Liu *et al. J Wireless Com Network* (2021) 2021:55
https://doi.org/10.1186/s13638-021-01936-0

EURASIP Journal on Wireless Communications and Networking

RESEARCH Open Access

Relation extraction for coal mine safety information using recurrent neural networks with bidirectional minimal gated unit

Xiulei Liu[1,2], Shoulu Hou[1,2], Zhihui Qin[1,2], Sihan Liu[3] and Jian Zhang[4*]

*Correspondence: jian.zhang-bistu@yandex.com
[4] School of Economics and Management, Beijing Information Science and Technology University, Beijing 100192, People's Republic of China
Full list of author information is available at the end of the article

Abstract

The data of coal mine safety field are massive, multi-source and heterogeneous. It is of practical importance to extract information from big data to achieve disaster precaution and emergency response. Existing approaches need to build more features and rely heavily on the linguistic knowledge of researchers, leading to inefficiency, poor portability, and slow update speed. This paper proposes a new relation extraction approach using recurrent neural networks with bidirectional minimal gated unit (MGU) model. This is achieved by adding a back-to-front MGU layer based on original MGU model. It does not require to construct complex text features and can capture the global context information by combining the forward and backward features. Evident from extensive experiments, the proposed approach outperforms the existing initiatives in terms of training time, accuracy, recall and *F* value.

1 Introduction

5G communication technology with low latency and great bandwidth accelerates the data transmission rate, enabling the use of real-time data to achieve the perception of Industrial Internet of Things (IIoT) environment [1, 2]. Coal mine production, one of the typical application scenarios of IIoT, has caused widespread concern about safety and efficiency. However, the number of major disaster accidents remains high, such as gas and fire explosion. The reason is that the automation and information systems in the coal mine are independent of each other and the data are not interconnected [3]. Another reason is that the coal mine safety field involves people, devices, environment and management, and the resulting data are massive, multi-source and heterogeneous, whereas insight and knowledge are hidden within these big data. How to extract and predict information from big data, support cross-system information sharing, and achieve diversified suggestions are of great significance for disaster prevention and emergency response in the process of coal mine production [4, 5, 6].

For the current information knowledge management system of coal mine, the production process is based on system integration technologies and the circulation process is based on digitization and information technologies [7], which are lack of intelligent reasoning capability. Based on the traditional technologies, existing knowledge

Springer Open

© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

representation approach [8] has coarse-granularity description, small representation range, and poor computational efficiency. Ontology technology has been increasingly adopted by the coal mine safety field [9], since it can describe knowledge in a standardized manner and realize the transfer, reuse and sharing of information. With the development of 5G communication and mobile edge computing technologies, the low latency and high bandwidth can provide more accurate services [10–12]. Relation extraction mines the association between concepts within a large-scale data, and is the key step in ontology construction.

With the development of deep learning technology [13], relation extraction has innovated and developed rapidly. In terms of the dependence on labelled data, automatic relation extraction approaches can be divided into four types: supervised learning, semi-supervised learning, unsupervised learning, and open extraction, respectively. The supervised learning approaches cannot capture the global context information, since it is based on original minimal gated unit (MGU) model which is unidirectional and processes data in one direction [14]. The semi-supervised learning approaches have a poor portability and are not suitable for coal mine safety data, since the accuracy of information extraction depends on the quality of the initial relation seed [15]. The unsupervised learning approaches need to analyse and post-process the extraction results, and the clustering threshold cannot be determined in advance [16]. The open extraction approaches map relation instances to texts by means of external knowledge bases such as DBPedia, OpenCyc and YAGO [17]. However, these bases rarely contain safety knowledge of coal mine, and the related research cannot apply to coal mine safety data. There is no mature relation extraction approach for coal mine safety information.

For the data in coal mine safety, this paper applies the recurrent neural networks (RNNs) with MGU to learn the high-dimensional attribute features and avoid complex feature selection problem. The key contributions of the paper can be summarized as follows.

1. We design an automatic relation extraction approach using RNNs with bidirectional minimal gated unit (Bi-MGU) model to capture the global context information. This is achieved by adding a back-to-front MGU layer based on original MGU model.
2. Based on the 2005 automatic content extraction (ACE2005), the experimental results show that the proposed approach has a higher accuracy, recall rate and F value, and a shorter training time, as compared to the existing initiatives.

The rest of this paper is organized as follows. In Sect. 2, the related works are reviewed. Section 3 presents the RNNs with Bi-MGU model to capture the global context information. The analytical comparison of the proposed model and existing alternatives is conducted in Sect. 4, followed by conclusions in Sect. 5.

2 Related work

Deep learning can learn the high-dimensional attribute features and reflect the semantic features of vocabularies well, and it is widely used for relation extraction.

