Neurocomputing 430 (2021) 104–111

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A knowledge graph method for hazardous chemical management: Ontology design and entity identification

Xue Zheng, Bing Wang, Yunmeng Zhao, Shuai Mao, Yang Tang *

The Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, Shanghai 200237, PR China

ARTICLE INFO

Article history:
Received 27 May 2020
Revised 27 August 2020
Accepted 25 October 2020
Available online 10 November 2020
Communicated by Zidong Wang

Keywords:
Knowledge graph
Ontology
Hazardous chemicals management
Named entity recognition

ABSTRACT

Hazardous chemicals are widely used in the production activities of the chemical industry. The risk management of hazardous chemicals is critical to the safety of life and property. Hence, the effective risk management of hazardous chemicals has always been important to the chemical industry. Since a large quantity of knowledge and information of hazardous chemicals is stored in isolated databases, it is challenging to manage hazardous chemicals in an information-rich manner. Herein, we prompt a knowledge graph to overcome the information gap between decentralized databases, which would improve the hazardous chemical management. In the implementation of the knowledge graph, we design an ontology schema of hazardous chemicals management. To facilitate enterprises to master the knowledge in the full lifecycle of hazardous chemicals, including production, transportation, storage, etc., we jointly use data from companies and open data from the public domain of hazardous chemicals to construct the knowledge graph. The named entity recognition task is one of the key tasks in the implementation of the knowledge graph, which is of great significance for extracting entity information from unstructured data, namely the hazardous chemical accidents records. To extract useful information from multi-source data, we adopt the pre-trained BERT-CRF model to conduct named entity recognition for incidents records. The model achieves good results, exhibiting the effectiveness in the task of named entity recognition in the chemical industry.

1. Introduction

Hazardous chemicals are widely used in almost every corner of the chemical industries and many other manufacturing fields. They not only build up the basis of the economy but also are closely related to our daily life. Due to the flammability, explosion, high toxicity, and high corrosivity [1], hazardous chemicals have an impact on all phases of their lifecycles, such as production, transportation, and storage. Risk management of hazardous chemicals is an important topic worldwide. Hazardous chemical management includes many aspects, including hazard identification, consequence analysis, probability analysis, and risk assessment, etc.

One important reason for the frequent occurrence of hazardous chemical related incidents is the lack of relevant knowledge and poor training. Due to enormous differences in physical and chemical properties, the management of hazardous chemicals involves a vast amount of interconnected information, especially when conducted in the full life cycle. Though nontrivial data sets have been established to facilitate the hazardous chemical management, these isolated islands of information hinder the comprehensive understanding and utilization rate of the knowledge of hazardous chemicals [2]. Some information technologies have been used in the management of hazardous chemicals, like the expert system [3] in the process of safety analysis. Although the expert system is suitable for hazard identification and inference in a specific field, it cannot cover the whole life cycle of hazardous chemicals. In the field of process knowledge engineering, the integrated ontology, OntoCAPE [4] was developed. OntoCAPE defined the overall schema of the process industry including the meta layer, the upper layer, the conceptual layer, and the application layer, etc. OntoCAPE is an important comprehensive ontology that is not applied to process data.

Knowledge graph, as a new type of graph database content retrieval method proposed by Google in 2012, has developed vigorously and effectively so far [5]. As a semantic network, knowledge graph has powerful expressive ability and modeling flexibility, and

* Corresponding author.
E-mail addresses: tangtany@gmail.com, yangtang@ecust.edu.cn (Y. Tang).

https://doi.org/10.1016/j.neucom.2020.10.095
0925-2312/

it can model entities, concepts, attributes, and their relationships [6,7]. The knowledge graph is promising in knowledge retrieval [8], question-answering [9], knowledge recommendation [10], knowledge visualization [11], and other applications. The knowledge graph has been applied in many fields and have achieved good results [12–14]. For instance, some scholars [15] have tried to apply knowledge graphs in the field of traditional Chinese medicine (TCM) health care and have expanded the scale of the knowledge graph. The platform supported by knowledge graph can provide non-professionals with knowledge services such as knowledge retrieval of traditional Chinese medicine [15]. Besides, the knowledge graph has been applied in the field of geological hazards. By constructing the knowledge graph of geological hazard documents, scholars have improved the utilization rate of literature information and provided knowledge services and knowledge bases for preventing and responding to geological hazards [16]. However, the use of a knowledge graph in the field of hazardous chemical management is challenged by the complex interconnections between various risk influencing factors as well as the overwhelming amount of interconnected information.

There are unstructured documents in the field of hazardous chemicals management, and the knowledge graph should be established on the embedded entities and relations within these documents. The named entity recognition is one of the key tasks in the implementation of the knowledge graph. Different entities need to be identified following an overall schema. The mainstream methods of identifying entities are divided into three categories: rule-based methods, learning-based methods, and hybrid methods [17]. The most basic of rule-based methods are dictionary-based entity recognition. For the good performance of deep learning in natural language processing, most of the learning-based methods are implemented using deep learning.

