Relation Extraction for Coal Mine Safety Information Using Recurrent Neural Networks with Bidirectional Minimal Gated Unit

Xiulei Liu
Beijing Information Science and Technology University
Shoulu Hou
Beijing Information Science and Technology University
Zhihui Qin
Beijing Information Science and Technology University
Sihan Liu
China National Software
Jian Zhang (jian.zhang-bistu@yandex.com)
Beijing Information Science and Technology University

Research

Keywords: Relation Extraction, Coal Mine Safety, Recurrent Neural Networks, Bidirectional Minimal Gated Unit

Posted Date: September 23rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-78037/v1

Version of Record: A version of this preprint was published on March 16th, 2021. See the published version at https://doi.org/10.1186/s13638-021-01936-0.

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 *]

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

[*]Correspondence: jian.zhang-bistu@yandex.com

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 on the basis of 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 rate 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 still 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]-[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 representation approaches [8] have 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 labeled 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 approches 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 approches have a poor portability and is 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 approches need to analyze 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 on the basis of 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.

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 labeled 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. 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. [22] designed a Snowball method based on vector representation to reduce the impact of manual intervention

on relation extraction. Chen et al. [23] proposed a semi-supervised extraction model based on graph strategy to improve the accuracy of relation extraction. Zhang et al. [24] proposed the BootProject algorithm based on random feature projection to achieve relation extraction. 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. [25] proposed a Chinese entity relation extraction model based on unsupervised learning, which achieved the relation extraction of large-scale unlabeled data. Shinyama et al. [26] proposed an unsupervised method based on multi-level clustering to achieve relation extraction based on reported articles. Gonzalez et al. [27] 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. [28] 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 [29]. Banko et al. [30] proposed a TextRunner-based approach to extract specific relations from the Web. Wu et al. [31] 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 first layer is the input layer, which is used to preprocess data. First extract the concepts in the sentence and delete the sentences that do not contain the concepts. Then, the data is divided into training data and test data. Next, each piece of data is represented as a <concept1 concept2 word spacing sentence relation type>. Finally, the training data is annotated with the assistance of relevant experts. The second layer is the word vector representation layer, which uses the vector model trained by word2vec to represent the words. Constructing a text word vector converts text information into a vector form. Each sentence is converted into a multidimensional matrix. The text features we use include the word itself and the word spacing. The third layer is a cyclic neural network, and the processed corpus is input into the Bi-MGU unit for training. The relation classification problem can be seen as a judgment of a multi-classification problem. In order to obtain the optimal model, we train the model by minimizing the negative log likelihood function. The fourth layer is the pooling layer, which uses the maximum pool operation to get the final vector representation of this input corpus. In order to make full use of the information of each sentence in the set, we introduce the attention mechanism to calculate the attention probability, so as to reflect the importance of a certain sentence in the set. After the pooling, the overall characteristics of the text are obtained. Finally, calculate new features that combine the overall and local features of the text. The fifth layer is the output layer, and the integrated SoftMax function is used to calculate the predictive relation category of the corpus. We fuse the local features of the text

with the overall features of the text to obtain new features. Then, the merged features are imported into the classifier for classification. Finally, the classification result is output.

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 [32]. 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 $W=\{w_1,w_2,...,w_m\}$, m is the number of words in the sentence S. The text feature set of the sentence S is $K=\{k_1,k_2,...,k_n\}$. n represents number of text features extracted from each sentence. The i-th text feature extracted from the t-th word is expressed as $\mathbf{w}_t^{k_i}(1\le i\le n)$.

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

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

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

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

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

$\mathbf{r}^{k_i}$ is the word vector representation of the i-thfeature. $\mathbf{W}^{k_i}$ is the eigenvector distribution of the i-th feature, $\mathbf{W}^{k_i}=\left(\mathbf{w}_1^{k_i},\mathbf{w}_2^{k_i},...,\mathbf{w}_m^{k_i}\right)$.

Vectorization of each word is connection of each vector. Vectorization of the t-th word is:

$$\mathbf{X}_t=\left[\mathbf{r}_t^w,\mathbf{r}_t^{k_1},\mathbf{r}_t^{k_2},...,\mathbf{r}_t^{k_n}\right] \tag{3}$$

The final text local feature is:

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

3.3 Bi-MGU model

Unidirectional MGU model can only process data in one direction. The proposed the Bi-MGU model to process data in two directions. In Fig. 2, the Bi-MGU neural network has a front-to-back MGU layer which captures the above feature information, and a back-to-front MGU layer which captures the following feature information. Then the global context information can be obtained by combining the forward and backward features. It is helpful for sequence modeling task to consider the global context information. Each training

sequence has two MGU units that move backward and forward, respectively. Both are connected to an output layer.

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

$$\vec{\mathbf{h}}_t = H\left(W_{x\vec{h}_t} x_t + W_{\vec{h}\vec{h}} \vec{\mathbf{h}}_{t-1} + b_{\vec{h}}\right) \tag{5}$$

$\vec{\mathbf{h}}_t$ is the state of the hidden layer from the front to the back at time t. $\vec{\mathbf{h}}_{t-1}$ 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. $W_{x\vec{h}_t}$ and $W_{\vec{h}\vec{h}}$ are weight matrices. $b_{\vec{h}}$ is bias term.