The supervised learning approaches use labelled data for model training. Vo et al. [18] proposed a relation extraction approach based on semantic information expansion

syntax tree to generate rules and accomplish relation extraction. Li et al. [19] proposed a classifier approach based on support vector machine to achieve relation extraction. Zheng et al. [20] introduced the word feature and proposed a relation extraction approach based on conditional random fields. Zhou et al. [21] proposed a relation extraction approach based on the kernel function by calculating the similarity of dependency tree. To tackle the relation classification task, Wu et al. [22] proposed a model that leverages the pretrained BERT language model and incorporates information from the target entities. Zeng et al. [23] proposed an end-to-end model based on sequence-to-sequence learning with copy mechanism, which jointly extracts relational facts from sentences of any of these classes. For relation extraction, Zhang et al. [24] proposed an extension of graph convolutional networks, which pools information over arbitrary dependency structures efficiently in parallel. Guo et al. [25] proposed an attention guided graph neural network for relation extraction tasks, which automatically learns how to selectively attend to the relevant sub-structures. To solve the natural language relational reasoning task, Zhu et al. [26] proposed graph neural networks with generated parameters to adapt graph neural networks. However, the mentioned approaches rely heavily on the linguistic knowledge of researchers and cannot make full use of contextual structure information. Additionally, it is difficult to extract large-scale data due to its slow training and testing speed.

The semi-supervised learning approaches reduce the dependence on manual annotation corpus by adding seed and iterative learning manually. Agichtein et al. [27] designed a Snowball method based on vector representation to reduce the impact of manual intervention on relation extraction. Chen et al. [28] proposed a semi-supervised extraction model based on graph strategy to improve the accuracy of relation extraction. Zhang et al. [29] proposed the BootProject algorithm based on random feature projection to achieve relation extraction. Wang et al. [30] proposed a label-free distant supervision approach, which makes use of the type information and the translation law derived from typical knowledge graph embedding model to learn embeddings for certain sentence patterns. However, these works have the semantic drift problems, and are easily affected by the quality of the initial relation seed.

The unsupervised learning approaches achieve the semantic extraction relations by learning entity context. Qin et al. [31] proposed a Chinese entity relation extraction model based on unsupervised learning, which achieved the relation extraction of large-scale unlabelled data. Shinyama et al. [32] proposed an unsupervised method based on multi-level clustering to achieve relation extraction based on reported articles. Gonzalez et al. [33] proposed a technique based on a probabilistic clustering model to achieve unsupervised relation extraction. However, these works lack clear boundaries and objective evaluation criteria, and have low accuracy. Also, the clustering threshold cannot be determined in advance.

The open extraction approaches have advantages in cross-domain and later expansion due to no constrains on the relation category and target text. Etzioni et al. [34] built the KnowItAll model and realized entity relation extraction by manually writing rule templates. The ever-increasing popularity of web APIs allows app developers to leverage a set of existing APIs to achieve their sophisticated objectives [35]. Banko et al. [36] proposed a TextRunner-based approach to extract specific relations from the Web. Wu

et al. [37] constructed an open extractor system to achieve relation extraction based on Wikipedia information. However, the existing works lack a recognized evaluation system and are unable to deeply explore the implicit relation between entities. Therefore, there is still a gap between the needs of coal mine safety field and the situation of relation extraction.

3 Method

3.1 Overall network structure

The overall network structure diagram elaborates the techniques used in each step of the relation extraction from a microscopic perspective, as shown in Fig. 1. The network structure can be divided into 5 layers, and the function of each layer is described as follows:

1. The input layer pre-processes data, extracts concepts in the sentence, and deletes the ones that do not contain any concepts. The resulting data is divided into training data and test data, where the training data is annotated with the assistance of relevant experts. Each piece of data is described as a tuple <*concept1, concept2, word spacing, sentence relation, type*>.
2. The word embedding layer constructs a text word vector to represent its corresponding words. This is achieved by converting text information into word vectors which are trained by word2vec. Each sentence is converted into a multidimensional matrix. The text features used in this paper include the word itself and the word spacing.
3. The cyclic neural network layer trains the corpus (processed data) that is input into the Bi-MGU unit. The train model tries to minimize the negative log likelihood function to obtain the optimal model. The relation classification is considered as a multi-classification problem.
4. The pooling layer uses a maximum pool operation to get the final vector representation of the input corpus. To make full use of the information in each sentence, we introduce the attention mechanism to calculate the attention probability, thereby reflecting the importance of a sentence in the set. The overall features of the text are obtained by using pooling operation.
5. The output layer calculates the predictive relation category of the corpus by using an integrated SoftMax function. The new features of the text are calculated by combin-