Knowledge and data of hazardous chemicals are usually stored in separate tables. We try to establish a new knowledge graph to build connections among a substantial volume of chemical companies, chemicals, hazards, accidents, and other types of related knowledge. One major challenge in constructing a knowledge graph for hazardous chemicals management is how to build a appropriate ontology structure. A critical part is defining categories, relations, and attributes. Another major challenge is how to make better use of textual information in related documents, namely efficient entity identification. In this article, we propose an ontology framework for hazardous chemicals management, which provides the foundation and solution for improving process safety. We also carry out named entity recognition based on deep learning methods to identify entities in textual information so that the information can be better utilized in the chemical industry.

The main contributions of this paper are as follows:

- The contribution of introducing the knowledge graph to the chemical industry is linking the corresponding knowledge in the unstructured data source together and establishing a knowledge network full of connections. Besides, the application of the knowledge graph is beneficial to the identification of risk sourcing and propagation.
- Another contribution of this article is improving the utilization of text information in chemical documents. We apply the natural language processing technology to chemical data to solve the problem of identifying entities in chemical documents.

The paper is organized as follows. In Section 2, a framework of the proposed ontology for hazardous chemical management is given. Section 3 describes the establishment of the ontology-based knowledge graph. Section 4 presents the method that we adopt to recognize named entities in the chemical industry. Some concluding remarks are finally given in Section 5.

2. Ontology development in a top-down manner

The first step is the design of ontologies for the knowledge graph of hazardous chemical management from accumulated data resources and human knowledge. Data resources can be divided into structured data and unstructured data [18]. As shown in Fig. 1, the overall architecture is a combination of top-down and bottom-up methods: at first, the ontology is proposed based on human knowledge and experiences, then rules are applied to form the early entity maps by integrating exiting structural data.

The most important task of constructing a knowledge graph is to design ontology [19,20]. Before designing the ontology, we first determined the scope. The hazardous chemical management involves chemical production, storage, transportation, usage, treatment, etc. The class hierarchy should consider the concepts of company, equipment, hazardous chemical, human, incident, risk, etc. The subject concept was divided into the company, equipment, hazardous chemical, person, and incident. The typical relations between the listed classes included chemical reaction, production, and consumption, etc.

Based on the provided concepts, we first designed seven top-level classes. And subclasses were then added to the top-level classes. The structure of classes in ontology is shown in Fig. 2. Secondly, we defined the relation between classes, namely, the object attribute. For example, we defined the object attribute "relatedIncChe" (related incident chemical) between the class Incident and Chemical, the domain of the object attribute was Incident, and Chemical was the range of the object attribute. More relationships between classes are shown in Fig. 3. At last, we defined the properties of each class, also known as data properties. Fig. 4 shows examples of data properties. On the left, we defined the following data attributes for the chemical class: "hasCASNumber", "hasFormula", "Hasupperhlrelimit", etc. The right subgraph in Fig. 4 shows the data attributes of the company class: "hasTaxFileNumber", "hasEmail", "has Address", etc.

After defining the object properties and data properties, we used the properties to add constraints to the classes in ontology. As shown in Fig. 5, the description of a class includes an existential quantifier description (some) and a full quantifier description (only). For example, "relatedIncOrg some company" indicating that the class has a relationship associated with the company, and "incident and (relatedIncChe only (hazardous chemical))" indicating hazardous chemicals incidents are incidents only associated with hazardous chemicals.

Fig. 1(a) shows six types of structured data resources in the project. Among them, chemical attribute data are open data obtained from the chemical registration catalog. Related companies, major hazards evaluations, and incidents data are data provided by enterprises. All datasets are arranged in CSV files or tables. Tables 2 and 4 show the size and type of the datasets. The content of the chemical registration dataset is shown in Table 1, and the content of data such as companies and hazardous source incidents is shown in Table 3. Some of the datasets describe the attributes of the entity, such as "Hazardous chemicals catalog 2015". Some datasets also include related information between entities, such as "Dataset of chemicals and enterprise related to hazard sources".

3. Completion of ontology in a bottom-up way

We adopted a rule-based method to map the structured data resources with a knowledge graph and establish a mapping relationship between concepts in the database and ontology in the knowledge graph. We used different extraction rules to achieve semi-automatic extraction of database entities, attributes, and relations from different data structures. The data sets in the field of hazardous chemicals have the following characteristics: many

