

Research on Network Prediction of Hidden Danger in Logistics Enterprises

Xiying Chen^{*}, Yufeng Zhuang, Lingyi Lu

Beijing University of Posts and Telecommunications, Beijing, China

Abstract

In view of the large amount of manually recorded safety inspection data accumulated in the current logistics industry, as well as the problem of hidden danger neglect caused by lack of experience of inspectors and other reasons, this paper constructs a knowledge graph to show the structure of hidden danger of logistics enterprises in the form of graph data, targeted analysis of various unsafe hidden factors in logistics enterprises, based on the original unstructured data, To realize the construction of the relationship network between enterprises and the hidden danger factors, and analyse the relationship between the hidden danger of logistics enterprises. At the same time, the time sequence information is integrated, the graph data of continuous time slices is integrated, and the time-varying rule is analysed to realize the dynamic prediction of the hidden danger network of logistics enterprises. The accuracy of prediction is increased by learning historical data, so as to provide more targeted and specific inspection focus for the hidden danger inspection of logistics enterprises, and become an important auxiliary tool in manual inspection.

Keywords

hidden danger prediction, knowledge graph, feature of time sequence

1. Introduction

With the development of express delivery in the logistics industry and the increase of the size of logistics enterprises, safety risks are also increasing. The safety of production activities of logistics enterprises is facing severe challenges. The failure of timely inspection and prevention of hidden risks often causes accidents, resulting in huge losses. In October 2018, 15 people died and 46 people were injured in a major accident on the Shenhai Expressway. According to the investigation, it was caused by the failure of the logistics enterprise to fulfill the main responsibility of production safety, the long-term neglect of the attached vehicles, and the substandard quality inspection of the vehicles. In November 2019, Shanghai Xinde Logistics Co., Ltd. caught fire due to combustible materials on the ground and ignition during welding operations. In the final analysis, the fire was caused by the company's inadequate implementation of housing safety management responsibilities. In July 2021, a logistics warehouse in Changchun also caught fire due to improper storage of combustible materials, resulting in 15 deaths and 25 injuries. In recent years, the state also attaches great importance to the investigation and inspection of the safety risks of enterprises, vigorously carry out the responsibility of the implementation of enterprise safety production and law enforcement inspection arrangements, the post inspection into the prevention and investigation in advance, really reduce the occurrence of accidents from the source.

In view of the important problems of accident hidden danger, knowledge graph has been applied in the field of emergency safety. Aditya Pingle et al. [1] proposed a system to create semantic triples of network security texts and extract possible relationships using deep learning methods. Security analysts can form decisions about cyber attacks from the knowledge graph. Celebi R et al. [2] used the knowledge map of the drug library to predict unknown drug interactions based on RDF2Vec. Based on

ICCEIC2022@3rd International Conference on Computer Engineering and Intelligent Control
EMAIL: *brianna_xiying@163.com (Xiying Chen)



© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

BiLSTM-CRF information extraction model, Sutphin C[3] identified adverse reactions and their causes from FDA drug labels. Elluri, L et al. [4] trained the custom named entity recognition (NER) model, constructed the network security knowledge graph (CKG) to infer the subjective association between the network security text and users, and generated the relevant features of the text. Wang Yibao et al. [5] drew a map of diversified, time-sharing and dynamic scientific knowledge of urban security research by using the core documents of urban security in CNKI database from 1993 to 2018.

From the research status of domain knowledge graph and its application in emergency security, it can be seen that there is no research on the construction of knowledge graph based on the relationship between hidden dangers in logistics enterprises, and most of them are still applied in the fields of network security and traffic management. In addition, the current construction of knowledge graph is often based on existing data, forming a triplet form to obtain static knowledge graph, which often ignores the time-varying law of hidden trouble data of logistics enterprises and fails to make full use of time information. Therefore, based on the existing hidden danger detection data and combined with the time series characteristics, this paper constructs the hidden danger knowledge map of logistics enterprises. In the form of graph data, the co-occurrence relationship of hidden dangers and the connection between enterprises and hidden dangers are analyzed, and the information of people, objects, environment and management is extracted to form a more comprehensive and relevant hidden dangers analysis.

