ORIGINAL ARTICLE

A domain semantics-enhanced relation extraction model for identifying the railway safety risk

Youwei Wang[1] · Chengying Zhu[2] · Qiang Guo[1] · Yangdong Ye[1]

Abstract

The identification of railway safety risk is important in ensuring continuous and stable railway operations. Most works fail to consider the important relation between detected objects. In addition, poor domain semantics directly degrades the final performance due to difficulty in understanding railway text. To solve these challenging issues, we introduce the triple knowledge from knowledge graph to model the railway safety risk with the knowledge interconnection mode. Afterward, we recast the identification of railway safety risk as the relation extraction task, and propose a novel and effective **D**omain **S**emantics-**E**nhanced **R**elation **E**xtraction (DSERE) model. Specifically, we design a domain semantics-enhanced transformer mechanism that automatically enhances the railway semantics from a dedicated railway lexicon. We further introduce piecewise convolution neural networks to explore the fine-grained features contained in the structure of triple knowledge. With the domain semantics and fine-grained features, our model can fully understand the domain text and thus improve the performance of relation classification. Finally, the DSERE model is used to identify the railway safety risk of south zone of China Railway, and achieves 81.84% AUC and 76.00% F1 scores on the real-world dataset showing the superiority of our proposed model.

Keywords Railway safety risk · Relation extraction · Domain semantics · Domain semantics-enhanced transformer

Introduction

The identification of railway safety risk is important in the intelligent transportation systems under industry 4.0 [48], which can guarantee stable operation and reduce the probability of incidents/accidents [8, 26]. The key is automatically detecting the safety risk from the "three-in-one" scenario of human, physical and technical defense. Recently, many promising deep neural networks (DNN) models [7, 13, 19,

28, 29, 34] have been proposed to detect safety risk in railway systems. However, these previous models suffer from two challenging problems.

Neglecting the importance of interconnection between railway objects. These existing models fail to explore the relation between railway objects and thus have to rely extra experiences or prior knowledge for incident/accident inference, as shown in Fig. 1a. For instance, the two-stage DNN model [49] is proposed to detect the abnormal state of a single railway object, but it ignores the relation between railway objects that is important for analyzing train collision accidents. In addition, introducing different experiences or prior knowledge to explore the relation can generate inconsistent results due to differences in human subjectivity.

Lacking domain semantics of railway. Almost all these DNN based models encounter with a severe semantic gap between the railway and universal domains, particularly in text mining. Existing models usually initialize the data representations with pre-trained models (e.g., BERT [9], GPT [32]) trained on the open corpus (that includes Wikipedia, Wall Street Journal, etc.), which lacks domain corpus, especially the Chinese railway corpus, as shown in Fig. 1b. For

✉ Yangdong Ye
 yeyd@zzu.edu.cn

  Youwei Wang
  youweiwang@zzu.edu.cn

  Chengying Zhu
  zhu_chengying@163.com

  Qiang Guo
  ieqguo@gs.zzu.edu.cn

1 School of Computer and Artificial Intelligence, Zhengzhou University, No. 100 Kexue Avenue, Zhengzhou 450001, Henan, China

2 School of Cyber Science and Engineering, Zhengzhou University, No. 100 Kexue Avenue, Zhengzhou 450001, Henan, China

Fig. 1 Challenges of existing models. **a** Existing models fail to mining the semantic relation between railway objects. **b** Pre-trained language models (e.g., BERT) can not well represent railway corpus due to absence of domain semantics
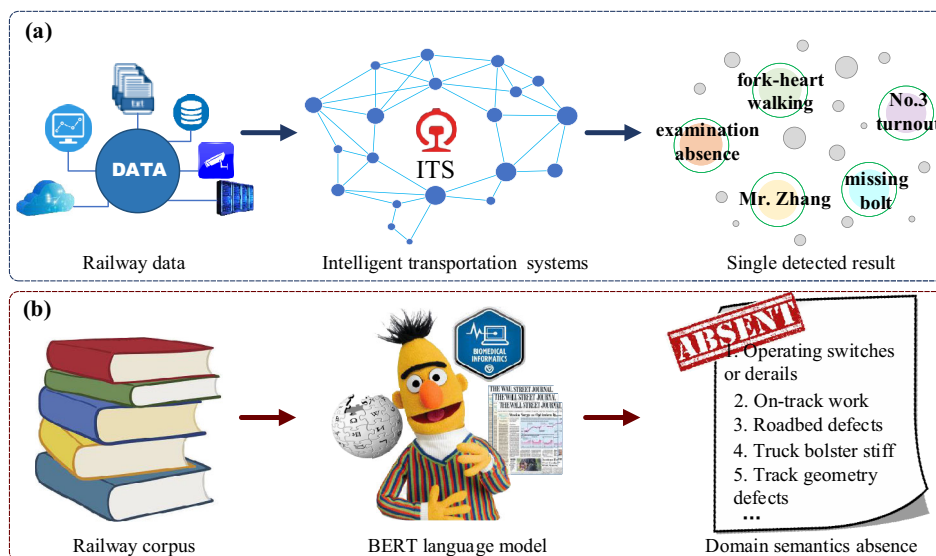
example, "turnout bolts" is an important and frequent phrase in the safety supervision log of the railway but missing in the open corpus, which do may do harm to understand railway text and thus impairs the final performance. It reveals that enhancing the domain semantics is important for understanding the railway text.

At present, only a few models attempt to address the aforementioned challenges. To our knowledge, a recent method [15] can be used to address the second challenge, which twice implements the pre-training on domain corpus to enrich domain semantics. However, two severe limitations are raised: (1) *How to effectively explore the semantic relation between railway objects?* and (2) *How to fully exploit domain semantics from limited Chinese railway corpus for a better result and extend the practical application?*

To solve these severe issues, this paper **recasts the identification of railway safety risk as the relation extraction task** and proposes a novel and effective **D**omain **S**emantics-**E**nhanced **R**elation **E**xtraction (DSERE) model. Specifically, we model railway safety risk as the triple knowledge $(e_i^{sbj}, R_i, e_i^{obj})$, where e_i^{sbj} and e_i^{obj} are the subject (or head entity) and object (or tail entity) of the railway safety risk interconnected by the relation R_i. The main advantage of introducing triple knowledge is that interconnection can be used for incident/accident inference. Afterward, our DSERE implements the relation extraction task by learning railway semantics simply and effectively. A domain semantics-enhanced transformer (DSET) mechanism is designed to obtain the domain semantics by classical N-Grams mechanism on the dedicated railway lexicon. The motivation behind it contains two aspects, for one thing, the scale of the railway lexicon is much smaller than the unsupervised corpus for twice pre-training [5, 18]. For the other thing, the domain semantics based on the statistical mechanism is more help-

ful in understanding the railway text. Then, we observe that the triple knowledge splits sentence into three segments, each contributing much to the nearest entity. Therefore, piece-wise convolutional neural networks (CNN) are introduced to capture fine-grained features. With the above mechanisms, the enhanced domain semantics is fully exploited to improve the performance of relation extraction to better identify the railway safety risk. Experimental results on real-world dataset show the effectiveness of our proposed model.

The main contributions of this paper are summarized as follows:

- A novel method for identifying the railway safety risk is proposed by extending the relation extraction task. To the best of our knowledge, this is the first work to project the railway safety risk into a knowledge interconnection space that can be used for incident/accident inference.
- In order to clearly express the interconnection of railway safety risk, we are the first to formally define three core concepts, that depict railway objects and their semantic relation from the perspective of knowledge.
- A novel DSERE model is proposed to enhance the domain semantics by designing an effective domain semantics-enhanced transformer mechanism, bridging the semantic gap between the railway and universal domains.
- Experiments on the real-world dataset verify the superiority and effectiveness of our proposed model. Our proposed DSERE has important practical values in promoting knowledge-driven safety risk management of China railway system.

The rest of this paper is organized as follows. The next section introduces fundamentals about identifying railway

safety risk, language representations and the practical application of relation extraction. The subsequent section presents the model for railway safety risk pattern and demonstrates the detail of the proposed DSERE. Experimental results and detailed analysis are shown in the penultimate section. The conclusion and future work are given finally.