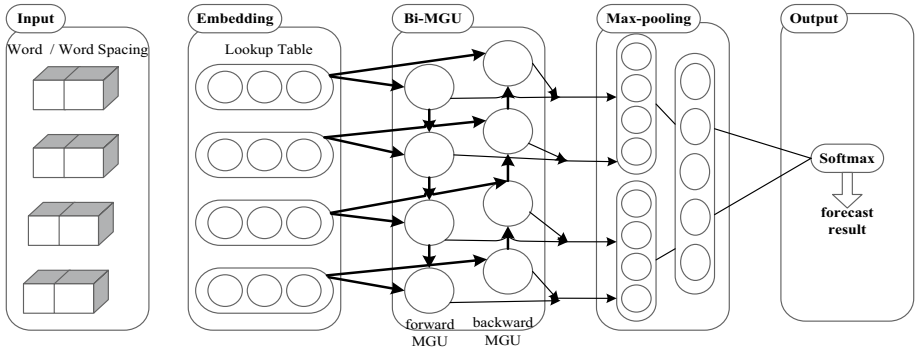

Fig. 1 An illustration on the network structure of the relation extraction

ing the overall features and local features. The resulting features are imported into the classifier for classification. Final classification result is output at the end of the layer.

3.2 Text vector representation

In order to process data by using a neural network model, the input data needs to be vectorized first. Different from the traditional one-hot representation, the word embedding based on neural network training contains rich context information, which can represent the semantic rules of the target words in the current text, and avoid the dimensional disaster [38]. We use the word2vec tool to train word embedding and choose the Skip-gram model as the training framework. Constructing a text word vector means converting text information into a vector form and each sentence is converted into a multidimensional matrix. Sentence S is given which contains the word set $W = \{w_1, w_2, \ldots, w_m\}$, m is the number of words in the sentence S. The text feature set of the sentence S is $K = \{k_1, k_2, \ldots, k_n\}$, n represents number of text features extracted from each sentence. The i-th text feature extracted from the t-th word is expressed as $w_t^{k_i} (1 \leq i \leq n)$.

The features used in this paper are the current word and word spacing. This paper performs word vectorization on text information:

$$r^w = W^{\text{word}} \times V^w \tag{1}$$

r^w is the word vector representation of the word w. W^{word} represents text word vector matrix, $W^{\text{word}} \in R^{l \times m}$. m indicates the number of words in the sentence. l represents the dimension of the word vector. V^w is the one-hot representation of the word w.

In the same way, word vectorization is performed on each text feature:

$$r^{k_i} = W^{k_i} \times V^w \tag{2}$$

r^{k_i} is the word vector representation of the i-th feature. W^{k_i} is the eigenvector distribution of the i-th feature, $W^{k_i} = \left(w_1^{k_i}, w_2^{k_i}, \ldots, w_m^{k_i}\right)$. Vectorization of each word is connection of each vector. Vectorization of the t-th word is:

$$X_t = \left[r_t^w, r_t^{k_1}, r_t^{k_2}, \ldots, r_t^{k_n}\right] \tag{3}$$

The final text local feature is:

$$e = \{X_1, X_2, \ldots, X_m\} \tag{4}$$

3.3 Bi-MGU model

Figure 2 shows the proposed Bi-MGU model. It consists of two layers: (1) the front-to-back MGU layer, which captures the above feature information; and (2) the back-to-front MGU layer, which captures the following feature information. By combining the forward and backward features, we can obtain the global context information which is helpful for sequence modelling task. Each training sequence has two MGU units that move backward and forward, respectively. Both layers are connected to an output layer. The proposed

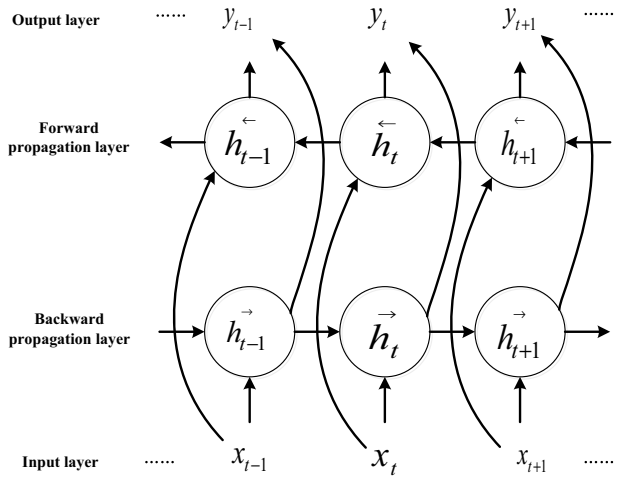

Fig. 2 The proposed Bi-MGU model

model overcomes the drawbacks of traditional unidirectional MGU model which captures data in only one direction.

