

Article

An Entity Relation Extraction Method Based on Dynamic Context and Multi-Feature Fusion

Xiaolin Ma , Kaiqi Wu *, Hailan Kuang * and Xinhua Liu 

Hubei Key Laboratory of Broadband Wireless Communication and Sensor Networks, School of Information Engineering, Wuhan University of Technology, Wuhan 430070, China; maxiaolin0615@whut.edu.cn (X.M.); liuxinhua@whut.edu.cn (X.L.)

* Correspondence: 252764@whut.edu.cn (K.W.); kuanghailan@whut.edu.cn (H.K.)

Abstract: Dynamic context selector, a kind of mask idea, will divide the matrix into some regions, selecting the information of region as the input of model dynamically. There is a novel thought that improvement is made on the entity relation extraction (ERE) by applying the dynamic context to the training. In reality, most existing models of joint extraction of entity and relation are based on static context, which always suffers from the feature missing issue, resulting in poor performance. To address the problem, we propose a span-based joint extraction method based on dynamic context and multi-feature fusion (SPERT-DC). The context area is picked dynamically with the help of threshold in feature selecting layer of the model. It is noted that we also use Bi-LSTM_ATT to improve compatibility of longer text in feature extracting layer and enhance context information by combining with the tags of entity in feature fusion layer. Furthermore, the model in this paper outperforms prior work by up to 1% F1 score on the public dataset, which has verified the efficiency of dynamic context on ERE model.

Keywords: mask selector; relation extraction; attention mechanism; threshold; multi-feature fusion



Citation: Ma, X.; Wu, K.; Kuang, H.; Liu, X. An Entity Relation Extraction Method Based on Dynamic Context and Multi-Feature Fusion. *Appl. Sci.* **2022**, *12*, 1532. <https://doi.org/10.3390/app12031532>

Academic Editor: Christos Bouras

Received: 11 January 2022

Accepted: 27 January 2022

Published: 31 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Nowadays, accompanying a boom in the Internet, massive amounts of heterogeneous information appear in our lives in the form of news articles, emails, blogs and Q&A. There is a large amount of information on science and technology security in these non-structural data. How to effectively analyze the important data by information technology, help the related experts obtain the significant information, learn the risk level of hot events, and clarify whether their occurrence can impact the national scientific and technological security has become an urgency suggesting being solved quickly. Therefore, it is necessary to master the method–information extraction technology, which can help experts gain the wanted information quickly and accurately. Entity relation extraction is an important branch of information extraction technology, the concrete task of which is to infer the semantic relationship between entity pairs from a given text corpus on the basis of semantic information. Entity relation extraction, a foundation for natural language processing, has been widely applied to information processing, automatic Q&A, automatic summarization and other fields, which has seen some initial achievements. With the rapid development of deep learning, researchers have applied neural networks to entity relation extraction tasks, which has brought some new breakthroughs for the mission. In other words, relation extraction not only has theoretical significance, but also has wide application prospect.

The methods of entity relation extraction are mainly divided into methods based on semantic rules, methods based on machine learning, and methods based on deep learning. In recent years, according to the research trend, it can be seen that the entity relation extraction method based on deep learning has become the mainstream idea. There have been many models of entity relation extraction based on deep learning, such as convolutional neural networks, recurrent neural networks, and Transformer. These models

can be used for relation extraction with some appropriate modifications. The convolutional neural networks were first applied to relation classification tasks in the paper [1]. Although its network structure was very simple, it was significantly improved compared to non-neural network methods. The bidirectional recurrent neural networks were used to handle sentences and capture more features in the paper [2] that detected word-level co-text information between entity pairs, because of which the final relation extraction effect was also improved. In addition, the paper [3] made use of the Bi-LSTM model in the relation extraction task, and achieved good results. It utilized attention mechanism to reduce the influence of words noisy in the sentence, and enhance the influence of keywords. At present, an end-to-end approach is often adopted for entity relation extraction tasks, which means identifying named entities and extracting semantic relations at the same time. The training methods are divided into pipeline style and joint learning. The pipeline learning method refers to the extracting relation of entities on the basis of the entity recognition. The joint learning method is mainly based on the end-to-end model of the neural network, sharing the word embedding representation, and jointly optimizing the entity recognition and ERE.

Existing models mostly rely on static context methods. A single vector representation is selected as the context of the sentence in the feature extraction layer, which leads to the problem of missing context, thereby limiting the effect of relation extraction. To address the problem, the model in this paper extends SPERT [4] model to achieve entity and relation joint task.

The main contributions of this paper can be summarized as follows:

- This paper proposes an entity relation extraction method based on dynamic context that can address the text context missing issue of existing models. With the help of dynamic context, the proposed method can efficiently extract context representation of sentences, yielding a wonderful result compared with prior studies on public datasets.
- This paper designs a threshold selector, which can (1) dynamically choose text context area with entity pairs position for sentences, and (2) suppress the negative influence of adjacent entity pairs.
- This paper proposes a Bi-LSTM combined with self-attention (Bi-LSTM_ATT). By using Bi-LSTM_ATT, learning long-term dependencies, and focusing on significant region of sentence, we greatly improve the compatibility for longer texts, while maintaining the effect of extraction for normal sentences.
- This paper uses multi-feature fusion (MFF) to enhance context representation. The entity tags are added into feature fusion layer, which is combined with others and fed into relation extraction module. Our experiments show the high performance of the proposed method on CoNLL04 and ADE.