Highlights

The main aspects of our work are shown as follows:

- Recast the identification of railway safety risk as the relation extraction task. Railway safety risk is formally defined from knowledge perspective, which is important for exploring the semantic relation between railway objects.
- Performance improvement on relation extraction. The proposed DSERE model includes a novel DSET mechanism enhancing the understanding of railway semantics from limited corpus, and thus greatly improves the result of relation extraction.
- Broad application of the proposed method. Since knowledge interconnection is important for mining semantic relation between railway objects for further incident/accident inference, the proposed DSERE model has broad applicability in scenarios of ITS, such as locomotive scenario and electrical scenario.

Related work

Identification of railway safety risk

Early identifying the railway safety risk significantly reduces the probability of incidents/accidents [1]. Recently, machine learning methods have been used to automatically identify railway safety risk from large-scale heterogeneous data instead of relying on manual detection [4, 24, 39]. Gabriel et al. [17] automatically detected the rail failure risk through an image processing approach to ensure the safety of rail transportation. Zhang et al. [53] proposed an MRSDI-CNN model to detect the rail surfaces, that satisfy the comprehensive, fast, and accurate detection process. Chen et al. [6] investigated using a data-driven method to identify the slight fault information from the signal features of the bogie. Gao et al. proposed the adaptive deep learning model to overcome the challenging task of inspecting the safety status of the swivel clevis, an important component in the high-speed railway. However, these existing studies fail to mine the relation between the railway objects and ignore the rich value contained in the text.

Language representations

Understanding abstract semantics is a severe challenge in the text mining fields [50, 54]. In the early pioneer work [27], a toolkit was developed to project words into continuous vectors through Continuous Bag-of-Words (CBOW) and Skip-gram (SKG) models. Afterward, Vaswani et al. [38] proposed the Transformer mechanism, a milestone to model sequence transduction entirely by the attention mechanism. The BERT [9], GPT [32] are proposed based on the encoder architecture and the decoder architecture of the Transformer. Yang et al. [47] proposed the XLNet that is a, a generalized autoregressive method trained in Transformer(-XL) architecture since the BERT model adopts a masking strategy that corrupts the dependency between masked positions. To fully explore the language universality for understanding low-resource languages, multilingual pre-training models [14, 35] spark interest and show the superiority of cross-lingual performance on the downstream tasks of the natural language processing community. The pre-trained models have already learned the intrinsic textual representations without prior knowledge, but they fail to capture specific domain semantics, especially the Chinese railway corpus.

Relation extraction

Relation extraction is the core process for constructing a knowledge graph, and has been used to address various practical problems in specific domain [42]. For example, Du et al. [10] proposed a relation extraction model to obtain knowledge in the manufacturing domains. Relation extraction is used in the biomedical domain to extract relevant information from text, such as research articles and electronic health records [21, 33]. Amin et al. [2] proposed an accurate benchmark for broad-coverage biomedical relation extraction, which can be used for extracting bio-molecular information extraction and identifying drug–drug interactions. In the webpage mining domain, Locard et al. [25] proposed a ZeroShotCeres model to extract structural triple knowledge through the "zero-shot" method from the webpages. More interestingly, Jie et al. [16] used relation extraction to learn deductive reasoning for the math word problem in the text. However, no relation extraction model is applied to identifying the railway safety risk from the Chinese railway text.

Table 1 provides an overview of related methodologies, including the recent methods of identifying railway safety risk, excellent language models and relation extraction work for solving domain challenges.

Complex & Intelligent Systems (2023) 9:6493–6507

Table 1 A survey of relevant papers

Year	References	Topic	Descriptions	Advantages/potential limitations
2017	[17]	▲	Two machine learning method based on DNN	Use DNN for wheel defect detection Neglect the interconnection between railway wagons
2018	[4]	▲	Use MLR, ANN, REPTree predict the risk level	Use supervised learning algorithm to predict risk Not applicable for railway domain
2020	[39]	▲	Use deep learning to model ITS issues	Connect the ML and ITS communities Neglect the interconnection between railway factors
2020	[6]	▲	Review on the FDD methods for high-speed trains	Use data-driven methods to implement FDDD Neglect the interconnection between faults
2022	[1]	▲	Overview the ITS (2000–2021)	Emphasize the importance of traffic risk prediction Neglect the interconnection between factors
2022	[24]	▲	Construct a BDA platform	Use deep belief network to analyze road safety Not applicable for railway domain
2022	[53]	▲	Propose a MRSDI-CNN model	Improve the detection performance by DNN Neglect the interconnection between factors
2022	[11]	▲	Propose a model combined adaptive SSN and local operators	Use DNN model detect swivel clevis defects of high-speed railway Neglect the interconnection between defects
2013	[27]	◆	Extend the Skip-gram model	Improve the quality of word vectors Ignore the context semantics
2017	[38]	◆	First propose the Transformer Mechanism	Solely using attention mechanisms Insensitive to the sequence features of language
2018	[32]	◆	First propose GPT model	Excellent performance in four NLU tasks Adopt unidirectional encoder
2019	[9]	◆	First propose BERT model	Deep represent unlabeled text Fail to represent Chinese railway text
2019	[47]	◆	Propose the XLNet model	Integrate autoregressive LMs and autoencoding LMs Train on universal open corpus
2021	[54]	◆	Propose a SDnet model	Use semantic prior to recover textures Not applicable for railway semantics
2021	[35]	◆	Propose a multilingual pretrained LMs	Combine the linguistic structure features with implicit semantic features Not applicable for railway semantics
2022	[14]	◆	Propose a MM-Deacon pre-trained language models	Use multilingual molecular embedding for chemical representation learning Not applicable for railway domain
2022	[50]	◆	Propose a joint pre-training framework JAKET	Taking the advantages of knowledge graph Not applicable for railway domain
2016	[33]	★	A domain invariant CNN model	Extract information from clinical texts Not applicable for railway domain
2020	[25]	★	Propose a ZeroShotCeres model	Use GNN based method to extract structured information from webpages Not applicable for railway domain
2021	[42]	★	Propose a NPORE model	Extract noun phrase-based RE in Chinese Not applicable for railway domain
2021	[43]	★	Propose a UNIRE model based on table encoder	Improve the performance of joint entity and relation extraction task Not applicable for single relation extraction task
2022	[10]	★	Propose a MKREM model	Use DNN to extract manufacturing knowledge Not applicable for railway domain
2022	[21]	★	Propose a FILR model	Extract relation facts from biomedical document Not applicable for railway domain
2022	[2]	★	Present a benchmark MedDISTANT19	Distantly supervised relation extraction for biomedical information extraction Not applicable for railway domain
2022	[16]	★	Propose a deductive reasoning model	Using relation extraction task to solve math word problem Not applicable for railway domain

The mark ▲ indicates work related with the identification of railway safety risk, ◆ indicates work related with language representations and ★ indicates work related with relation extraction

Springer

Table 2 Notations and explanations

Notation	Explanation
S	An input sentence
x_i	The ith word in S
e_i^{sbj}	The subject of the ith railway safety risk
e_i^{obj}	The object of the ith railway safety risk
R_i	The semantic relation between e_i^{sbj} and e_i^{obj}
E_i^{sbj}	The set of e_i^{sbj}
E_i^{obj}	The set of e_i^{obj}
D	The railway lexicon
s_g	The railway terminology
c_i	The ith word of s_g
Q	The query matrix
K	The key matrix
V	The values matrix
$\mathbf{P}_i$	Result of the ith convolutional operation

The proposed method

Problem formulation

The railway safety risk is the unsafe state of railway systems, including human, infrastructure, environmental and management safety risks. This study models railway safety risks to interconnect railway objects as a triple knowledge $(e_i^{\text{sbj}}, R_i, e_i^{\text{obj}})$. Table 2 presents the notations used in this paper. This paper introduces three core concepts, defined as follows.

Definition 1 (*Subject*) It describes the subject of the railway safety risk in the text, and let E^{sbj} be a set that includes the subject of human safety risk E_1^{sbj}, the subject of infrastructure safety risk E_2^{sbj}, the subject of environment safety risk E_3^{sbj}, and the subject of management safety risk E_4^{sbj}.

