], data changes will influence the decision model [17], Research on the update mechanism of the knowledge reasoning model is particularly important, when dealing with complex contexts, relative to the above state for enriching knowledge base and feature databases. Some attempt to model evolution are researched. [18] used a comprehensive flood ontology with a scalable structure to develop a network-based emergency preparedness and response knowledge system that embodies concepts/rules to update by establishing a number of extensible interfaces. [19] designed a framework for the simultaneous involvement of users and experts in the design process of geospatial data risk identification to avoid risk about improper use of spatial data. [20] used ontology-based weighted data normalized transduction neural fuzzy reasoning to combine personal portraits with existing ontology to establish a personal diabetes risk model, and vice versa [21]. The individualized modeling data of individuals on the impact of chronic disease ontology structure will be the research focus. [22] construct an ontology that represents the temporal relationship between semantic details and text elements in the knowledge domain, combining with SVD technology, the strength of association rules could change with time. And [23] also provides a good case for evolution of rules for the model.

To summarize, there are two problems remain:

- The lack of multi-heterogeneous data management makes it difficult to form knowledge for decision support. It's necessary to apply mathematical models to conceptualize heterogeneous data, especially for complex railway management scenario.
- The new concepts/rules are separated from the original knowledge structure model. But knowledge modeling and automatically updating are necessary, especially for the potential risk mining in continuous railway operation. New knowledge should be integrated into the previous reasoning model to support automatic evolution.

### **3 Ontology Reasoning Framework**

This chapter introduces the scenario-risk-accident chain (SRAC) ontology model that was constructed in the previous period. The research domain of this paper is specified by this ontology reasoning framework. In heavy haul railway domain, the occurrence

of major accidents is often in the form of accident chains. Different accidents in the chain can be traced back to different risk sources. There are often multiple risk sources with interaction in accident chains. Figure 2 is the classification of risk factors.

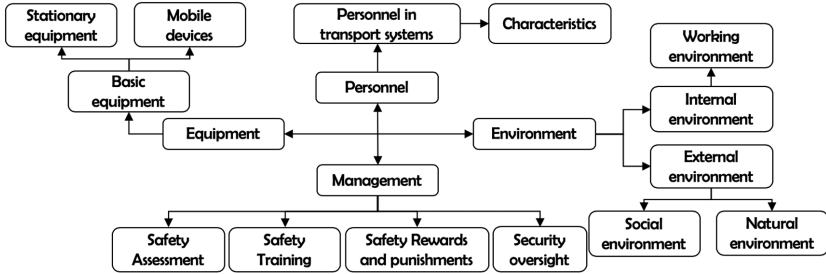


Fig. 2. Concept classification in the risk ontology

Among the risk sources, their accumulation is important implicit knowledge, e.g. “Personnel-personal skills” and “Management-safety training” are relevant, “equipment-aging” and “Environment - extreme environment” also. Define a model to describe the accumulation relationship is important. For the purpose of mining potential risk sources and evaluating risk levels through context knowledge in collaborative railway accident scenarios, we construct a SRAC model in Fig. 3. The model is obtained by the integration of risk ontology and context ontology. The risk ontology is built following Fig. 2. In this model, we use reasoning rules to describe the “produce”, “accumulate” among context, risk sources, and accidents.

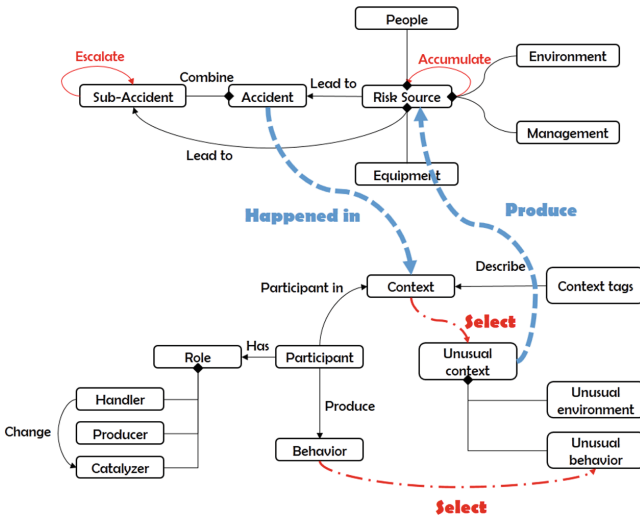


Fig. 3. Scenario-risk-accident chain model (top level structure)