2. Related Work

In ERE mission, traditional methods achieve entity detection and relation extraction subtasks through separate models. Li et al. [5] propose a joint model based on artificial feature extraction, which performs the subtasks of named entity recognition and relation extraction at the same time. The methods rely on the use of NLP tools or the characteristics with manual design, which leads to additional complexity. Neural networks have been used to solve the problem of feature design, which usually involve RNNs and CNNs.

Specifically, Miwa and Bansal [6] and Li et al. [7] use bidirectional tree structure RNNs to obtain grammatical information in different contexts. Guo et al. [8] propose a novel method combining CNN and RNN to extract text features and classify relations, then adopt word-level and sentence-level attention to enhance performance. Adel and Schütze [9] achieve entity classification instead of named entity recognition, and the context around the entity and entity pairs are input to the relation extraction layer. Katiyar and Cardie [10] carefully learn recurrent neural networks, but ignore that the relation labels were not mutually exclusive. Bekoulis et al. [11] extract one relationship at a time by joint model based on LSTM, which increased the complexity of the named entity recognition. However, the paper [12] takes joint entity recognition and relation extraction task as

a multi-head selection problem using CRF (Conditional Random Fields) layer, without extracting feature manually. Li et al. [13] regards relation extraction task as a multi-turn Q&A problem which transforms the extraction to the mission of selecting correct spans from sentence. Geng et al. [14] propose a method based on long short-term memory to extract feature. The method relies on the syntactic dependency analysis tree to enhance extraction performance, and the bidirectional LSTM combining attention is used to obtain word-level information such as the position of entity pair in sentence. Then these feature representations are fused to classify relation. Xiao et al. [15] use a hybrid deep neural network model for ERE tasks to solve the noisy labeling problem. The model capable of filter out noisy, includes embedding module based on Transformer that is used to suppress negative influence of noisy data, an LSTM-based entity detection module and a relation extraction module founded on reinforcement learning.

In addition, a representation iterative fusion based on heterogeneous graph neural networks for ERE [16] is proposed, within which relations and words are taken as nodes on the graph. Meanwhile, to obtain nodes representation suitable for ERE, both of semantic nodes are iteratively fused via the message transmission mechanism. Zhang et al. [17] used the Bi-LSTM model to extract relations, combining with the information around the current word, which verified the effectiveness of Bi-LSTM on relation extraction. Fan et al. [18] apply attention-based Bi-LSTM network to ERE mission concerned with the usage of dietary supplements in clinical text, which achieves a fine performance. Cui et al. [19] introduce a hybrid Bi-LSTM siamese network using two word-level Bi-LSTM to learn similarity of sentences and classify relations.

In the past few years, the attention mechanism has been seen in the paper of natural language processing, playing an important role. The attention mechanism was first applied in the field of image processing. There are many studies in this field and a long history. The attention mechanism has applied in many image processing models, which give more attention to the target area needing to be focused and adjust the focus. On the natural language processing, Bahdanau et al. [20] applied the attention mechanism to machine translation tasks for the first time, achieving good results. Shang et al. [21] proposed a novel ERE model to settle the existing Transformer's difficulty to identify relational patterns in sentences, within which a special self-attention network is adopted to detect relational pattern in form of end-to-end. Christou et al. [22] put forward a ERE approach based on Transformer to obtain a extensive relation set via entity instance and label by using pre-trained BERT.

With the in-depth study of deep learning in natural language processing, more and more researchers have begun to concentrate on the fusion of various feature information on text. Han et al. [23] propose a Chinese named entity recognition model based on multi-feature adaptive fusion. The model uses a bidirectional long and short-term memory neural network to extract strokes and basic features, and adaptively fuse two sets of features with weighted cascade method, which work well to improve the entity recognition ability of model. Being motivated by the finds: the event attributes can be used for relation classification, a novel method [24] for ERE is came up with, with which representation and extraction of the feature are remodeled. As usual, it also utilizes the bidirectional long short-term memory to learn relation of entity pairs and attention mechanism to focus on more information. To improve performance of ERE tasks, Zhang et al. [25] adopt a multi-feature fusion model based on character embedding, integrating a amount of word-level feature into neural network. Additionally, Bekoulis et al. [26] use adversarial training with adding small perturbations in initial data to improve the performance and robustness of relation extraction model.

It is noted that more and more scholars concentrate on the span-based method that is utilized to detect all entity candidates for the ERE tasks. In particular, the method has been used for ERE by Dixit et al. [27], with which they use fixed ELMo for contextual word embedding and character embedding through Bi-LSTM over it. Luan et al. [28] introduce a DyGIE model for information extraction tasks, which can capture span representation by

constructed span graphs dynamically. Eberts et al. [4] propose SpERT to improve the effect of ERE tasks, which takes BERT as word-embedding and detects all entity candidates produced by span-based distribution. Specially, SpERT adopts negative sampling to improve its robustness.