We generate the subset E_1^{sbj} with railway workers such as line workers, guard workers and flaw detection workers to explore the subject semantics. For the subset E_2^{sbj}, we focus on the infrastructure such as rails, ballast beds and roadbeds. When considering the environmental factors that seriously affect the railway operation, we present the subset E_3^{sbj} as a set of extreme weather (e.g., strong winds, floods and earthquakes.). The subject of management safety risk E_4^{sbj} mainly includes management rules and regulations, including the "Rail Traffic Reliability, Availability, Maintainability and Safety Specifications and Examples"and "Railway Subgrade Major Maintenance Rules".

Definition 2 (*Object*) It describes the object of the railway safety risk in the text, and let E^{obj} be a set that includes the

object of human safety risk E_1^{obj}, the object of infrastructure safety risk E_2^{obj}, the object of environment safety risk E_3^{obj}, and the object of management safety risk E_4^{obj}.

This paper explores object semantics by mining the object corresponding to the subject. Specifically, E_1^{obj} represents unsafe behavior of workers, such as not wearing reflective tape, not implementing "hand-to-mouth", and not getting off the road in time, which is vulnerable to induce personnel injuries. The E_2^{obj} means the unsafe state of infrastructure that will induce potential dangers, for example, cracks, subsidence, siltation and collapse. For E_3^{obj}, extreme weather can cause safety hazards and thus induce the unsafe states of humans or infrastructure. The E_4^{obj} represents the human (e.g., grass-roots workers) or infrastructure who violates those management rules and regulations in E_4^{sbj}.

Definition 3 (*Relation*) It describes the interconnected relation between the subject and object of the railway safety risk. Let $R = \{R_1, R_2, R_3, R_4\}$ be a set representing the types of the relation, where $R_i \in R$ represents the relation between E_i^{sbj} and E_i^{obj}.

Based on the above definitions, identifying railway safety risk is recast as extracting the relation R between E^{sbj} and E^{obj} from the input sentence S. Unlike the existing models that regard railway objects as independent, our relation extraction method interconnects the subject and object of railway safety risk by the relation, and thus generates a space of knowledge interconnection [40]. This space can be viewed as a knowledge graph for analyzing incidents/accidents by knowledge inference, since the relation fully explores the chain of incident/accident evolution. As Fig. 2 shows, (No.3 turnout, R_2, missing bolt) is extracted from context "missing bolt is occurred on the east point heel of No.3 turnout", which expresses the safety risk of *turnout* induced by a *missing bolt*. When analyzing the accident depicted as "a train derails beside the No.3 turnout, causing the worker Mr. Zhang to be injured", the knowledge graph can be searched to get the triple knowledge (No.3 turnout, R_2, missing bolt) and (Mr. Zhang, R_1, fork-heart walking) that show the direct cause of the train derailment accident. Furthermore, to analyze the "Mr. Zhang" node, we discover a triple knowledge (Mr. Zhang, R_4, examination absence), indicating that the above-mentioned accident's root cause is not strictly implemented in the management measures.

Modeling the railway safety risk as the triple knowledge interconnects the railway objects opens a new dimension for identifying of the railway safety risk. The next subsection explores the new and uncultivated areas that even were considered as a place that cannot be probable. Additionally, we show how to construct dedicated neural networks to automatically extract relation in railway systems.
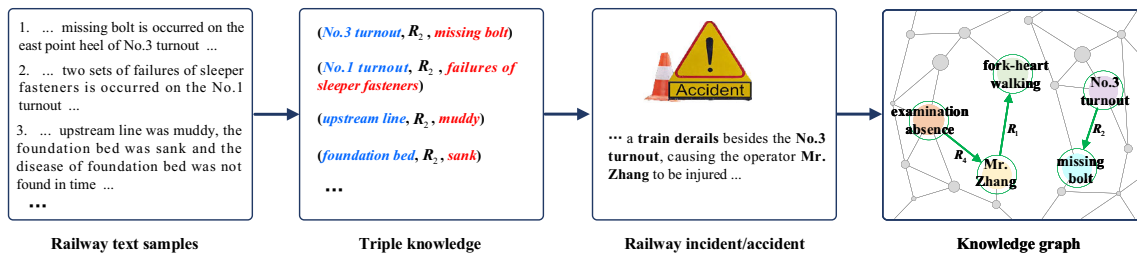
Fig. 2 A novel method to identify the railway safety risk by relation extraction. This method can project the railway safety risk into a knowledge interconnection space and support the incident/accident inference

DSERE model

The most challenging problem when extracting the relation from railway text is how to deeply understand abstract railway semantics. A common method is to twice pretrain the pre-trained model on large unsupervised domain corpus, which reshapes the adaptability of the deep neural networks to enrich the domain semantics. However, this method consumes computing resources and introduces a new problem of large-scale unsupervised railway corpus shortage. This study proposes a novel DSERE model to deeply understand railway semantics with a simple and effective model to solve this issue, as shown in Fig. 3. The DSET mechanism is designed to fuse railway and universal semantics, where the important railway semantics is extracted from a dedicated railway lexicon with railway domain terminologies through the N-Grams mechanism. Furthermore, a piece-wise CNN is used to extract the fine-grained features in the triple knowledge structure. With the above mechanisms, the proposed model can fully understand the railway semantics and extract relation using the softmax classifier.

Specifically, let $S = \{x_1, x_2, \ldots, x_{k_1}\}$ be the input sentence, and D be the railway lexicon with railway terminologies. The DSET mechanism aims to obtain enhanced domain semantics to improve the performance of our DSERE model. We introduce the N-Grams mechanism to explore word boundary features from the statistical perspective. Let $s_g = \{c_i^g | c_i^g \in D, i \in [1, n]\}$ be a railway terminology, and the joint probability of N-Grams is expressed as follows:

$$p(c_1^g, c_2^g, \ldots, c_n^g) = \prod_{i=1}^{n} p(c_i^g | c_{i-N+1}^g, \ldots, c_{i-1}^g), \qquad (1)$$

the N-Grams matching matrix $\mathbf{M} \in \mathbf{R}^{k_1 \times k_2}$ is generated, where k_2 means numbers of extracted N-Grams. Then the n-gram embedding is fused with word embedding of S through $\mathbf{M}$, and utilizes the transformer architecture to encode it as follows:

$$\begin{cases} u_j^{l+1} = MHA(Q = u_j^l, K = V = U^l) \\ v_i^* = v^l + \sum_k u_{i,k}^l \\ \mathbf{V}^{l*} = \mathbf{V}^l + \mathbf{M} \times U^l, \end{cases} \qquad (2)$$

where $MHA(\cdot)$ means the multi-head self-attention layer, u_j^l is used to query the vector Q that calculates the attention score of other N-Grams elements in the l layer, and U^l is a matrix representing the characteristics of all N-Grams elements in the l layer, providing information for K and V in MHA. The v^l and $u_{i,k}^l$ represent the ith word of l layer and the kth N-Grams element associated with the word. $\mathbf{V}^l$ is the embedding matrix of all words and v_i^{i*} is the embedding result of the next input layer.

Capturing only the railway semantics is not capable of identifying the relation, we have to explore the structural features of triple knowledge. Recurrent neural networks (RNNs) are usually used to extract temporal features in the text, but they are hard to train and accumulate redundant information in the last hidden unit [3, 45]. To alleviate this issue, CNN-based models are proposed and perform comparably to RNNs on sequence modeling [12]. In addition, the subject $e^{sbj} \in E^{sbj}$ and the object $e^{obj} \in E^{obj}$ split the input sentence S into three segments, and each segment contributes a lot to the nearest entity. Therefore, a piece-wise max-pooling operation is introduced in each segment to capture fine-grained features instead of implementing a single max-pooling over the entire sentence. Formally, the ith convolutional operation is defined as follows:

$$\begin{cases} \text{COV}_i = \tanh(Wq + b_1)_i \\ \mathbf{P}_i = [\text{MaxP}(\text{COV}_{i1}), \text{MaxP}(\text{COV}_{i2}), \text{MaxP}(\text{COV}_{i3})], \end{cases} \qquad (3)$$

