Reinforcement Learning-based Distant Supervision Relation Extraction for Fault Diagnosis Knowledge Graph Construction Under Industry 4.0

Chong Chen[a], Tao Wang[a, b*], Yu Zheng[c], Ying Liu[d], Haojia Xie[a], Jianfeng Deng[a, b] and Lianglun Cheng[a]

[a] *Guangdong Provincial Key Laboratory of Cyber-Physical System, Guangdong University of Technology, Guangzhou 510006, China*

[b] *School of Automation, Guangdong University of Technology, Guangzhou 510006, China.*

[c] *School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*

[d] *Department of Mechanical Engineering, School of Engineering, Cardiff University, Cardiff CF24 3AA, UK*

Abstract

Fault diagnosis is the key concern in the operation and maintenance of industrial assets. A fault diagnosis knowledge graph (KG) can provide decision support to the engineers to efficiently conduct maintenance tasks. However, as a type of domain KG, it would be time-consuming to manually label the corpus collected from the multi-source including the maintenance log, handbook and article. Meanwhile, the existence of the noisy sentence in the multi-source corpus jeopardises the performance of relation extraction modelling. In order to address this issue, this paper proposes a distant supervision relation extraction (DSRE)-based approach to construct a fault diagnosis KG. In this approach, the ontology of the fault diagnosis KG is firstly designed. Subsequently, a DSRE algorithm named relation-aware-based sentence-level attention enhanced piecewise convolutional neural network with reinforcement learning strategy (PCNN-ATTRA-RL) is proposed. The algorithm can effectively lower the impact of noisy sentences and accurately label the relation of different entities when the labelled data is insufficient. In this algorithm, PCNN-ATTRA is designed as the DSRE classifier to effectively extract the relation between entity pairs. RL is conducted to remove the noisy sentence so as to further improve the performance. An experimental study based on the multi-source corpus collected from the real world reveals that the proposed approach shows merits in comparison with the state-of-the-art algorithms. Meanwhile, a fault diagnosis KG, which can greatly support the decision-making of the engineers in the fault diagnosis, is established via the proposed approach.

Keywords: Distant supervision relation extraction; Knowledge graph; Fault diagnosis; Reinforcement Learning.

1. Introduction

With the arrival of Industry 4.0, advanced technologies such as the industrial Internet of things (IIoT), cyber-physical systems (CPS) and artificial intelligence (AI) have provided opportunities for the smart maintenance of industrial assets [1, 2]. The industrial assets in the Industry 4.0 tend to be automated, intelligent, and complex, which poses a challenge in fault diagnosis [3, 4]. The fault diagnosis is a key issue in the operation management of industrial assets. The unplanned stoppage of the assets may lead to the interruption of the production line for a variety of reasons, which would result in economic loss and casualties [5]. The troubleshooting and maintenance can be quickly conducted if the location and the health status of the failure parts can be diagnosed accurately [6, 7]. Therefore, to ensure the safe and stable operation of production, it is necessary to study the intelligent fault diagnosis for the industrial asset.

The existing research on fault diagnosis heavily relies on data analytics based on the condition monitoring data [4, 8-10]. Taking industrial robots as an example, researchers have investigated fault diagnosis modelling based on altitude data [11] and feedback current data [12, 13]. However, in the actual industrial environment, fault diagnosis modelling is hard to conduct since the run-to-failure data is hard to collect and the failure and maintenance records are not always reliable. Even though there are various studies on fault diagnosis modelling, most of them focus on the data from the simulation or testing environments, which applicability in the actual industry still needs to be further investigated. Meanwhile, some of the results of the data-driven fault diagnosis model are short of explainability, which impedes its deployment in the actual scenario. Besides the fault diagnosis based on condition monitoring data, another path toward fault diagnosis is the knowledge-based approach [14]. The knowledge-based fault diagnosis approach aims to locate the faulty part and explore the failure's root cause via the pattern matching to an established knowledge base. The fault-related knowledge is stored in various forms including operational handbooks, maintenance logs and expert systems. With the recent advance in knowledge graph technology, the knowledge from different sources can be integrated to establish a knowledge base so as to provide valuable insights into the fault diagnosis of the industrial asset. In comparison with the data-driven approach, the knowledge-based approach can make full use of the existing knowledge and provide insights to the engineers with explainability. However, the knowledge-driven fault diagnosis has not gained sufficient attention. Hence, it is worthwhile to investigate the construction of KG for the fault diagnosis.

KG can be classified into two types which are general KG and domain KG. The general KG such as Freebase [15] and DBpedia [16] can be constructed based on a large sum of the comprehensive and reasonable corpus. Because of the large available dataset, intelligent entity recognition and relation extraction approaches based on deep learning can greatly boost the development of general KG. However, the general KG tends to obtain well-known knowledge in different fields. However, industry knowledge has a complex hierarchy and relationships, which cannot be fully covered by general KG [17, 18]. Different from the general KG, the domain KG only focuses on a specific domain and its ontology is different from the general KG [19-21]. The entity and relations of the domain KG can be greatly different from that of the general KG [17]. As a type of domain KG, the KG for fault diagnosis has not gained sufficient attention. In the construction of KG for fault diagnosis, the main challenge is the relation extraction of the entities. Relation Extraction is the core task in KG construction. If the relation of entities cannot be correctly extracted, the quality of the knowledge graph will be damaged. In relational extraction, supervised learning-based approaches have shown their effectiveness. Nevertheless, in domains with large amounts of data, their performance is often hindered by the lack of labelled data.

Recently, researchers have focused on DSRE to ease the pressure and cost of manually annotating corpora by the means of automatic construction of the labelled dataset for relation extraction tasks. With the rapid development of deep learning, deep learning has been widely studied in relation extraction modelling [22]. However, the modelling performance can be jeopardised when the labelled data is insufficient. Hence, DSRE which can handle both labelled and unlabelled data has recently gained increasing attention for their ability to enhance modelling performance [23]. Deep learning algorithms such as Transformers [24], generative adversarial network (GAN) [25] and piecewise convolutional neural network (PCNN) [22, 26] have been investigated in DSRE. Although deep learning algorithms offer effective solutions to automatic feature extraction, they suffer from wrong labelling issues in DSRE when applied to real-world datasets [27]. In order to reduce the impact of wrong labelling from the dataset, researchers have studied reinforcement learning (RL) [28, 29] and attention mechanisms [30, 31] to enhance the robustness of the algorithms. RL can leverage the algorithm performance by continuously training the metrics on the dataset to reallocate the samples, while the attention mechanism can locate the key features relevant to the relation. In DSRE, two widely used attention mechanisms are sentence-level attention and bag-level attention. The sentence-level attention approach can maximise the use of information from other examples in the bag to enhance the bag representation [32], while it is not specialised in noise removal. In contrast, the bag-level attention-based approach is effective in addressing the wrong labelling issue [33], while its performance can be significantly jeopardised when the knowledge base cannot be perfectly aligned with the unannotated corpus. For the construction of an industrial domain KG, the corpus is hard to meet the alignment criterion of bag attention. In order to address this challenge, this study proposes a DSRE approach that takes advantage of sentence-level attention and RL to achieve better alignment via the effectively removal of noisy sentences.

In this study, a DSRE-based approach to construct a fault diagnosis KG is proposed. The main contributions of this study are three-fold: (1) a DSRE-enabled KG construction approach for fault diagnosis without a large number of annotated data is proposed; (2) A DSRE algorithm called PCNN-ATTRA-RL is proposed to effectively restrain the impact of the noisy sentences and achieve satisfactory performance; (3) An experimental study was implemented based on the corpus collected from the real-world industry, which reveals that the proposed approach shows merits in comparison with the state-of-the-art methods. The rest of the paper is organised as follows: Section 2 reviews the related works on recent advances in the construction of KG and the latest research on distant supervision relation extraction. Section 3 presents the methodology of this paper. Section 4 introduces the experimental setup and the experimental results are demonstrated and discussed in Section 5. Finally, Section 6 concludes.

2. Literature review

2.1. Recent advances in domain knowledge graph construction

The domain KG plays a core role to generate context-awareness and proactive services for the optimization of serviceability and productivity of the industry. Recently, researchers have studied the construction of KG for different industrial scenarios. The introduction of the smart product-service system can boost the competitiveness of industrial companies. Wang et al. [34] propose a novel data-driven graph-based requirement elicitation framework in the smart product-service system, aimed at assisting engineering designers to improve the design concept generation process in a closed-loop manner. Lyu et al. [20] proposed a generic crowdsourcing approach for continuously evolving the industrial KG (IKG) to achieve knowledge as a service in IIoT-driven smart manufacturing environment. In this study, IKG-enabled systems demonstrate the ability to utilize knowledge as a kind of service rather than a kind of resource through IKG continuous enrichment. Zheng et al. [21] proposed an IKG-based multi-agent reinforcement learning approach for the Self-X cognitive manufacturing network. In this approach, a graph neural network-based embedding algorithm is proposed to establish a KG based on the massive multimodal data. Subsequently, a multi-agent reinforcement learning-based decentralized system is introduced to achieve the optimisation of the manufacturing process. Wan et al. [35] proposed a KG based approach to achieve multi-faced modelling for cold rolling process. In this study, the ontology was firstly built according to the domain knowledge. Then, a KG was generated based on the multi-faceted data before it is further fed into a graph convolutional network to further learn the hidden patterns of KG. Lin et al. [36] proposed an intelligent development environment and software knowledge graph to realise data aggregation, analysis and intelligent assistance for software development. In order to address the issue that geographic information is not explicit and structured, Nizzoli et al. [37] first deployed Geo-Semantic-Parsing (GSP) for enriching text documents with structured geographic information. Firstly, the geographic location and coordinates are extracted from the general knowledge graphs. In order to efficiently traverse the knowledge graph, an expansion step is then conducted. Finally, GSP solves a regression task to select the best entity to geotag the input document with. Liu et al. [38] studied a KG-based data representation approach for

IIoT-enabled cognitive manufacturing. A multi-layer manufacturing KG was established based on the data collected from manufacturing process. This paper also developed a cognition-driven approach based on a dual system of perception and cognition that achieves perception analysis and cognition decision-making in the allocation of manufacturing resources.