3. Methodology

SPERT-DC extends the joint entity recognition and relation extraction model based on SPERT in structure, whose framework is shown in Figure 1. It performs entity relation extraction subtasks while completing the named entity recognition. The model is divided into word embedding layer, entity recognition layer, and relation classification layer. The word embedding layer convert the text sentence to vector with BERT to get the representation of text $(e_1, e_2, \dots, e_n, \text{cls})$, where the last token cls represents the word vector of the special token CLS. All token subsequences among sentence are used for detecting entities in the entity recognition layer. For example, the sentence (you are the most beautiful) maps to the subsequences (you), (you, are), (the, most, beautiful), etc. For possible entities, its representation, width embedding and CLS context are input into entity recognition module. At last, the entity type is obtained by softmax classifier. The relation is extracted from some entity pairs in relation classification layer, which also relies on text context. The model chooses the local context between the entity pair instead of the token CLS as text context. Then the entity pair representation is input to the relation extraction module, and relation class is obtained by activation of sigmoid.

The baseline creates a static mask matrix with the start and end position of entity pair in sentence, where $\text{mask}[\text{start}:\text{end}] = 1$, with which the baseline selects the corresponding area of sentence representation as text context. It can be seen that the static mask only focus on the limited information between entity pair, which will lead to lack of context when the distance between entity pair is small.

This paper propose SPERT-DC to solve the limitation of baseline on selecting text context for the research of relation extraction. According to position of entity, the method corrects text context with dynamic mask to improve the relation extraction effect of model at last.

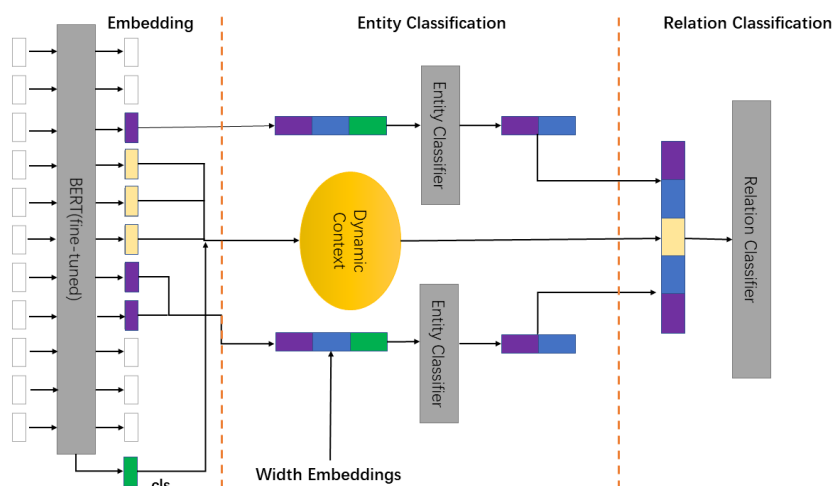


Figure 1. There are three sections: embedding, entity classification and relation classification. Then, all words are converted to token representation through BERT. All entities in the sentence are detected. All entity pairs fusing their dynamic context are classified into relations.

3.1. Joint Entity and Relation Extraction Model

The pre-trained BERT model in this paper is utilized to perform *embedding* in word Embedding layer, whose character set is used to convert text into vector, yielding the

representation of sentence doc_k . What's more, the feature will be used in entity recognition and relation extraction model. It is formalized as follow:

$$doc_k = (e_1, e_2, \dots, e_n, cls) = Embedding_{BERT}(w_1, w_2, \dots, w_n). \quad (1)$$

Among the equation, cls refers to handle vector of sentence. Then all possible entities will be detected with various windows in the subsequences, each of which can be expressed as:

$$entity_k = c(e_i, e_{i+1}, \dots, e_{i+k}), \quad (2)$$

$$ent_k = MaxPooling(entity_k). \quad (3)$$

Then the length embedding w_k and the word vector CLS are input into the fusion function f to obtain the final representation which is fed into softmax classifier layer, yielding entity type at last. The formula is as follow:

$$x^k = f(ent_k, w_k, cls), \quad (4)$$

$$y^k = softmax(W^k * x^k + b^k). \quad (5)$$

Not only does it extract entity feature, but also relies on the classification of the text context in relation extraction module. Generally, the CLS representation is often chosen as text context. The baseline selects the text representation which is handled by max-pooling between entities instead. It is expressed as $c(e_1, e_2)$, and the entity pair can be shown as follow:

$$rel = f(ent_1, ent_2, c(e_1, e_2)). \quad (6)$$

Then, it is input to the linear layer and activated by sigmoid. The sum of the loss of entity recognition L^e and the loss of relation extraction L^r calculate the loss of model, which is as follow:

$$y^r = \sigma(W^r * rel + b^r), \quad (7)$$

$$Loss = L^e + L^r. \quad (8)$$

In the relation classification process, a threshold δ is defined to compare with y^r to confirm the entity pair triplet at last.

3.2. A Method For Selecting Context Based on Threshold

The core of the method for relation extraction based on threshold is how to select text context correctly. The CLS representation of sentence or local context between entity pair can be chosen as text context. The single static text context is always used in the most of existing models, which often select CLS or local context between entity pair. There is no doubt that the local context between entity pair is applied in the baseline. Although the local context performs better than CLS representation in the baseline, it leads to the phenomenon of missing of text context. Therefore, this paper chooses corresponding text context for entity pair dynamically with a method based on dynamic context, which draws on the idea of the threshold classification. This paper calculate the ratio r of the length of sentence and the distance of entity pair $content_size$ obtained by the start and end positon, which compares ratio r with threshold T to create mask matrix. The text representation comes into being with the combination of mask and H, which is expressed as:

$$r = \frac{abs(end - start)}{content_size}, \quad (9)$$

$$f(r) = \begin{cases} CLS, & r \leq T \\ MID, & r > T \end{cases} \quad (10)$$

Among the formula above, the *MID* indicates selecting local context between entity pair, otherwise selecting CLS representation. The flowchart of the method is shown in Figure 2.