where $q = x_{i-l_1+1:i}$ ($1 \le i \le k_1 + l_1 - 1$, l_1 is the convolution kernel size) is the concatenation of the word vector sequence in the ith sliding window, W is the convolution kernel matrix and b is the bias vector and $\tanh(\cdot)$ is the activation function. The COV_{ij} corresponds to the convolutional result of the segment split by e^{sbj} and e^{obj}, and the $\text{MaxP}(\cdot)$ is the max pooling operation.
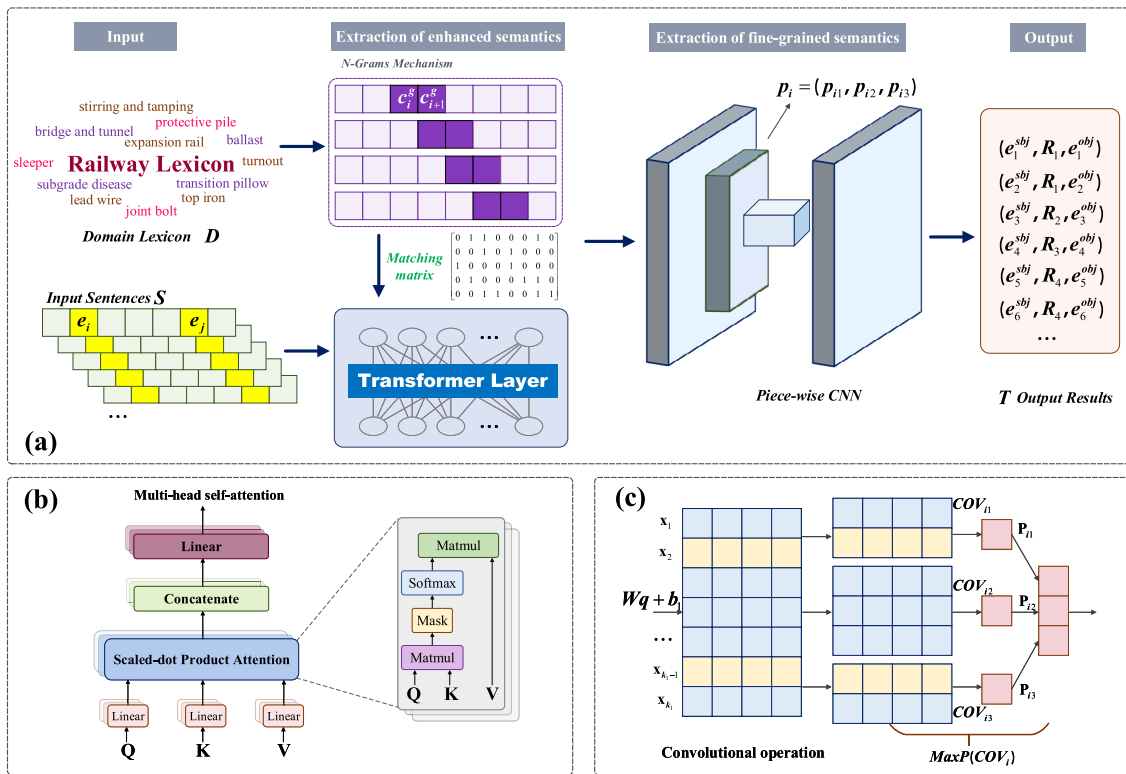
Fig. 3 Overview of DSERE model. **a** The DSERE model includes two important components—the domain semantics-enhanced transformer mechanism and piece-wise CNN. **b** The encoder details of the Transformer layer aim to bridge the gap between domain and universal semantics. **c** The encoder details of piece-wise CNN aim to extract fine-grained features contained in the triple knowledge. The convolution window slides from $\mathbf{x}_1$ to $\mathbf{x}_{k_1}$, and its size is closely related to the input sentence

The output of the piece-wise CNN is a score matrix **O** that associates with the types of relation. Then for each sentence S, we can calculate probability of the relation type by a softmax operation as follows:

$$p(r_i|S) = \frac{\exp(\mathbf{O}_i)}{\sum_{j=1}^{k_1} \exp(\mathbf{O}_j)}. \tag{4}$$

Optimization

Our DSERE model aims to automatically enhance the domain semantics and capture fine-grained features for extracting the relation. The objective function of the proposed model is defined as follows:

$$\mathbf{J}(\theta) = \sum_{S} \log p(r_i|S, \theta), \tag{5}$$

where θ is a parameter set learned by deep neural networks and $r_i \in R$ represents the real label of relation types. Following previous work [36], we adopt gradient descent algorithm to optimize this objective function. Algorithm 1 summarizes the complete procedure for our DSERE model.

Algorithm 1 The DSERE model for extracting the relation

1: **Input:** Input sentence S, entity pairs labels (e^{sbj}, e^{obj}), learning rate ϵ, dropout hyper-parameter λ.
2: **Initialization:** Generate initial representations of S by pre-training language model, e.g. BERT.
3: Construct a railway lexicon D.
4: **repeat**
5:     Matching matrix $\mathbf{M} \leftarrow$ N-Grams mechanism captures the railway semantics via Eq. (1)
6:     Enhanced representations of $S \leftarrow$ Merge the railway semantics and universal semantics via Eq. (2)
7:     Score matrix $\mathbf{O} \leftarrow$ Extract fine-grained structural features via Eq.(3)
8:     $p(k_i|s) \leftarrow$ Calculate the probability of relation r_i via Eq.(4)
9:     Obtain the overall objective function via Eq.(5)
10: **until** converge
11: **Output:** Final classification result of relation type r_i.

Experiments

This section presents the extensive experiments on DSERE using a real-world dataset. These experiments aim to verify that the proposed DSERE model can identify railway safety risks through relation extraction. We compare our model with the classical models to show its priority in

understanding railway semantics. Additionally, an ablation experiment is presented to verify the effectiveness of the DSET mechanism. We also explore how the fine-grained features contained in the triple knowledge structure improve the proposed model's performance. The overall performance of our model is validated by comparing with recent derivation models. A case study demonstrates that the proposed DSERE greatly enhances the railway semantics, which helps to classify types of the relation.

Experimental settings

Dataset. The railway maintenance safety risk (RMSR) dataset is the safety supervision log from the South zone of the China Railway System. RMSR depicts the unsafe production status of track maintenance recorded by safety inspectors. It is important to identify the railway safety risk from the RMSR dataset, since the railway maintenance section is the core department. This will ensure the safe operation of the overall railway system. Introducing relation extraction is a novel and intelligent approach to mining the RMSR dataset, showing its practical values and potential values of extending to other departments.

Six months were spent manually annotating the RMSR dataset, and three master volunteers annotated each sentence. According to Definition 1, Definition 2 and Definition 3, four types of railway safety risk relations and eight types of entities were identified. Detailed information on the RMSR dataset is presented in Table 3. Each training input sentence is truncated if the length exceeds 512 tokens. This preprocessing operation helps to speed up the acquisition of railway semantics since controlling the sequence length reduces the training cost of the DSET mechanism.

Evaluation metrics. Following previous studies [31, 43], the held-out evaluation is used to verify the performance of our proposed model. Specifically, 80% of the RMSR dataset was randomly selected as the train set and 20% as the test set In our experiments, Precision@N (P@N), AUC, F1 score, and precision/recall curves were adopted to verify the effectiveness of DSERE.

Baselines. To better understand the domain semantics especially the Chinese railway semantics, we adopt dynamic language representations to initialize the RMSR dataset. In our experiments, we compare the DSERE with different

semantics to verify the effectiveness of enhanced domain semantics. Meanwhile, we choose classical DNN models as competitors to demonstrate the importance of fine-grained features for results of relation classification. DSERE was also compared with recent derivation models to further validate the overall performance of our proposed model.

Language models:

W2V: This is a state-of-the-art toolkit of distributed word vector representations proposed in [27], which learns semantics through the two-layer neural networks. Following previous work [23], we represent it as W2V in the subsequent experiments.

ELMo: This is a state-of-the-art language model based on Bi_LSTM neural networks proposed by Matthew et al. [30], which extracts sequence features in the text to obtain context-dependent semantics.

GPT: This model [32] has an excellent semantic understanding ability for text generation tasks. It extracts semantics through the decoder of Transformer mechanism, and adopts a unidirectional left-to-right encoder.

BERT: Devlin et al. [9] proposed this state-of-the-art language model, which improves the natural understanding ability by the deep bidirectional encoder of transformers. The model fully exploits the contextualized features and has been successfully applied in various natural language understanding tasks, such as machine reading comprehension and natural language inference tasks.