Zhou et al. [39] investigated an event link reasoning over an IKG embedding time series data. The ontology was firstly constructed through a specific manufacturing process. The semantic information related to product quality prediction was derived using deep learning algorithms based on time series data collected from the devices. Subsequently, the prediction result was added as a specified node in the KG, which enables the KG to achieve a dynamic update. Since there are various rules and constraints in the data collected from discrete manufacturing, Zhou et al. [40] proposed a KG-driven approach to optimise resource allocation. The workshop resource KG was first constructed based on the multi-source data in the workshop. Then, an algorithm to mine implicit resource information for real-time updating of the workshop resource KG using a distributed knowledge representation was developed. To construct text-based KGs and tabular information-based causality event evolutionary graphs, Zhou et al. [19] proposed an end-to-end methodology for extracting joint entity relationships from engineering documents. An entity alignment approach based on neighbourhoods sample graphs is proposed to integrate the knowledge graphs into a unified knowledge base. Moreover, a graph-driven Q&A approach with translation is designed to assist in problem tracing and cause analysis.

For the constriction of fault diagnosis, Shi et al. [41] studied an information integration approach to spacecraft fault diagnosis. A knowledge graph is established to integrate the information about the entire lifecycle of a spacecraft, which is then applied to develop a question-answering system for fault diagnosis. Deng et al. [14] proposed an event logic KG construction approach for robot transmission system fault diagnosis. Firstly, the event arguments and the ontology of the fault diagnosis were defined. Then, the bidirectional long short-term memory network with conditional random field (Bi-LSTM-CRF) was adopted to realize entity recognition. Thirdly, a Bidirectional graph convolutional network (Bi-GCN) was used to achieve relation extraction. In order to reduce the uncertainties and local–global fault interpretation perception for practical physical meaning, Ren et al. [42] proposed a multilevel KG approach for fault detection. In this study, a multilevel KG was constructed by combining domain knowledge with monitoring data. The variables of the data were adopted as the nodes in the multilevel KG. When the node status exceeds the pre-set thresholds, the corresponding fault can be detected.

2.2. The research on distant supervision relation extraction

As one of the most basic tasks in natural language processing, relation extraction aims to extract the relationships between entities. In comparison with the existing relation extraction methods, supervised learning-based methods are more reliable at capturing relationships between two entities, but they are limited by a lack of large-scale manually labelled data. To address this issue, a method known as distant supervision has been presented by Mintz et al. [43], which uses an unstructured corpus of training data to generate training data automatically by aligning it with existing knowledge. In spite of the fact that this approach can ease the challenge of limited labelled data, it is greatly limited by the issue of wrong labelling, since the assumption is difficult to satisfy. To alleviate the impact caused by wrong labelling, Riedel et al. [44] reported an approach that treats relation classification as a multi-instance learning problem, which extracts relations from bags of sentences rather than a single sentence.

In recent years, deep learning techniques have been prevailing in the extraction of distant supervision relationships by researchers. Zeng et al. [26] combined PCNN and multi-instance learning to automatically extract sentence feature information and selected the most reliable sentence as the representation for the bag. This approach has been extensively studied in the DSRE in recent years. The attention mechanism is an effective tool to boost the performance of deep learning. For example, a sentence-level attention mechanism was proposed by Lin et al. [32], which weights the sum of all sentence representations inside the bag to calculate the bag representation. With the consideration of the importance of keywords at the word level, researchers have studied word-level attention to improve the modelling performance [45, 46]. Both studies deployed word-level attention to effectively identify the keywords and the noise. Inspired by word-level attention, a cross-bag attention mechanism was presented by Yuan et al.

[47], in which the weighted sum of all bag representations in the group was adopted as the group representation. In order to minimize the impact of noise data, Ye et al. [48] combined intra-bag and inter-bag attentions. Zhou et al. [31] adopted convolutional layers and a self-attention mechanism to encode instances which can get better semantic vector representation. Subsequently, the correctly labelled instances are given a higher weight based on the correlation between them, and the weighted sum of the instances in the bag was adopted to get a bag vector representation.

In order to address the negative effect of the noisy labelling problem, Chen et al. [49] proposed a method called Neural Instance Selector (NIS) to utilize rich supervision information to filter the less informative sentences. Since the standard Transformer network lacks the capability of capturing global patterns. Shang et al. [50] developed a specific-designed pattern-aware self-attention network to capture both local and global patterns. The method assumed that an adjacent token's correlation indicates a shared pattern. Then, a novel self-attention network is designed to generate a probability distribution for all patterns in a sentence on the basis of this assumption. In the first Transformer layer, the probability distribution serves as a constraint to guide attention heads towards relational patterns. Finally, by enhancing fine-grained pattern information, global dependencies are captured in the pre-trained Transformer.

RL has been studied in noise data removal. Qin et al. [29] proposed an RL strategy to identify the false positive samples and reallocate them to the negative sample set, which is helpful to improve the performance of DSRE. The performance can be evaluated via a relation extraction classifier. To lower the impact of wrong labelling, Chen et al. [28] proposed a deep-Q-Network based label-denoiser that can select the most reliable labels for the unlabeled data. Subsequently, a relation extraction classifier was adopted to get the reward metrics.

2.3. A brief review

Constructing a fault diagnosis KG is mainly based on the corpus collected from various sources including the operational handbooks, maintenance logs and articles. The existing studies of domain KG construction reveal that it is challenging to build a domain KG for specific industrial cases due to the limited scale and the lack of continuous enrichment. Meanwhile, the corpus collected can be noisy and most of them are unlabelled. DSRE can make full use of the labelled and unlabelled data for relation extraction modelling. In order to establish fault diagnosis KG based on the unsatisfactory corpus, DSRE is a potential solution. The existing DSRE approaches based on attention mechanisms are capable to lower the impact of noise. When the number of noisy data far outweighs the positive instances, the modelling performance is hard to leverage using the methods reported in the state-of-the-art. RL and sentence-level attention mechanisms are two mainstream methods of addressing the noisy issue of DSRE. The noise level in the multi-source corpus collected from the industry can be higher than that of the corpus for general KG. Hence, the challenges for constructing a fault diagnosis KG are: (1) When performing the relation extraction task, the labelled corpus data is insufficient; (2) The existing DSRE approach may not able to address the corpus with heavy noise. Therefore, it is worthwhile to investigate a DSRE approach that takes advantage of both RL and sentence-level attention to achieve effective noise removal.

3. Methodology

In this section, the overall flow of constructing a fault diagnosis KG and the proposed PCNN-ARRTRA-RL algorithm are elaborated. The overall flowchart of the construction of KG for fault diagnosis is shown in Figure 1. Firstly, the ontology of the KG is designed via seven steps approach. Then the name entity recognition task is performed via the entity recognition tool. In the next stage, the relation of the entity is labelled via the proposed PCNN-ATTRA-RL algorithm. Finally, the KG is visualised and used for fault diagnosis.