2. Hidden Danger Prediction of Logistics Enterprise Based on Knowledge Graph

2.1. Overview of hidden danger prediction methods for logistics enterprises

This paper mainly includes two parts, one is to build the knowledge map of hidden dangers in the field of logistics, so as to realize the targeted study of hidden dangers in logistics enterprises; the other is to carry out the dynamic prediction of hidden dangers network of logistics enterprises in time sequence, reduce the risk of accidents, and convert the after-repair into pre-prevention. By predicting the hidden dangers that may occur in slice logistics enterprises in the next time, guide logistics enterprises to inspect the hidden dangers of safety production. The time knowledge graph is constructed in a top-down way. The pattern layer is designed first, and then the fact representation of the data layer is realized. The ontology design is used to standardize the actual data, which ensures the rationality of the knowledge graph. The process of prediction is to embed entity and time with the Diachronic Embedding (DE) [6] model and realize the training of model combined with the scoring function of complEx[7] model. The design of knowledge graph of logistics enterprise is shown in Figure 1.

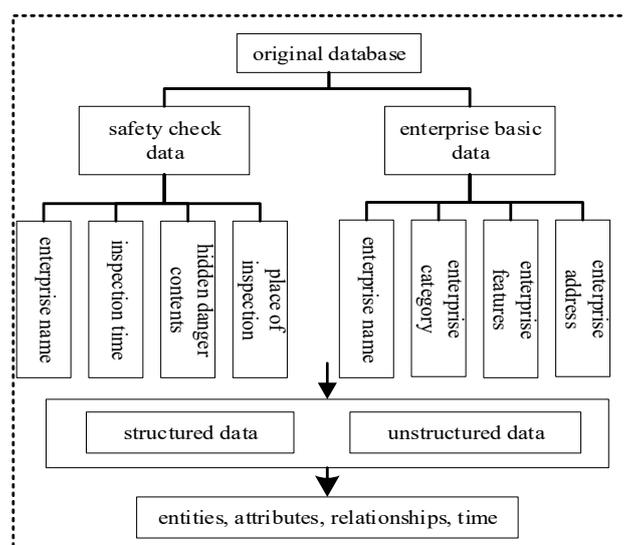


Figure 1 Design of knowledge graph of logistics enterprise

Entity, attribute, relation and time are extracted from the data, and the quadruples are constructed by "entity-relationship-entity-time" and "entity-attribute-attribute value-time". According to the time information, the data is divided into multiple time windows, then the corresponding triplets in each time window can form a static knowledge graph G^τ (τ represents the τ th time window). Temporal knowledge graph is the set of knowledge graph under different time windows, which is defined as $G = \{G^1, G^2, \dots, G^\tau\}$. E^τ is the set of all entities under this time window, R^τ is the set of all relations, and Q^τ is the set of all quadruples (h, r, t, τ) . The hidden danger network prediction of logistics enterprises is shown in Figure 2.

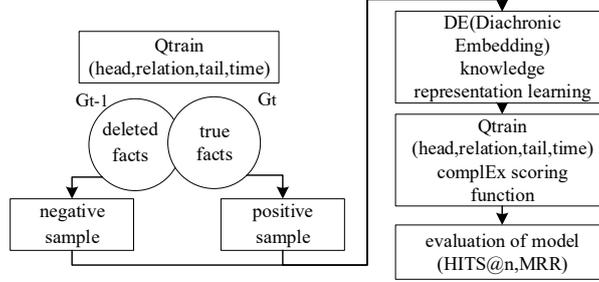


Figure 2 Process of hidden danger prediction

This paper employs DE as the basic model of linkage prediction of the sequential knowledge graph, and combines the complEx scoring function for model training. DE has strong generalization ability. Any static KG embedded model can be extended to TKGC by using DE, and complete expression of temporal knowledge graph can be realized by using DE and complEx. E represents the set of all entities, and R represents the set of all relations. In the process of entity embedding, DE takes both entity and time as input and outputs entity embedding vector that integrates time information. In the same way, relationship embedding is completed.