For the relationships in the model, since the model mines potential risk sources based on context knowledge, the “produce” relationship must be automatically derived. And the “accumulate” shows collaborative relations among risk sources. It is constructed by expert rules, e.g. correlating risk sources occurring at the same place or at the same time; correlating ones in frequent item sets using Apriori algorithm. The rules and relations are important foundations for knowledge graph implementation in Sect. 5.

In summarize, the SRAC model constructs a reasoning chain between participants’ unusual behavior/context and accident knowledge, intermediating for risk sources which includes self-association.

## 4 Ontology Evolution

Within the scope of the SRAC model, research is conducted on both the risk ontology extension and the potential accident level assessment. To archive a quantitative risk assessment model, it is necessary to give weight for different risk factors. In this section, this paper draws on the railway safety indicators architecture constructed by other scholars, and analyzes its similarity and knowledge heterogeneity with the risk ontology in this paper. Then this section defines concept mapping rules and concept updating mechanism. This mechanism helps to add weight attribute and other knowledge to achieve ontology evolution.

### 4.1 Heterogeneous Knowledge Analysis

In order to introduce weight attribute for risk evaluation, we can refer to some existing index system and weight analysis research work. But different risk indicator system has knowledge heterogeneity problems.

In view of the construction of a safety impact factor indicators system [24] in Fig. 4, the relevant scholars build a three-level indicator system from the perspective of people, equipment, environment and management. This indicator architecture is mainly based on the universal safety theory of high-speed railway, it collects risk, fault and accident data in railway-related operations. It builds a safety knowledge analysis table based on some calculation such as factor reduction, conditional attribute ratio, etc. Some key factors are calculated to compare their weights.

This paper introduces those weight attributes into SRAC from the perspective of ontology evolution and concept updating, thus form a new method of risk assessment in a more quantitative level.

The concept in the safety impact indicator system (Fig. 4) and the risk sources in the accident-risk ontology (Fig. 2) are obviously heterogeneous. If the weight of the safety impact factors is introduced into the ontology as the attribute of risk concepts, it

Attribute		2-level Attribute		3-level Attribute	
Indicator	Weight	Indicator	Weight	Indicator	Weight
People	0.352	Physiological quality	0.103	visual & acoustic	0.065
		Psychological quality	0.089	.....	.....
		Personal skills	0.077	Rapid response	.....
		.....	.....	.....	.....
Equipment	0.212	Line and tracks	0.067	carrying capacity	0.011
		Power supply	0.042	bridge intensity	.....
		.....	.....	.....	.....
Environment	0.122	.....	.....	Man-made destruction	.....
Management	0.314	.....	.....	.....	.....

Rough Set Calculation
Rough Set Calculation
Rough Set Calculation

Fig. 4. Safety impact factor indicators architecture - a 3 level example

can indicate the impact degree on the potential accident in the knowledge reasoning process. The prerequisite is to achieve interoperability between safety indicator knowledge and risk source knowledge through inter-concept mapping. A necessary process is to analyze the concept similarity of different systems. The mapping rules between the management and personnel elements is shown in Fig. 5. By similar concept recognition, this paper can introduce those weight attributes into SRAC from the perspective of concept updating.

**Rule 1-1.** People: Sum (Railway Safety. Influence Factors. People) → Accident Risk Ontology. People. Characteristics.

**Rule 2-1.** Safety Assessment: Railway Safety. Influence Factors. Management. Safety Assessment → Accident Risk Ontology. Management. Safety Assessment.

**Rule 2-2.** Safety Training: Railway Safety. Influence Factors. Management. Safety Training → Accident Risk Ontology. Management. Safety Training.