XLNet: The core component of this model [47] is the two-stream attention mechanism that takes advantages of auto-regressive and auto-encoding mechanisms to learn semantics. It employs a permutation operation to capture high-order, long-range dependency for downstream tasks.

Classical DNNs:

Bi_LSTM: This model is proposed in the literature [46] to reduce the reliance on handicraft features. Using bidirectional recurrent neural networks, it captures dependency features between entity pairs and context.

CNN: Instead of using Bi_LSTM to capture structural features of the triple knowledge, CNN is proposed in literature [52] to implement relation classification. This network architecture can extract the lexical and syntactic features from the raw sentence by the convolution operations.

PCNN: This is an improved version of CNN architecture that splits the pooling layer into three segments according to the positions of entity pairs. This operation can fully extract fine-grained features for identifying the types of relation.

Table 3 Details of RMSR dataset

Element	Train	Test
Sentences	21,476	5369
Entities	19,841	6029
Relation	8746	1741

Recent derivation models:

SeG: This is a light-weight self-attention neural framework [22], which designs a pooling-equipped gate instead of a selective attention mechanism.

MVC: To avoid the impact of external parsers, Amir et al. [41] proposed to extract features from structure view and semantic view. The mutual information is introduced to encourage these two views consistencies.

FIFRE: This model is proposed in literature [55] to extract relation by a representation iterative fusion mechanism based on the heterogeneous graph.

A-GCN: Avoid being affected by the noise of dependency tree, Tian et al. [37] proposed the attention graph convolutional network for the relation extraction task.

FastRE: This model is proposed to improve the efficiency of relation extraction task [20] by taking advantages of the convolutional encoder and the improved cascade binary tagging framework.

TACNN: Geng et al. [12] integrate the CNN model and target attention mechanism for extracting semantic relation to fully explore the information of word embeddings and position embeddings.

Hyperparameter settings

In maintaining a fair comparison, this study chooses the TensorFlow framework to implement the proposed model and tune hyperparameters. In the DSET module, we adopt a bidirectional Transformer networks that consist of 12-layer Transformer blocks and 12 self-attention heads. The dimension of word embedding (or hidden size) is set to 768. In the PCNN module, we use one convolutional layer and a piecewise maximum pooling layer, where the window size is set to 3 and the hidden size is set to 230 followed the previous work [51]. The batch size is 256, mainly considering better optimization when computing resources are satisfied. We tune the learning rate in {1e-5,3e-5,1e-4,3e-4,1e-3,3e-3,1e-2}, and show the optimal hyper-parameters in Table 4. We conduct all experiments on a server equipped with Intel(R) Xeon(R) Silver 4216 CPU @ 2.10 GHz on 4 NVIDIA Tesla V100 GPU (32GB for each) and Ubuntu 14.04 operating system.

Performance analysis

Comparison With the SOTA models

To evaluate the performance of our proposed DSERE, we compare it with competitive models on the RMSR dataset. Table 5 shows the performance of the three metrics, and

Table 4 Hyper-parameters settings

| Description of parameters	Value
The size of window	3
The hidden size of convolution layer	230
The dimension of word embedding	768
Batch size	256
Learning rates	0.001

some comparison results are marked to better demonstrate our interesting findings. The rows in the table are models that combine various semantic features with different knowledge structural features. There are interesting observations as follows:

- The P@N metric calculates the average precision of top N test data, and the higher values are, the better performance obtained by the model. It can be seen that our DSERE model can always get better results than other models by a large margin especially on P@200 and P@300 metrics, which shows the effectiveness of our model. Although the DSERE model on the value of the P@100 metrics is slightly lower than XLNet+PCNN, it surpasses the XLNet+PCNN on values of P@200, P@300 and average accuracy with the increase in the amount of dataset. This indicates that our proposed model better understands the railway semantics and thus helps to identify the relation between the subject and object of the railway safety risk.

- It is observed that models based on the enhanced domain semantics outperform other models on the AUC metric. For example, observing the CNN-based models and Bi_LSTM-based models, we find that the DSET+CNN model and the DSET+Bi_LSTM model respectively outperform 6.18% and 5.75% with those optimal models based on same structural features. For the PCNN-based models, our DSERE model achieves the best performance, meaning that enhancing domain semantics enables models to deeply understand the Chinese railway text.

- The proposed model achieves the 76.00% result on F1 score, which is superior to other models on this metric. The DSERE model achieves 2.05% absolute gain in the F1 score compared with the state-of-the-art model XLNet+PCNN. On one hand, this is because the enhanced domain semantics can bridge the gap between railway semantics and universal semantics, breaking through the challenge problem of neural networks in mining the Chinese railway text. On the other hand, the PCNN module can capture the fine-grained structural features of the triple knowledge simply and effectively, promoting the performance of identifying types of relation.

Table 5 Performance comparisons of different models

Method	P@N				AUC	F1
	100	200	300	Mean		
W2V+Bi_LSTM	91.09	87.56	85.71	88.13	55.72	58.09
W2V+CNN	87.03	87.56	84.72	86.47	56.85	58.17
W2V+PCNN	92.08	89.55	85.72	89.12	57.11	58.32
ELMo+Bi_LSTM	80.20	81.09	78.74	80.01	55.10	57.69
ELMo+CNN	91.09	89.55	86.71	89.12	57.02	58.24
ELMo+PCNN	91.09	89.55	85.71	88.79	57.09	58.30
GPT+Bi_LSTM	89.10	91.04	89.04	89.73	69.31	67.58
GPT+CNN	85.14	86.07	86.04	85.75	69.97	68.14
GPT+PCNN	90.09	88.06	88.04	88.73	71.95	69.69
BERT+Bi_LSTM	84.16	87.06	87.04	86.09	57.22	59.38
BERT+CNN	84.16	81.09	81.40	82.22	60.96	60.30
BERT+PCNN	91.09	90.05	91.36	90.83	75.25	71.69
XLNet+Bi_LSTM	91.09	83.08	80.73	84.97	55.61	57.79
XLNet+CNN	72.28	74.13	77.08	74.49	66.33	66.91
XLNet+PCNN	**96.03**	94.05	93.68	94.73	79.44	73.95
DSET+Bi_LSTM	92.08	91.04	89.70	90.94	75.06	71.34
DSET+CNN	87.13	88.06	87.71	87.63	76.15	73.01
DSERE	95.04	**94.53**	**95.02**	**94.86**	**81.84**	**76.00**

Bold values are marked to better highlight the superiority of the DSERE model
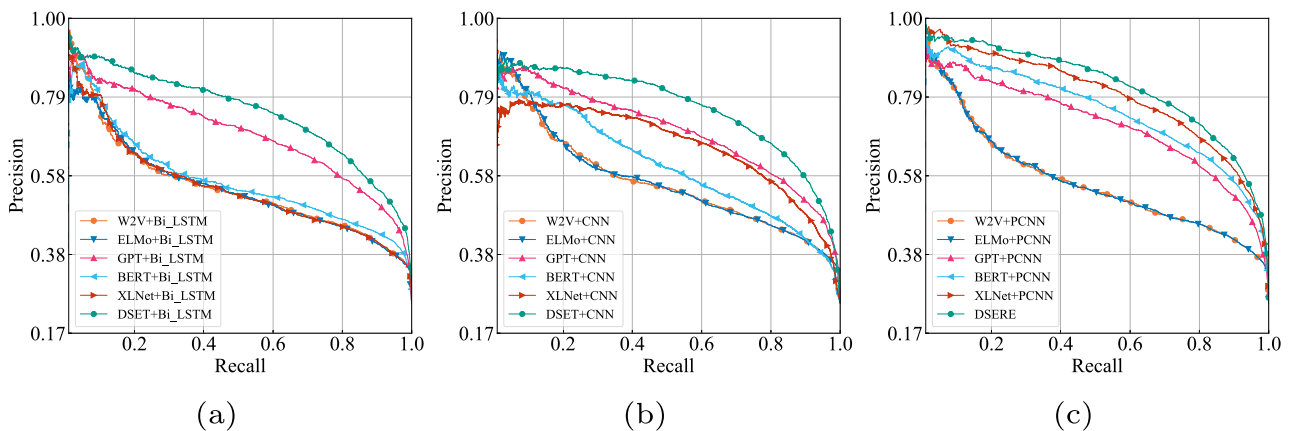
Fig. 4 The precision–recall comparisons of domain semantics-enhanced analysis

Furthermore, we use aggregate precision/recall curves to intuitively compare the impact of enhanced domain semantics and universal semantics on the real RMSR dataset and show them in Fig. 4. It is observed that enhanced domain semantics-based models consistently outperform universal semantics-based models, demonstrating that enhancing railway semantics improves the performance of identifying relation. The permutation operation of the XLNet architecture that addresses the data corruption problem still fails to address the insufficient domain semantics, especially the Chinese railway semantics. This means that the proposed model automatically enhances railway semantics and demonstrates its potential to support practical applications in railway systems.