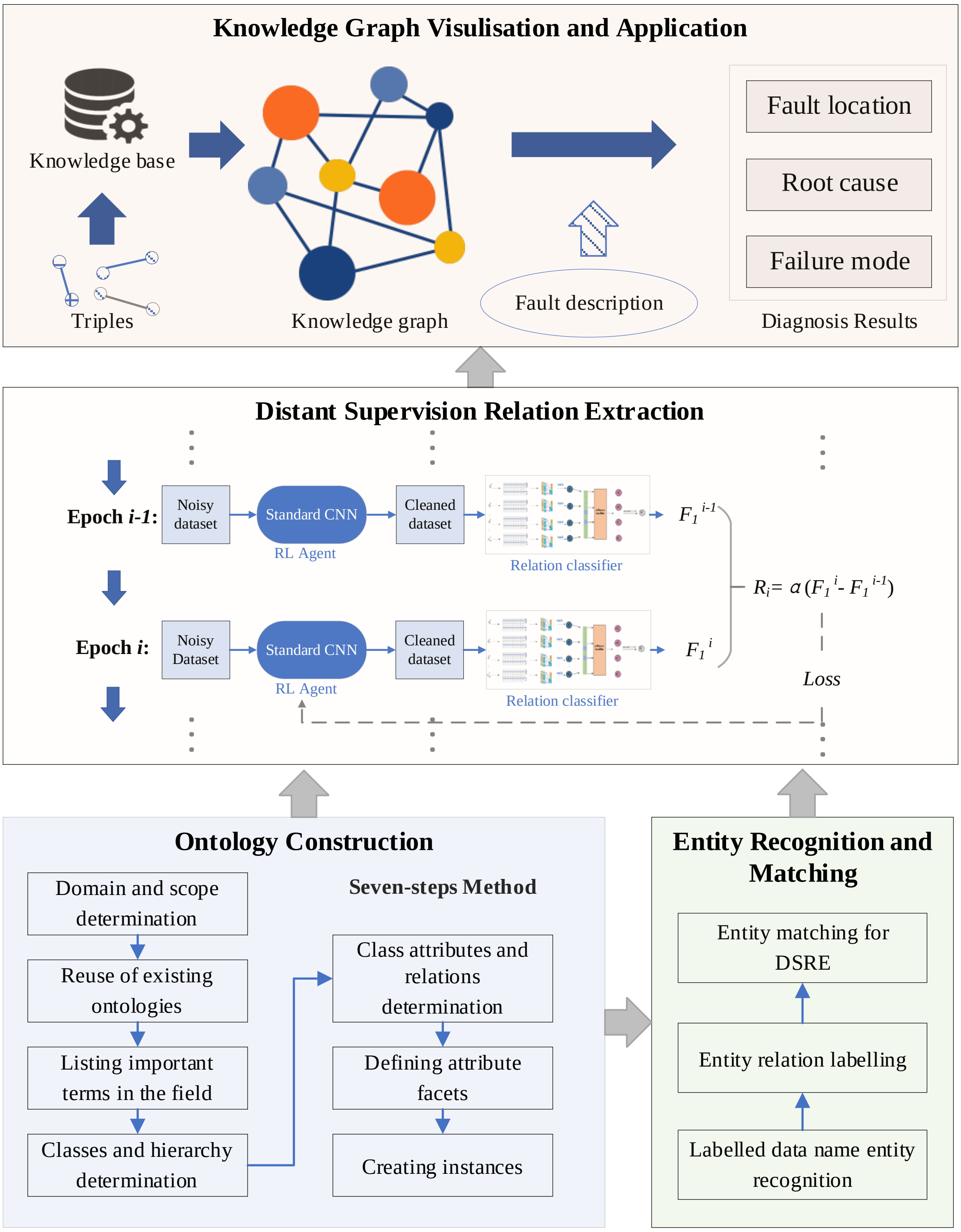

Figure 1. The flowchart of fault diagnosis KG

3.1. The ontology design and name entity processing

The ontology is a key part of KG, which defines knowledge concepts and their relationships. In this study, the fault diagnosis ontology is firstly determined via the seven-step method. Specifically, the domain knowledge of fault diagnosis is represented by event logic. The seven-step method begins with defining the domain and scope, reusing existing ontologies, defining classes and their hierarchy, defining class attributes and relations, and defining attribute facets. Firstly, the domain and scope of the ontology is fault diagnosis. Secondly, it defines the object (O), the trigger word (T), and the status (S) of the fault diagnosis event argument, before defining the class concept and class relation of event arguments. The class concept describes the definition of argument class and ontology class concept construction of fault diagnosis events, while class relation describes the definition of event trigger word relations and logical relations between events, construction of ontology relation concepts. For the fault diagnosis event argument definition, it can be expressed as follows:

$$e =< O,T ,S > \tag{1}$$

where O is the faulty part, corresponding to equipment words; T is the event trigger element, which corresponds to the event trigger word; S is the state expression of equipment, which is the fault state word.

Based on the defined event and event argument, the conceptual knowledge model of fault diagnosis events and event argument is shown in Figure 2. There are two types of fault diagnosis events: fault phenomenon events and fault cause events. The relationship between the fault case and the fault phenomenon is defined as ‘lead_to’. Event 1 and event 2 are two event instances, which contain the corresponding fault equipment objects and equipment fault status.

After the events and event arguments are modelled, the fault diagnosis event argument class and relation need to be defined. There are three classes defined in the fault diagnosis ontology model, which are fault equipment structure, equipment attributes and equipment state values. Furthermore, equipment fault attributes contain three types of subclasses, which are fault attribute value, fault equipment structures including sub-equipment, component subclasses, and part. In the next stage, the corresponding class relation between each conceptual class needs to be defined. There are four types of event argument class relations defined in the fault diagnosis ontology, which are *consist_of*, *lead_to*, *has_attribute*, and *appear*. After the classes and class relations are defined, the fault diagnosis ontology model is constructed via the Protégé tools, which are shown in Figure 3. After the ontology and the labelled strategy are determined, the entity and the relation of the corpus can be labelled accordingly.

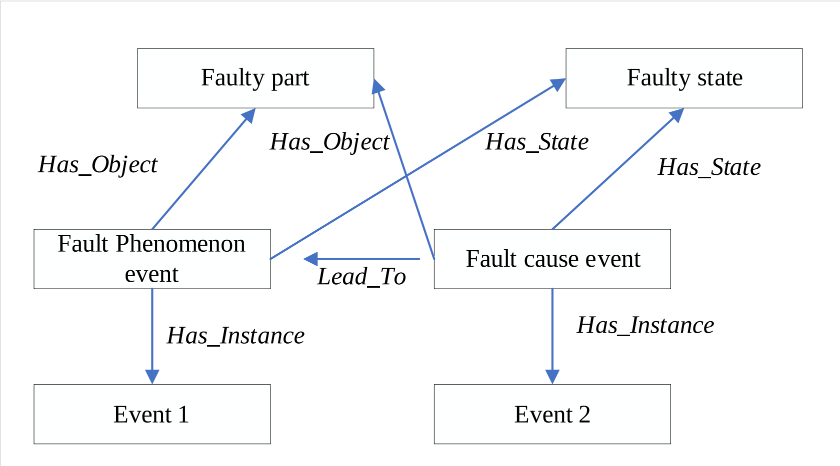

Figure 2. A conceptual knowledge model of fault diagnosis events and event arguments

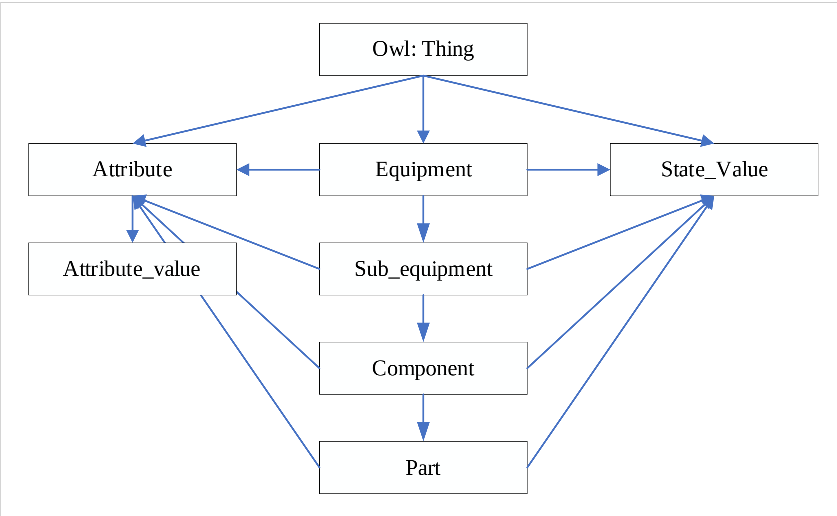

Figure 3. The ontology structure of fault diagnosis

3.2. Distant Supervision Relation Extraction

In order to address the challenge that the labelling process of the multi-source corpus is monotonous and time-consuming, we propose PCNN-ATTRA-RL algorithm that can effectively remove the noisy sentences in the DSRE modelling and uncover more hidden relationships between the entities. In this approach, RL is conducted to remove the noisy sentence so to improve performance. The relation-aware-based sentence-level attention enhanced PCNN is designed as the relation classifier to accurately extract the relation from the corpus. A relation classifier is used to identify relationships between entities in a KG. It can be used to predict the probability of a relationship between two entities based on the label data available in the knowledge graph. When the labelled data is insufficient, the performance of distant supervision relation extraction tends to yield better performance than the supervised relation extraction. Relation classifier in distant supervision relation extraction is used to extract structured information from unstructured text and classify those relationships into predefined categories. The flow chart of the PCNN-ATTRA-RL for DSRE is presented in Figure 4.

