ISBN: 978-981-18-5852-9
WCSE 2022 Spring Event: 2022 9th International Conference on Industrial Engineering and Applications
doi: 10.18178/wcse.2022.04.177

Medical Relation Extraction from Electronic Medical Records

Ben Li [1], Jungang Han [1, 2], Xiaowei Wang [1] and Xiaoying Pan [1, 2]

[1] Xi’an University of Posts & Telecommunications

[2] Shaanxi Key Laboratory of Network Data Analysis and Intelligence Processing

Abstract. In the field of natural language processing, relation extraction (RE) aims to identify the relationship between entities from text and is a vital step for subsequent tasks such as question answering and Knowledge Graph (KG). Comparing to the general domain text the relationship extraction from electronic medical records (EMRs) is more challenging due to the high-density distribution of entities of EMRs. In addition, relationship extraction usually suffers from the long-tail problem. The training data mainly focuses on a few types of relations, leading to the lack of sufficient annotations for the remaining types of relations, which is more severe in electronic medical records. To address the above problems, this paper proposes a segmented attentional Convolutional Neural Network (CNN+SegAtt) for medical relationship extraction and combines it with modified typed entity representations, which significantly improve the semantic learning ability from electronic medical records. A loss function with category weights is also proposed to facilitate the extraction of long-tail relationships. Experiments are conducted on the publicly available medical corpus 2010 i2b2/VA. The results show that the value of F1 reached 73.13%, better than that of existing methods, and the values of F1 for the small category samples are also significantly improved.

Keywords: relation extraction, electronic medical record, convolutional neural network, long-tail relations.

1. Introduction

Electronic medical records (EMRs) contain all the information of paper medical records, which are unstructured. It is necessary to analyze and mine the text of EMRs using natural language processing (NLP) [1]. As one of the basic tasks of natural language processing, relation extraction (RE) plays an important role in applications such as semantic understanding, question-answering, and knowledge graphs. At present, the research on relationship extraction for general domains is progressing relatively fast. Limited by the medical knowledge and scarce of open datasets in the medical domain, the research on medical relationship extraction from electronic medical record text still faces some difficulties and challenges.

Relationship extraction aims to identify the relationship between two entities in a given text from a predefined set of relationships, often expressed in the form of a triple <E1, R, E2>, where E1 and E2 denote entities and R denotes the semantic relationship between the entities. Take the 2010 i2b2/VA dataset as an example, as shown in Fig. 1, The sentence contains two types of three entities DVT (problem), PE (problem), warfarin (treatment), which produce three relationships, namely <DVT, TrAP, warfarin>, <DVT, PIP, PE>, <PE, TrAP, warfarin>. Table 1 lists the specific interpretation of the relationships. The example in Fig. 1 shows that a short sentence in a medical record text contains a high-density distribution of entities and each entity is repeatedly involved in different relationships. Contrasting to general domain public dataset such as SemEval2020_task8 which contains an average of 2 different entities every 23 characters and produces only 1 set of relations on average, statistics of the 2010 i2b2/VA corpus show that an average of 4 different entities appears and an average of 6 sets of inter-entity relations occur in every 18 characters. With such a high density of entity distribution, entity representation is particularly important, Peng and Gao et al [2] pointed out that both textual context and entity mentions provide key information to the relational model, where the type information in entity mentions is emphasized; Zhong and Chen et al [3] pointed out that unlike traditional methods incorporating entity types in the input layer of the relational model is very important[4], and this explicit inclusion of entity category information in the input layer can better improve the model performance. Inspired by. this idea we propose a variant of typed entity marker to improve the existing entity representation techniques without introducing special labels. In addition, the local features extracted by CNN can also help to express the high density entity distribution [5]. The proposed "CNN+ SegAtt" network architecture with a segmented attention mechanism can take the advantage of entity location information and achieve deeper mining and learning of semantic features of electronic medical record text.

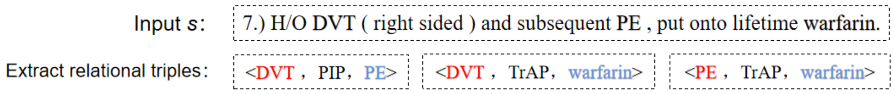

Fig. 1: An example of the 2010 i2b2/VA dataset.