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

$$\overleftarrow{\mathbf{h}}_t = H\left(W_{x\vec{h}_t} x_t + W_{\overleftarrow{h}\overleftarrow{h}} \overleftarrow{\mathbf{h}}_{t+1} + b_{\overleftarrow{h}}\right) \tag{6}$$

$\overleftarrow{\mathbf{h}}_t$ is the state of the hidden layer from the front to the back at time t. $\overleftarrow{\mathbf{h}}_{t+1}$ 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. $W_{x\vec{h}_t}$ and $W_{\overleftarrow{h}\overleftarrow{h}}$ are weight matrices. $b_{\overleftarrow{h}}$ is bias term.

The cumulative result of the two MGU layers that is input into hidden layer can be calculated as follows.

$$y_t = W_{\vec{h}y} \vec{\mathbf{h}}_t + W_{\overleftarrow{h}y} \overleftarrow{\mathbf{h}}_t + b_y \tag{7}$$

where y_t is the output at time t, and b_y is bias term.

Each node in Fig. 2 is a 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 long short-term memory network (LSTM) with three gated structures and the gated recurrent unit (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 the Fig. 3, we have

$$\mathbf{f}_t = \sigma\left(W_f\left[\mathbf{h}_{t-1}, \mathbf{x}_t\right] + b_f\right) \tag{8}$$

$$\dot{\mathbf{h}}_t = \tanh\left(W_h\left[\mathbf{f}_t \text{ e } \mathbf{h}_{t-1}, \mathbf{x}_t\right] + b_h\right) \tag{9}$$

$$\mathbf{h}_t = \left(1 - \mathbf{f}_t\right) \text{e } \mathbf{h}_{t-1} + \mathbf{f}_t \text{ e } \dot{\mathbf{h}}_t \tag{10}$$

$\mathbf{h}_{t-1}$ and $\mathbf{h}_t$ are the states of the hidden layer at time $t-1$ and t, respectively. $\mathbf{x}_t$ is the input at time t. $\mathbf{f}_t$ is the activation function of the gated structure at time t. $\dot{\mathbf{h}}_t$ is short-term memory item. W_f and W_h are weight matrices. b_f and b_h are bias terms. e is the component-wise product between two vectors.

3.4 Attention mechanism and pooling

In relation extraction, the relation set used for classification differs in the importance of words in sentences. Therefore, we use the word-level attention weight matrix to capture the information associated with the target relations in the sentence. The advantage of the Attention Mechanism is that it can automatically adjust the weights so that the deep learning model can focus on the more important parts of the task goal. Its weight calculation for:

$$\mathbf{a}_t = \frac{\exp(f(\mathbf{y}_t, \mathbf{n}))}{\sum_{k=1}^{l} \exp(f(\mathbf{y}_k, \mathbf{n}))} \tag{11}$$

Among them, $\mathbf{a}_t$ the weight of the vector $\mathbf{m}_t$ automatically calculated in the attention mechanism. f is a function that connects the vector $\mathbf{m}_t$ that needs to be calculated with the vector $\mathbf{n}$ corresponding to the factors that affect the weight l the number of vectors that need to be assigned weights. $\mathbf{a}_t$ use softmax to normalize it. The function f has many forms, the article used is:

$$f(\mathbf{m}_t, \mathbf{n}) = \mathbf{v}_a^T \tanh(W_a \mathbf{y}_t + U_a \mathbf{n}) \tag{12}$$

Among them, $\mathbf{v}_a$ is the weight vector, 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, the details as follows:

$$f(\mathbf{y}_t, \mathbf{n}) = \mathbf{v}_a^T \tanh(W_a \mathbf{y}_t + U_a \mathbf{n}) \tag{13}$$

$$\mathbf{a}_t = \frac{\exp(f(\mathbf{y}_t, \mathbf{n}))}{\sum_{k=1}^{l} \exp(f(\mathbf{y}_k, \mathbf{n}))} \tag{14}$$

$$\mathbf{y} = \sum_{k=1}^{l} \mathbf{a}_t \mathbf{y}_t \tag{15}$$

Among them, $\mathbf{y}_t$ is the output of the t-th step of the hidden layer, $\mathbf{n}$ is the vector corresponding to the factors that affect the weight, l is the sentence length, and $\mathbf{y}$ 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 $\mathbf{F} = \{\mathbf{F}_1, \mathrm{L}\ , \mathbf{F}_m\}$. Then, use the largest pooling operation to get the most significant feature representation.

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

Among them, $\mathbf{d}$ 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 analyzed 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. In addition to the above 6 relations, all relations are all labeled 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 the concepts. Then, each piece of data is denoted by <concept1 concept2 word spacing sentence relation type>.

4.2 Results and discussion

We compare the proposed Bi-MGU approach with the current benchmarks in the literature [32,34], in terms of training time, accuracy, recall rate and F value.