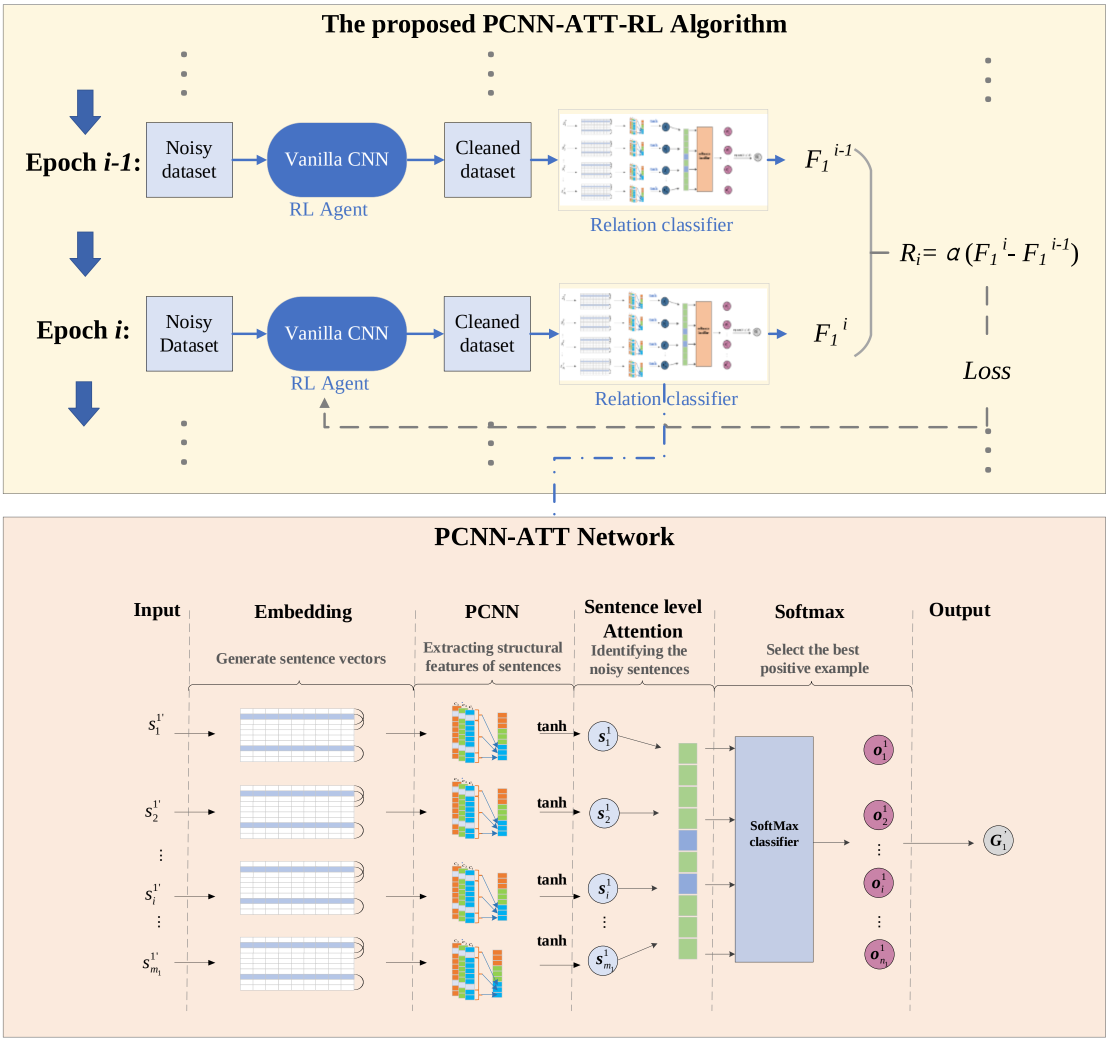

Figure 4. The demonstration of the PCNN-ATTRA-RL algorithm

3.2.1. *PCNN-ATTRA*

In the relation-aware-based sentence-level attention enhanced PCNN network for DSRE, we denote $g_j = \{b_1^j, b_2^j, \dots, b_n^j\}$ as a set of bags which have the same relation label n, and n_j is the number of bags within this group. All sentences in the bag b_i is denoted as: $b_i = \{s_1^{i'}, s_2^{i'}, \ldots, s_m^{i'}\}$, where m is the number of sentences in this bag. PCNN is adopted as the sentence encoder to extract information about the feature structure of the sentence. After using the sentence encoder to obtain all sentence feature vector representations, ATTRA is introduced to calculate the confidence of different sentences to the corresponding relationship.

To better extract the semantic and structural information of sentences, the word2vector model is used to convert words into word vectors with rich semantic information. We first convert each word into a d_w-dimensional word embedding. Word representations are encoded by column vectors in an embedding matrix $V \in \mathbb{R}^{d_w \times |V|}$ where V is a fixed-sized vocabulary. In addition, to indicate the position of the entity pair in the sentence, we add position feature information for each word. For each word w_i, its position features are mapped to two feature vectors pl_i and pr_i of d_p dimensions. Then, the word vector w_i, is concatenated with its corresponding two position feature vectors $pl_i$and pr_i to obtain the final word representation which can be expressed as;

$$w_i' \in \mathbb{R}^d (d = d_w + d_p * 2) \tag{1}$$

PCNN [26] is an effective tool in relation extraction. It is introduced in this study to capture the local features of each sentence after a representation of the sentences is obtained. Taking sentence $s_j^{i'}$ as an example. Firstly, the convolution operation is deployed to the matrix of its word vector representation $x \in \mathbb{R}^{l_{ij} \times d}$. l_{ij} is the number of words of the sentence $s_j^{i'}$. Subsequently, Feature maps generated by convolutional layers are further extracted using piecewise max-pooling to get salient feature representations of sentences $\{s_1^i, s_2^i, \ldots, s_m^i\}$ ($s_j^i \in \mathbb{R}^{3n}$).

The relation-aware-based sentence-level attention proposed in [24] has shown its merits in the bag representations, while its performance can be damaged when the noise level is high. Since each sentence in the bag b_i expresses the relation k to a different degree, a weight vector β is allocated to each sentence to reveal the sentence's impact on various relation types. The weight vector $\{\beta_1^i, \beta_2^i, \ldots, \beta_m^i\}$ of each sentence for a certain relation type *k* is defined as follows:

$$\beta_{kj}^i = \frac{\exp\left(e_{kj}^i\right)}{\sum_{j'=1}^{m_i} \exp\left(e_{kj'}^i\right)}, 1 \le k \le h \tag{2}$$

where h is the number of relationship types; β_{kj}^i is the weight value obtained after the SoftMax normalization of e_{kj}^i; e_{kj}^i is the confidence between j^{th} sentence in bag b^i and k^{th} relation query, which can be obtained by the following formula:

$$e_{kj}^i = q_k s_j^{iT} \tag{3}$$

where q_k is the k^{th} row of the relation embedding matrix R.

3.2.2. *Reinforcement Learning Strategy*

In order to achieve better performance when the noise level in the corpus is high, the RL strategy is introduced. The false positive samples in the positive set of DSRE are challenging to be removed by the attention mechanism. The identification of the false positive samples can be helpful to leverage the performance of DSRE. RL has shown its merits in identifying and removing false positive samples [29]. The corpus used for the construction of fault diagnosis is collected from various sources, which may contain a large number of noisy samples. Hence, in this study, RL is introduced to further improve the performance of DSRE. In the RL strategy, an agent is adopted for distant supervision relation extraction, which can effectively remove the noisy sentences according to the performance change of the PCNN-ATTRA network.

In order to perform RL for DSRE, the external environment firstly needs to be modelled as a Markov decision process. The state s contains the information from the current sentence and the removed sentences from earlier states. The vector generated by word embedding and position embedding is used to represent the current state as the concatenation of the current sentence vector and the average vector of the sentences that have been removed from the early stages. Furthermore, to magnify the dominant impact of the current sentence information on the decision of action, the relatively large weights to the vector of the current sentence are allocated. For the action of the agent, there are two types of actions which are retaining and removing the noisy sentence. The agent is expected to identify and remove the noisy sentences via the policy network. Subsequently, the performance change in each step is used to determine the award of the agent. The F1 score is adopted as the evaluation metric, which can be expressed as follows:

$$R_i = \alpha(F_i - F_{i-1}) \tag{4}$$

where F_i is the F1 score in the i^{th} step. α is used to convert the F1 scores change into a range from -1 to 1.

After the action and reward are determined, the policy network needs to be designed. The policy network is used to determine whether a sentence should be removed. Hence, a vanilla CNN is used to achieve this target. The vanilla CNN is the policy network in reinforcement learning. The policy network evaluates each input sentence to determine if it expresses the target relation type, and remove any that are not relevant. Thus, the policy network is analogous to a binary relation classifier. CNN is widely used in relation classification since it can effectively learn the useful patterns in each sentence [29]. Since the task of policy network is just binary classification, the vanilla CNN is therefore adopted. A CNN whose window size is c_w and the kernel size is c_k is adopted to model the policy network $\pi(s;\theta)$.

In the training stage of the policy network, the reward is calculated after the entire positive set of this relation type is processed. Meanwhile, the pre-train strategy is adopted to improve the robustness of the modelling. After the pre-training, the model can identify those obvious noisy sentences, while it is still challenging to make the decision on the ambiguous sentences. Hence, the agent needs to be retrained with rewards. In the retraining stage, the distantly-supervised positive dataset is firstly split as the training positive set and the validation positive set. Similarly, the training negative set and the validation negative set are constructed based on the DS negative dataset. In every epoch in the retraining stage, the agent removes a noisy sample from the training positive set according to the stochastic policy $\pi(a|s)$. After the loss of the policy network converged, a clean dataset for DSRE can be obtained. The overall loss function can be expressed as follows:

$$Loss(\theta) = \sum^{\Omega_i} log\pi(a|s;\theta)\, R + \sum^{\Omega_{i-1}} log\pi(a|s;\theta)\,(-R) \tag{5}$$

where Ω_i is the removal instances in the i^{th} epoch; R is the reward of the policy network.

3.3.Knowledge Graph Visualisation and Application

After the relation of the unlabelled entity pairs is obtained via the proposed PCNN-ATTRA-RL algorithm, the triples can be obtained. The obtained triples are then stored in a database. The most widely used graph database is Neo4j [51], which is mainly stored in the form of network structure with nodes and relations as objects. In this study, Neo4j is used to store the relevant entities and relations of the fault diagnosis knowledge. After that, the association between nodes in the knowledge graph can be visualised.

The established KG can then be used for fault diagnosis. When a fault happens, engineers can generate a fault description, which is then sent to the KG for knowledge matching. Then the KG can output the diagnosis results such as the fault location, fault root cause and the faulty mode, which can be beneficial for troubleshooting and maintenance.

4. Experimental Setup

4.1. Data collection

In this study, in order to establish a fault diagnosis KG, a dataset about robotic fault diagnosis event description cases as the corpus comes from various sources including operation and maintenance logs, books and articles. From the corpus, over 900 sentences are manually labelled according to the labelling strategy detailed in Section 3.1. Specifically, the labelled sentence records the historical event of robot transmission faults, the causes of the faults and the associated fault phenomenon. Besides the labelled data, there are over 17,000 unlabelled data in the corpus. The category of the corpus is shown in Figure 5. There are 11 categories of entity relation in the corpus. In the labelled data, the event argument entities and the entity relations were labelled according to the labelling strategy mentioned in Section 3.1. Finally, there are 5860 event argument entities and 5205 semantic relationships in the multi-source corpus database. Meanwhile, there are over 17000 sentences in the dataset. The entities in the unlabelled sentences are then recognised by matching the entities in the labelled dataset via the name entity recognition tool. Finally, the processed data was used for DSRE modelling.