$$m_e^\tau[n] = \begin{cases} a_e^\tau[n]\sigma(\omega_e[n]\tau + b_e[n]), & 1 \leq n \leq \gamma d \\ a_e[n], & \gamma d < n \leq d \end{cases}$$

m_e^τ is the output vector of the entity embedded at time τ , and $m_e^\tau[n]$ represents the n th element of the vector. Where $e \in E$, τ is the corresponding time, $a_e \in \mathbb{R}^d$, $\omega_e, b_e \in \mathbb{R}^{\gamma d}$ are all vectors with learnable parameters associated with a particular entity or relationship, $\sigma(*)$ is the activation function. It can be seen from the formula that the first γd elements of the vector are used to capture time sequence features changing with time, while the remaining $(1 - \gamma)d$ elements are used to capture static features. γ is between 0 and 1, which is the hyperparameter controlling the percentage of time sequence features.

ComplEx model introduces complEx space and uses complex $a + bi$ to represent entities and relations in the static knowledge graph, which can better model the asymmetric relations.

$$\begin{aligned} f(h, r, t) &= \text{Re}(\langle \mathbf{m}_r, \mathbf{m}_h, \overline{\mathbf{m}_t} \rangle) \\ &= \langle \text{Re}(\mathbf{m}_{rd}), \text{Re}(\mathbf{m}_h), \text{Re}(\mathbf{m}_t) \rangle \\ &+ \langle \text{Re}(\mathbf{m}_{rd}), \text{Im}(\mathbf{m}_h), \text{Im}(\mathbf{m}_t) \rangle \\ &+ \langle \text{Im}(\mathbf{m}_{rd}), \text{Re}(\mathbf{m}_h), \text{Im}(\mathbf{m}_t) \rangle \\ &- \langle \text{Im}(\mathbf{m}_{rd}), \text{Im}(\mathbf{m}_h), \text{Re}(\mathbf{m}_t) \rangle \end{aligned}$$

$\mathbf{m}_h, \mathbf{m}_t \in \mathbb{C}^k$ is the complex vector corresponding to the head and tail entities, $\mathbf{m}_{rd} \in \mathbb{C}^k$ is the relation vector, $\bar{\cdot}$ means take the conjugate vector, $\text{Re}(\cdot)$ means take the real part, $\text{Im}(\cdot)$ means take the imaginary part, $\langle \cdot, \cdot, \cdot \rangle$ represents the product sum of vector elements. The function is antisymmetric when \mathbf{m}_r has only an imaginary part, and symmetric when \mathbf{m}_r has only a real part.

In the process of model training, deleted facts are also used to construct negative samples, and positive and negative samples are combined. On the one hand, the prediction effect and stability of the model are improved. On the other hand, the prediction accuracy of the model for the facts that were correct in the past but are wrong now (i.e. deleted) is improved.

$$Q_{trainpositive} = \{(h, r, t, \tau) | (h, r, t, \tau) \in Q_{train}^\tau \cap (h, r, t, \tau) \notin Q_{train}^{\tau-1}\}$$

$$Q_{trainnegative} = \{(h, r, t, \tau) | (h, r, t, \tau) \notin Q_{train}^{\tau} \cap \exists (h, r, t, \tau') \in Q_{train}^{\tau-1}\}$$

2.2. Evaluation indicators of predictive effectiveness

In terms of the selection of evaluation indexes, same evaluation indexes as the static knowledge graph are used. For the quintuples in the test set, two types of candidate quintuples are constructed by replacing the head or tail entities, and the scores of these entities are output in order.