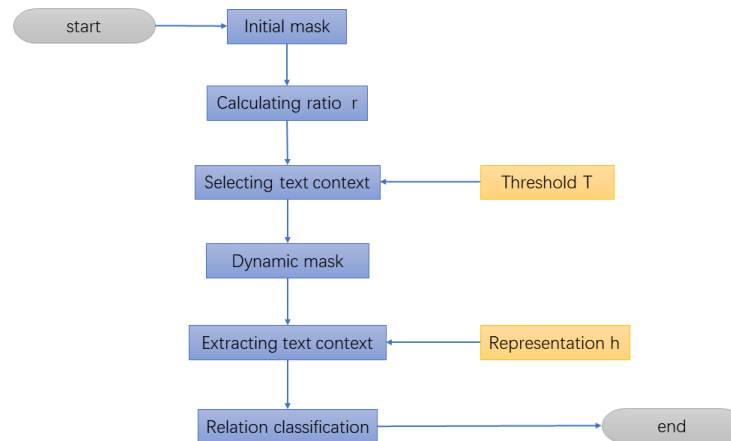


Figure 2. The flow chart of method based on threshold. This method is applied in dynamic context module for relation extraction mission.

3.3. A Text Feature Extraction Network Based on Bi-LSTM_ATT

In the previous section, this paper takes the local context between entity pair or the CLS representation as text context. In fact, it can be seen that the effect of extraction decreases as the length of sentence increases in experiment process. Comparing with the whole, the longer texts have a poorer performance, that is, the compatibility of model for longer texts needs to be improved. Bi-LSTM, a special kind of RNN, is a two-way LSTM network, capable of learning long-term dependencies, which is suitable for modeling long texts and learning context features. In order to improve the compatibility of model, this paper proposes Bi-LSTM fused with self-attention mechanism (Bi-LSTM_ATT). The network help to learn the context around long texts. In the meantime, it maintains the relation extraction effect of the short text to advance the robustness of the model. Different from the conventional self-attention module, this paper focus more on the the local context representation between entity pair, which in order to improve the model's performance by learning local context and enhancing text feature extraction.

The structure of the *Bi-LIST_ATT* is shown in Figures 3 and 4. The selected area of text context is determined by martrix output by the threshold selector, which is converted into a mask matrix. The sentence embedding is fed into Bi-LSTM network in order to learn long-term dependencies and yield text representation. The combination of mask and text representation is input into maximum pooling layer to obtain text context representation, which fuses length embedding and is fed into self-attention layer together for next operation. The process can be formalized as:

$$D_k = [d_1, d_2, \dots, d_n] = BiLSTM([l_1, l_2, \dots, l_n]), \quad (11)$$

$$DC_k = f(D_k, mask), \quad (12)$$

$$w_{ik} = \frac{Q_i K_k^T}{\sqrt{d_k}}, \quad (13)$$

$$S_i = \sum_{k=1}^N w_{ik}(D_k, len_k). \quad (14)$$

Among the formulas above, the d_k indicates the dimensions of Q , K and V , which is used for reducing the dot product QK^T . It is aimed to prevent the model from difficulty of convergence when the dot product is too large.

3.4. Enhancing Context Model Based on MFF

The Bi-LSTM_ATT is proposed to improve the model’s compatibility for longer texts in the previous section. In order to further enhance model, multi-feature fusion (MFF) method is utilized to advance context information representation, which fuses the previous text context and the entity tag. In fact, resulting from their coming from different domains, there is a no direct fusion. The diagram is shown in Figure 3.

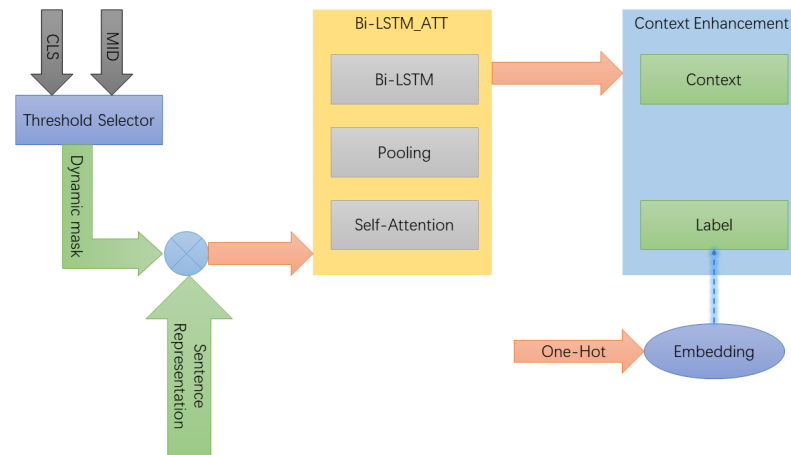


Figure 3. There are three sections in dynamic context module: Threshold Selector, Bi-LSTM_ATT and Context Enhancement. Then, all context are dynamically determined by selector. The Bi-LSTM_ATT is used for improving compatibility. Fusing entity tag is to enhance context representation.