The state update of the front-to-back MGU layer is given by:

$$\overrightarrow{h_t} = H\left(W_{x\overrightarrow{h_t}}x_t + W_{\overrightarrow{hh}}\overrightarrow{h_{t-1}} + b_{\overrightarrow{h}}\right) \tag{5}$$

$\overrightarrow{h_t}$ is the state of the hidden layer from the front to the back at time t.$\overrightarrow{h_{t-1}}$ is the state of the hidden layer from the front to the back layer at time $t-1$. x_t is the input at time t. $W_{x\overrightarrow{h_t}}$ and $W_{\overrightarrow{hh}}$ are weight matrices. $b_{\overrightarrow{h}}$ is a bias term.

The state update of the back-to-front MGU layer is:

$$\overleftarrow{h_t} = H\left(W_{x\overrightarrow{h_t}}x_t + W_{\overleftarrow{h}\,\overleftarrow{h}}\overleftarrow{h_{t+1}} + b_{\overleftarrow{h}}\right) \tag{6}$$

$\overleftarrow{h_t}$ is the state of the hidden layer from the front to the back at time t. $\overleftarrow{h_{t+1}}$ is the state of the hidden layer from the front to the back layer at time $t+1$. x_t is the input at time t. $W_{x\overrightarrow{h_t}}$ and $W_{\overleftarrow{h}\,\overleftarrow{h}}$ are weight matrices. $b_{\overleftarrow{h}}$ is bias term.

The cumulative results of the two MGU layers are input into the hidden layer, which is calculated as follows.

$$y_t = W_{\overrightarrow{hy}}\overrightarrow{h_t} + W_{\overleftarrow{h}\,y}\overleftarrow{h_t} + b_y \tag{7}$$

where y_t is the output at time t, $W_{\overrightarrow{hy}}$ and $W_{\overleftarrow{h}\,y}$ are weight matrices, and b_y is a bias term.

Each node in Fig. 2 is an MGU unit. The MGU has only one gated structure that combines the input gate (reset gate) with the forgotten gate (update gate). Compared to the LSTM with three gated structures and GRU with two gated structures, the structure of MGU is simpler and contains less parameters, as shown in Fig. 3.

As can be seen from Fig. 3, we have

$$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right) \tag{8}$$

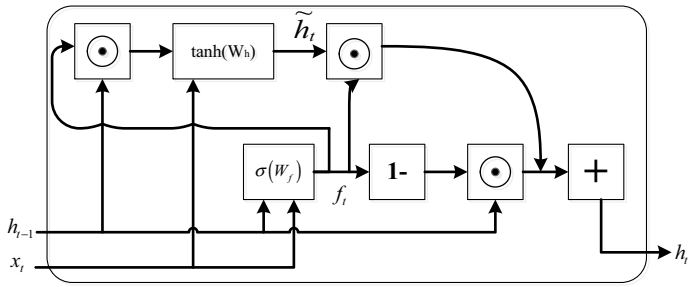

Fig. 3 An illustration on the structure of MGU

$$\widetilde{h_t} = \tanh\left(W_h[f_t \odot h_{t-1}, x_t] + b_h\right) \tag{9}$$

$$h_t = \left(1 - f_t\right) \odot h_{t-1} + f_t \odot \widetilde{h_t} \tag{10}$$

h_{t-1} and h_t are the states of the hidden layer at time $t-1$ and t, respectively. x_t is the input at time t. f_t is the activation function of the gated structure at time t. $\widetilde{h_t}$ is short-term memory item. W_f and W_h are weight matrices. b_f and b_h are bias terms. $\odot$ is the component-wise product between two vectors.

3.4 Attention mechanism and pooling

On the basis that the relation set for classification differs in the importance of words in sentences, this paper adopts a word-level attention weight matrix to capture the information associated with the target relations. Since the attention mechanism can automatically adjust the weights, the deep learning model can focus on the more important parts of the task goal. The weight is calculated as:

$$a_t = \frac{\exp\left(f\left(y_t, n\right)\right)}{\sum_{k=1}^{l} \exp\left(f\left(y_k, n\right)\right)} \tag{11}$$

where a_t is the weight of the vector and can be normalized by using Softmax, y_t is automatically calculated in the attention mechanism, and l the number of vectors to be assigned weights. f is a function of vector y_t and vector n, and has different forms. This paper adopts the following:

$$f(m_t, n) = V_a^T \tanh\left(W_a y_t + U_a n\right) \tag{12}$$

where V_a is the weight vector, and W_a and U_a are the weight matrices.

This paper uses formula (12) to link the output of each step of the hidden layer of the bidirectional MGU model with influencing factors, then the output of each step of the hidden layer is weighted to obtain the representation of the sentence, as follows:

$$f\left(y_t, n\right) = \mathrm{v}_a^T \tanh\left(W_a y_t + U_a n\right) \tag{13}$$

$$a_t = \frac{\exp\left(f\left(y_t, n\right)\right)}{\sum_{k=1}^{l} \exp\left(f\left(y_k, n\right)\right)} \tag{14}$$

$$y = \sum_{k=1}^{l} a_t y_t \tag{15}$$

where y_t is the output of the t-th step of the hidden layer, n is the vector corresponding to the factors that affect the weight, l is the sentence length, and y is the final output, which is used as the representation of the sentence.