LSTM model is a kind of time-recurrent neural network. Compared with the traditional RNN, it adds a processor called "cell" to determine whether the information is useful. A cell contains input, forget and output gates. If the information is judged to be useless, it is forgotten through the forget gate.

As a variant of LSTM, GRU model [33] is simpler and contains only two gate structures: update gate whose function is similar to forget gate and reset gate whose function is similar to input gate. GRU model removes the output gate of LSTM, and mixes cell state and hidden state.

MGU model is a RNN with minimal gate structure, and it uses only one gate structure [34]. Based on the GRU model, MGU combines reset and forget gates. Compared with LSTM and GRU, MGU has a simpler structure and fewer parameters.

Table 1Comparison on training time of different models

Model Name	LSTM	GRU	Bi-MGU
Training Time (s)	10260	8600	6850

From Table 1, it can be seen that the training time of the LSTM, GRU, and MGU models gradually decreased. This proves that the simpler the model structure and the fewer the training parameters, the less training time is required. Compared with the traditional LSTM

and GRU models, the performance of relation extraction for each type of relations are compared according to the evaluation indicators, as shown in Fig. 4, Fig. 5, and Fig.6, respectively.

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 rate and F value 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 values 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 of time. In 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 rate and F value of different types of relations. The specific results are shown in Fig. 7.

From Fig. 7, the occurrence relation has a good performance for each model. By analyzing 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 rate. 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 rate and F value 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 on the basis of 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 rate and F value.

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

Availability of data and materials

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

Competing interests

The authors declare that they have no competing interests.

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 & Technology University (2020); financially supported by Construction project of innovative scientific research platform for edge computing (No.2020KYNH105).

Authors' contributions

Xiulei Liu proposed the idea and revised this paper. Xiulei Liu, Shoulu Hou and Zhihui Qin wrote the manuscript and participated in the experiment. Sihan Liu and Jian Zhang gave some suggestions and participated in the paper revision. All authors have contributed to this research work. Allauthors have read and approved the final manuscript.

Acknowledgments

Not applicable.

Author details

Not applicable.

References

1. S. Vitturi, C. Zunino, T. Sauter. Industrial communication systems and their future challenges: next-generation Ethernet, IIoT, and 5G. Proceedings of the 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 of Things Journal. 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. Applied geochemistry.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 Transactions on Computational Social Systems. 1-12 (2020).
5. L. Qi, C. Hu, X. Zhang, R. Mohammad, Khosravi, S. Sharma, S. Pang, T. Wang, Privacy-aware Data Fusion and Prediction with Spatial-Temporal Context for Smart City Industrial Environment. IEEE Transactions on Industrial Informatics. 1-1 (2020).
6. L. Wang, X. Zhang, R. Wang, C. Yan, H. Kou, L. Qi, Diversified Service Recommendation with High Accuracy and Efficiency. Knowledge-Based Systems. 204, 106196 (2020).
7. L. Xue, C. Zhang, H. Ling, X. Zhao, Risk Mitigation in Supply Chain Digitization: System Modularity and Information Technology Governance. Journal of Management Information Systems. 30(1), 325-352(2013).
8. O. Ozturk, An Ontology Based Approach for Knowledge Representation in Oil, Gas and Mining Industry. BilişimTeknolojileriDergisi. 12(2), 147-158 (2019).

9. Y. Cui, Application and Research on Shared Ontology Model of Coal Mine Emergency Cases. conference on industrial electronics and applications. 113-121 (2015).
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 Transactions on Cloud Computing. 1-1 (2019).
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. Computer Communications. 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 of Things Journal. 1–1 (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 Journal of Embedded Systems and Applications. 15(2), 51-60 (2020).
14. G. Zhou, J. Wu, C. Zhang, Z. Zhou, Minimal gated unit for recurrent neural networks. International Journal of Automation and Computing. 13(3), 226-234(2016).
15. H. Gan, L. Guo, S. Xia, T. Wang, A hybrid safe semi-supervised learning method. Expert Systems with Applications. 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. Geophysical Research Letters.e2020GL088353(2020).
17. R. Glauber, D. B. Claro, A systematic mapping study on open information extraction. Expert Systems with Applications. 112, 372-387 (2018).
18. D. T. Vo, E. Bagheri, Open information extraction. Encyclopedia with Semantic Computing and Robotic Intelligence.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 (2016).
22. E. Agichtein, L. Gravano, Snowball: Extracting relations from large plain-text collections. acm international conference on digital libraries. 85-94 (2000).
23. J. X. Chen, D. H. Ji, Graph-based semi-supervised relation extraction. Journal of Software. 19(11), 2843-2852(2008).
24. Z. Zhang, Weakly-supervised relation classification for information extraction. conference on information and knowledge management. 581-588 (2004).
25. Q. Bing, L. An, Unsupervised Chinese open entity relation extraction. Journal of computer research and development. 52(5), 1029 (2015).
26. Y. Shinyama, S. Sekine, Preemptive Information Extraction using Unrestricted Relation Discovery. language and technology conference. 304-311 (2006).
27. E. Gonzalez, J. Turmo, Unsupervised Relation Extraction by Massive Clustering. international conference on data mining. 782-787 (2009).
28. O. Etzioni, M. Cafarella, D. Downey, A. Popescu, T. Shaked, S. Soderland, D. S. Weld, Alexander Yates. Unsupervised named-entity extraction from the web: An experimental study. Artificial intelligence. 165(1), 91-134 (2005).
29. L. Qi, Q. He, F. Chen, X. Zhang, W. Dou, Q. Ni, Data-Driven Web APIs Recommendation for Building Web Applications. IEEE Transactions on Big Data. 1-1 (2020).
30. O. Etzioni, M. Banko, S. Soderland, S. D. Weld, Open information extraction from the web. Communications of the ACM. 51(12), 68-74 (2008).
31. F. Wu, D. S. Weld, Open Information Extraction Using Wikipedia. meeting of the association for computational linguistics. 118-127 (2010).
32. C. Zhang, Z. Tang, B. Yu, Y. Xie, K. Pan, Deep heterogeneous network embedding based on Siamese Neural Networks. Neurocomputing. 388, 1-11 (2020).
33. 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).
34. A. Dong, Z. Du, Z. Yan, Round trip time prediction using recurrent neural networks with minimal gated unit. IEEE Communications Letters. 23(4), 584-587 (2019).