Ablation study of DSET mechanism

The DSET mechanism is an important module of our proposed model, that aims at enhancing the domain semantics for relation classification. In this section, we conduct ablation experiments to explore how DSET improves the performance of our DSERE model. Through pre-trained models capture features of a universal language framework, they fail to fully understand the specific domain semantics, especially Chinese railway semantics is hidden in a scarce corpus. As Table 6 shows the ablation results of DSERE model on the real-word dataset, we can easily see that the DSET mechanism significantly improves the performances of our proposed model. For example, on the Mean of the P@N metric, the

Table 6 Ablation analysis of domain enhanced semantics

Setting	P@N				AUC	F1
	P@100	P@200	P@300	Mean		
w/o DSET	91.09	90.05	91.36	91.36	75.25	71.69
w/o N-Grams	93.02	92.54	91.69	92.41	79.47	74.19
DSERE	95.04	94.53	95.02	94.86	81.84	76.00
Improve	2.02	1.51	3.33	2.43	2.37	1.84

DSERE model increases by 2.43% than the DSERE model without the N-Grams mechanism, which is significant in the practical identification of the railway safety risk. In addition, we observe that without DSET mechanism, our model degenerates BERT+PCNN model only considering universal semantics and fine-grained structural features of the triple knowledge. The performance of the DSERE model without the *N*-Grams mechanism is superior to the DSERE model without DSET, since the plain transformer can capture more features for the final classification. However, both of these two compared models fail to enhance the domain semantics and show the limitations of neural networks in understanding the Chinese railway text. Our DSET mechanism can fully learn the railway semantics for identifying the types of relation in the practical scenes of China railway system.

Fine-grained feature analysis

Three classical neural networks were compared based on the same semantics to verify the effectiveness of fine-grained feature extraction, as shown in Fig. 5. This figure shows that PCNN-based models consistently outperform Bi_LSTM and CNN-based models, consistent with the values in Table 5. This phenomenon verifies that capturing fine-grained features can improve the performance of relation classification, since the entity pair splits the input sentence into three
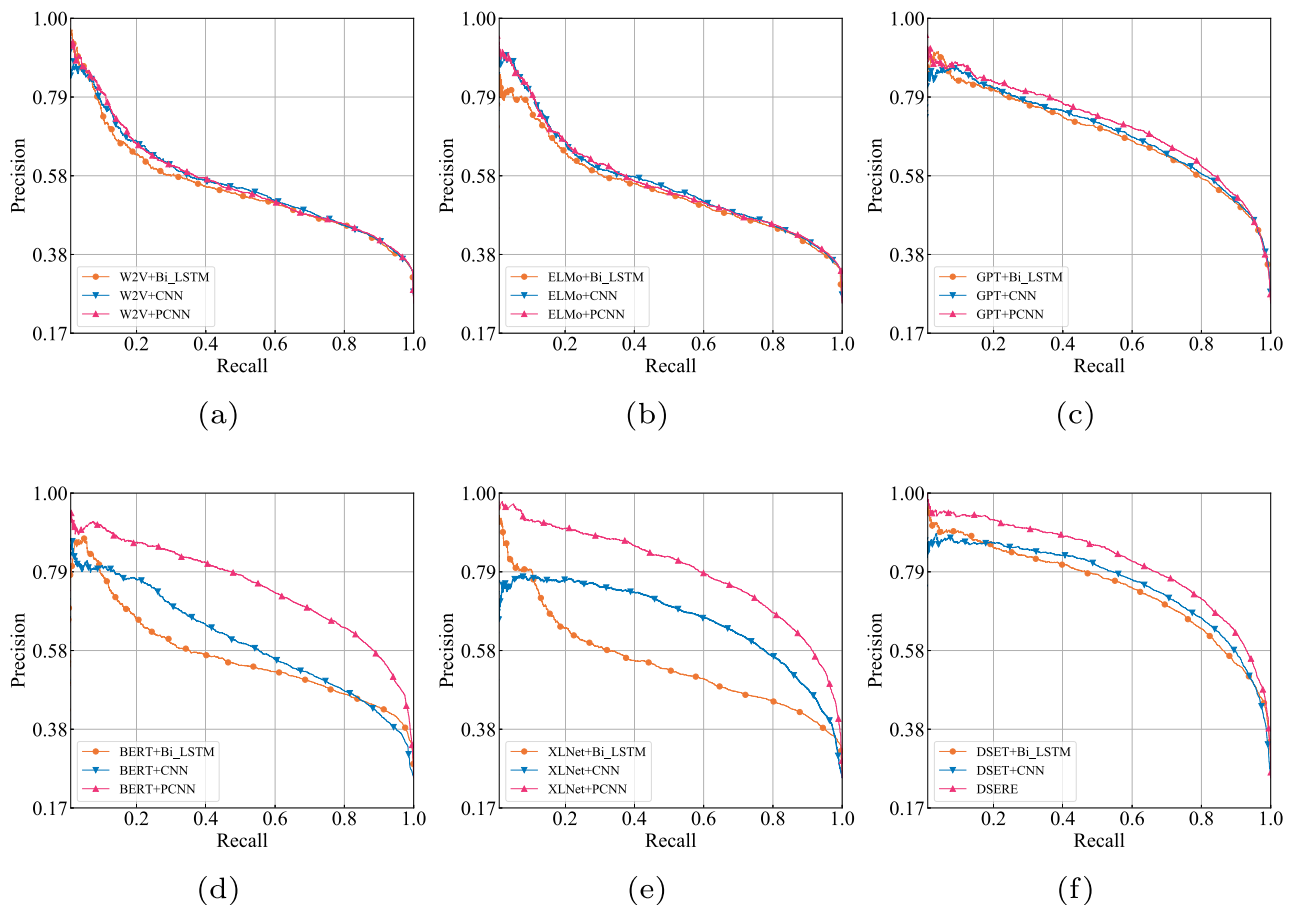

Fig. 5 The precision–recall comparisons of knowledge structural features analysis

Table 7 Performance comparisons of recent derivation models

Model	Semantics	Solution	AUC	F1
SeG (AAAI'20)	W2V	Self-attention mechanism with selective gate mechanism	60.34	71.24
MVC (AAAI'20)	W2V	ON-LSTM model with self-attention mechanism	57.77	69.98
FIFRE (KBS'21)	BERT	Heterogeneous graph neural networks	58.21	71.86
A-GCN (ACL'21)	BERT	Attentive graph neural networks	60.03	75.43
FastRE (IJCAI'22)	Glove	Convolutional encoder with improved cascade binary tagging framework	54.18	67.12
TACNN (IS'22)	W2V	Target attention convolutional neural networks	55.76	68.29
DSERE (Ours)	DSET	Piece-wise convolutional neural networks	**81.84**	**76.00**

Bold values are marked to better highlight the superiority of the DSERE model

segments and each segment contributes a lot to the nearest entity. The unique contribution of each segment may decrease if we use only a max-pooling operation on the whole sentence. Furthermore, the PCNN module is a convenient and efficient network architecture that captures fine-grained features through the piece-wise pooling operation without introducing extra parameters and complex network layers. This convenience enables the DSERE model to be more practical in real scenes of the China railway systems.

Comparison with the recent derivation models

The comparison results with recent competitive models are shown in the Table 7. We summarize the semantics, solution, metrics of AUC and F1 scores to evaluate the overall performance. The performance of competitors is inferior to our proposed model, because the compared models only use the universal semantics but ignore the domain semantics. This reveals that only capturing universal semantics and then constructing complex networks of relation extraction networks may be ineffective in addressing the challenges of identifying railway safety risk. In contrast, our proposed model can fully explore and utilize the railway semantics for performance improvement. It is also worth mentioning that our model can capture meaningful fine-grained features with a convenient and effective PCNN module.