**Rule 2-3.** Safety Rewards and Punishment: Railway Safety. Influence Factors. Management. Rewards and Punishment → Accident Risk Ontology. Management. Rewards and Punishment.

**Rule 2-4.** Safety Supervision: Railway Safety. Influence Factors. Management. Safety Supervision → Accident Risk Ontology. Management. Safety Supervision.



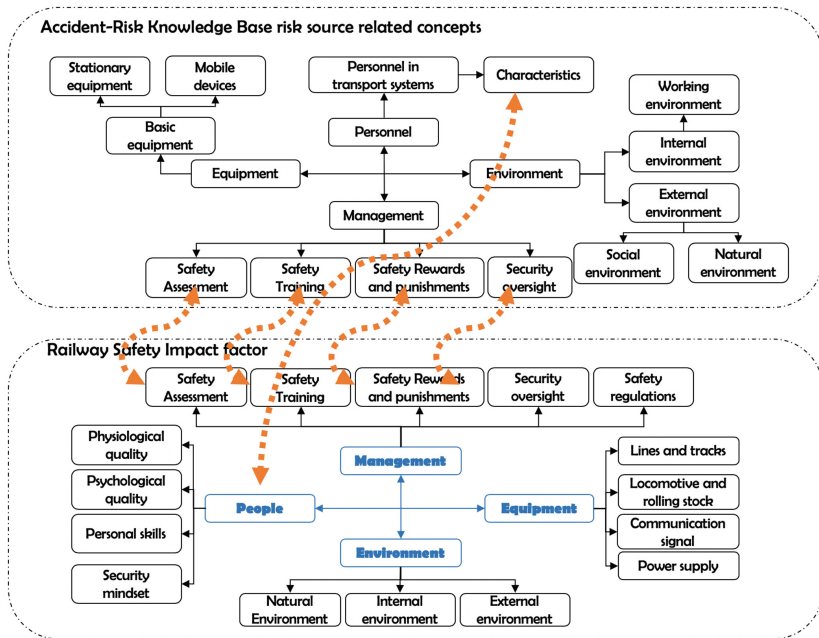
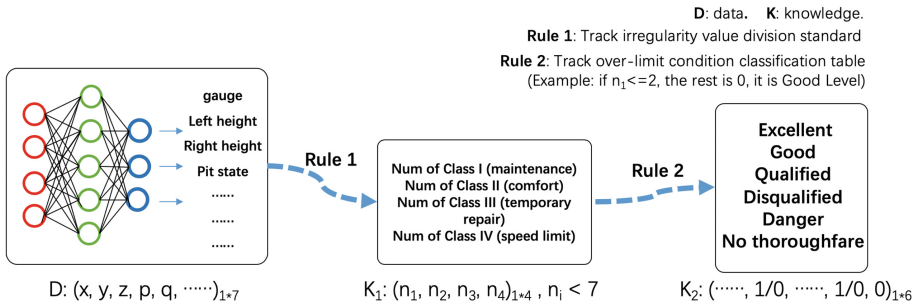


Fig. 5. Mapping rules for different systems

## 4.2 Knowledge Integration

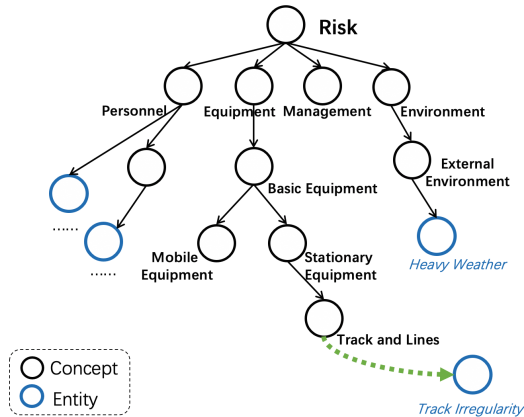
In Sect. 4.1, this paper updates the attributes of some concepts in SRAC through the similar concept mapping. This update process does not introduce new business knowledge. But with the progress of the railway operation business and risk control tasks, new business knowledge is produced continuously. Based on the detection data of the irregularity of railway tracks, this section analyzes how to integrate “track irregularity” concepts into existing risk ontology and complete concept updating. The track irregularity in geometry can cause the vibration of the rolling stock and the force of the wheel-rail action, which is the source of the disturbance of the wheel-rail system. Through analysis and prediction of the track irregularity in Fig. 6, it can effectively grasp the trend of its state change and provide a scientific basis for the track maintenance and repair work.