In order to consider more contextual semantic associations and obtain features that are more relevant to relation classification tasks, this paper uses the pooling method of attention mechanism. First, the sentence vector after the Bi-MGU layer is multiplied by the attention weight matrix to obtain the corresponding output features $F = \{F_1, F_2, \ldots, F_m\}$. Then, use the largest pooling operation to get the most significant feature representation.

$$d = \max(F) \tag{16}$$

where d is the overall characteristics of the text after pooling. Since the feature dimension after pooling is fixed, the problem of different lengths of text sentences can be solved. Finally, the SoftMax classifier is used to predict the relation category labels.

4 Performance evaluation

4.1 Experimental description

All the neural network models are carried out in the Google open-source deep learning framework TensorFlow v1.2 (Windows 10, 64 bit). The performance of the Bi-MGU model proposed in this paper is analysed by comparing the performance of LSTM model, GRU model and MGU model in training time, relation extraction accuracy, recall rate and F value.

Based on ACE2005 standard and the marked corpus, this experiment extracts 7 types of relations: location, causality, occurrence, responsibility, part-whole, possession and others relations. Among them, the location relation describes the geographical location; the causality relation describes causal connection or mutual influence between concepts; the occurrence relation means the fact that has occurred; the responsibility relation usually exists in concepts such as personnel and institution; the part-whole relation represents a hierarchical structure of two concepts; and the possession relation generally includes usage, adoption and so on. Except for the above 6 relations, the remaining relations are labelled as others relation. Next, the dataset is divided into training corpus and test corpus, where 16,544 items are taken as training corpus and 3,496 items are taken as test corpus.

The corpus we used here is the coal mine accident case and coal mine accident analysis reports, which are crawled from coal mine safety net, coal mine accident net and safety management network. First, delete the corpus that does not contain any concepts. Then, each piece of data is described as a tuple <*concept1, concept2, word spacing, relation type, sentence*>.

4.2 Results and discussion

RNN is a key technology for processing time-series data in deep learning. However, there exists the problem of vanishing gradient in RNN. The problem increases the difficulty

Table 1 Comparison on training time of different models

Model name	LSTM	GRU	Bi-MGU
Training time (s)	10,260	8600	6850

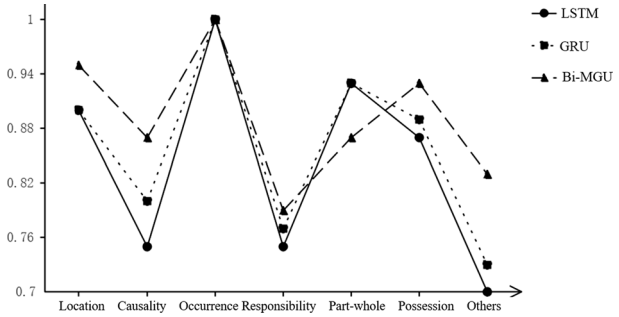

Fig. 4 Comparison on accuracy of different relation extraction approaches, where each type of relation is used as a data point to tease out the effectiveness of the proposed model

that the network learns long-distance dependence, and restricts the practical application of RNN. The long short-term memory network (LSTM), gated recurrent unit (GRU) and minimal gated unit (MGU) are the widely used RNN variants, which can alleviate the problem of vanishing gradient of RNN through gating mechanism [39, 40]. However, LSTM and GRU have complex internal structures, i.e. LSTM and GRU have three and two gated structures, respectively. MGU can achieve the same function as LSTM and GRU by using one gated structure. We choose the MGU model due to its simpler structure and fewer parameters. Additionally, unidirectional MGU model can only capture data in one direction, i.e. the above feature information. To obtain the global context information, this paper proposes a Bi-MGU model by adding a back-to-front MGU layer to capture the following feature information.

We compare the proposed Bi-MGU approach with the current benchmarks in the literature [38, 40] in terms of training time, accuracy, recall and *F*-Score. Table 1 shows that the proposed Bi-MGU model has the minimum training time. This confirms that the simpler the model structure and the fewer the training parameters, the less training time is required. The performance of different is evaluated on different types of relations, as shown in Figs. 4, 5 and 6.