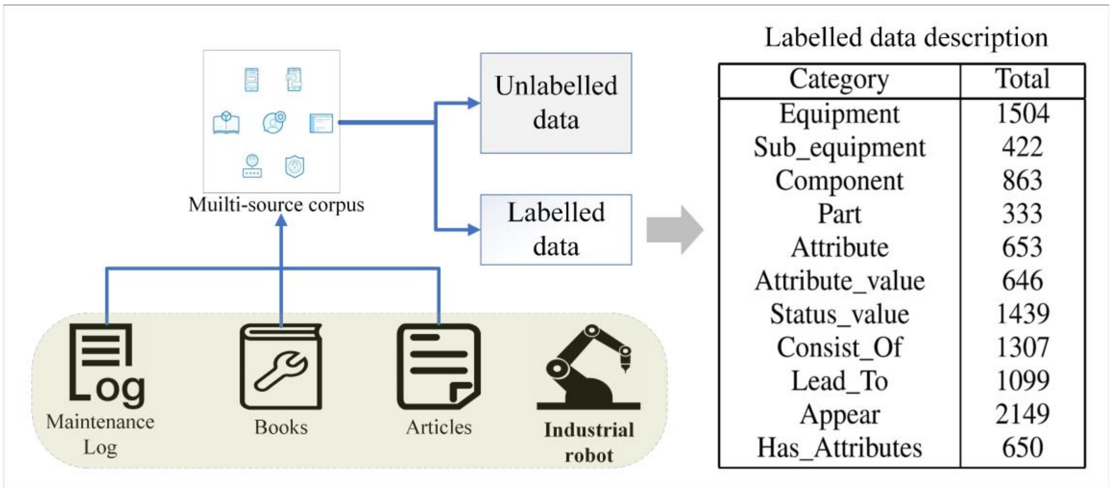

Figure 5. The collected corpus of fault diagnosis

4.2. Benchmarking experiments

Triple validation was used to select all hyperparameters on the training set. All hyperparameters adopted in our experiments are listed in Table 1. Since the parameters of the model affect the algorithm performance of the PCNN-ATTRA, a grid search was employed to determine the optimal model parameters. Table 1 shows the optimal parameters combination of the PCNN-ATTRA.

Table 1. The optimal parameters combination of the PCNN-ATTRA

Component	Parameter	Value
Word embedding	dimension	50
Position feature	max relative distance dimension	±30
	kernel size	3
Network	strides	1
	filter number	230
	learning rate	0.1
	$batch\ size\ epoch_p$	50
Optimization	optimizer	SGD
	batch size epoch t	10
	group size n	5
Dropout	dropout rate	0.5

In this study, there are three experiments were set up. Firstly, the ablation experiment was conducted. In order to evaluate the effeteness of PCNN-ATTRA-RL, three baseline algorithms which are PCNN, PCNN-ATT and PCNN-RL were adopted for comparison. Secondly, the benchmarking experiment was conducted. Four prevailing DSRE algorithms were adopted. The details of the benchmarking algorithms are shown as follows:

1. APCNN-NIS [49]: This algorithm adopts a neural instance selector to filter the less informative sentences via multilayer perceptron and logistic classification and a selective attention module is adopted to select the important sentences.
2. PCNN-ATT-NIS [49]: Based on APCNN-NIS, sentence-level attention is further introduced to lower the impact of noisy sentences.
3. PCNN-ONE [26]: This algorithm introduced the strategy that electing the best sentence in each bag for training.
4. PCNN-BAGATT [48]: This algorithm adopts both intra-bag and inter-bag attention into modelling. The intra-bag representation is used to get the relationaware bag representations. Then, the representation of a group of bags is obtained by weighting similarity-based inter-bag attention.

For the evaluation, the experiments were conducted in alignment with the existing studies [26, 31, 48, 49]. Three testing scenarios which are denoted as One, Two and All were set up to comprehensively reveal the algorithm performance. The details of the three evaluation scenarios are shown as follows:

1. **One:** Only one sentence is randomly selected in each testing entity pair, and then is used to predict the relation.
2. **Two:** Two sentences are randomly selected in each testing entity pair, and then are used to predict the relation.
3. **All:** All the sentences are randomly selected in each testing entity pair, and then are used to predict the relation.

4.3.Evaluation metrics

Consistent with previous studies [26, 31, 48, 49], we also evaluated our model on the held-out evaluation. Precision recall (PR) curve, the area under curve (AUC) value and the precision

@N (P@N) value which is the N^{th} largest precision value in the PR curve were adopted as evaluation metrics in our experiment. PR curve is a prevailing metric that has been widely used to measure the classifier's performance. P@N is the precision when the classification threshold is set to exactly *N* positive cases, which is a metric to observe how the model performs at different thresholds. In the testing stage, the P@100, P@200, P@300 and the mean of them for each model were adopted for evaluation. All numerical results of the experiments are the average of 10 times modelling.

5. Experimental Results and Discussion

5.1. Ablation Experiment

In the ablation experiment, three benchmarking algorithms which are PCNN, PCNN-RL, and PCNN-ATTRA were introduced to reveal the impact of different components on the algorithm performance of PCNN-ATTRA-RL. Figures 6-8 demonstrated the comparison of the algorithm performance in terms of precision in the evaluation scenarios of 'One', 'Two' and 'All'. In Figures 6-8, the x-axis represents the *N* value in the P@N and the y-axis represents the precision. It can be seen from Figure 6 (a) that the proposed algorithm achieved the highest precision in the P@100 category, which is 0.84. The precisions of PCNN-ATTRA and PCNN-RL are approximately 0.8, while the precision of PCNN is mere 0.64. In the category of p@200 and p@300, the precision of the PCNN-ATTRA-RL is lower than the rest algorithms. In the mean category, the difference between PCNN-ATTRA-RL and PCNN-ATTRA is marginal, which is less than 0.02. However, the performance of PCNN-ATTRA-RL is obviously higher than that of PCNN. Figure 6 (b) illustrates the comparison results of different algorithms in scenario 'Two'. It is evident that PCNN-RL achieves comparative results in comparison to PCNN-ATTRA-RL. In the results of P@100, it can be seen that the precision of the proposed algorithm is obviously higher than the other three algorithms, while PCNN-ATTRA shows worse performance than that of PCNN. The results of P@200 reveal that the difference in the precision of all the algorithms is marginal, and PCNN-ATTRA-RL achieves the highest precision. In the results of P@300, the precision of PCNN-RL surpasses PCNN-ATTRA-RL. In the mean category, the algorithm performance of PCNN-ATTRA-RL is slightly higher than that of

PCNN-RL.

Different from the results in the scenario ‘One’ and ‘Two’, Figure 6 (c) reveals that PCNN-ATTRA-RL achieved the highest precision in all the stages. According to the results of P@100, PCNN-ATTRA-RL and PCNN-ATTRA achieve the same precision, which is 0.95. This result is evidently higher than that of the PCNN and PCNN-RL. In the results of P@200 and P@300, the results are different to that in scenarios ‘One’ and ‘Two’, PCNN-ATTRA-RL shows better precision than other algorithms. In the mean category, PCNN-ATTRA-RL achieves a precision of 0.82, which is 0.04 better than the second-highest result. Compared to scenarios 'One' and 'Two', the overall performance achieved in this scenario is higher.

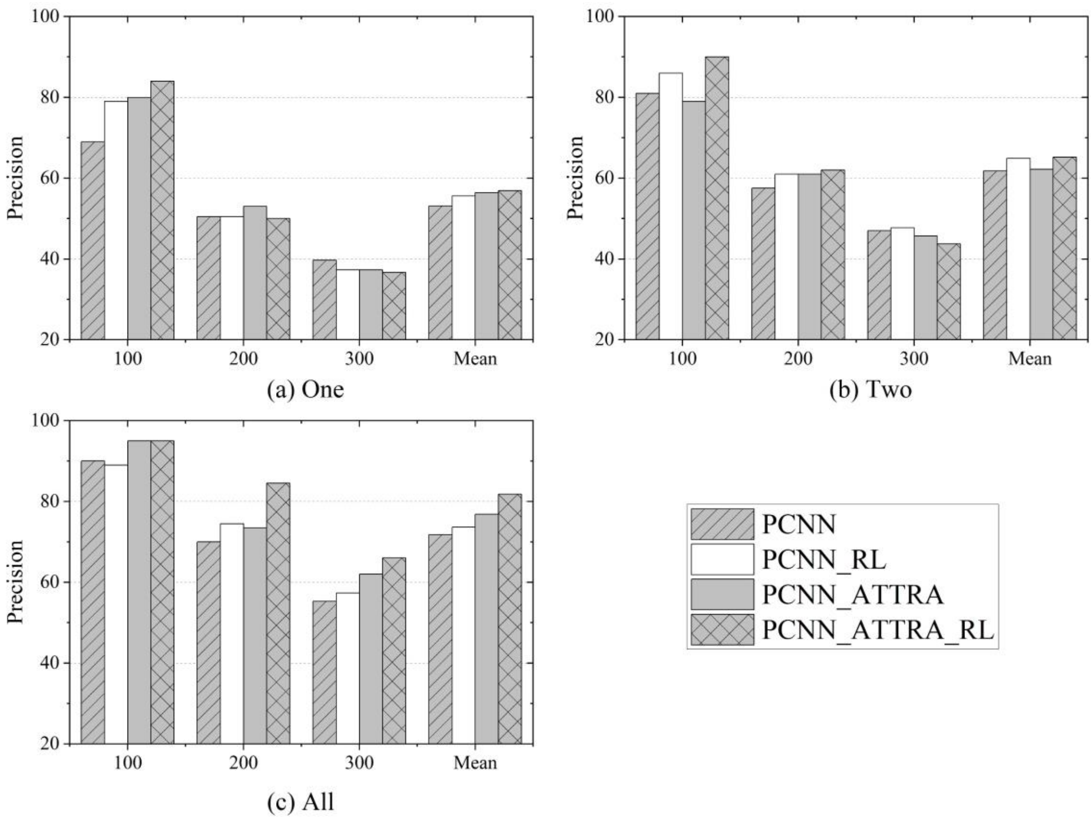

Figure 6. The comparison of different algorithms in the scenario ‘One’, ‘Two’, and ‘All’

The PR curves of different algorithms are shown in Figure 7. The PR curve of the proposed algorithm is always higher than other algorithms. The difference between the algorithms in the first half is not evident. PR curves, while the difference between each curve is obvious in the second half. The performance of PCNN, PCNN-ATTRA and PCNN-RL are similar when the recall is lower than 0.2. After that, PCNN-RL gradually surpassed than other two algorithms. It also can be seen that the curve of PCNN drops rapidly after the recall reaches 0.2 before it gets the worst performance in the second half of the figure. The comparison of the AUC for different algorithms is demonstrated in Table 2. The experimental results were obtained in the scenario ‘All’. It can be seen that PCNN can only reach 0.430. With the introduction of ATTRA and RL, the performance can be leveraged by 0.28 and 0.38, respectively. When both ATTRA and RL are jointly introduced into PCNN, the AUC of PCNN+ATTRA+RL reaches 0.491.