Combined with the nonlinear mapping ability of the neural network, the BP neural network can be used to predict the state of the track irregularity. The input of neural network is a large amount of dynamic track inspection data generated by the track detection vehicle during the inspection process, the output is some prediction values of track parameters in next month. Through expert rules of the “track irregularity classification standard” and “over-limit condition coding rule”, the knowledge of line condition can be reasoned from the prediction values of track parameters.



**Fig. 6.** Early warning mechanism of track irregularity

In view of the new knowledge structure of “track irregularity”, this paper considers adding it to the risk ontology, realizes a method for the automatic updating of concepts to reduce excessive reliance on domain experts. After the risk ontology is updated, the track irregularity concept, as a new risk source concept, join the self-learning of the risk-related knowledge base. The other personnel, management, and environmental risk factors associated with “track irregularity” can be mined and related, which are all implicit knowledge.



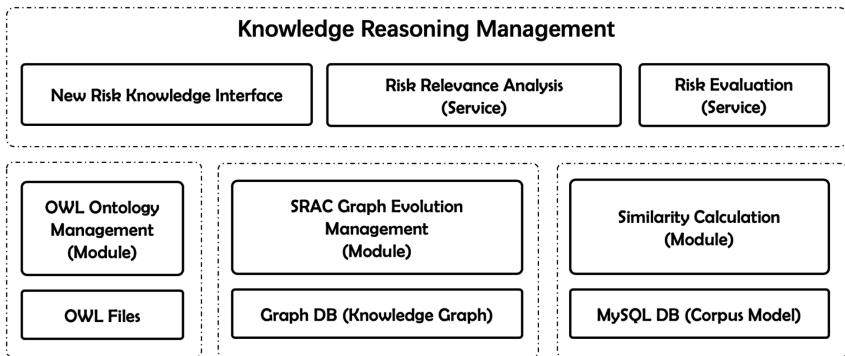
**Fig. 7.** Add track irregularity to risk concepts tree

Integrating new concepts into the concept tree (risk ontology) requires calculating the similarity between the new concept and the existing ones, and selecting the most appropriate parent concept as an insertion position. In Fig. 7, “track irregularity” is inserted behind the concept “track and lines”, using the semantic similarity algorithm based on word2vec model, which has detailed introduction in Sect. 5. This method has achieved good results on more concrete concepts and instance layers, and the addition operation can be performed. However, this algorithm does not accurately mine specific

semantic features of abstract concepts such as “Management” because words contained in abstract ones have strong universality. Therefore, the dynamic change of the ontology structure still requires further model and algorithm design.