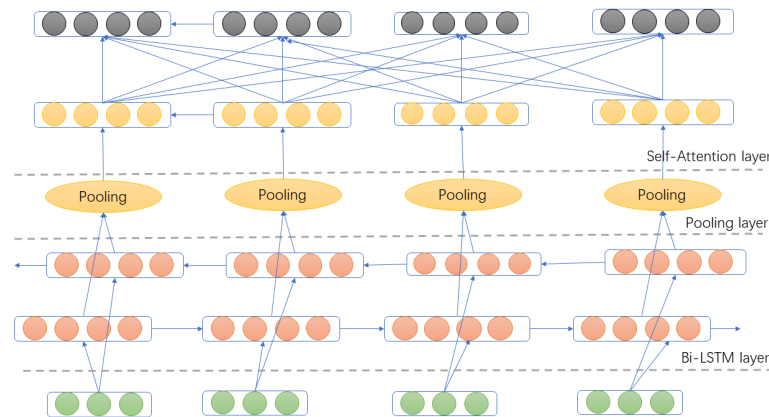


Figure 4. This is the framework of the Bi-LSTM_ATT module, which mainly contains three neural network layers: Bi-LSTM layer, Pooling layer and Self-Attention layer.

The all entities tag are fed into the module in the form of one-hot, whose representation is obtained through the tag embedding layer. According to entity position, the corresponding representation is extracted from them. These feature representation such as entity, entity width, text context and entity tag, is input into relation extraction module. The process can be expressed as:

$$label_i = Embedding([0, i, 0 \dots, 0]), \tag{15}$$

$$label = (label_h, label_t) = Index(label_i, index_{head}, index_{tail}), \tag{16}$$

$$rel_{input} = f((ent_h, ent_t), wid, ctx, label). \tag{17}$$

Among the formulas above, the Index function is position indicator. The wid infers to the width embedding. The ctx is the text context, and the f is a fusion function.

4. Experiments

4.1. Dataset and Evaluation

For the proposed method in this work, it is described with lots of detail, which needs to be proved the effectiveness on relation extraction tasks. Thus, this paper utilizes the CoNLL04 dataset and the ADE dataset for training and evaluation process. Furthermore, the two datasets contain plenty of entities composed of English words and entity-relation-entity tuple, which are widely adopted by models focusing on English entity relation extraction tasks.

In this paper, a series of experiments are carried out on English sentences from the CoNLL04 dataset and the ADE dataset to evaluate the significant superiority and performance of the proposed method compared with the existing results. The detail of dataset is shown in Table 1 below:

Table 1. The detail in some datasets. A series of experiments are carried out on them.

Dataset	Train	Dev	Test
CoNLL04	922	231	288
ADE	2563	804	905

The CoNLL04 dataset contains texts with named entities and relations extracted from news articles, which has been applied for the research of entity relation extraction in recent years. It includes four entity classes, such as Location, Organization, People and Other, and five relation types, such as Work-For, Kill, Organization-Based-In, Live-In and Located-In.

The ADE dataset comes from some medical reports describing adverse drug reactions, including a single relation type Adverse-Effect and the two entity types Adverse-Effect and Drug.

There are some evaluation indicators for experiments such as precision, recall and F1-scores. To compare our method with other researches, this paper adopts the micro-average F1-scores as evaluation metric in the following experiments. The computing method of F1-scores is shown:

$$pre = \frac{TP}{TP + FP'} \quad (18)$$

$$rec = \frac{TP}{TP + FN'} \quad (19)$$

$$F1 = \frac{2 \times pre \times rec}{pre + rec} \quad (20)$$

4.2. Experimental Settings

This paper uses English cased version of BERT as word embedding. The batch size for training is set to 2 and testing is set to 1. The dropout of model is set to 0.1 and width embedding layer count to 25. For the Bi-LSTM_ATT module, it contains 4 layers LSTM and one max-pooling layer, then the input size is set to 768 and dropout to 0.5. The span width threshold is same as the baseline, which is set to 10.

4.3. Overall Results

In this paper, a series of experiments are carried out on the CoNLL04 dataset and the ADE dataset. The data is randomly divided. Sixty percent of them is used as training part, so the development and test part share the remains. To verify the effect of proposed method based on dynamic context, this paper explores performance of some similar model.

As shown in Table 2, this part makes a comparison with some competing results to evaluate whether the dynamic context has an enhancement to relation extraction model. The first column refers to the applied datasets. In the second column, some ERE methods are exhibited, i.e., previous study and proposed method. The results of the relation extraction

are presented in last three column, i.e., precision, recall and F1. In addition, the best results for all datasets are marked with bold font.

Table 2. The detail in result of experiments. The best results for all datasets are marked with bold font.

Dataset	Model	P	R	F1
CoNLL04	paper [12]	72.99	63.37	67.01
	paper [13]	69.20	68.20	68.90
	paper [26]	-	-	67.99
	Baseline	73.04	70.00	71.47
	SPERT-DC	75.05	70.21	72.55
ADE	paper [7]	67.50	75.80	71.40
	paper [12]	72.10	77.24	74.58
	paper [26]	-	-	75.52
	Baseline	78.09	80.43	79.24
	SPERT-DC	76.82	84.08	80.28

Comparing with the methods cited and baseline comprehensively, the SPERT-DC has a best performance on two datasets, the results of which are presented in Table 2. On the CoNLL04 dataset, the SPERT-DC outperforms the model [13] which also relies on BERT as word embedding by up to 3.6%. In addition, SPERT-DC improves on both precision and recall. This paper has a better performance compared with the model [26] which uses adversarial training to enhance robustness for ERE tasks on two datasets. On the ADE dataset, the method in this paper also has a greater improvement in F1-scores than the model [12] based on Bi-LSTM-CRF, indicating that the BERT surpasses traditional network in feature extraction, such as LSTM, RNN et al. Meanwhile, SPERT-DC outperforms the baseline model by about 1%.