The extraction accuracy of location, causality, occurrence, and others relations with Bi-MGU model is higher than that with LSTM and GRU models. However, the extraction accuracy of the part-whole relation is not as high as the LSTM and GRU models. Compared with the LSTM and GRU models, the Bi-MGU model has a higher recall and *F*-Score in extracting causality, part-whole and possession relations. However, the recall rate of extracting the others relation is not as high as the LSTM and GRU models. These three models have similar *F*-Score when extracting causality, responsibility, possession relations. In summary, the Bi-MGU model we propose has a good performance. This is because the Bi-MGU model has the simplest structure. As the length of the sequence increases, the Bi-MGU is more likely to achieve the desired result in a shorter period. In

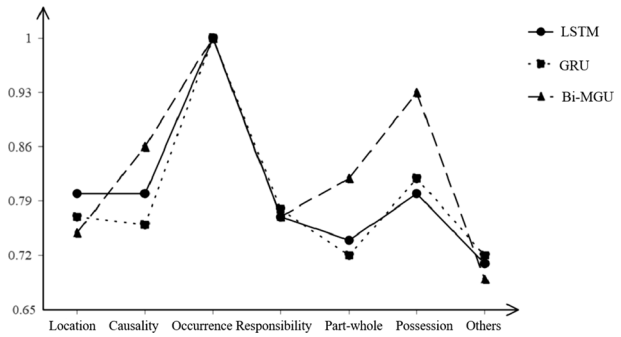

Fig. 5 Comparison on recall of different relation extraction approaches, where each type of relation is used as a data point to tease out the effectiveness of the proposed model

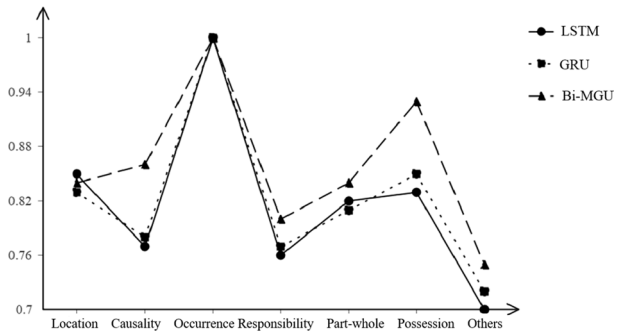

Fig. 6 Comparison on *F*-Score of different relation extraction approaches, where each type of relation is used as a data point to tease out the effectiveness of the proposed model

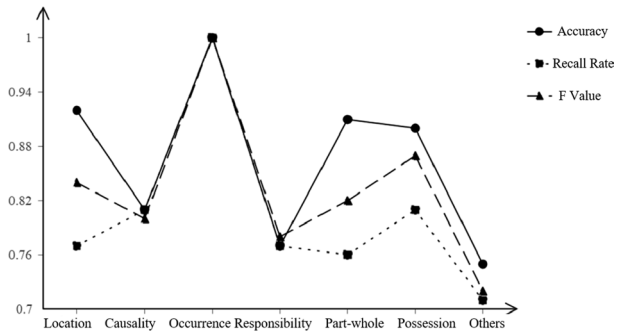

Fig. 7 Comparison on average accuracy, recall and *F*-Score

addition, the reverse layer allows the Bi-MGU model to process the above information. This not only exploits richer semantic information, but also makes full use of contextual information.

Next, we compare the average extraction accuracy, recall and *F*-Score of different types of relations. From Fig. 7, the occurrence relation has a good performance for each model. By analysing the corpus, it is found that the point with the occurrence relation has high-frequency vocabularies, such as coal mines and accidents. At the same time, the structure of the sentences with the occurrence relation is relatively simple, and the features

are more accurate and reliable. The average extraction accuracy of location, part-whole and possession relations is much higher than the recall. This shows that these three relations are more likely to be misjudged as the remaining relations, and the rest of the relation types are rarely misjudged as these relations. This is because the number of the three relation types in the dataset is small, and occurrence, responsibility and causality relations occur frequently. The average extraction accuracy, recall and *F*-Score of other relations are relatively low, because the location and sentence structure are not fixed, and the concepts within the relations are irregular, leading to the unobvious features.

5 Conclusions

This paper proposed a new relation extraction approach based on Bi-MGU model to extract information from big data to achieve disaster precaution and emergency response during the production of coal mine. This is achieved by adding a back-to-front MGU layer based on original MGU model. The proposed approach does not require complex text features and can capture the global context information by combining the forward and backward features. Based on ACE2005 standard and the marked corpus, the experimental results show that our approach outperforms the existing initiatives in terms of training time, accuracy, recall and *F*-Score.

This paper is mainly based on the pre-defined relations, and has a limited ability to extract undefined relations in text. In the future, we will focus on the research of open relation extraction. In addition, an open relation is determined by the core verb, and there exists the problem of polysemy. The problem leads to semantic uncertainty of the extracted relation in the practical application. Therefore, how to disambiguate the relation is another our future work.

Abbreviations

MGU: Minimal gated unit; IIoT: Industrial Internet of Things; RNNs: Recurrent Neural Networks; Bi-MGU: Bidirectional minimal gated unit; ACE2005: 2005 Automatic Content Extraction; LSTM: Long short-term memory network; GRU: Gated recurrent unit.