Fig. 1 An Illustration on The Network Structure for Relation Extraction. The network structure consists of input layer, embedding layer, Bi-MGU, Max-pooling layer and output layer.

Fig. 2 The proposed Bi-MGU Model. The Bi-MGU neural network includes a front-to-back MGU layer and back-to-front MGU layer, which capture the above and following feature information, respectively.

Fig. 3 An Illustration on The Internal Structure of MGU Unit. MGU model is a RNN and uses only one gate structure.

Fig. 4 Comparison on Accuracy of Relation Extraction. The accuracy in location, causality, occurrence, responsibility, part-whole, possession and other relations of the three models.

Fig. 5 Comparison on Recall Rate of Relation Extraction. The recall rate in location, causality, occurrence, responsibility, part-whole, possession and other relations of the three models.

Fig. 6 Comparison on F Value of Relation Extraction. The F value in location, causality, occurrence, responsibility, part-whole, possession and other relations of the three models.

Fig. 7 Comparison on Average Accuracy, Recall rate and F value. The average accuracy, recall rate and F value on the terms of location, causality, occurrence, responsibility, part-whole, possession and other relations.

Figures

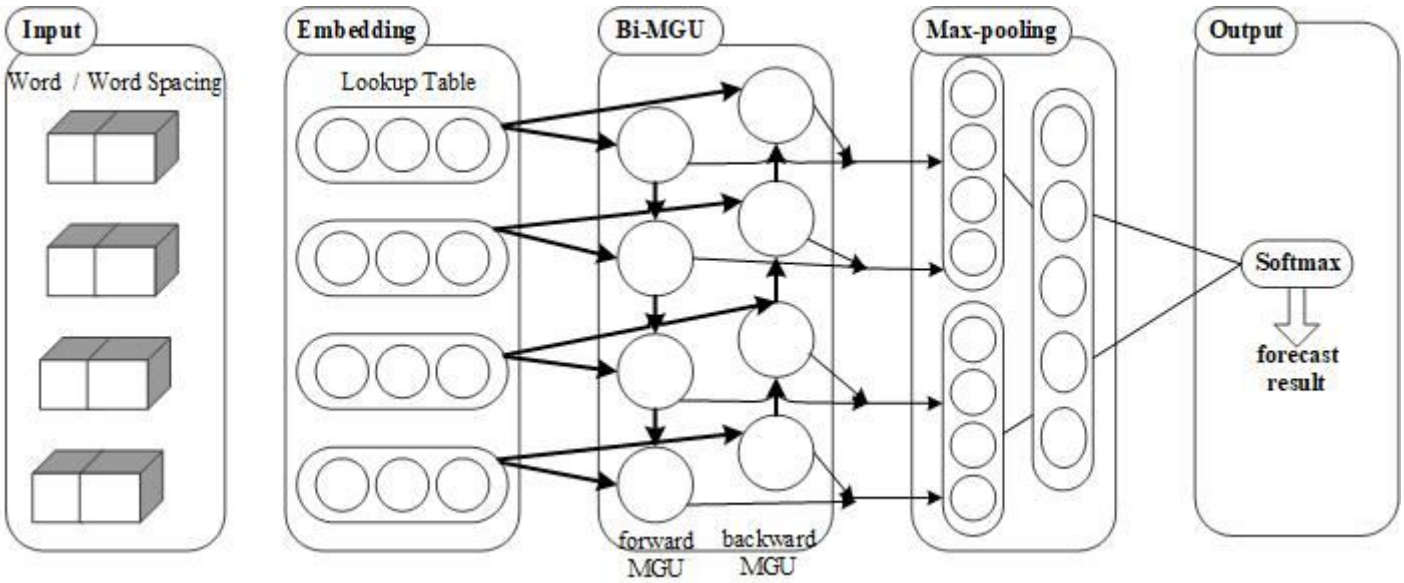

Figure 1

An Illustration on The Network Structure for Relation Extraction. The network structure consists of input layer, embedding layer, Bi-MGU, Max-pooling layer and output layer.