In our experiments, we also test other configurations to set optimal coefficients. When we change the structure of the DSET mechanism to larger networks with 24-layer Transformer blocks, 1024 hidden units and 16 self-attention heads, the performance of model has no gain in the metrics (in some cases even declines at least 3%). This is because the domain semantics from the railway lexicon overfit in the deep networks, resulting in obtained semantics being close to the universal semantics. For the PCNN architecture, we try to set the window size as 1 and 5; however, this operation does

not improve the experimental results. This is because the convolutional operation fails to capture temporal and syntactic features of input sentence with the window size 1. When window size is 5, the convolutional operation cannot effectively obtain phrase features, since the length of the railway phrase is usually 3. Therefore, we present the DSERE architecture as the final model for identifying the railway safety risk.

Case study

To further compare the effectiveness of our proposed DSERE model, we randomly select three samples from test set of RMSR to report some micro-level case studies in Table 8. Specifically, we show each sample in three lines, where the top line is the ground truth, the second line is the result of compared model and the last line is the result of our proposed model. Moreover, we use the "[]"symbol with a subscript to express the entity pairs and their relation types. Inspecting these cases, we have the following observations: (1) Pre-trained model BERT is trained on the open domain corpus that lacks railway words, e.g., "tamping rod", "switch rail", and thus fails to understand the semantics of domain words. Then the BERT+Bi_LSTM model mistakes the relation type that influenced by the dependency of the word "artificially", and incorrectly identifies the relation types of S_1 as R_1 representing the relation of human safety risk. (2) The PCNN module can capture fine-grained structural features to improve the performance of identifying the relation. For example, in the instance S_3, the ELMo+CNN recognizes the relation between "Mr. He" and "absented from the business examination" as R_1 (the relation of human safety risk). At the same time, the fine-grained features from "employee educational ledger" indicate the ground truth is R_4 means the relation of management safety risk. From the above analysis, our proposed DSERE model can correctly identify railway safety risk based on enhanced domain semantics and fine-grained features of knowledge structure.

Table 8 Examples from the RMSR test set for comparison

Models	Output
Ground Truth	S_1: ... 送修 [捣固棒]$_{R_2}$, 发现 人为 造成 捣固棒 [缸体 破裂]$_{R_2}$, 导致 养路 机械 报废 ...
	..., when the [tamping rod]$_{R_2}$ was sent for repair, it was found that the cylinder body of tamping rod was [broken]$_{R_2}$ artificially, resulting in the scrapping of the maintenance machinery, ...
BERT+Bi_LSTM	... 送修 [捣固棒]$_{R_1}$, 发现 人为 造成 捣固棒 [缸体 破裂]$_{R_1}$, 导致 养路 机械 报废 ...
DSERE	... 送修 [捣固棒]$_{R_2}$, 发现 人为 造成 捣固棒 [缸体 破裂]$_{R_2}$, 导致 养路 机械 报废 ...
Ground Truth	S_2: ... 发现 XX 号 复交 [道岔]$_{R_2}$ 前 转辙 甲股 尖轨跟 [接头 错牙]$_{R_2}$ 2mm ...
	..., it is found that the joint of the a-strand [switch rail]$_{R_2}$ and the joint in front of the switch of No. XX double crossing are [staggered]$_{R_2}$ by 2mm, ...
GPT+Bi_LSTM	... 发现 XX 号 复交 [道岔]$_{R_1}$ 前 转辙 甲股 尖轨跟 [接头 错牙]$_{R_1}$ 2mm ...
DSERE	... 发现 XX 号 复交 [道岔]$_{R_2}$ 前 转辙 甲股 尖轨跟 [接头 错牙]$_{R_2}$ 2mm ...
Ground Truth	S_3: ... 维修工区 职工 教育 台账 , 发现 工区 员工 [何某]$_{R_4}$ X 月份 业务学习 [考试 缺考]$_{R_4}$ ...
	... on employee educational ledger of maintenance work area, it is found that [Mr. He]$_{R_4}$, an employee in the maintenance work area, [absented from business examination]$_{R_4}$ in one of months, ...
ELMo+CNN	... 维修工区 职工 教育 台账 , 发现 工区 员工 [何某]$_{R_1}$ X 月份 业务学习 [考试 缺考]$_{R_1}$ ...
DSERE	... 维修工区 职工 教育 台账 , 发现 工区 员工 [何某]$_{R_4}$ X 月份 业务学习 [考试 缺考]$_{R_4}$ ...

Discussion

Strengths of the proposed method

This section mainly discusses the strengths of our proposed method from four aspects.

- This paper uses relation extraction to explore a new and uncultivated area to identify the railway safety risk. The method projects the railway safety risk into a knowledge interconnection space, supporting the incident/accident inference in practical applications.
- Our proposed DSERE model can bridge the gap between the railway and universal semantics, then obtain better representations.
- This study captures fine-grained features by a simple and effective PCNN architecture instead of complex encoder networks.
- The extensive experimental results show that our DSERE model is more suitable for identifying railway safety risk in the practical scenario. The promising performance indicates that the proposed model has potential to be a convenient and intelligent assistant for the knowledge-driven management of railway safety risk.

Weakness of the proposed method

Two limitations of the proposed method are still worth solving. First, the ambiguity of unimodal semantics may occur with a small probability that would affect the model's understanding of the input sentence. A multi-modal semantic enhanced relation extraction is worth exploring in future work. Second, the design of the proposed model is only for the railway scenario, in which the domain semantics cannot support understanding the other transportation scenarios, such as road transport, air transport and water transport. For example, roads are the key to maintenance in road transport [44], but the semantics of railway (e.g., rail) cannot be directly applied to road scenarios. In the future, we will focus more on identifying safety risk in various transportation scenarios.

Conclusion

In this paper, we introduce the triple knowledge of knowledge graph to model the railway safety risk with a knowledge interconnection mode, and recast its identification as the

relation extraction task. Afterward, a novel and effective DSERE model is proposed to extract the relation between the subject and object of the railway safety risk. Instead of twice pre-trained models, a simple and effective domain semantic enhanced transformer mechanism is designed to bridge the semantic gap between Chinese railway and universal domains. Furthermore, our DSERE model capture the fine-grained features contained in the structure of triple knowledge through a convenient piece-wise pooling operation. Experimental results on the real-world dataset verify the effectiveness of our proposed model and the potential to be an intelligent assistant for the knowledge-driven management of railway safety risk.

In the future, we will concentrate on solving the two limitations that discussed in the subsection of Discussion part. In addressing the first limitation, we plan to integrate image semantics of entity pairs to alleviate the ambiguity of natural language. While for the second limitation, we plan to investigate the similarities and differences between the railway and other transportation scenarios, and then reshape the domain lexicon to improve the transferability of the DSERE model.

Data availability The authors do not have authority to make the dataset publicly available.

Declarations

References

1. Abduljabbar R, Dia H (2022) A bibliometric overview of IEEE transactions on intelligent transportation systems (2000–2021). IEEE Trans Intell Transp Syst 23(9):14066–14087

2. Amin S, Minervini P, Chang D, et al (2022) Meddistant19: towards an accurate benchmark for broad-coverage biomedical relation extraction. In: Proceedings of the 29th International Conference on Computational Linguistics. International Committee on Computational Linguistics: 2259–2277

3. Bai S, Kolter JZ, Koltun V (2018) Convolutional sequence modeling revisited. In: Proceedings of 6th International Conference on Learning representationss. OpenReview.net, pp 0–20

4. Boodhun N, Jayabalan M (2018) Risk prediction in life insurance industry using supervised learning algorithms. Complex Intell Syst 4(2):145–154

5. Cai X, Liu S, Yang L et al (2022) Covidsum: a linguistically enriched scibert-based summarization model for COVID-19 scientific papers. J Biomed Inform 127(103):999

6. Chen H, Jiang B (2020) A review of fault detection and diagnosis for the traction system in high-speed trains. IEEE Trans Intell Transp Syst 21(2):450–465

7. Chen C, Li K, Cheng Z et al (2022) A hybrid deep learning based framework for component defect detection of moving trains. IEEE Trans Intell Transp Syst 23(4):3268–3280