Relational extraction usually encounters the long-tail problem, i.e., the data imbalance problem, for example, the incidence of common diseases is higher than that of rare diseases and this will in turn result in an unbalanced distribution of electronic medical records data. As shown in Fig. 2, nearly 62% of relations in 2010 i2b2/VA are long-tail relations. Most studies of long-tail relations in recent years are based on NYT datasets [6], and Xu [7] and Zhang [8] et al. exploited the hierarchical structure of the relations labels, e.g., /location/province/capital, which can naturally transfer relations knowledge from data-rich and semantically similar head-class relationships to tail-class relations. But this approach does not apply to other datasets where relations are not marked hierarchically. In order to overcome the data imbalance problem, we propose a loss function with category weights to improve the relational classification accuracy for small category samples.

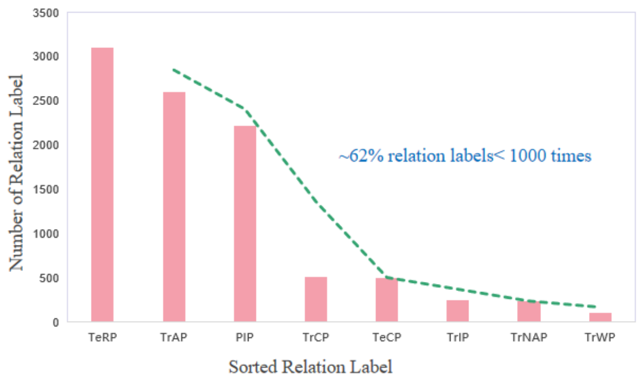

Fig. 2: Label distribution in the 2010 i2b2/VA dataset (except NA).

Based on the above studies, this paper proposed a CNN-based method for recognizing medical concept relations. Its contributions are as follows:

- A variant of typed entity marker is proposed to improve existing entity representation while not introducing special markup.
- The segmented attention mechanism is proposed and combining with the location information of entities to learn the semantic features of sentences efficiently.
- A loss function with class weights is proposed to improve the performance of a CNN-based classifier for relational extraction from imbalanced data of electronic medical records.
- Experimenting on the publicly available 2010 i2b2/VA medical dataset, the feasibility and effectiveness of the proposed method is evaluated and achieves the performance that outperforms the existing methods.

2. Related Work

Before the popularity of deep learning technology, most relational extraction approaches use statistical learning. In both the general and medical domain, many researchers focused on feature-based and kernel-based approaches [9-11], which are limited by conditions such as manual feature engineering and reliance on existing NLP toolkits.

Neural network models are now widely used for relation extraction. These neural network models can accurately capture textual relations without explicit linguistic analysis [12-13]. Various neural network models have been proposed for relational classification in general domains, including recurrent neural networks (RNN) and convolutional neural networks (CNN). Liu [14] et al. first proposed a CNN-based relationship extraction model using a simple convolutional neural network structure; Zeng [15] also used a convolutional neural network for relationship classification while designing features based on the location of entities in the text. In addition to supervised learning, distant supervision has flourished. Minrz [16] et al. first proposed the distant supervision for relation extraction, by mainly using labeled, unlabeled public domain data, with good scalability and generalizability.

Several neural networks have been applied to the classification of relationships in biomedical and electronic medical record texts. Liu et al. [17] used CNN for drug-drug interaction (DDI) extraction, Quan et al. [18] proposed a multichannel convolutional neural network (MCCNN) to accomplish this task, and He [19] et al. proposed CNN-Multi, which introduced a multi-pool operation in convolutional neural networks and presented a loss function with a class-level constraint matrix. BLSTM is one of the successful applications of recurrent neural networks for this task [20], and the authors used bidirectional recurrent neural networks for feature extraction and modeling of textual information. CRNN [21] sequentially used RNN with CNN to learn global and local contexts in sentences, which were then used for relation classification. In addition, several methods based on attention mechanisms [22-23] were proposed to automatically extract medical relationships.