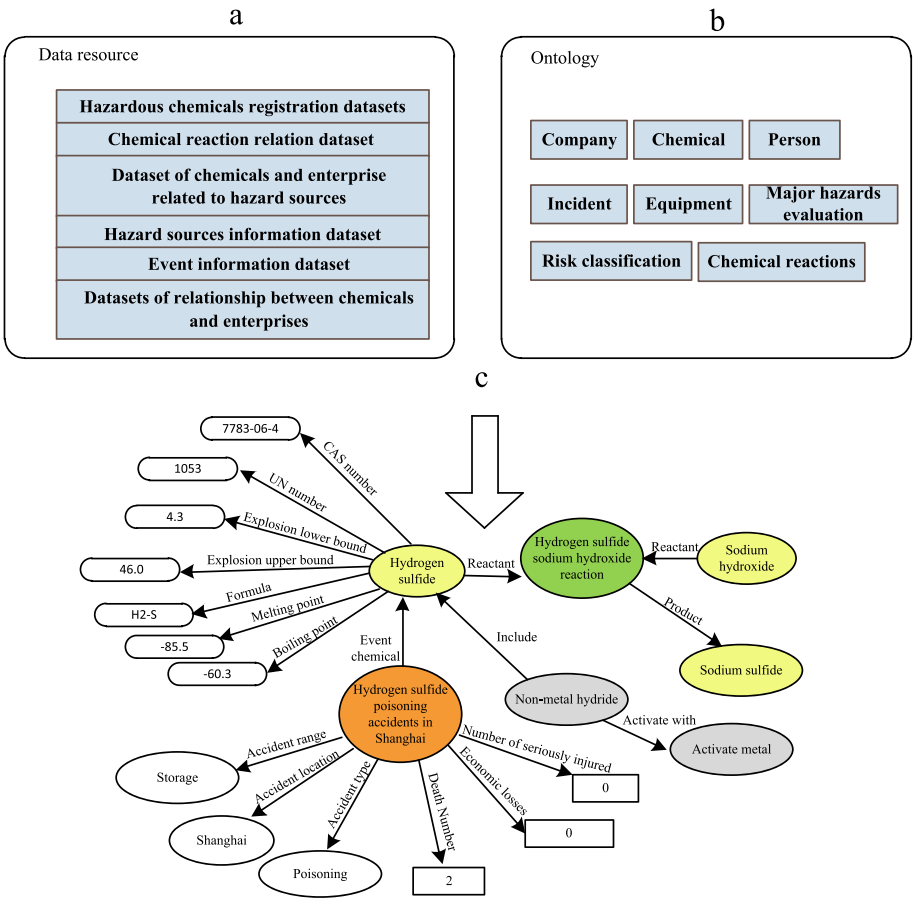

Fig. 1. Technical framework for building a knowledge graph of hazardous chemicals management.

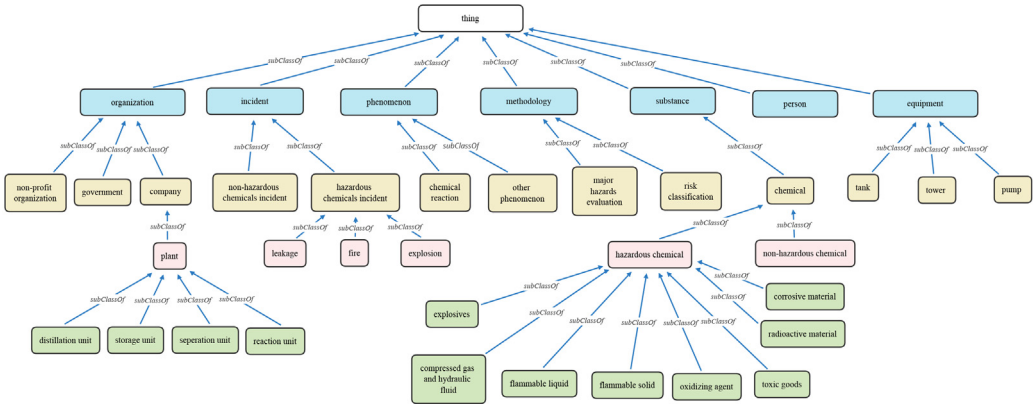

Fig. 2. The hierarchy of classes in the hazardous chemical ontology.

aliases and common names of chemicals, inconsistent names of chemicals in different data sets, and error and missing record in CAS number file. The above characteristics lead to the information island problem in the chemical industry. Compared to the non-exclusive names, the CAS number is an important basis for determining chemicals. Thus, to solve the above problems, we created and updated the CAS number file of chemicals, and we used the chemical name in the CAS number file as the standard chemical name. Also, we created and updated the chemical alias set.

With the aid of the integration of the structured data, we obtained a knowledge graph of hazardous chemicals management, which contained 124,593 attributes, 66,184 entities, and 223,640 relationships. However, the current graph construction had the following deficiencies:

- The amount of data was small, and the external chemical resources were not fully utilized. Various information such as chemical names, devices, processes, and accident types involved in the chemical process were included in the text, requiring more effort to identify safety-related entities.
- The graph did not completely describe the contents of the table, and the utilization rate of the table information could be further improved.
- The manual construction of the graph was inefficient.