## 5 Implementations

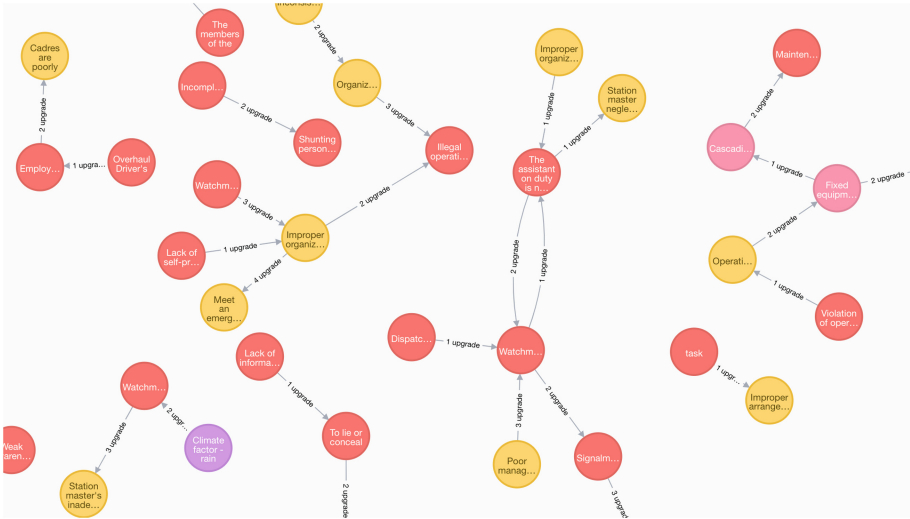
The scenario-risk-accident chain ontology includes custom knowledge reasoning rules. In order to ensure the stability and adapt to ontology reasoning ability with big data magnitude, this paper use Neo4j graph database to support ontology construction and knowledge reasoning process. This chapter migrates the risk knowledge base originally represented by OWL to graph database to support efficient reasoning and custom search. This paper tends to build a knowledge reasoning evolution platform (KREP) based on the Django framework and Python. A brief architecture is shown in Fig. 8. Encapsulating the risk knowledge reasoning services is for decision support. The new knowledge interface is for knowledge base updating.



**Fig. 8.** A brief architecture of the core of risk knowledge reasoning evolution platform

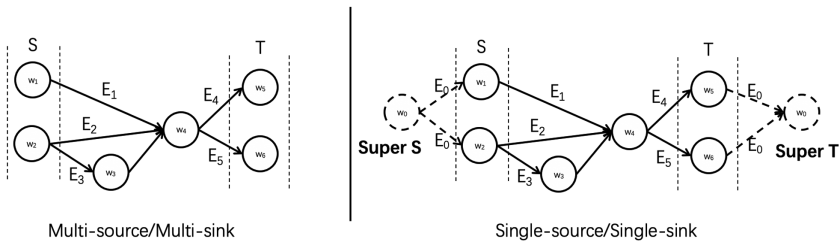
As shown in Fig. 9, after risk source entities are divided into four categories according to personnel, equipment, management, and environment, the risk entities are introduced into the risk knowledge base, and the n-level cascade effect relationship is established, the cascade number is about 1 to 4 times. The Cypher query in Neo4j can search the knowledge graph and complete some simple reasoning tasks about risk escalation or association.

The part of knowledge graph shows the upgrade relation among risk sources. The upgrade relation is one of the forms of “accumulate”. From this result of cypher query, we could find many sub-graphs that describe the potential relationships among the risk sources. On this basis, each risk entity needs to complete the updating of the weight attributes according to the mapping rules established in Sect. 4.1, which is implemented by Python programming. After adding weights to risk factors in the risk knowledge graph, the risk level of potential accidents can be quantified.



**Fig. 9.** Part of knowledge graph about risk entities

When one risk occurs, the other risks associated with it can be extracted according to the association mining algorithm, which is equivalent to extracting a directed subgraph from the knowledge graph shown in Fig. 9. The risk source contained in the subgraph could infer a potential accident based on the SRAC model. The problem level prediction for the potential accident is transformed into the problem of maximum flow for graph with multiple source/multiple sinks. The value of the maximum flow is the predicted value of the level of the potential accident.



**Fig. 10.** Equivalent flow problem of potential accident level prediction

The left side of Fig. 10 represents the risk subgraph consisting of risk (node) and risk associations (edge). To measure the maximum level of potential accidents in this subgraph, that is, to find the maximum flow in the graph. This paper defines the weights as  $W_i$ , which are calculated using the rough set from Fig. 4. The similarity between the risk nodes is defined as  $E_i$ . The purpose is to find the link in the graph that can make  $\sum |S_i + E_i|$  the largest. Since the subgraph extracted from the risk knowledge

map may have multiple source points  $S$  and multiple sink points  $T$ , the problem can be transformed into a single source/single sink problem on the right side of Fig. 10, thereby turning the problem into a classic maximum flow problem in Graph theory. Let the  $W$  vector in the graph be  $(0.3, 0.2, 0.05, 0.1, 0.3, 0.2)$ , and the  $E$  vector be  $(0.2, 0.05, 0.1, 0.2, 0.2)$ , then the link corresponding to the largest stream in the risk graph is  $(w_1, w_4, w_5)$ , the reference value of the potential accident level is 1.1. To evaluate the risk level and integrate new knowledge to the original ontology both needs the value of concept similarity and entity similarity. For an appropriate model, we use word2vec model to evaluate similarity values in Table 1.