There are few studies on long-tail relation extraction, among which Gui et al. [24] proposed an explanation-based approach, while Lei et al. [25] utilized external knowledge (logic rules); later studies focused on datasets with hierarchical structures, such as NYT, Han et al. [26] proposed a hierarchical attention mechanism scheme for long-tail relations in relation extraction, Zhang et al. [27] combined the embedding representation obtained after pre-training by graph convolutional networks [28] into the hierarchical structure of relations, but this class of methods does not work well on other datasets, so this paper uses the traditional approach, i.e., mitigating the data imbalance problem by increasing the loss of small class samples.

Compared with the existing relationship extraction models, the proposed "CNN+SegAtt" model achieves better performance by making full use of the entity location information and the modified entity representation techniques. In addition, the proposed loss function with category weights helps to solve the problem caused by long-tail relationships.

3. Methodology

In this section, the task of medical relationship extraction in EMRs is first formally defined in Section 3.1, Then the proposed model architecture and entity representation techniques are introduced in Sections 3.2 and 3.3 Finally, the loss function based on category weights is described in Section 3.4.

3.1. Problem Definition

Our study focuses on the Relation Extraction of sentence level. Specifically, given a sentence and an entity pair (e_1, e_2), the relation r between e_1 and e_2 is predicted from R∪{NA}, where R is a predefined set of relations of interest, and the entity pairs that do not belong to the set R are denoted as NA.

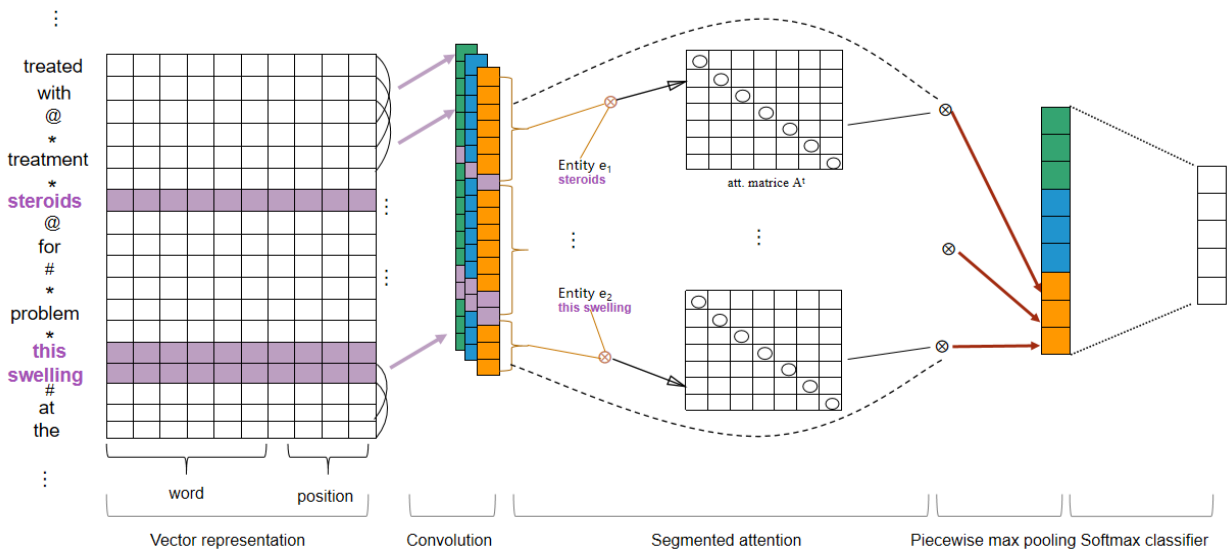

Fig. 3: CNN+SegAtt network structure.

3.2. Model Architecture

Fig. 3 depicts the network architecture of the proposed CNN-based medical relation classification model. With "she was treated with [steroids]$_{e1}$ for [this swelling]$_{e2}$ at the outside hospital, and these were continued." as an example input. The model learns the distribution representation of each relation sample, generates a feature vector to represent the example s containing two entities, and the final score can be obtained by the relation type representation. Further details are discussed in the following subsections.

1) Embedding Layer

Most of the previous works have demonstrated that taking the combination of two features: word embedding [29] and positional embedding [15] as the representation of a sentence is crucial for relation extraction [16-18], To enhance the ability of the presentation further we embed the typed entity marker into the input layer which will be described in detail in Section C.