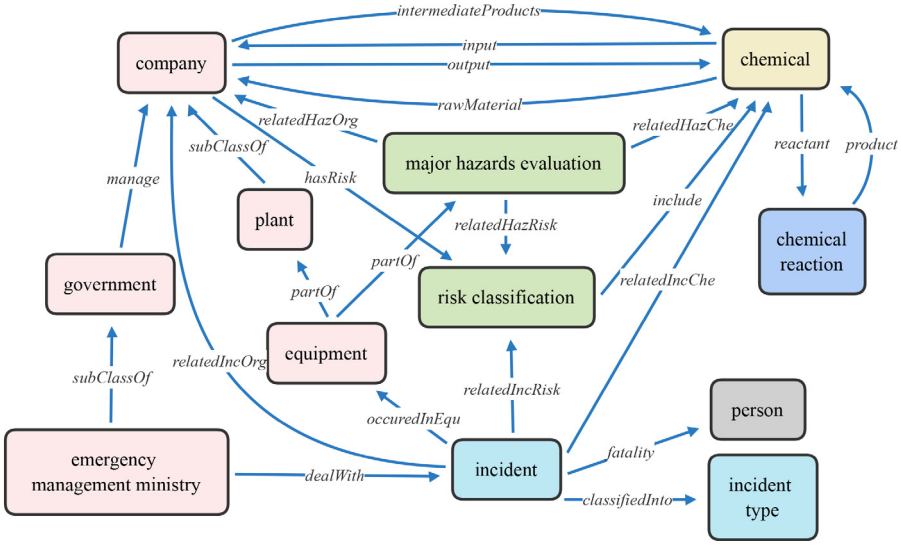

Fig. 3. The example of relations between classes in the hazardous chemical ontology.

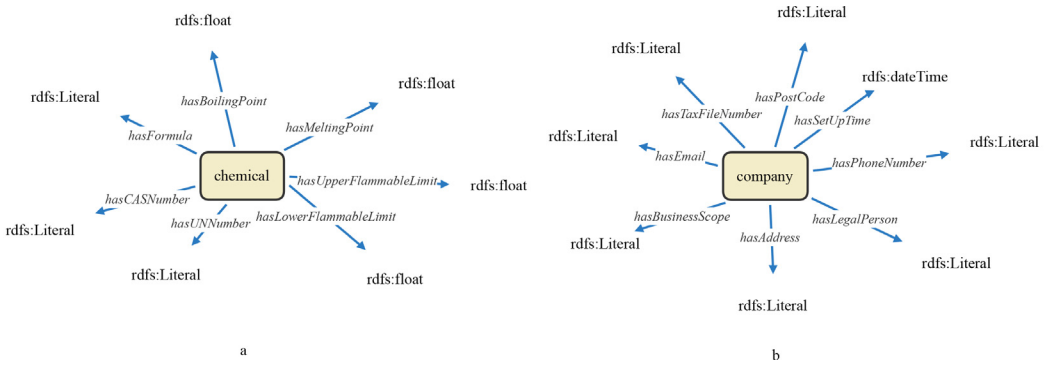

Fig. 4. Two examples of data properties of the class.

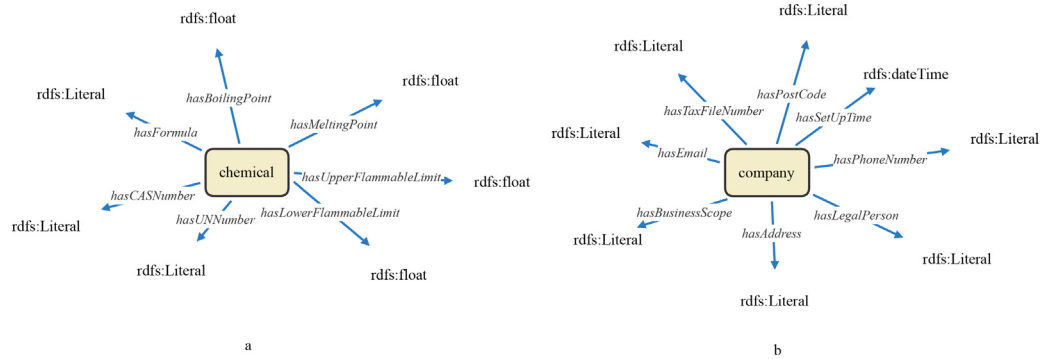

Fig. 5. Two examples of axioms in defining the class.

To overcome these deficiencies, the automation and efficiency of the construction of knowledge graphs need to be improved. Information extraction can automatically [21,22] or semi-automatically extract useful triples from the text, improving the speed and efficiency of constructing knowledge graphs. The information extraction task is divided into three steps: firstly, the named entities are correctly identified; secondly, the relationships between concepts are described; finally, the relationships are nicely classified. Therefore, the accuracy of the information extraction task depends on the correct named entity [23,24]. Named entity recognition is an important step in the information extraction task. Our goal in the next section is to train named entity recognizers to prepare for subsequent tasks such as relationship extraction and knowledge question answering. Specific identification methods will be developed in the next section.

4. Deep learning-powered named entity recognition

The representation of chemical knowledge includes structured, unstructured, semi-structured data. The relational database is the

Table 1
Hazardous chemicals registration datasets and content.

Hazardous chemicals registration dataset	Chemical attributes in related Database
Hazardous chemicals catalog 2015	CAS number, Chemical Chinese name
Dangerous goods list	UN dangerous goods number, Chemical English name, Chemical Chinese name, Category, Special dangers
Occupational disease classification and catalog	Chemical Chinese name, Occupational disease
List of hazardous chemicals under key supervision	CAS number, Chemical Chinese name
List of highly toxic substances 2003	CAS number, Chemical name, Chemical Chinese name, Chinese alias
Catalog of categories and varieties of precursor chemicals	Chinese name of chemical, Category
Catalogue of hazardous chemicals under key environmental management 2014	CAS number, Chinese alias,Chinese name of chemical
Explosive hazardous chemicals list 2011	name, CAS number, explosive danger category, un_num, chemId, classification

Table 2
The size and type of hazardous chemicals registration datasets and content.