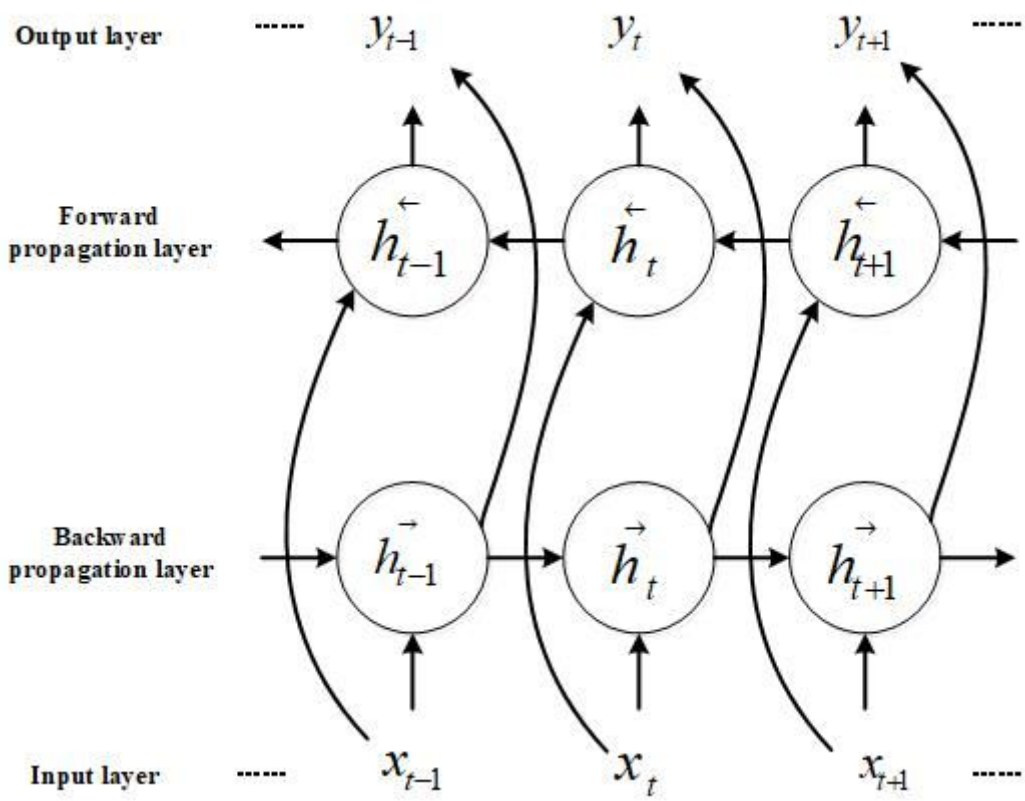

Figure 2

An Illustration on The Network Structure for Relation Extraction. The network structure consists of input layer, embedding layer, Bi-MGU, Max-pooling layer and output layer.

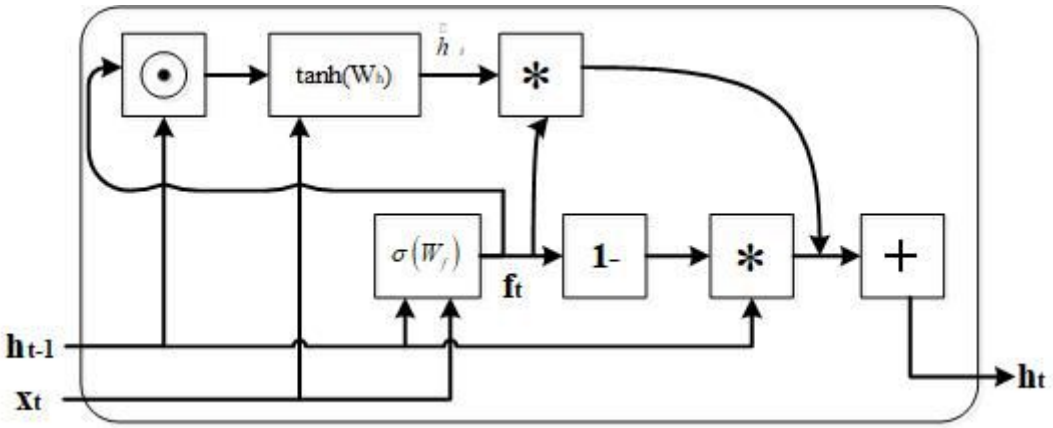

Figure 3

An Illustration on The Internal Structure of MGU Unit. MGU model is a RNN and uses only one gate structure.