A sentence $s = \{w_1, \cdots, w_n\}$ is a sequence of n words. To construct the data representation for the encoding layer, each word in the sentence is converted into a k_w-dimensional vector by word2vec [30]. For each word w_i $(1 \leq i \leq n) \in s$, the relative distances from w_i to the given two entities are embedded into two k_p-dimensional vectors respectively. The final embedding representation $x_i \in R^{ki}$ $(k_i = k_w + k_p \times 2)$ of each word is obtained by connecting word embedding and two position embeddings.

2) Encoding Layer

The encoding layer aims to compose the input embeddings of the given instance into its corresponding instance embedding. Since a sentence contains multiple entities in an electronic medical record, each of which may consist of several words, the location features of entities are important for the medical relationship classification. PCNN [31], as the most commonly used network model for distant supervision, can capture structural information between two entities by considering their positions due to PCNN has a piecewise pooling layer [15] which is the key part that makes it different from 1DCNN.

All of the segments (a piece of information) in an input representation of a sentence might provide useful information for the RE task. However, not all the segments contribute equally to the sentence meaning for different relations. It is beneficial to integrate an entity-aware attention layer into the PCNN network. Therefore. we propose an additional attention mechanism called segmented attention, which built upon segment-level representations to capture crucial segments in each sentence and thus improve the performance of the medical RE task.

First, the PCNN uses a convolutional kernel with sliding window size m on the input sequence $\{x_1, \cdots, x_n\}$ to obtain the k_h-dimensional hidden embedding.

$$h_i = CNN(x_{i-\frac{m-1}{2}}, \ldots, x_{i+\frac{m-1}{2}}), \quad (1)$$

Two diagonal matrices A^t are introduced, $A^t_{i,i} = f(e_t, h_i)$ to represent the link between entities and h_i, and the evaluation function f is calculated as the inner product between the respective embeddings of h_i and e_t and is parameterized into the network for updating A^t during the training process. Given A^t, we define the relevance of h_i to the t-th entity ($t \in \{1,2\}$) as follows:

$$\alpha^t_i = \frac{exp(A^t_{i,i})}{\sum_{i'=1}^{n} exp(A^t_{i',i'})}, \quad (2)$$

Next, we apply the attention mechanism at the segment level to assign a different weight α_i to each hidden vector h_i, and pay more attention to the informative segment. The specific segmented calculating formulas are as follows,

$$\begin{aligned} \left[r^{(1)}\right]_j &= h_i \alpha^1_i, \quad 0 \leq i \leq i_1 \\ \left[r^{(2)}\right]_j &= h_i \frac{\alpha^1_i + \alpha^2_i}{2}, \quad i_1 \leq i \leq i_2 \\ \left[r^{(3)}\right]_j &= h_i \alpha^2_i, \quad i_2 \leq i \leq n \end{aligned} \quad (3)$$

where $[\bullet]_j$ is the j-th value of a vector, i_1 and i_2 are the entity positions.

Then, a piecewise max-pooling is applied over the hidden embeddings,

$$
\begin{aligned}
\left[s^{(1)}\right]_j &= max\left\{\left[r^{(1)}\right]_j\right\},\\
\left[s^{(2)}\right]_j &= max\left\{\left[r^{(2)}\right]_j\right\},\\
\left[s^{(3)}\right]_j &= max\left\{\left[r^{(3)}\right]_j\right\},
\end{aligned} \tag{4}
$$

At last, the three pooling results are concatenated to obtain the final instance embedding s,

$$s = \left[s^{(1)}; s^{(2)}; s^{(3)}\right]. \tag{5}$$

3.3. Entity Representation

Both the span and the type of the entity are necessary to be presented in the input, and the recent works [32,3] have shown that the appropriate use of them can improve the performance of relation extraction. To select proper representation on the 2010 i2b2/VA dataset, the following representation is introduced and evaluated.

Entity mask (Zhang et al. [4], 2017): This method replaces two entities in the original sentence using the special token [SUBJ-NER] or [OBJ-NER], where NER is the corresponding entity type.

Entity marker (Baldini Soares et al. [32], 2019): This method introduces the special tokens [E1], [/E1], [E2], [/E2] to indicate the entity span and modifies the sentence to "[E1] SUBJ [/E1]... [E2] OBJ [/E2]".