8. Dai X, Zhao H, Yu S, et al (2021) Dynamic scheduling, operation control and their integration in high-speed railways: A review of recent research. IEEE Trans Intell Transp Syst: 1–17

9. Devlin J, Chang M, Lee K, et al (2019) Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. ACL, pp 4171–4186

10. Du K, Yang B, Wang S et al (2022) Relation extraction for manufacturing knowledge graphs based on feature fusion of attention mechanism and graph convolution network. Knowl Based Syst 255(109):703

11. Gao S, Kang G, Yu L et al (2022) Adaptive deep learning for high-speed railway catenary swivel clevis defects detection. IEEE Trans Intell Transp Syst 23(2):1299–1310

12. Geng Z, Li J, Han Y et al (2022) Novel target attention convolutional neural network for relation classification. Inform Sci 597:24–37

13. Goerlandt F, Li J, Reniers G (2021) Virtual special issue: mapping safety science—reviewing safety research. Saf Sci 140(105):278

14. Guo Z, Sharma PK, Martinez A, et al (2022) Multilingual molecular representation learning via contrastive pre-training. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, pp 3441–3453

15. Gururangan S, Marasovic A, Swayamdipta S, et al (2020) Don't stop pretraining: Adapt language models to domains and tasks. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, pp 8342–8360

16. Jie Z, Li J, Lu W (2022) Learning to reason deductively: math word problem solving as complex relation extraction. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. ACL, pp 5944–5955

17. Krummenacher G, Ong CS, Koller S et al (2017) Wheel defect detection with machine learning. IEEE Trans Intell Transp Syst 19(4):1176–1187

18. Lee J, Yoon W, Kim S et al (2020) Biobert: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 36(4):1234–1240

19. Li J, Goerlandt F, Reniers G (2021) An overview of scientometric mapping for the safety science community: methods, tools, and framework. Saf Sci 134(105):093

20. Li G, Chen X, Wang P, et al (2022a) Fastre: Towards fast relation extraction with convolutional encoder and improved cascade binary tagging framework. In: Proceedings of the Thirty-First Inter-

national Joint Conference on Artificial Intelligence. ijcai.org, pp 4201–4208

21. Li L, Lian R, Lu H, et al (2022b) Document-level biomedical relation extraction based on multi-dimensional fusion information and multi-granularity logical reasoning. In: Proceedings of the 29th International Conference on Computational Linguistics. International Committee on Computational Linguistics, pp 2098–2107

22. Li Y, Long G, Shen T, et al (2020) Self-attention enhanced selective gate with entity-aware embedding for distantly supervised relation extraction. In: Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence. AAAI Press, pp 8269–8276

23. Lin Y, Shen S, Liu Z, et al (2016) Neural relation extraction with selective attention over instances. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. ACL, pp 1–10

24. Liu Y, Zhang Q, Lv Z (2022) Real-time intelligent automatic transportation safety based on big data management. IEEE Trans Intell Transp Syst 23(7):9702–9711

25. Lockard C, Shiralkar P, Dong XL, et al (2020) Zeroshotceres: zero-shot relation extraction from semi-structured webpages. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, pp 8105–8117

26. Martínez P, LMartínez J, Segura-Bedmar I, et al (2016) Turning user generated health-related content into actionable knowledge through text analytics services. Comput Ind 78:43–56

27. Mikolov T, Sutskever I, Chen K, et al (2013) Distributed representationss of words and phrases and their compositionality. In: Proceedings of 27th Annual Conference on Neural Information Processing Systems, pp 3111–3119

28. M H, Kenk MA, Khan M, et al (2021) Vehicle detection and tracking in adverse weather using a deep learning framework. IEEE Trans Intell Transp Syst 22(7):4230–4242

29. Ni X, Liu H, Ma Z, et al (2021) Detection for rail surface defects via partitioned edge feature. IEEE Trans Intell Transp Syst

30. Peters ME, Neumann M, Iyyer M, et al (2018) Deep contextualized word representationss. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, pp 2227–2237

31. Qin H, Tian Y, Song Y (2021) Relation extraction with word graphs from n-grams. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, pp 2860–2868

32. Radford A, Narasimhan K, Salimans T, et al (2018) Improving language understanding by generative pre-training. OpenAI

33. Sahu SK, Anand A, Oruganty K, et al (2016) Relation extraction from clinical texts using domain invariant convolutional neural network. In: Proceedings of the 15th Workshop on Biomedical Natural Language Processing. ACL, pp 206–215

34. Silka J, Wieczorek M, Wozniak M (2022) Recurrent neural network model for high-speed train vibration prediction from time series. Neural Comput Appl 34(16):13305–13318

35. Sun K, Li Z, Zhao H (2021a) Multilingual pre-training with universal dependency learning. In: AProceedings of the 34th Annual Conference on Neural Information Processing Systems, pp 8444–8456

36. Sun K, Zhang R, Mensah S, et al (2021b) Progressive multi-task learning with controlled information flow for joint entity and relation extraction. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol 35. AAAI, pp 13851–13859

37. Tian Y, Chen G, Song Y, et al (2021) Dependency-driven relation extraction with attentive graph convolutional networks. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, pp 4458–4471

38. Vaswani A, Shazeer N, Parmar N, et al (2017) Attention is all you need. In: Proceedings of the 30th Annual Conference on Neural Information Processing Systems, pp 5998–6008

39. Veres M, Moussa M (2020) Deep learning for intelligent transportation systems: a survey of emerging trends. IEEE Trans Intell Transp Syst 21(8):3152–3168

40. Verma S, Bhatia R, Harit S, et al (2022) Scholarly knowledge graphs through structuring scholarly communication: a review. Complex Intelligent Systems pp 1–37

41. Veysseh APB, Dernoncourt F, Thai MT, et al (2020) Multi-view consistency for relation extraction via mutual information and structure prediction. In: Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence. AAAI Press, pp 9106–9113

42. Wang C, He X, Zhou A (2021) Open relation extraction for chinese noun phrases. IEEE Trans Knowl Data Eng 33(6):2693–2708

43. Wang Y, Sun C, Wu Y, et al (2021b) Unire: a unified label space for entity relation extraction. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. ACL, pp 220–231

44. Wozniak M, Zielonka A, Sikora A (2022) Driving support by type-2 fuzzy logic control model. Expert Syst Appl 207(117):798

45. Wozniak M, Wieczorek M, Silka J (2023) Bilstm deep neural network model for imbalanced medical data of iot systems. Fut Gen Comput Syst 141:489–499

46. Xiao M, Liu C (2016) Semantic relation classification via hierarchical recurrent neural network with attention. In: Proceedings of the 26th International Conference on Computational Linguistics. ACL, pp 1254–1263

47. Yang Z, Dai Z, Yang Y, et al (2019) Xlnet: generalized autoregressive pretraining for language understanding. In: Proceedings of 32th Advances in Neural Information Processing Systems, pp 5754–5764

48. Yang F, Gu S (2021) Industry 4.0, a revolution that requires technology and national strategies. Complex Intell Syst 7(3):1311–1325

49. Ye T, Zhang X, Zhang Y et al (2021) Railway traffic object detection using differential feature fusion convolution neural network. IEEE Trans Intell Transp Syst 22(3):1375–1387

50. Yu D, Zhu C, Yang Y, et al (2022) JAKET: joint pre-training of knowledge graph and language understanding. In: Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence. AAAI Press, pp 11630–11638

51. Zeng D, Liu K, Chen Y, et al (2015) Distant supervision for relation extraction via piecewise convolutional neural networks. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. ACL, pp 1753–1762

52. Zeng D, Liu K, Lai S, et al (2014) Relation classification via convolutional deep neural network. In: Proceedings of the 25th International Conference on Computational Linguistics. ACL, pp 2335–2344

53. Zhang H, Song Y, Chen Y et al (2022) MRSDI-CNN: multi-model rail surface defect inspection system based on convolutional neural networks. IEEE Trans Intell Transp Syst 23(8):11162–11177

54. Zhang S, Ren W, Tan X, et al (2021) Semantic-aware dehazing network with adaptive feature fusion. IEEE Trans Cybern:1–14

55. Zhao K, Xu H, Cheng Y et al (2021) Representation iterative fusion based on heterogeneous graph neural network for joint entity and relation extraction. Knowl Based Syst 219(106):888