Acknowledgements

Not applicable.

Authors' contributions

XL proposed the idea and revised this paper. XL, SH and ZQ wrote the manuscript and participated in the experiment. SL and JZ gave some suggestions and participated in the paper revision. All authors have contributed to this research work. All authors read and approved final manuscript.

Funding

This work was supported by National Key R&D Program of China (No. 2018YFC0830202); was granted by Qin Xin Talents Cultivation Program, Beijing Information Science and Technology University (2020); financially supported by Construction project of innovative scientific research platform for edge computing (No.2020KYNH105).

Availability of data and materials

Data sharing not applicable to this article as no data sets were generated or analysed during the current study.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

[1] Beijing Advanced Innovation Center for Materials Genome Engineering, Beijing Information Science and Technology University, Beijing 100101, People's Republic of China. [2] Laboratory of Data Science and Information Studies, Beijing Information Science and Technology University, Beijing 100101, People's Republic of China. [3] China National Software, Beijing 100081, People's Republic of China. [4] School of Economics and Management, Beijing Information Science and Technology University, Beijing 100192, People's Republic of China.

Received: 18 September 2020 Accepted: 2 March 2021
Published online: 16 March 2021

References

1. S. Vitturi, C. Zunino, T. Sauter, Industrial communication systems and their future challenges: next-generation Ethernet, IIoT, and 5G. Proc. IEEE **107**(6), 944–961 (2019)
2. Y. Chen, N. Zhang, Y. Zhang, X. Chen, Dynamic computation offloading in edge computing for internet of things. IEEE Internet Things J. **6**(3), 4242–4251 (2019)
3. C.K. Hwang, J. Cha, K. Kim, H. Lee, Application of multivariate statistical analysis and a geographic information system to trace element contamination in the Chungnam Coal Mine area, Korea. Appl. Geochem. **16**(11), 1455–1464 (2001)
4. L. Wang, X. Zhang, T. Wang, S. Wan, G. Srivastava, S. Pang, L. Qi, Diversified and scalable service recommendation with accuracy guarantee. IEEE Trans. Comput. Soc. Syst. (2020). https://doi.org/10.1109/TCSS.2020.3007812
5. L. Qi, C. Hu, X. Zhang, M.R. Khosravi, S. Sharma, S. Pang, T. Wang, Privacy-aware data fusion and prediction with spatial-temporal context for smart city industrial environment. IEEE Trans. Ind. Inf. (2020). https://doi.org/10.1109/TII.2020.3012157
6. L. Wang, X. Zhang, R. Wang, C. Yan, H. Kou, L. Qi, Diversified service recommendation with high accuracy and efficiency. Knowl.-Based Syst. (2020). https://doi.org/10.1016/j.knosys.2020.106196
7. L. Xue, C. Zhang, H. Ling, X. Zhao, Risk mitigation in supply chain digitization: system modularity and information technology governance. J. Manag. Inf. Syst. **30**(1), 325–352 (2013)
8. O. Ozturk, An ontology based approach for knowledge representation in oil. Gas Min. Ind. **12**(2), 147–158 (2019). https://doi.org/10.17671/gazibtd.469637
9. Y. Cui, Application and research on shared ontology model of coal mine emergency cases, in *Proceedings of the AASRI International Conference on Industrial Electronics and Applications*, pp. 113–121 (2015). https://doi.org/10.2991/iea-15.2015.99
10. Y. Chen, N. Zhang, Y. Zhang, X. Chen, W. Wu, X. Shen, TOFFEE: task offloading and frequency scaling for energy efficiency of mobile devices in mobile edge computing. IEEE Trans. Cloud Comput. (2019). https://doi.org/10.1109/TCC.2019.2923692
11. W. Zhong, X. Yin, X. Zhang, S. Li, W. Dou, R. Wang, L. Qi, Multi-dimensional quality-driven service recommendation with privacy-preservation in mobile edge environment. Comput. Commun. **157**, 116–123 (2020)
12. Y. Chen, Y. Zhang, Y. Wu, L. Qi, X. Chen, X. Shen, Joint task scheduling and energy management for heterogeneous mobile edge computing with hybrid energy supply. IEEE Internet Things J. **7**(9), 8419–8429 (2020)
13. B. Subedi, J. Yunusov, A. Gaybulayev et al., Development of a low-cost industrial OCR system with an end-to-end deep learning technology. IEMEK J. Embed. Syst. Appl. **15**(2), 51–60 (2020). https://doi.org/10.14372/IEMEK.2020.15.2.51
14. G. Zhou, J. Wu, C. Zhang, Z. Zhou, Minimal gated unit for recurrent neural networks. Int. J. Autom. Comput. **13**(3), 226–234 (2016)
15. H. Gan, L. Guo, S. Xia, T. Wang, A hybrid safe semi-supervised learning method. Expert Syst. Appl. **149**, 113–295 (2020)
16. C.W. Johnson, Y. Ben-Zion, H. Meng, F. Vernon, Identifying different classes of seismic noise signals using unsupervised learning. Geophys. Res. Lett. **47**(15), e2020GL088353 (2020). https://doi.org/10.1029/2020GL088353
17. R. Glauber, D.B. Claro, A systematic mapping study on open information extraction. Expert Syst. Appl. **112**, 372–387 (2018)
18. D.T. Vo, E. Bagheri, Open information extraction. Encycl. Semant. Comput. Robot. Intell. **1**(1), 2529–7376 (2017)
19. F. Li, M. Zhang, G. Fu, D. Ji, A neural joint model for entity and relation extraction from biomedical text. BMC Bioinformatics **18**(1), 198–198 (2017)
20. 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**, 59–66 (2017)
21. H. Zhou, H. Deng, L. Chen, Y. Yang, C. Jia, D. Huang, Exploiting syntactic and semantics information for chemical–disease relation extraction. Database **16**, baw048 (2016)
22. S. Wu, Y. He, Enriching pre-trained language model with entity information for relation classification, in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 2361–2364 (2019).
23. X. Zeng, D. Zeng, S. He et al., Extracting relational facts by an end-to-end neural model with copy mechanism, in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pp. 506–514 (2018)
24. Y. Zhang, P. Qi, C.D. Manning, Graph convolution over pruned dependency trees improves relation extraction, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (Association for Computational Linguistics, Brussels, 2018), pp. 2205–2215
25. Z. Guo, Y. Zhang, W. Lu, Attention guided graph convolutional networks for relation extraction, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (Association for Computational Linguistics, Florence, 2019), pp. 241–251
26. H. Zhu, Y. Lin, Z. Liu et al., Graph neural networks with generated parameters for relation extraction, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (Association for Computational Linguistics, Florence, 2019), pp. 1331–1339
27. E. Agichtein, L. Gravano, Snowball: extracting relations from large plain-text collections, in *Proceedings of the ACM International Conference on Digital Libraries*, pp. 85–94 (2000)
28. J.X. Chen, D.H. Ji, Graph-based semi-supervised relation extraction. J. Softw. **19**(11), 2843–2852 (2008)
29. Z. Zhang, Weakly-supervised relation classification for information extraction, in *Proceedings of the 2004 ACM CIKM International Conference on Information and Knowledge Management*, pp. 581–588 (2004)