Entity marker (punct) (Wang et al. [33], 2020; Zhou et al. [34], 2021): This approach is a variant of Entity marker, which marks entity spans by modifying the original sentence to "@ SUBJ @"... # OBJ #" to mark entity spans.

Typed entity marker (Zhong and Chen et al. [3], 2020): This method also introduces special tokens, which modifies the original sentence to "<S:NER> SUBJ </S:NER>... <O:NER> OBJ </O:NER>", where NER is the corresponding entity type.

Table 1 shows the performance of the different entity representation methods. For each method, an example of processed sentences is provided. It can be observed from the experimental results that Typed marker (punct) significantly outperforms the other markers. In particular, the simple PCNN model achieves an F1 score of 69.59% using Typed marker (punct), which is closed to the most updated result of 69.7% (He et al. [19], 2019), while the proposed "CNN+SegAtt" model achieved an F1 of 72.15%, The experimental results indicate that entity information and segmental attention mechanism can further improve performance of the relation extraction task.

Table 1: Test f1 (%) for different entity representation methods on 2010 i2b2/va. For each method, a sample sentence "on examination, he is afebrile." is also provided as a processing input. Typed marker (punct) is clearly superior to the other methods

Method	Input Example	PCNN(F1%)	CNN+SegAtt(F1%)
Original	On examination , he is afebrile .	65.41	66.23
Entity Mask	On [SUBJ-TEST] , he is [OBJ-PROBLEM] .	62.27	61.87
Entity Marker	On [E1] examination [/E1] , he is [E2] afebrile [/E2] .	65.76	65.95
Entity Marker (punct)	On @ examination @ , he is # afebrile # .	67.58	68.47
Typed Marker	On <S:TEST> examination </S:PROBLEM> , he is <O:PROBLEM> afebrile </O:PROBLEM> .	67.76	68.83
Typed Marker (punct)	On @ * test * examination @ , he is # * test * afebrile # .	69.59	72.15

3.4. Loss Function with Category Weights

In this paper, a Softmax classifier is used to probabilistically map the hidden layer vectors to determine the relation between entities. the common loss function used by the Softmax classifier is the cross-entropy loss function,

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} t_i \, log(\hat{y_i}) + \lambda\|\theta\|_F^2 \tag{6}$$

Where $t \in R^c$ stands for the one hot encoding of the category to which the current sample belongs, and c is the total number of categories, $\hat{y}_i$ indicating that the Softmax layer maps the hidden layer vector to the probability values of each category. The last term in Eq. (6) represents the *L2* canonical term.

To solve the data imbalance problem, the traditional sampling method changes the distribution of the original data. The oversampling process may introduce additional noise to the dataset, and the undersampling process may result in the loss of original data. Therefore, this section adjusts the loss of different category samples of the EMRs text by introducing category weights $W^{wgt} = <w^{c_1}, w^{c_2}, ..., w^{c_c}>$ in the loss function to increase the importance of small category samples in the training process. The calculation of category weights relies on the original distribution of each relational category sample, as shown in Eq. (7).

$$W^{c_i} = \frac{N^{all} - N_i}{(c-1)N^{all}} \tag{7}$$

We define the total number of samples as $N_{\text{all}} = \sum_{i=1}^{c} N_i$, where N_i ($i \in R^c$) represents the number of samples of class i in the training set.

We update t_i in Eq. (6) to $t_i w^{\text{wgt}}$ and obtain Eq. (8), which is the loss function with weights for learning proposed in this section.

$$J_{\text{wgt}}(\theta) = -\frac{1}{m}\sum_{i=1}^{m} t_i w^{wgt} \log(\hat{y}_i) + \lambda \|\theta\|_F^2 \tag{8}$$

4. Experiments

The proposed method will be evaluated by comprehensive experiments in this section.

4.1. Dataset

We evaluate our model on the 2010 i2b2/VA [35] dataset, which is one of the recognized datasets for entity relation reviews in electronic medical records and has been widely used in recent works. The dataset, derived from three hospital discharge summaries contains eight relation types: treatment improve or cure medical problem (TrIP), treatment worsen medical problem (TrWP), treatment caused medical problems (TrCP), treatment administered medical problem (TrAP), treatment was not administered because of medical problem (TrNAP), test reveal medical problem (TeRP), test conducted to investigate medical problem (TeCP), and medical problem indicates medical problem (PIP); The three types of entities: problem, test, treatment. The problem is a phrase that describes the physical or mental abnormality of the patient as observed by the patient or the doctor, the test is a phrase that describes the further physical examination to obtain the patient's symptoms and health condition, and the treatment is a phrase that describes the treatment taken to treat the patient.