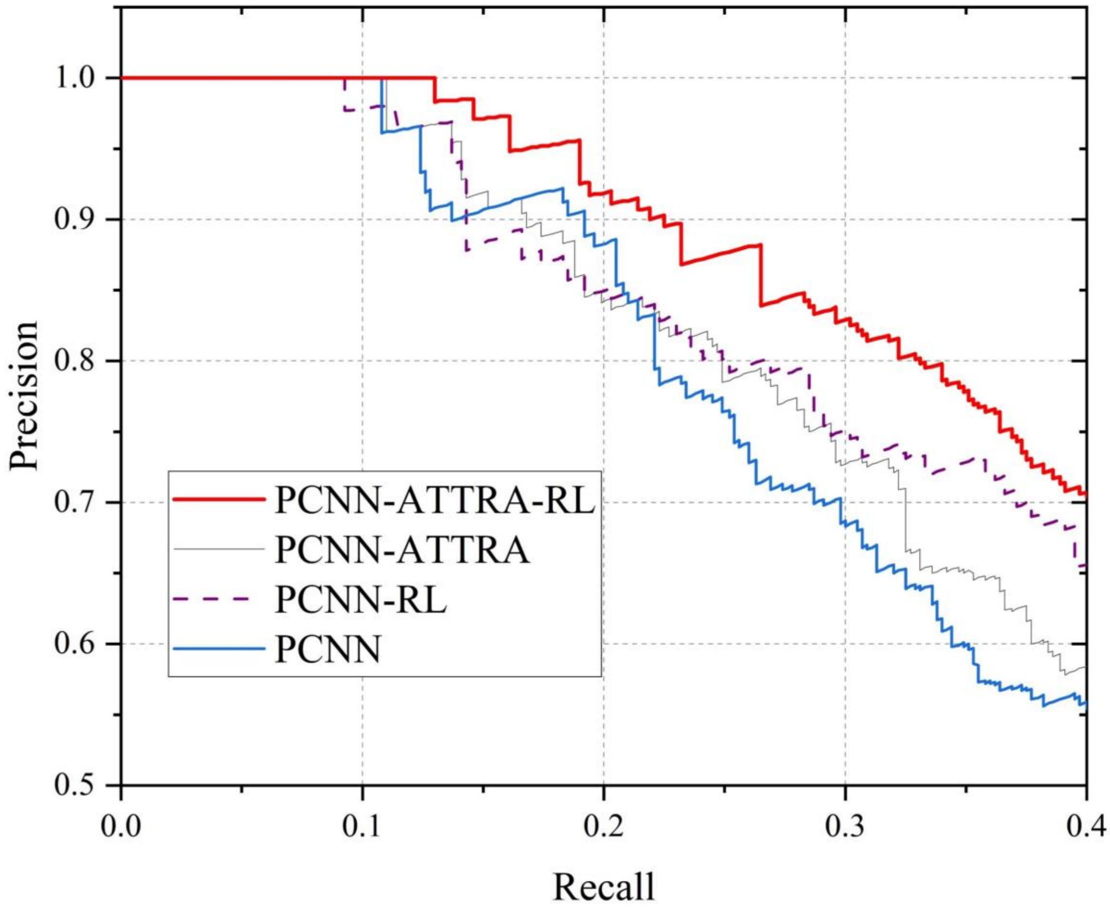

Figure 7. The PR curves of different algorithms in ablation experiment

Table 2. The AUC of different algorithms in the ablation experiment

Algorithms	PCNN	PCNN+ATTRA	PCNN+RL	PCNN+ATTRA+RL
AUC	0.430	0.458	0.468	**0.491**

5.2. Benchmarking Experiments with Prevailing Algorithms

The benchmarking results in terms of P@N are shown in Table 3. It can be seen that the proposed PCNN-ATTRA-RL achieves the best result in most situations. In the testing scenario 'One', the performance of PCNN-BAGATT outperforms the proposed algorithm. The P@100 of PCNN-ATTRA-RL are the same as the PCNN-BAGATT, which is 84.0%. In the testing scenario 'Two', the performance of PCNN-ATTRA-RL is higher better than other benchmarking algorithms in P@100, P@200 and mean, while the P@200 is slightly lower than that of PCNN-BAGATT. The results of testing scenario 'All' indicates that PCNN-ATTRA-RL achieves the best performance except the P@300 is slightly lower than that of PCNN-ATT-NIS.

Table 3. The P@N of algorithms in different testing scenarios

Testing scenarios	One				Two				All			
P@N (%)	100	200	300	Mean	100	200	300	Mean	100	200	300	Mean
PCNN-ONE [26]	71.0	43.0	29.7	47.9	73.0	41.0	28.7	47.5	70.0	72.5	65.7	69.4
PCNN-ATT-NIS [49]	**86.0**	49.0	38.0	57.8	80.0	45.5	31.6	52.4	81.0	45.5	32.3	52.9
PCNN -BAGATT [48]	84.0	**53.0**	**38.0**	**58.3**	86.0	60.5	**46.0**	64.2	92.0	76.5	64.0	77.5
APCNN-NIS [49]	83.0	45.5	31.3	53.9	73.0	41.0	28.6	47.5	87.0	64.0	48.0	66.3
PCNN-ATTRA-RL	84.0	50.0	36.7	56.9	**90.0**	**62.0**	43.7	**65.2**	**95.0**	**84.5**	**66.0**	**81.8**

The PR curves of different algorithms are plotted in Figure 8. It is obvious that the precisions achieved by the proposed PCNN-ATTRA-RL are the highest at all stages. PCNN-BAGATT ranks second at all stages, which performance is obviously better than the rest three algorithms. The APCNN-NIS shows the comparative performance when the recall is lower than 0.2, followed by a steep fall in the range from 0.15 to 0.3. The PCNN-ATT-NIS decreases rapidly in the beginning stage, followed by a gradual decline. The performance of PCNN-ONE fluctuates before the recall reaches 0.1. Then the precision of PCNN-ONE raised by 0.15 in the middle stage before it declines steadily in the final stage. The comparisons of the AUC are listed in Table 4. It can be seen that the proposed algorithm achieved the highest AUC which is 0.491. The PCNN -BAGATT and PCNN-ATT- NIS ranks in the second and third position, which AUC are 0.458 and 0.443 respectively. Meanwhile, the AUC of PCNN-ONE is the lowest in this experiment.

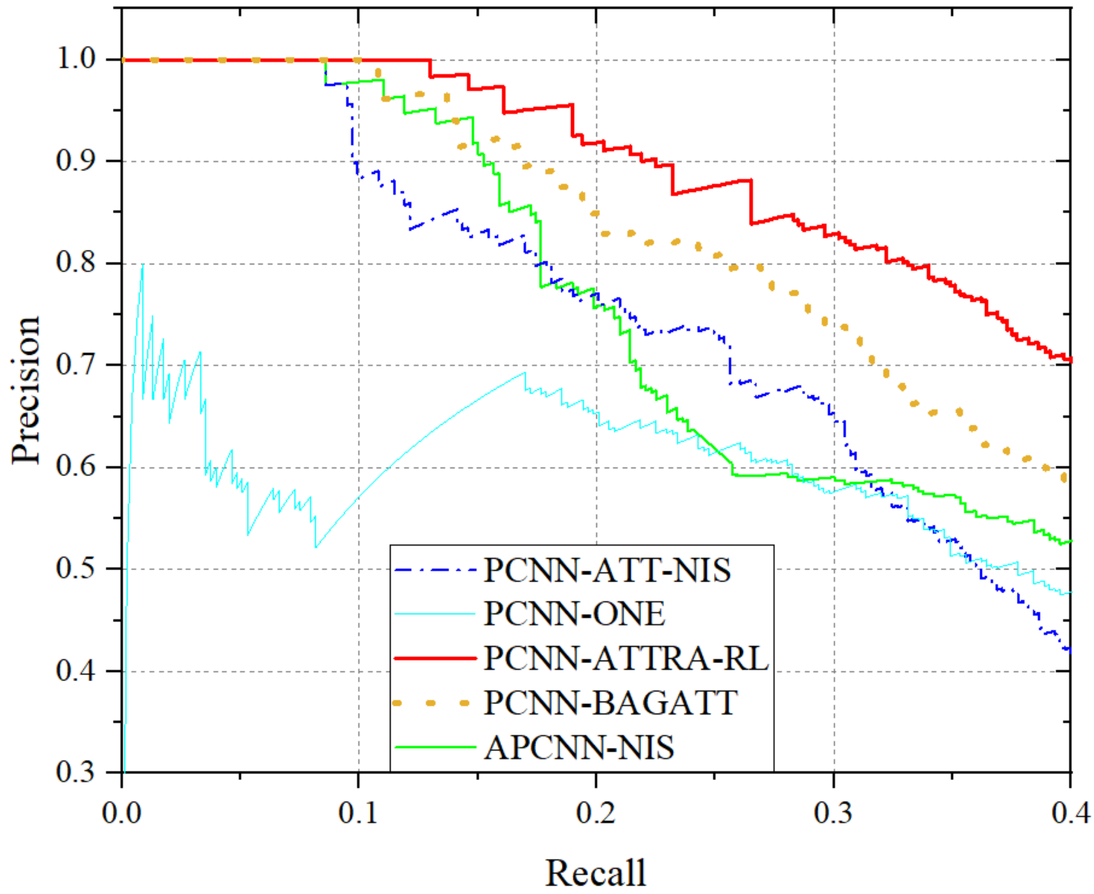

Figure 8. The PR curve of different algorithms

Table 4. The AUC of different algorithms in the ablation experiment

Algorithms	PCNN-ONE [26]	PCNN-ATT-NIS [49]	PCNN -BAGATT [31]	APCNN-NIS [49]	PCNN-ATTRA-RL (Proposed)
AUC	0.349	0.443	0.458	0.427	**0.491**

5.3. Knowledge Graph Visualisation

The relation of the unlabelled data was predicted via the trained PCNN-ATTRA-RL. The triples with the predicted relation were adopted for visualisation. The KG was constructed via the proposed ontology and labelling strategy mentioned in Section 3.1. The visualisation of the fault diagnosis KG was conducted using the Neo4j software. The visualisation of KG is shown in Figure 9. Ten triples were adopted as examples to illustrate the relation prediction and visualisation. It can be seen that the PCNN-ATTRA-RL model that was trained by both labelled and unlabelled data can achieve 7 out of 10 correct predictions. The relation between 'Motor' and 'Fuse' should be 'consist_of'. The relation between 'Motor' and 'Main_circuit' should be 'consist_of'. The relation between 'Grounded' and 'Burn out' should be 'Lead_to'. Besides the three wrong prediction cases, the relation between other entities was correctly predicted. The established fault diagnosis KG can be used for information retrieval. For example, the engineer can search the reason for fuse burnout. The KG can locate two relevant entities which are 'grounded' or 'short circuit'. Subsequently, the engineers can quickly conduct troubleshooting based on the retrieved information.