This paper also replaces the CLS representation by full-context, i.e., the a max pooling over the sentence’s all tokens to conduct a host of studies, the results of which are shown in Figure 5. It is found that the cls-mid combination is higher than the full-mid combination on F1-scores, indicating that the CLS representation is more suitable for text context.

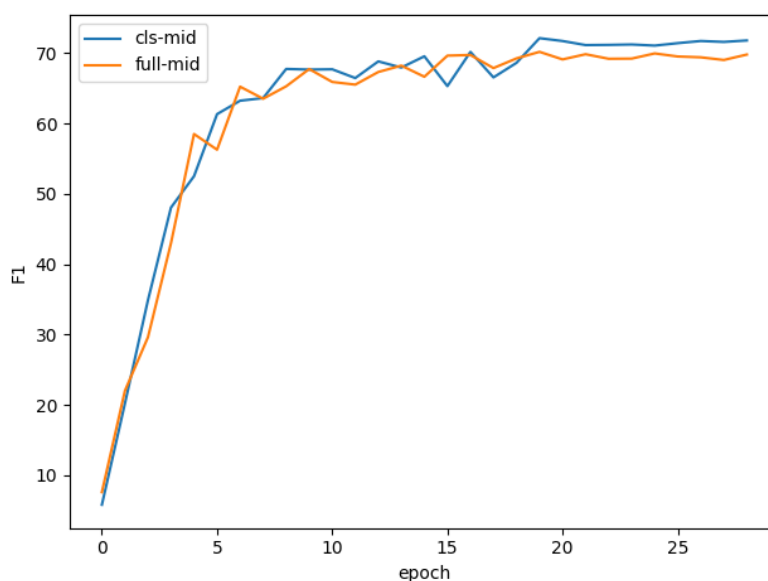


Figure 5. The tendency of F1-scores of two context combination pattern. The cls-mid (blue) performs better than the full-mid (orange).

Some thresholds in 0–1 range are taken to test, result of which is shown in Figure 6 and reveals a performance in form of a broken line. A good result is seen while the threshold T

is in 0–0.1 range and the F1-score peaks in case of zero. Therefore, this paper takes 0 as the final threshold for follow-up study.

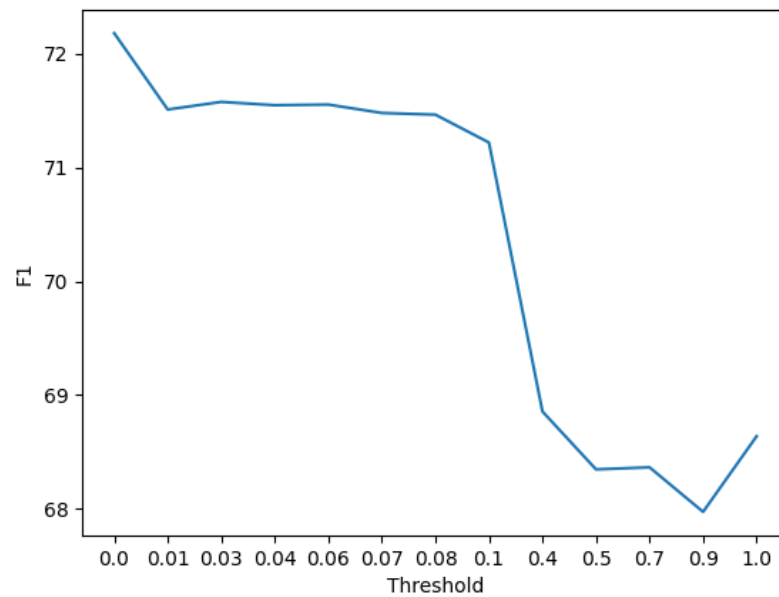


Figure 6. F1-scores of relation extraction on the CoNLL04 dataset when using different threshold. The wonderful vaule occurs in 0–0.1 threshold range, particular at 0.

As shown in Figure 7, in order to evaluate the performance of model, this paper conducts some experiments on many sentences varying in length. Before the Bi-LSTM_ATT is added, the experiment on longer texts has a poor result. The downward trend slows down after it is add, which leads to significant improvements.