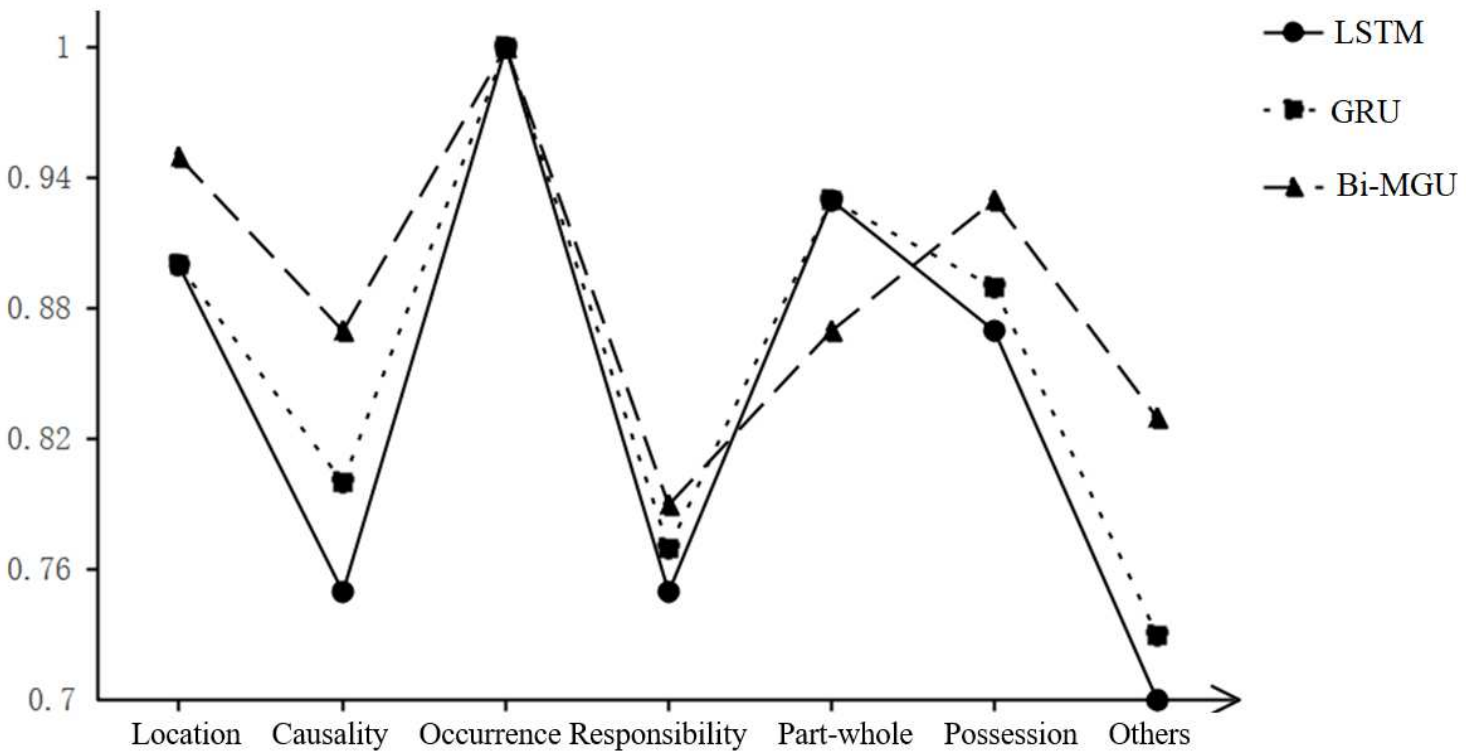

Figure 4

Comparison on Accuracy of Relation Extraction. The accuracy in location, causality, occurrence, responsibility, part-whole, possession and other relations of the three models.