Hazardous chemicals registration dataset	Size	Type
Hazardous chemicals catalog 2015	2997 items	csv
Dangerous goods list	2505 items	csv
Occupational disease classification and catalog	136 items	csv
List of hazardous chemicals under key supervision	76 items	csv
List of highly toxic substances 2003	54 items	csv
Catalog of categories and varieties of precursor chemicals	41 items	csv
Catalogue of hazardous chemicals under key environmental management 2014	84 items	csv
Explosive hazardous chemicals list 2011	99 items	csv

Table 3
Datasets and content provided by enterprises.

Datasets provided by the enterprise	Chemical attributes in related Database
Enterprise information	User enterprise ID, Unit name, Address, Province ID, City ID, County ID, Nature units, Post code, Business license number, Manager name, Phone number, Fax number, Mail, Unit code, Set-up time, Production range, Representative, etc.
Datasets of relationship between chemicals and enterprises	Unit name, Chemical code, Chemical property, Name, Alias name, English name, English alias name, CAS number, UN number, Formula
Dataset of chemicals and enterprise related to hazard sources	Hazard sources company, Danger source ID, Hazard source chemicals
Hazard sources information	Company ID, Hazard source name, Hazard source address, Hazard source level, Hazard source R value, Hazard source scale, Safety distance, People number, etc.
Event information Involved chemicals, Hazardous chemicals, Accident type, Accident level, etc.	Accident name, Domestic or international, Accident location,

Table 4
The size and type of datasets provided by enterprises.

Datasets provided by the enterprise	Size	Type
Enterprise information	47067 items	csv and unstructured data
Datasets of relationships between chemicals and enterprises	246562 items	csv
Dataset of chemicals and enterprise related to hazard sources	25541 items	csv
Hazard sources information	11295 items	csv
Event information	25522 items	csv and unstructured data

most common structured data format, and web documents are common semi-structured data types such as XML, JSON, etc. However, human-readable documents are one of the main data sources in the hazardous chemical industry. In Section 2 and Section 3, the ontology and knowledge graph have been constructed based on the structured data by combining the top-down and bottom-up approaches. For unstructured data in the chemical industry, such as position statement, operating procedures, and accident information, the identification of the entity is essential, and a named entity reorganization model is trained based on ontology definitions.

In this work, the original corpus contains five types of entities: chemicals, event type, enterprise organization, chemical equipment, and chemical operation system, all of which are derived from available documents such as job data, operating procedures, and accident information. For the available documents, we first performed sentence segmentation processing. We collected a total of 12,689 original corpus sentences. The training set contained a

Table 5
The number of entities in datasets.

Datasets	Chemicals	Accident type	Enterprise organization	Chemical equipment	Chemical operation system	Total number
Training dataset	7314	8263	4276	4171	1742	25,766
Validation dataset	1325	1326	731	672	302	4347
Testing dataset	1138	1369	722	604	309	4142

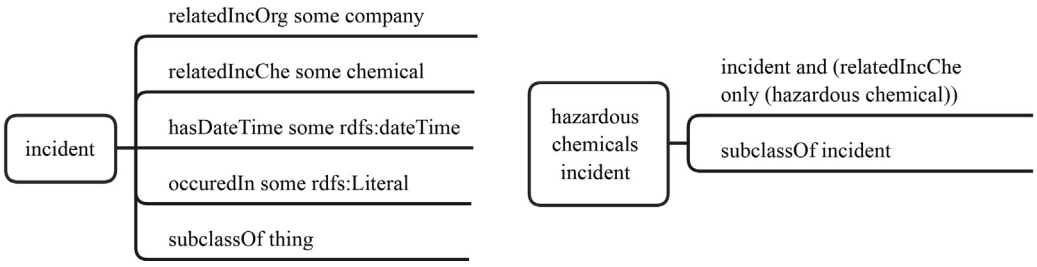

Fig. 6. Picture of the process of data preprocessing.

total of 9649 sentences, the verification set contained a total of 1577 sentences, and the test set contained a total of 1463 sentences. The dataset statistics are shown in Table 5.

4.1. Data preprocessing

The labeling strategies for named entity recognition include BIO mode, BIOE mode, and BIOS mode. We used the BIO labeling strategy, where B represents the beginning of the entity, I represents the non-starting part of the entity, and O represents the part that is not the entity in the sentence. When predicting the entity boundary, it is necessary to predict the entity type at the same time. So there are eleven types of labels to be predicted, namely "O", "B-CHE", "I-CHE", "B-ETP", "I-ETP", "B-ORG", "I-ORG ", "B-EQU", "I-EQU", B-SYS, I-SYS. CHE, ETP, ORG, EQU and SYS indicate the corresponding chemicals, event type, enterprise organization, chemical equipment and chemical operation system, respectively. The entire data processing diagram is shown in Fig. 6. The named entity identification in the hazardous chemicals industry was regarded as a sequence labeling problem. We organized the available documents related to hazardous chemicals, such as operating procedures and accident records. Then we preprocessed the documents into a training standard format through sentence segmentation and character splitting. The classic BIO labeling scheme was used to manually label industrial corpus sentences in the available documents. An example of labeling the original corpus data is shown in the right half of Fig. 6. The first part of each line is a word, and the second part is a word label. The word and the word label are separated by spaces. Sentences are separated by blank lines.