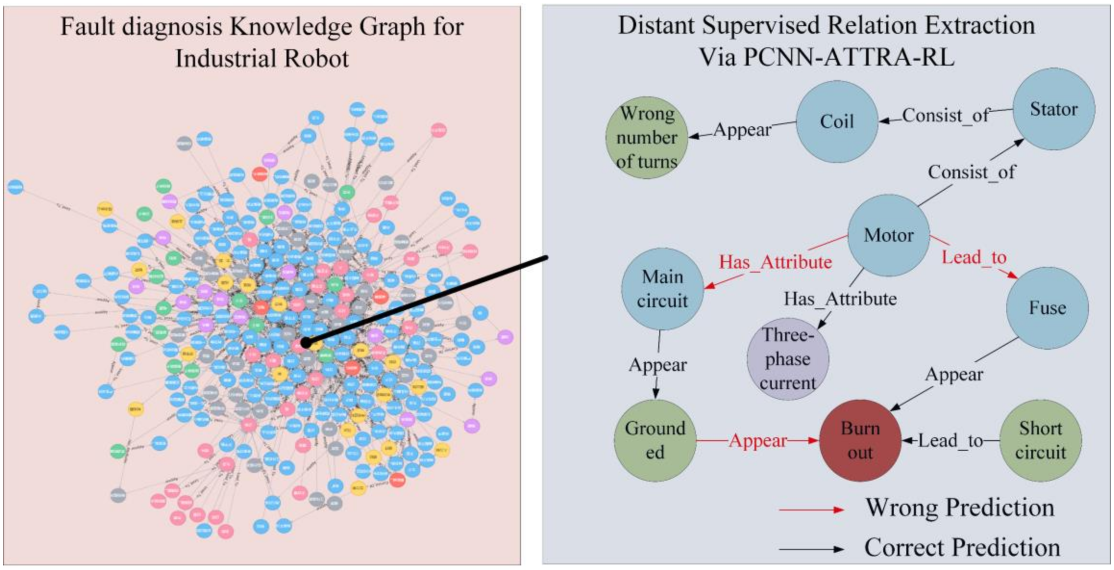

Figure 9. The visualisation of the fault diagnosis KG and the prediction of DSRE

5.4. Discussions

In the construction of a fault diagnosis KG, relation extraction is an essential task. In comparison with the pure supervised learning methods, DSRE can achieve accurate and reliable relation prediction in comparison when the labelled data is limited. The result of the ablation experiment suggested that both ATTRA and RL are effective in boosting the performance of PCNN in the DSRE task. In all the testing scenarios, PCNN-ATTRA-RL shows merits. Especially in the testing scenario 'All', the advantage of the proposed algorithm is obvious. The results of AUC indicate that ATTRA and RL can leverage the AUC by 6.5% and 8.8% respectively. When they were jointly adopted, the prediction can be improved by 14.2%, which is dramatically higher than other algorithms. This improvement indicates the effectiveness of the PCNN-ATTRA-RL. The benefits that both components of the PCNN-ATTRA-RL are different. The PR curves reveal that the PCNN showed better performance in the beginning stage of the curve, while the curve of PCNN-RL was evidently higher than PCNN and PCNN-ATTRA. In the benchmarking experiment, the P@N results of the proposed algorithm are slightly worse than that of PCNN-BAGATT in the testing scenario 'One'. In all three testing scenarios, the condition of the testing scenario 'One' is the harshest and the condition of the testing scenario 'All' is the easiest. In testing scenario 'One', the proposed algorithm is not satisfactory. The reason can be that the entity pairs in the randomly selected sentence are low in frequency which is hard to represent all the sentences in the bag. It can be seen that when all the sentences were selected, the performance of PCNN-ATTRA-RL shows merits in comparison with other algorithms. When multiple sentences are selected in testing scenarios 'Two' and 'All', the performance of the proposed algorithm shows merits because of its capability in identifying and ignoring the noisy sentence. This indicates that the proposed algorithm is able to get satisfactory performance for the noisy corpus. Besides the P@N values, the PR curve reveals that PCNN-ATTRA-RL shows merits in comparison with those prevailing algorithms. The AUC score of the proposed algorithm is 0.491, which is evidently higher than the benchmarking algorithms. The benchmarking algorithms mainly deployed different attention mechanisms to improve the performance of DSRE. In our proposed algorithm, the introduction of reinforcement learning to further process the noisy corpus, which leads to better modelling performance. In the KG graph visualisation, it can be seen that PCNN-ATTRA-RL can correctly predict the relation of most entities. The constructed fault diagnosis KG can be used to assist the engineers in the maintenance of industrial assets.

With the deployment of the proposed PCNN-ATTRA-RL algorithm, the relation in the unlabelled corpus can be correctly predicted, which can greatly reduce the manual efforts in sentence labelling. In this study, the spatial patterns of the unlabelled data have not been considered. In future works, the spatial patterns of the unlabelled data will be studied to achieve better DSRE performance. Since graph neural network is a type of algorithm that can effectively extract spatial patterns, it will be further investigated in the DSRE in the fault diagnosis KG. In the future, the fault diagnosis KG also needs to be further investigated. Firstly, the KG established in this study is based on the multi-source corpus. Besides these data, the sequential data such as sensor data and historical maintenance data collected from the industry is also rich in the hidden knowledge of the fault diagnosis for industrial assets. In the context of Industry 4.0, data integration is a great challenge. How to extract the semantic information from the sequential data and introduced it into the fault diagnosis will be studied. Furthermore, with the increasing requirement for the maintenance of industrial assets, the application of the fault diagnosis KG in real industrial scenarios also needs to be investigated. In the future, a Q&A system, which can provide insights into troubleshooting and repairing, can be developed based on the KG constructed in this study. The knowledge reasoning that can provide suitable knowledge according to the maintenance requirement will be further researched.

6. Conclusions

With the rapid growth of complex and automated assets in the modern industry, fault diagnosis KG can bring tangible benefits to smart asset maintenance. However, it is challenging to perform relation extraction modelling from the corpus when labelled data is limited. Bearing the nature that the labelled data is insufficient and the impact of the noisy data is non-negligible, this study proposed an RL-based DSRE approach to effectively lower the impact of the noisy sentences and achieve satisfactory DSRE performance. The experimental study based on the real-world multi-source corpus indicated that our proposed approach can effectively predict the relations of the unlabelled data and it shows merit in comparison with the prevailing benchmarking algorithms. Meanwhile, a fault diagnosis KG based on the multi-sources corpus was established. In future works, the spatial relationship of the unlabelled data and the application of the fault diagnosis KG in the actual industry will be further researched.

Acknowledgement

Our work is supported by multiple funds in China, including the Key Program of NSFC-

Guangdong Joint Funds (U2001201), the Natural Science Foundation of Guangdong Province (2020A1515010890), National Natural Science Foundation (52075338), Industrial core and key technology plan of ZhuHai City (ZH22044702190034HJL). Our work is also supported by Guangdong Provincial Key Laboratory of Cyber-Physical System (2020B1212060069).