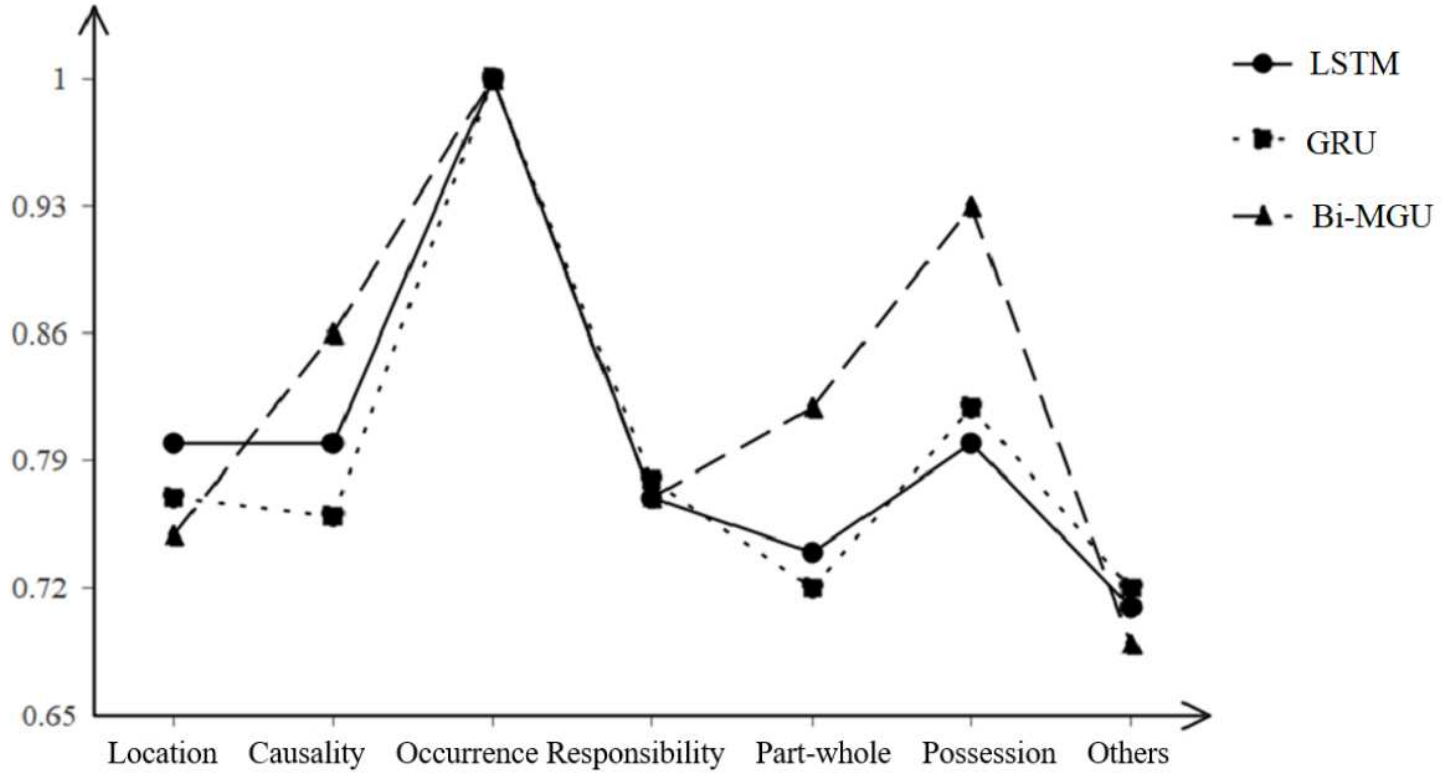

Figure 5

Comparison on Recall Rate of Relation Extraction. The recall rate in location, causality, occurrence, responsibility, part-whole, possession and other relations of the three models.