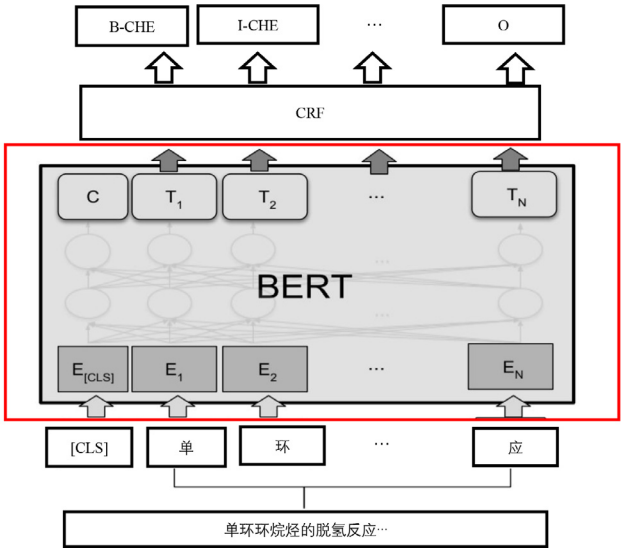

Fig. 7. This figure describes the structure of the BERT + CRF model, and the part in red box comes from the google [25].

4.2. Method

In the research of chemical named entity recognition, we took up a hybrid method to identify entities, and this model is based on bidirectional encoder representation from transformers (BERT) model and conditional random field (CRF) model. First, a deep neural network was used to obtain a deep representation of sentence semantics, and then through the constraint function of the CRF layer, the maximum probability sequence was output.

The overall structure is shown in Fig. 7. The entire model is divided into two parts. Among them, the BERT model [25] is a pre-trained language model that is trained from a large number of text corpora by using unsupervised training methods. The conditional random field [26] has been widely used in sequence labeling for a long time, and it is an undirected probability graph discriminant model. In this work, Chinese sentences were first split into single characters, each character was mapped to a character id by the dictionary that came with the BERT model, and then the character id was transformed into an embedding vector with complex semantics through the embedding layer. Then the word vector sequence was input to the CRF layer. The CRF layer [26] can learn the constraint conditions of the sentence and improve the accuracy of the prediction sequence.

4.3. Experimental parameters and results

Google provides two pre-trained language models: BERT-Base and BERT-Large. The network structures of these two models are the same, with only some parameters being different. It takes more graphics card memory to train BERT-Large. In contrast, BERT-Base model requires less memory and the accuracy of training has met our needs. Thus, we adopted the BERT-Base model. BERT-Base model had a total of 12 layers, and the hidden layer was 768 dimensions. We adopted a 12-head mode with a total of 110M

Table 6
The comparison of results on test datasets between different models. "Avg." means the average score.

Model	Evaluation	CHE	ETP	ORG	EQU	SYS	Avg.
Rule-based method	P	0.5382	0.3478	0.4212	0.6372	0.7480	0.4850
	R	0.4785	0.2655	0.8486	0.6527	0.9194	0.5309
	F	0.5066	0.3011	0.5630	0.6449	0.8249	0.4924
CRF [28]	P	**0.9643**	0.9039	0.9590	**0.9397**	**0.9605**	0.9395
	R	0.9500	0.9338	0.9500	0.9579	0.9419	**0.9452**
	F	**0.9571**	0.9186	0.9545	**0.9488**	**0.9511**	0.9423
BiLSTM + CRF [29]	P	0.9326	0.9171	0.8829	0.9052	0.8984	0.9123
	R	0.9500	**0.9396**	0.8865	0.9257	0.9129	0.9292
	F	0.9412	0.9282	0.8847	0.9153	0.9056	0.9206
BERT + CRF	P	0.9500	**0.9455**	**0.9623**	0.9190	0.8830	**0.9411**
	R	0.9508	0.9248	**0.9557**	**0.9586**	**0.9618**	0.9450
	F	**0.9504**	**0.9350**	**0.9590**	0.9384	0.9207	**0.9428**

parameters. The maximum sequence length was 230, the train batch size parameter was 32, the learning rate was 1e−5, the droop out rate parameter was 0.5.

We adopted precision (P), recall (R), and F indicators as our evaluation criteria, the three indicators can be calculated by true positives (TP), false positives (FP), and false negatives (FN). The specific calculation formula is as follows:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{3}$$

Simply, the precision is computed by our prediction results. It indicates the percentage of positive samples that are true positive samples. The recall rate indicates the percentage of positive examples in the sample that are predicted correctly. The F value combines the results of accuracy and recall [27].