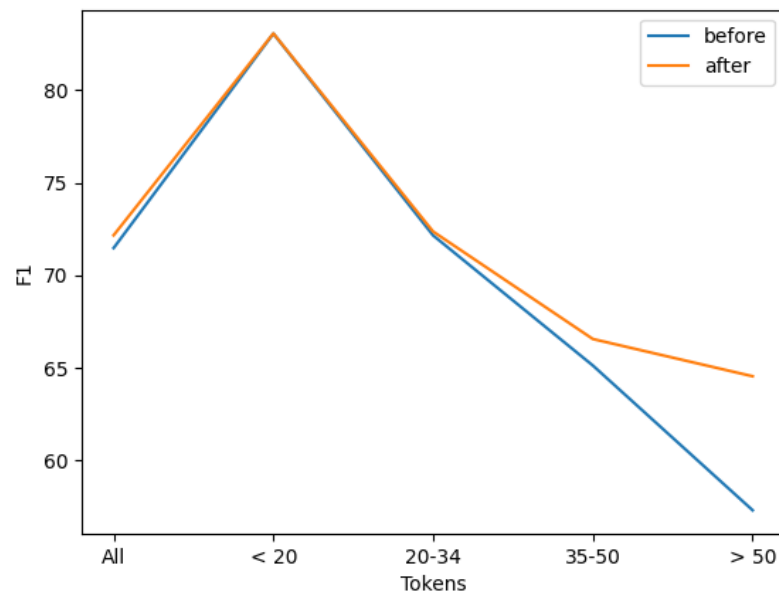


Figure 7. The changes of F1-scores before and after added Bi-LSTM_ATT network when using different tokens. Some improvements are seen after it is added (orange).

As shown in Figure 8, the ROC curve for each model of the two datasets is determined in order to further proves the superiority of proposed method on entity relation extraction. What’s more, the SPERT-DC shows the higher ROC scores compared to others.

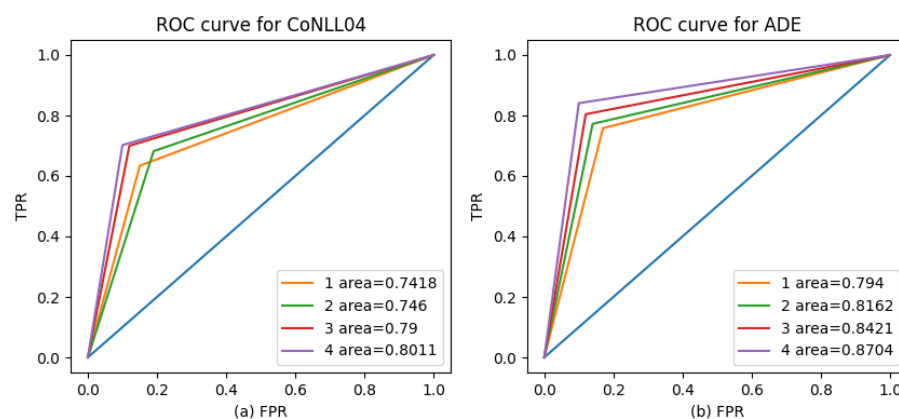


Figure 8. Receiver operating characteristic curves of the method compared to other model. (a) is the ROC curve for CoNLL04 dataset. (b) is ROC curve for ADE dataset. The number “1” and “2” in left image represent the model in paper [12] and paper [13] respectively, unlike it represent the model in paper [7] and paper [12] respectively. The number “3” refers to baseline and “4” refers to our method.

Based on the results and data of experiment above, the proposed method based on dynamic context takes the lead in two datasets, illustrating the effectiveness of SPERT-DC on relation extraction.

5. Conclusions and Discussion

Aiming at the missing feature of static text context of existing models, the proposed method used threshold to choose context area dynamical in threshold selector layer. The Bi-LSTM_ATT network was utilized to improve compatibility for the longer texts in feature extraction layer. To enhance text context representation, the MFF method was picked up in feature fusion layer. Making a comparison with prior studies, the SPERT-DC shares F1-score outperformance over theirs on the public datasets, the practicability of which was proved on ERE tasks. More attention will be paid to dynamic selection of the context representation and other languages in the future work. What is focused on is not only the existing feature area but also semantic representation, making a further advancement on the ERE model. It is noted that we only carry out experiments on English sentence dataset. We will also try to study the performance of our method on other languages such as Chinese in next research.

In this paper, a inspiring method based on dynamic context is proposed for entity relation extraction task, which can dynamically extract feature information. According to the experimental result on two datasets, it is found that the proposed method bears fruit theoretically. In reality, the entity relation extraction is aimed at convert unstructured information to structural data that will be saved in database to be widely applied to natural language processing applications such as search engines and question answering systems. Due to the rapid increase of knowledge in the network, people prefer to obtain information quickly and accurately, rather than manually annotating the relation between entity pairs. In order to add abundant knowledge to the knowledge graph as timely and accurately as possible, researchers are constantly striving to explore technical methods for efficiently and automatically acquiring knowledge, that is, to continuously improve the effect of entity relation extraction, which also is precisely the purpose of this paper and the future application filed.