1. MRR (Mean Reciprocal Ranking) is the average proportion of quadrilateral ranking in the prediction result. The larger the MRR value is, the higher the triplet ranking with correct prediction will be, and the better the prediction effect of the model will be.

$$MRR = \frac{1}{|S|} \sum_{i=1}^{|S|} \frac{1}{rank_i}$$

$$= \frac{1}{|S|} \left(\frac{1}{rank_1} + \frac{1}{rank_2} + \dots + \frac{1}{rank_{|S|}} \right)$$

2. HITS@n refers to the average proportion of triples ranked less than n in the link prediction.

$$HITS@n = \frac{1}{|S|} \sum_{i=1}^{|S|} \mathbb{I}(rank_i \leq n)$$

|S| is the number of triples, $rank_i$ refers to the ranking of link predictions for the i th triplet. $\mathbb{I}(\ast)$ is the indicator function. The value of the function is 1 if the condition is true, and 0 otherwise. In general, n is equal to 1, 3, or 10. The larger HITS@n is, the higher the probability that the predicted results rank less than n, and the better the link prediction effect is.

3. Construction and Prediction of Knowledge Graph

3.1. Basic information of data

This paper takes the inspection records of security risks in a city in a certain year as the core data, with a total of 512840 pieces of data content, including 169 fields such as the enterprise, time, place, place, and specific content related to the inspection of potential risks. Delete the fields with more than 80% missing values, and the remaining 68 fields are divided into 19 indicators related to the company, 27 indicators related to hidden dangers, and 22 indicators of other categories.

According to statistics, the top ten feature words of hidden dangers in logistics enterprises are security, fire extinguisher, logo, clutter, production, warehouse, use, channel, time and line. The top ten feature words of all enterprises in the city are security, fire extinguisher, logo, clutter, distribution box, exit, cover, record, training and use. As can be seen from the statistics, logistics enterprises and all enterprises have both common problems and great differences. If all analysis is done using enterprise data, considerations of operational safety, workplace, and problem urgency will inevitably be ignored, thereby affecting the reliability and value of real-world applications.

3.2. Hidden danger network construction and sample setting

Using the method of fuzzy matching and enterprise classification, 14,831 pieces of data of logistics enterprises were selected, involving 4042 logistics enterprises. According to the top-level structure design of knowledge graph, 11 fields are extracted and six types of quad pattern are designed, which is shown in Table 1.

Table 1 Basic quadruple form and quantity

head	relation	tail	t
enterprise features	happen	hidden danger category	t
enterprise category	happen	hidden danger category	t

industry classification	happen	hidden danger category	t
hidden danger category	appear	address	t
hidden danger category	appear	place	t
hidden danger category	co-occurrence	hidden danger category	t

According to the concept layer's quadruple pattern, the knowledge graph under different time windows can be constructed. As shown in Figure 3, the time sequence changes of corresponding facts in the knowledge graph can be used to complete the construction of positive and negative samples.

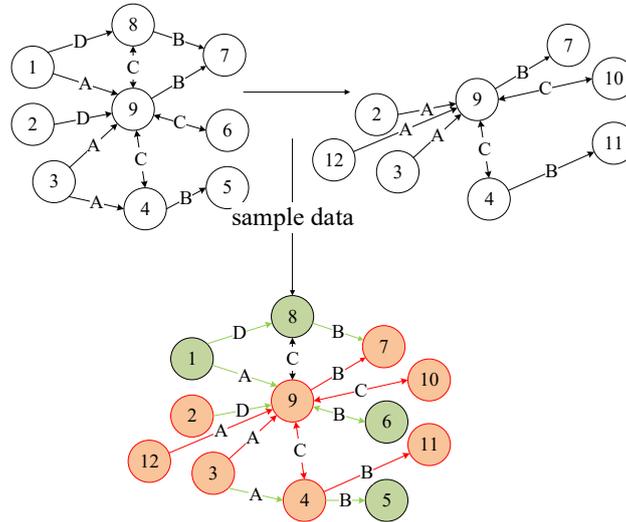


Figure 3 Hidden danger network and positive and negative sample settings

Where, the number represents the entity, the letter represents the relationship. And the green indicates that the current time window is deleted relative to the previous time window, that is, the negative sample; and red represents the real sample under the current time window, that is, the positive sample.