**Table 1.** Examples of similarity with “track irregularity”

ID	Entity/concept name	Similarity (“track irregularity”)
1	Track and lines	0.61
2	Interruption on lines	0.43
3	Wrong operation by workers	0.18
4	People’s weak business ability	0.02
5	.....	.....

We select the 10-year heavy haul railway accident analysis reports in 2006–2016 for training. These reports are prepared for the corpus. The training results of word2vec (word embedding) - word vectors, is a good way to measure the similarity between the words. But the form of risk source is usually phrases or sentences. In this paper, we use the word segmentation tools to transform the sentences into word sequence vectors. After that, this paper calculates the similarity between the word sequence of the target concept and the sequence of the original concepts. Then the result is the concept similarity. The process of integrating new risk knowledge with original risk concept tree is equivalent to find an appropriate subclass for a new risk entity. we consider the conceptual similarity and attribute similarity to evaluate the similarity of two knowledge nodes in graph database. For it involves the similarity comparison of texts. Therefore, this task requires corpus training in special scenarios, the quality of corpus and word segmentation has a great influence on the results.

## 6 Conclusion

In the Scenario-Risk-Accident chain framework, this paper uses ontology and knowledge graph to formalize accident cause mechanisms in collaboration scenarios. This paper discusses attribute and concept updating with multi-dimensional heterogeneous risk knowledge in heavy haul railway accident-handling collaboration processes. We establish a risk ontology evolution mechanism using conceptual semantics. By introducing weight attribute of risks, a maximum link search of knowledge graph is completed on risk relevance basis. A new quantitative evaluation method of potential accident level is proposed. There are two parts of future work:

- Ontology updating method based on semantic similarity has achieved good results on entity sets. However, due to the abstraction of ontology concepts, similar features between concepts are difficult to extract through semantics, which still needs artificial participation. Therefore, the model for extracting more detailed features to support automatic updates of ontology concepts will be the focus of future research.
- The ontology evolution mechanism is for adding elements, but does not involve reconstruction of the ontology structure. In order to enhance the flexibility of the ontology, it is necessary to do fuzzification on the ontology concept definition and set fuzzy reasoning rules. Thereby the adaptability to complex collaborative scenarios could be enhanced.

**Acknowledgments.** The presented research works have been supported by “the National Science Foundation for Young Scientists of China” (61703032, Context based Multi-dimension ontology modeling and alignment). The authors would like to thank the project partners for their advice and comments.