Author Contributions: Conceptualization, X.M. and K.W.; methodology, X.M.; software, K.W.; validation, H.K.; formal analysis, K.W. and H.K.; writing—original draft preparation, H.K. and X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (No. 61772088).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liu, C.; Sun, W.; Chao, W.; Che, W. Convolution neural network for relation extraction. In *International Conference on Advanced Data Mining and Applications*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 231–242.
2. Zhang, R.; Meng, F.; Zhou, Y.; Liu, B. Relation classification via recurrent neural network with attention and tensor layers. *Big Data Min. Anal.* **2018**, *1*, 234–244.
3. Zhou, P.; Shi, W.; Tian, J.; Qi, Z.; Li, B.; Hao, H.; Xu, B. Attention-based bidirectional long short-term memory networks for relation classification. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Berlin, Germany, 7–12 August 2016; pp. 207–212.
4. Eberts, M.; Ulges, A. Span-based joint entity and relation extraction with transformer pre-training. In Proceedings of the 24th European Conference on Artificial Intelligence (ECAI), Santiago de Compostela, Spain, 31 August–4 September 2020; pp. 2006–2013.
5. Li, Q.; Ji, H. Incremental joint extraction of entity mentions and relations. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Baltimore, MD, USA, 22–27 June 2014; pp. 402–412.
6. Miwa, M.; Bansal, M. End-to-end relation extraction using lstms on sequences and tree structures. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, 7–12 August 2016; pp. 1105–1116.
7. Li, F.; Zhang, M.; Fu, G.; Ji, D. A neural joint model for entity and relation extraction from biomedical text. *BMC Bioinform.* **2017**, *18*, 198. [[CrossRef](#)] [[PubMed](#)]
8. Guo, X.; Zhang, H.; Yang, H.; Xu, L.; Ye, Z. A single attention-based combination of CNN and RNN for relation classification. *IEEE Access* **2019**, *7*, 12467–12475. [[CrossRef](#)]
9. Adel, H.; Schütze, H. Global normalization of convolutional neural networks for joint entity and relation classification. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 7–11 September 2017; pp. 1723–1729.
10. Katiyar, A.; Cardie, C. Going out on a limb: Joint extraction of entity mentions and relations without dependency trees. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, BC, Canada, 30 July–4 August 2017; pp. 917–928.
11. Bekoulis, G.; Deleu, J.; Demeester, T.; Develder, C. An attentive neural architecture for joint segmentation and parsing and its application to real estate ads. *Expert Syst. Appl.* **2018**, *102*, 100–112. [[CrossRef](#)]
12. Bekoulis, G.; Deleu, J.; Demeester, T.; Develder, C. Joint entity recognition and relation extraction as a multi-head selection problem. *Expert Syst. Appl.* **2018**, *114*, 34–45. [[CrossRef](#)]
13. Li, X.; Yin, F.; Sun, Z.; Li, X.; Yuan, A.; Chai, D.; Zhou, M.; Li, J. Entity-Relation Extraction as Multi-Turn Question Answering. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 1340–1350.
14. Geng, Z.; Chen, G.; Han, Y.; Lu, G.; Li, F. Semantic relation extraction using sequential and tree-structured LSTM with attention. *Inf. Sci.* **2020**, *509*, 183–192. [[CrossRef](#)]
15. Xiao, Y.; Tan, C.; Fan, Z.; Xu, Q.; Zhu, W. Joint entity and relation extraction with a hybrid transformer and reinforcement learning based model. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 9314–9321.
16. Zhao, K.; Xu, H.; Cheng, Y.; Li, X.; Gao, K. Representation iterative fusion based on heterogeneous graph neural network for joint entity and relation extraction. *Knowl.-Based Syst.* **2021**, *219*, 106888. [[CrossRef](#)]
17. Zhang, S.; Zheng, D.; Hu, X.; Yang, M. Bidirectional long short-term memory networks for relation classification. In Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation, Shanghai, China, 30 October–1 November 2015; pp. 73–78.
18. Fan, Y.; Zhou, S.; Li, Y.; Zhang, R. Deep learning approaches for extracting adverse events and indications of dietary supplements from clinical text. *J. Am. Med. Inform. Assoc.* **2021**, *28*, 569–577. [[CrossRef](#)] [[PubMed](#)]
19. Cui, Z.; Pan, L.; Liu, S. Hybrid bilstm-siamese network for relation extraction. In Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, Montreal, QC, Canada, 13–17 May 2019; pp. 1907–1909.
20. Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015; Conference Track Proceedings.
21. Shang, Y.M.; Huang, H.; Sun, X.; Wei, W.; Mao, X.L. A pattern-aware self-attention network for distant supervised relation extraction. *Inf. Sci.* **2022**, *584*, 269–279. [[CrossRef](#)]
22. Christou, D.; Tsoumakas, G. Improving Distantly-Supervised Relation Extraction Through BERT-Based Label and Instance Embeddings. *IEEE Access* **2021**, *9*, 62574–62582. [[CrossRef](#)]

23. Han, X.; Zhou, F.; Hao, Z.; Liu, Q.; Li, Y.; Qin, Q. MAF-CNER: A Chinese Named Entity Recognition Model Based on Multifeature Adaptive Fusion. *Complexity* **2021**, *2021*, 6696064. [[CrossRef](#)]
24. Zhang, J.; Hong, Y.; Zhou, W.; Yao, J.; Zhang, M. Interactive learning for joint event and relation extraction. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 449–461. [[CrossRef](#)]
25. Zhang, J.; Hao, K.; Tang, X.S.; Cai, X.; Xiao, Y.; Wang, T. A multi-feature fusion model for Chinese relation extraction with entity sense. *Knowl.-Based Syst.* **2020**, *206*, 106348. [[CrossRef](#)]
26. Bekoulis, G.; Deleu, J.; Demeester, T.; Develder, C. Adversarial training for multi-context joint entity and relation extraction. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 2830–2836.
27. Dixit, K.; Al-Onaizan, Y. Span-level model for relation extraction. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 5308–5314.
28. Luan, Y.; Wadden, D.; He, L.; Shah, A.; Ostendorf, M.; Hajishirzi, H. A general framework for information extraction using dynamic span graphs. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MN, USA, 2–7 June 2019; pp. 3036–3046.