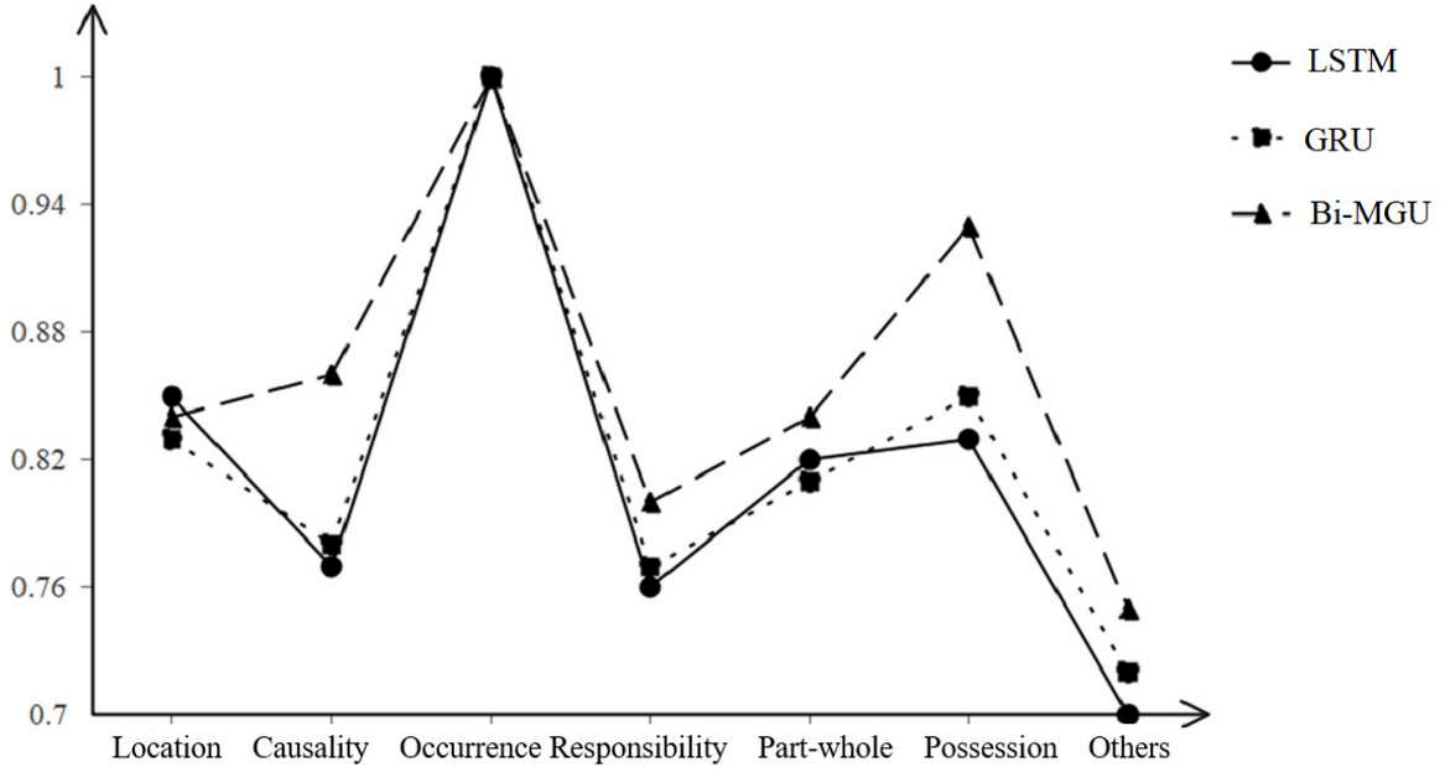

Figure 6

Comparison on Recall Rate of Relation Extraction. The recall rate in location, causality, occurrence, responsibility, part-whole, possession and other relations of the three models.

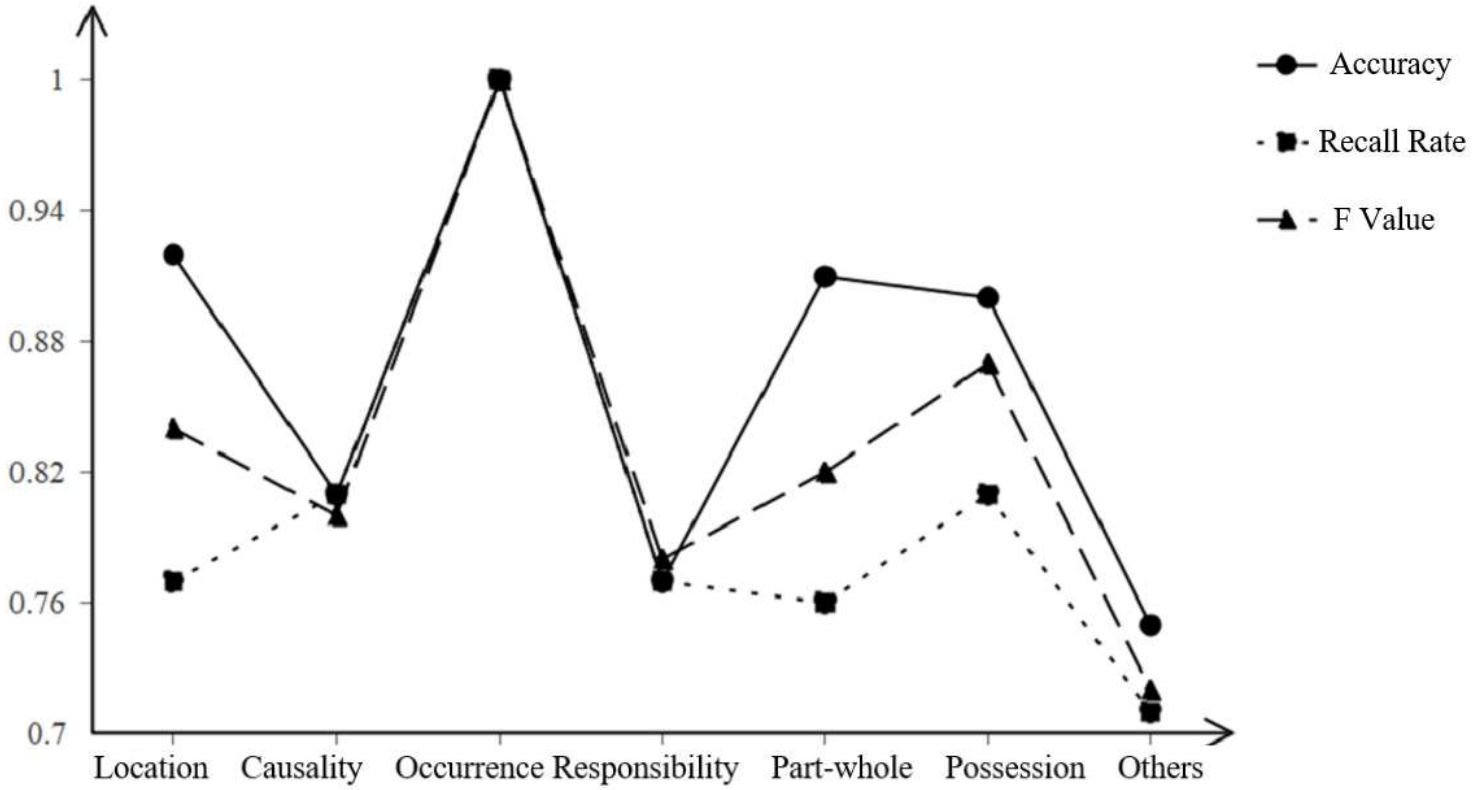

Figure 7

Comparison on Average Accuracy, Recall rate and F value. The average accuracy, recall rate and F value on the terms of location, causality, occurrence, responsibility, part-whole, possession and other relations.