3.3. Prediction process and result analysis

In this paper, DE is used as the embedded representation method of timing knowledge graph, the time information is embedded into the representation of entity and relationship, and the model is trained by combining the complEx scoring function. In this paper, 70% data is taken as the training set and 30% data as the verification set. The learning rate of the model is set as 10^{-3} , and the embedding size of representation learning is set as 128. When using positive and negative samples for model training, ensure that the total number of samples for each training is less than 1024. Through the performance test and comparison of the verification set, 100 samples are taken as the negative sampling rate in this paper, that is, 100 negative samples are set for each prediction fact, including 50 for head entity replacement and 50 for tail entity replacement. The experiment found that this ratio achieved an appropriate trade-off between task performance and training time.

When the loss value of the model is no longer reduced, the average ranking of the real entity is between 2 and 3. The evaluation results on the test set are shown in Table 2.

Table 2 Evaluation indicators of model predictions

Hit@10	Hit@3	Hit@1	MRR
0.87	0.59	0.33	0.51

It can be seen that Hit@10 of this model has achieved a good result, that is, the link prediction ranking corresponding to most correct facts is less than 10, and the link prediction ranking corresponding to more than half of correct facts is less than 3, which indicates that the ranking conforms to the facts well.

The entity importance under different time windows is analysed by using the measurement method of complex network centrality. Take point-degree centrality as an example, the five most important hidden danger categories under the three continuous time Windows are shown in Table 3.

Table 3 Entity importance under different time windows

t=17	t=18	t=19
safety equipment and facilities	emergency rescue plan and implementation	safety in production
operating environment	auxiliary system equipment and facilities	other security management
auxiliary system equipment and facilities	material	safety equipment and facilities
safety signs and identifiers	operation behavior of the practitioner	education and training of practitioners
other equipment and facilities	safety equipment and facilities	operating environment

According to the importance of hidden trouble entities under different time windows, it can be seen that the emphasis of key hidden trouble under different time windows is different, and the attention of the same type of hidden trouble has time sequence change. For example, the indexes of safety equipment and facilities are all in the top five hidden danger categories in the three time windows, indicating their universality, but their importance is different in different time windows. The job environment class ranked the top five in t=17, and after t=18 was improved, t=19 again attracted attention.

4. Summary

Combined with the previous experimental analysis, we found that the time-series knowledge graph can be used to predict the hidden dangers of logistics enterprises. The time-series knowledge graph can be combined with various information to learn in the way of graph structured data, and the triplet relationship between data can be used to improve the ability of hidden dangers prediction. The method in this paper provides effective data support for daily inspection and hidden danger prevention of logistics enterprises, improves the comprehensiveness of hidden trouble investigation, and greatly reduces the probability of hidden trouble occurrence.

5. References

- [1] Aditya Pingle, Aritran Piplai, Sudip Mitta et al. Relext: Relation extraction using deep learning approaches for cybersecurity knowledge graph improvement[M]. Vancouver, BC, Canada: Association for Computing Machinery, Inc,2019: 879-886.
- [2] Celebi R, Uyar H, Yasar E, et al. Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings[J]. BMC Bioinformatics, 2019,20(1).
- [3] Sutphin C, Lee K, Yepes A J, et al. Adverse drug event detection using reason assignments in FDA drug labels.[J].Journal of biomedical informatics, 2020,110:103552.
- [4] Elluri L, Nagar A, Joshi K P. An Integrated Knowledge Graph to Automate GDPR and PCI DSS Compliance[M]. Abe N, Liu H, Pu C, et al. IEEE International Conference on Big Data. 2018:1266-1271.
- [5] Wang Yibao, Yang Tinghui. Visual analysis of Knowledge graph of Urban security Research [J]. Urban Development Studies,2019,26(03):116-124.
- [6] Rishab Goel, Seyed Mehran Kazemi, Marcus Brubaker, Pascal Poupart. Diachronic Embedding for Temporal Knowledge Graph Completion[J]. Proceedings of the AAAI Conference on Artificial Intelligence,2020,34(04).
- [7] Trouillon T , Welbl J , Riedel S , et al. Complex Embeddings for Simple Link Prediction[J]. JMLR.org, 2016.