References

[1] E. Zio, Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice, Reliability Engineering & System Safety, 218 (2022) 108119.
[2] C. Chen, Y. Liu, X. Sun, C. Di Cairano-Gilfedder, S. Titmus, An integrated deep learning-based approach for automobile maintenance prediction with GIS data, Reliability Engineering & System Safety, 216 (2021) 107919.
[3] T. Zhang, J. Chen, F. Li, K. Zhang, H. Lv, S. He, E. Xu, Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions, ISA transactions, 119 (2022) 152-171.
[4] W. Li, X. Zhong, H. Shao, B. Cai, X. Yang, Multi-mode data augmentation and fault diagnosis of rotating machinery using modified ACGAN designed with new framework, Advanced Engineering Informatics, 52 (2022) 101552.
[5] J. Long, J. Mou, L. Zhang, S. Zhang, C. Li, Attitude data-based deep hybrid learning architecture for intelligent fault diagnosis of multi-joint industrial robots, Journal of Manufacturing Systems, 61 (2021) 736-745.
[6] P. Lyu, K. Zhang, W. Yu, B. Wang, C. Liu, A novel RSG-based intelligent bearing fault diagnosis method for motors in high-noise industrial environment, Advanced Engineering Informatics, 52 (2022) 101564.
[7] X. Li, H. Shao, S. Lu, J. Xiang, B. Cai, Highly Efficient Fault Diagnosis of Rotating Machinery Under Time-Varying Speeds Using LSISMM and Small Infrared Thermal Images, IEEE Transactions on Systems, Man, and Cybernetics: Systems, (2022).
[8] Y. Xiao, H. Shao, S. Han, Z. Huo, J. Wan, Novel Joint Transfer Network for Unsupervised Bearing Fault Diagnosis From Simulation Domain to Experimental Domain, IEEE/ASME Transactions on Mechatronics, (2022).
[9] S. Yan, H. Shao, Y. Xiao, B. Liu, J. Wan, Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noises, Robotics and Computer-Integrated Manufacturing, 79 (2023) 102441.
[10] L.-C. Zhao, H.-X. Zou, Z.-Y. Wu, Q.-H. Gao, G. Yan, F.-R. Liu, K.-X. Wei, W.-M. Zhang, Dynamically synergistic regulation mechanism for rotation energy harvesting, Mechanical Systems and Signal Processing, 169 (2022) 108637.
[11] J. Long, Y. Qin, Z. Yang, Y. Huang, C. Li, Discriminative feature learning using a multiscale convolutional capsule network from attitude data for fault diagnosis of industrial robots, Mechanical Systems and Signal Processing, 182 (2023) 109569.
[12] Y. Liu, C. Chen, T. Wang, L. Cheng, An attention enhanced dilated CNN approach for cross-axis industrial robotics fault diagnosis, Autonomous Intelligent Systems, 2 (2022) 1-11.
[13] K. Lu, C. Chen, T. Wang, L. Cheng, J. Qin, Fault diagnosis of industrial robot based on

dual-module attention convolutional neural network, Autonomous Intelligent Systems, 2 (2022) 1-12.
[14] J. Deng, T. Wang, Z. Wang, J. Zhou, L. Cheng, Research on Event Logic Knowledge Graph Construction Method of Robot Transmission System Fault Diagnosis, IEEE Access, 10 (2022) 17656-17673.
[15] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, J. Taylor, Freebase: a collaboratively created graph database for structuring human knowledge, Proceedings of the 2008 ACM SIGMOD international conference on Management of data, 2008, pp. 1247-1250.
[16] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, Z. Ives, Dbpedia: A nucleus for a web of open data, The semantic web, Springer2007, pp. 722-735.
[17] X. Li, M. Lyu, Z. Wang, C.-H. Chen, P. Zheng, Exploiting knowledge graphs in industrial products and services: a survey of key aspects, challenges, and future perspectives, Computers in Industry, 129 (2021) 103449.
[18] B. Lai, W. Zhao, Z. Yu, X. Guo, K. Zhang, A Multi-Domain Knowledge Transfer Method for Conceptual Design Combine With FBS and Knowledge Graph, International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, 2022, pp. V002T002A059.
[19] B. Zhou, B. Hua, X. Gu, Y. Lu, T. Peng, Y. Zheng, X. Shen, J. Bao, An end-to-end tabular information-oriented causality event evolutionary knowledge graph for manufacturing documents, Advanced Engineering Informatics, 50 (2021) 101441.
[20] M. Lyu, X. Li, C.-H. Chen, Achieving Knowledge-as-a-Service in IIoT-driven smart manufacturing: A crowdsourcing-based continuous enrichment method for Industrial Knowledge Graph, Advanced Engineering Informatics, 51 (2022) 101494.
[21] P. Zheng, L. Xia, C. Li, X. Li, B. Liu, Towards Self-X cognitive manufacturing network: An industrial knowledge graph-based multi-agent reinforcement learning approach, Journal of Manufacturing Systems, 61 (2021) 16-26.
[22] W. Li, Q. Wang, J. Wu, Z. Yu, Piecewise convolutional neural networks with position attention and similar bag attention for distant supervision relation extraction, Applied Intelligence, 52 (2022) 4599-4609.
[23] L. Deng, B. Yang, Z. Kang, S. Yang, S. Wu, A noisy label and negative sample robust loss function for DNN-based distant supervised relation extraction, Neural Networks, 139 (2021) 358-370.
[24] D. Christou, G. Tsoumakas, Improving distantly-supervised relation extraction through bert-based label and instance embeddings, IEEE Access, 9 (2021) 62574-62582.
[25] P. Li, X. Zhang, W. Jia, H. Zhao, GAN driven semi-distant supervision for relation extraction, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019, pp. 3026-3035.
[26] D. Zeng, K. Liu, Y. Chen, J. Zhao, Distant supervision for relation extraction via piecewise convolutional neural networks, Proceedings of the 2015 conference on empirical methods in natural language processing, 2015, pp. 1753-1762.
[27] A. Smirnova, P. Cudré-Mauroux, Relation extraction using distant supervision: A survey, ACM Computing Surveys (CSUR), 51 (2018) 1-35.
[28] T. Chen, N. Wang, M. He, L. Sun, Reducing wrong labels for distantly supervised relation

extraction with reinforcement learning, IEEE Access, 8 (2020) 81320-81330.
[29] P. Qin, W. Xu, W.Y. Wang, Robust distant supervision relation extraction via deep reinforcement learning, arXiv preprint arXiv:1805.09927, (2018).
[30] G. Ji, K. Liu, S. He, J. Zhao, Distant supervision for relation extraction with sentence-level attention and entity descriptions, Proceedings of the AAAI conference on artificial intelligence, 2017.
[31] Y. Zhou, L. Pan, C. Bai, S. Luo, Z. Wu, Self-selective attention using correlation between instances for distant supervision relation extraction, Neural Networks, 142 (2021) 213-220.
[32] Y. Lin, S. Shen, Z. Liu, H. Luan, M. Sun, Neural relation extraction with selective attention over instances, Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2016, pp. 2124-2133.
[33] Z.-X. Ye, Z.-H. Ling, Distant Supervision Relation Extraction with Intra-Bag and Inter-Bag Attentions, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019, pp. 2810-2819.
[34] Z. Wang, C.-H. Chen, P. Zheng, X. Li, L.P. Khoo, A novel data-driven graph-based requirement elicitation framework in the smart product-service system context, Advanced engineering informatics, 42 (2019) 100983.
[35] Y. Wan, Z. Chen, F. Hu, Y. Liu, M. Packianather, R. Wang, Exploiting knowledge graph for multi-faceted conceptual modelling using GCN, Procedia Computer Science, 200 (2022) 1174-1183.
[36] Z.-Q. Lin, B. Xie, Y.-Z. Zou, J.-F. Zhao, X.-D. Li, J. Wei, H.-L. Sun, G. Yin, Intelligent development environment and software knowledge graph, Journal of Computer Science and Technology, 32 (2017) 242-249.
[37] L. Nizzoli, M. Avvenuti, M. Tesconi, S. Cresci, Geo-semantic-parsing: AI-powered geoparsing by traversing semantic knowledge graphs, Decision Support Systems, 136 (2020) 113346.
[38] M. Liu, X. Li, J. Li, Y. Liu, B. Zhou, J. Bao, A knowledge graph-based data representation approach for IIoT-enabled cognitive manufacturing, Advanced Engineering Informatics, 51 (2022) 101515.
[39] B. Zhou, X. Shen, Y. Lu, X. Li, B. Hua, T. Liu, J. Bao, Semantic-aware event link reasoning over industrial knowledge graph embedding time series data, International Journal of Production Research, (2022) 1-18.
[40] B. Zhou, J. Bao, J. Li, Y. Lu, T. Liu, Q. Zhang, A novel knowledge graph-based optimization approach for resource allocation in discrete manufacturing workshops, Robotics and Computer-Integrated Manufacturing, 71 (2021) 102160.
[41] H.-B. Shi, D. Huang, L. Wang, M.-Y. Wu, Y.-C. Xu, B.-E. Zeng, C. Pang, An information integration approach to spacecraft fault diagnosis, Enterprise Information Systems, 15 (2021) 1128-1161.
[42] H. Ren, Z. Chen, Z. Jiang, C. Yang, W. Gui, An Industrial Multilevel Knowledge Graph-Based Local–Global Monitoring for Plant-Wide Processes, IEEE Transactions on Instrumentation and Measurement, 70 (2021) 1-15.
[43] M. Mintz, S. Bills, R. Snow, D. Jurafsky, Distant supervision for relation extraction without labeled data, Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and

the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2009, pp. 1003-1011.
[44] S. Riedel, L. Yao, A. McCallum, Modeling relations and their mentions without labeled text, Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2010, pp. 148-163.
[45] J. Qu, D. Ouyang, W. Hua, Y. Ye, X. Li, Distant supervision for neural relation extraction integrated with word attention and property features, Neural Networks, 100 (2018) 59-69.
[46] X. Zhang, T. Liu, P. Li, W. Jia, H. Zhao, Robust neural relation extraction via multi-granularity noises reduction, IEEE Transactions on Knowledge and Data Engineering, 33 (2020) 3297-3310.
[47] Y. Yuan, L. Liu, S. Tang, Z. Zhang, Y. Zhuang, S. Pu, F. Wu, X. Ren, Cross-relation cross-bag attention for distantly-supervised relation extraction, Proceedings of the AAAI conference on artificial intelligence, 2019, pp. 419-426.
[48] Z.-X. Ye, Z.-H. Ling, Distant supervision relation extraction with intra-bag and inter-bag attentions, arXiv preprint arXiv:1904.00143, (2019).
[49] Y. Chen, H. Liu, C. Wu, Z. Yuan, M. Jiang, Y. Huang, Distant supervision for relation extraction with neural instance selector, CCF International Conference on Natural Language Processing and Chinese Computing, Springer, 2018, pp. 209-220.
[50] Y.-M. Shang, H. Huang, X. Sun, W. Wei, X.-L. Mao, A pattern-aware self-attention network for distant supervised relation extraction, Information Sciences, 584 (2022) 269-279.
[51] J. Webber, A programmatic introduction to neo4j, Proceedings of the 3rd annual conference on Systems, programming, and applications: software for humanity, 2012, pp. 217-218.