The performance of these four methods on the test set is shown in Table 6. Among them, the rule-based method was based on the dictionary, which consisted of entities that appear in the training set. Besides, the external chemical dictionary was also used to identify chemical entities. The experiment result proved that the dictionary-based method was not effective in the test set of the chemical industry. The main reason is that the dictionary in the chemical industry is not complete. The other three models achieved good results. The test results of the best BERT model were as follows: the precision rate was 0.9411, the recall rate was 0.9450, and the F1 value was 0.9428. From the perspective of the model effect, the BERT model performed better than the BiLSTM model in the chemical dataset. In some entity types, the CRF model achieved superior results. For example, in terms of chemical entities, the CRF model was more accurate than the BERT model. However, in the training process, the CRF model needed to carefully select the feature template. In general, the BERT model did not require manual feature selection and performed well in named entity recognition tasks in the chemical industry.

In this section, we applied the methods in natural language processing to the chemical industry. Aiming at the large number and constant update of chemical names in the chemical industry, we used a combination of a pre-trained model and a probability graph model to identify entities in the text of the chemical industry. The experimental verification showed that the method had a significant effect on the test set.

5. Conclusion

Our work aims to extract chemical-related named entities from the considerable body of the unstructured document and build a hazardous chemical management knowledge graph. In this paper, an ontology for risk management of hazardous chemicals is designed. Besides, a deep learning method is adopted to identify named entities in the chemical industry, which greatly improves the effectiveness of named entity recognition. The method achieves the highest precision of 0.9411, recall rate of 0.9450, and an F1 score of 0.9428.

Compared with traditional data storage methods, the knowledge graph connects chemical industry-related datasets, laying a foundation for knowledge services in the chemical industry. The proposed hazardous chemicals ontology framework can provide basic support for information integration and inference. Named entity recognition lays the foundation for chemical corpus processing and chemical knowledge graph question and answer application, etc., leading to a significant improvement in the utilization of chemical-related document information.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported in part by the National Key Research and Development Program of China under Grant 2018YFC0809302, the National Natural Science Foundation of China under Grants 61988101, 61751305, 61673176 and the Programme of Introducing Talents of Discipline to Universities (the 111 Project) under Grant B17017.

References

[1] B. Wang, C. Wu, G. Reniers, L. Huang, L. Kang, L. Zhang, The future of hazardous chemical safety in China: opportunities, problems, challenges and tasks, Sci. Total Environ. 643 (2018) 1–11.
[2] A. Menon, N.B. Krdzavac, M. Kraft, From database to knowledge graph—using data in chemistry, Curr. Opin. Chem. Eng. 26 (2019) 33–37.
[3] J. Zhao, L. Cui, L. Zhao, T. Qju, B. Chen, Learning hazop expert system by case-based reasoning and ontology, Comput. Chem. Eng. 33 (1) (2009) 371–378.
[4] J. Morbach, A. Yang, W. Marquardt, Ontocape—a large-scale ontology for chemical process engineering, Eng. Appl. Artif. Intell. 20 (2) (2007) 147–161.
[5] W. Liu, J. Liu, M. Wu, S. Abbas, W. Hu, B. Wei, Q. Zheng, Representation learning over multiple knowledge graphs for knowledge graphs alignment, Neurocomputing 320 (2018) 12–24.
[6] W. Li, R. Peng, Y. Wang, Z. Yan, Knowledge graph based natural language generation with adapted pointer-generator networks, Neurocomputing 382 (2020) 174–187.