## References

1. Zhang, C., He, L., Mao, Y., Xiao, B.: Knowledge discovery of network public opinion in the concept of smart city. In: *Industrial Electronics and Applications*, pp. 1197–1202 (2015)
2. Chowdhury, T., Chakraborty, S., Setua, S.K.: Knowledge extraction from big data using MapReduce-based parallel-reduct algorithm. In: *International Conference on Computer Science and Network Technology*, pp. 240–246 (2017)
3. Miao, Q., Meng, Y., Zhang, B.: Chinese enterprise knowledge graph construction based on linked data. In: *IEEE International Conference on Semantic Computing*, pp. 153–154 (2015)
4. Li, Y., Thomas, M.A., Osei-Bryson, K.-M.: Ontology-based data mining model management for self-service knowledge discovery. *Inf. Syst. Front.* **19**, 925–943 (2017)
5. Wang, X., Liu, J., Sheng, F.: Analysis of hotspots in the field of domestic knowledge discovery based on co-word analysis method. *Cybern. Inf. Technol.* **14**, 145–158 (2014)
6. Soundappan, S.J., Sugumar, R.: Optimal knowledge extraction technique based on hybridisation of improved artificial bee colony algorithm and cuckoo search algorithm. *Int. J. Bus. Intell. Data Min.* **11**, 338 (2016)
7. Liu, H.C., You, J.X., Li, Z.W., Tian, G.: Fuzzy Petri nets for knowledge representation and reasoning. *Eng. Appl. Artif. Intell.* **60**, 45–56 (2017)
8. Ligęza, A., Potempa, T.: Artificial intelligence for knowledge management with BPMN and rules. In: Mercier-Laurent, E., Boulanger, D. (eds.) *AI4KM 2012. IAICT*, vol. 422, pp. 19–37. Springer, Heidelberg (2014). [https://doi.org/10.1007/978-3-642-54897-0\\_2](https://doi.org/10.1007/978-3-642-54897-0_2)
9. Ha, Y.M., Park, S., Kim, S.W., Won, J.I., Yoon, J.H.: A stock recommendation system exploiting rule discovery in stock databases. *Inf. Softw. Technol.* **51**, 1140–1149 (2009)
10. Abu Muntalib, S., Sidi, F., Jabar, M.A., Ishak, I.: Big data trend: knowledge discovery on the unstructured data (2014)
11. Hao, J., Yan, Y., Gong, L., Wang, G., Lin, J.: Knowledge map-based method for domain knowledge browsing. *Decis. Support Syst.* **61**, 106–114 (2014)
12. Lee, W.L.J., Parra, D., Munoz-Gama, J., Sepúlveda, M.: Predicting process behavior meets factorization machines. *Expert Syst. Appl.* **112**, 87–98 (2018)
13. Huanyu, Z., Xin, L.: Research on the application of knowledge discovery in digital library service (2017)

14. Kazemzadeh, R.S., Sartipi, K.: Interoperability of data and knowledge in distributed health care systems. In: IEEE International Workshop on Software Technology & Engineering Practice (2005)
15. Schmidt, C., Sun, W.N.: Synthesizing agile and knowledge discovery: case study results. *J. Comput. Inf. Syst.* **58**(2), 142–150 (2016)
16. Allarakhia, M.: Evolving models of collaborative drug discovery: managing intellectual capital assets. *Expert Opin. Drug Discov.* **13**(6), 473–476 (2018)
17. Usai, A., Pironti, M., Mital, M., Mejri, C.A.: Knowledge discovery out of text data: a systematic review via text mining. *J. Knowl. Manag.* **22**, 1471–1488 (2018)
18. Sermet, Y., Demir, I.: An intelligent system on knowledge generation and communication about flooding. *Environ. Model Softw.* **108**, 51–60 (2018)
19. Grira, J., Bédard, Y., Roche, S., Devillers, R.: Towards a collaborative knowledge discovery system for enriching semantic information about risks of geospatial data, vol. 40, pp. 53–58 (2013)
20. Verma, A., Fiasché, M., Cuzzola, M., Morabito, F.C., Irrera, G.: Knowledge discovery and risk prediction for chronic diseases: an integrated approach. In: Engineering Applications of Neural Networks - INNS EANN-SIG International Conference, EANN 2011 and IFIP WG 12.5 International Conference, AIAI 2011, Corfu, Greece, 15–18 September 2011, Proceedings, pp. 270–279 (2011)
21. Qu, M., Chang, Y., Reddy, Y.V., Reddy, S.: A behavior centric service discovery model for building knowledge advantage machine. In: Advanced Information Technology, Electronic and Automation Control Conference, pp. 630–634 (2017)
22. Woszezenki, C.R., Goncalves, A.L., de Souza, J.A.: A knowledge discovery model based on semantic and temporal associations between textual elements. *IEEE Lat. Am. Trans.* **16**, 1243–1249 (2018)
23. Chen, Q., Fan, Z., Kaleshi, D., Armour, S.: Rule induction-based knowledge discovery for energy efficiency. *IEEE Access* **3**, 1423–1436 (2015)
24. Peng, L.Y., Feng, Y.Q., Gou, J.Q., Li, X.W.: Heavy haul railway safety factors selection and weight analysis. *China Saf. Sci. J.* **28**(2), 116–121 (2018)