30. G. Wang, W. Zhang, R. Wang et al. Label-free distant supervision for relation extraction via knowledge graph embedding, in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 2246–2255 (2018)
31. Q. Bing, L. An, Unsupervised Chinese open entity relation extraction. J. Comput. Res. Dev. **52**(5), 1029 (2015)
32. Y. Shinyama, S. Sekine, Preemptive information extraction using unrestricted relation discovery, in *Proceedings of the human language technology conference of the NAACL*, pp. 304–311 (2006)
33. E. Gonzalez, J. Turmo, Unsupervised relation extraction by massive clustering, in *2009 Ninth IEEE International Conference on Data Mining*, pp. 782–787 (2009). https://doi.org/10.1109/ICDM.2009.81
34. O. Etzioni, M. Cafarella, D. Downey, A. Popescu, T. Shaked, S. Soderland, D.S. Weld, A. Yates, Unsupervised named-entity extraction from the web: an experimental study. Artif. Intell. **165**(1), 91–134 (2005)
35. L. Qi, Q. He, F. Chen, X. Zhang, W. Dou, Q. Ni, Data-driven web APIs recommendation for building web applications. IEEE Trans. on Big Data (2020). https://doi.org/10.1109/tbdata.2020.2975587
36. O. Etzioni, M. Banko, S. Soderland, S.D. Weld, Open information extraction from the web. Commun. ACM **51**(12), 68–74 (2008)
37. F. Wu, D.S. Weld, Open information extraction using Wikipedia, in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 118–127 (2010)
38. C. Zhang, Z. Tang, B. Yu, Y. Xie, K. Pan, Deep heterogeneous network embedding based on Siamese Neural Networks. Neurocomputing **388**, 1–11 (2020). https://doi.org/10.1016/j.neucom.2020.01.012
39. X. Li, X. Xu, J. Wang, J. Li, S. Qin, J. Yuan, Study on prediction model of HIV incidence based on GRU neural network optimized by MHPSO. IEEE Access. **8**, 49574–49583 (2020). https://doi.org/10.1109/ACCESS.2020.2979859
40. A. Dong, Z. Du, Z. Yan, Round trip time prediction using recurrent neural networks with minimal gated unit. IEEE Commun. Lett. **23**(4), 584–587 (2019)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com