[7] Q. Wang, Y. Hao, ALSTM: An attention-based long short-term memory framework for knowledge base reasoning, Neurocomputing (2020).
[8] X. Zhang, X. Hou, X. Chen, T. Zhuang, Ontology-based semantic retrieval for engineering domain knowledge, Neurocomputing 116 (116) (2013) 382–391.
[9] S. Zhu, X. Cheng, S. Su, Knowledge-based question answering by tree-to-sequence learning, Neurocomputing 372 (2020) 64–72.
[10] Z. Sun, J. Yang, J. Zhang, A. Bozzon, L.-K. Huang, C. Xu, Recurrent knowledge graph embedding for effective recommendation, in: Proceedings of the 12th ACM Conference on Recommender Systems, 2018, pp. 297–305.
[11] X. He, R. Zhang, R. Rizvi, J. Vasilakes, X. Yang, Y. Guo, Z. He, M. Prosperi, J. Bian, Prototyping an interactive visualization of dietary supplement knowledge graph, in: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2018, pp. 1649–1652.
[12] Y. Liang, F. Xu, S.-H. Zhang, Y.-K. Lai, T. Mu, Knowledge graph construction with structure and parameter learning for indoor scene design, Comput. Visual Media 4 (2) (2018) 123–137.
[13] H. Weng, Z. Liu, S. Yan, M. Fan, A. Ou, D. Chen, T. Hao, A framework for automated knowledge graph construction towards traditional Chinese medicine, in: International Conference on Health Information Science, Springer, 2017, pp. 170–181.
[14] P. Zhu, W. Zhong, X. Yao, Auto-construction of course knowledge graph based on course knowledge, Int. J. Perform. Eng. 15 (8) (2019).
[15] T. Yu, J. Li, Q. Yu, Y. Tian, X. Shun, L. Xu, L. Zhu, H. Gao, Knowledge graph for TCM health preservation: design, construction, and applications, Artif. Intell. Med. 77 (2017) 48–52.
[16] R. Fan, L. Wang, J. Yan, W. Song, Y. Zhu, X. Chen, Deep learning-based named entity recognition and knowledge graph construction for geological hazards, ISPRS Int. J. Geo-Inf. 9 (1) (2020) 15.
[17] M. Xiaofeng, W. Wei, X. Aiping, Incorporating token-level dictionary feature into neural model for named entity recognition, Neurocomputing 375 (2020) 43–50.
[18] Y. Jia, Y. Qi, H. Shang, R. Jiang, A. Li, A practical approach to constructing a knowledge graph for cybersecurity, Engineering 4 (1) (2018) 53–60.
[19] J. Dou, J. Qin, Z. Jin, Z. Li, Knowledge graph based on domain ontology and natural language processing technology for Chinese intangible cultural heritage, J. Visual Lang. Comput. 48 (2018) 19–28.
[20] M. Pietranik, N.T. Nguyen, A multi-attribute based framework for ontology aligning, Neurocomputing 146 (2014) 276–290.
[21] S. Zheng, Y. Hao, D. Lu, H. Bao, J. Xu, H. Hao, B. Xu, Joint entity and relation extraction based on a hybrid neural network, Neurocomputing 257 (2017) 59–66.
[22] J. Zhang, Y. Zhang, D. Ji, M. Liu, Multi-task and multi-view training for end-to-end relation extraction, Neurocomputing 364 (2019) 245–253.
[23] A. Goyal, V. Gupta, M. Kumar, Recent named entity recognition and classification techniques: a systematic review, Comput. Sci. Rev. 29 (2018) 21–43.
[24] M.W. Alnabki, E. Fidalgo, E. Alegre, L. Fernandezrobles, Improving named entity recognition in noisy user-generated text with local distance neighbor feature, Neurocomputing 382 (2020) 1–11.
[25] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).
[26] A. Chen, F. Peng, R. Shan, G. Sun, Chinese named entity recognition with conditional probabilistic models, in: Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing, 2006, pp. 173–176.
[27] W. Li, W. Song, X. Jia, J. Yang, Q. Wang, Y. Lei, K. Huang, J. Li, T. Yang, Drug specification named entity recognition base on BiLSTM-CRF model (2019) 429–433.
[28] N. Sobhana, M. Pabitra, S. Ghosh, Conditional random field based named entity recognition in geological text, Int. J. Comput. Appl. 1 (02 2010). doi:10.5120/72-166.
[29] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, C. Dyer, Neural architectures for named entity recognition, in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, San Diego, California, 2016, pp. 260–270. doi:10.18653/v1/N16-1030. https://www.aclweb.org/anthology/N16-1030.

Xue Zheng received the B.S. degree in the school of control science and engineering from Qingdao University, in 2018. She is currently pursuing an M.A. degree from East China University of Science and Technology. Her research interests include knowledge graph, natural language processing, and their applications in chemical engineering.

Bing Wang received the B.S. and Ph.D. degree in chemical engineering from Tsinghua University, Beijing China in 2011 and 2016, respectively. From 2014 to 2015, he was a Visiting Scholar with the Mary Kay O'Connor Process Safety Center in Texas A&M University, College Station, United States. After graduation, he has become an assistant professor in East China University of Science and Technology, Shanghai. Dr. Bing Wang's research interests includes atmospheric dispersion modeling, quantitative risk assessment to chemical spills, source term estimation for air pollution dispersion, knowledge base on chemical process safety.

Yunmeng Zhao received his B.S. in Chemistry from Nanjing University in 2014, and his PhD in Chemical Engineering from Monash University in 2019. He is currently an assistant professor at East China University of Science and Technology. His research interests are soft electronics, wearable sensors, and machine learning.

Shuai Mao received the B.S. degree in school of control science and engineering from East China University of Science and Technology, in 2017. He is currently pursuing the Ph.D. degree from East China University of Science and Technology. His research interests include multi-agent systems, distributed optimization and their applications in practical engineering.

Yang Tang received the B.S. and Ph.D. degrees in electrical engineering from Donghua University, Shanghai, China, in 2006 and 2010, respectively. From 2008 to 2010, he was a Research Associate with The Hong Kong Polytechnic University, Hong Kong. From 2011 to 2015, he was a Post-Doctoral Researcher with the Humboldt University of Berlin, Berlin, Germany, and with the Potsdam Institute for Climate Impact Research, Potsdam, Germany. Since 2015, he has been a Professor with the East China University of Science and Technology, Shanghai. His current research interests include distributed estimation/control/optimization, cyber-physical systems, hybrid dynamical systems, computer vision, reinforcement learning and their applications. Prof. Tang was a recipient of the Alexander von Humboldt Fellowship and the ISI Highly Cited Researchers Award by Clarivate Analytics from 2017. He is a Senior Board Member of Scientific Reports, an Associate Editor of IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Emerging Topics in Computational Intelligence and IEEE Systems Journal, etc.