If there are more than two entities in a sentence, one instance is created for each pair of entities. Since the part of 394 original training documents available for download contains only 170 training sets and 256 test sets, our preprocessing adopts the preprocessing steps used by Raj D. et al [21]. all training and test instances were merged, and then the training and test sets were redistributed in the ratio of 8:2. Table 2 describes in detail the meaning of their relationship categories and the relevant statistical information after redistributing the training and test sets.

Table 2: The number of training and testing instances for each relation type in the i2b2 dataset

Class	Definition	Train size	Test size
TrIP	treatment improve or cure medical problem	162	41
TrWP	treatment worsen medical problem	106	27
TrCP	treatment caused medical problems	420	106
TrAP	treatment administered medical problem	2093	524
TrNAP	treatment was not administered because of medical problem	139	35
TeRP	test reveal medical problem	2442	611
TeCP	test conducted to investigate medical problem	403	101
PIP	and medical problem indicates medical problem	1762	441
NA	none of the above	44637	11160
Total	/	52164	13046

4.2. Experimental Setting

Evaluation metrics. As shown in Table 2, there are eight positive relation types (predefined) and one negative relation type (in addition to the defined relationship). The performance of each positive relation type was evaluated using the precision, recall, and F1-measure. According to the official evaluation metrics [36], the performance of the model is defined based on the micro-averaged F1 scores of all positive relation types.

Parameter settings. A 5-fold cross-validation was performed on the training set, and the key parameters of the model are detailed in Table 3. the word2vec tool was used to train the initial word representations as inputs to the word embedding layer involved in model training.

Table 3: Key parameters of the model

Word embedding	Position embedding	Batch size	Window size	LR
128	60	512	[3,4,5]	0.001

Experimental environment. The experiments are based on Linux operating system, model coding is based on python 3.5, and Tensorflow deep learning framework. The machine used to train the model is NVIDIA dgx-l with 8 Tesla P100 GPUs, each with 16 GB of memory.

4.3. Results

Table 4 shows the results obtained for the proposed model on the 2010 i2b2/VA dataset and the comparison with the previous methods. In the table, the typed marker (punct) and the weighted loss function are abbreviated as Tmp and Wgt. From the results in the table, it is clear that both proposed methods are effective in improving the F1 value. In particular, the proposed "CNN+SegAtt "+Tmp+Wgt model achieves 73.13%, which is better than all the compared methods.

Table 4: Class-level performance of various models on i2b2 dataset (based on f1 score)

Class	TrIP	TrWP	TrCP	TrAP	TrNAP	TeRP	TeCP	PIP
CNN-Max	21.74	10.26	41.25	57.59	10.68	73.16	31.82	54.95
LSTM-Max	0	0	35.48	61.07	0	76.50	23.20	53.79
LSTM-Att	13.33	0	41.01	60.38	12.29	79.12	38.46	60.99
CRNN-Max	25.71	16.67	43.18	67.39	36.36	80.32	39.46	58.04
CRNN-Att	34.48	9.52	47.66	63.94	18.60	76.31	39.76	55.53
CNN-Multi	3.82	0	48.14	72.83	5.21	83.02	33.41	63.63
"CNN+SegAtt"+Tmp	51.02	40.43	59.97	74.30	38.63	82.61	63.82	66.89
"CNN+SegAtt"+Tmp+Wgt	71.43	60.00	64.97	79.30	58.33	86.91	66.00	65.99

Table 5 shows the performance (F1 value) of different approaches on the positive relation types, and combined with the statistics of the class distribution of the dataset in Table 2, it can be found that the classification ability of "CNN+SegAtt "+Tmp+Wgt for small category samples is significantly improved compared to the baseline model, e.g., TrIP, TrWP, TrNAP, with the largest improvement achieved on the TrWP class. For categories with a large sample size (TrAP, TeRP), the model also improved to a large extent.

Table 5: Model performance comparison

Model	Precision/%	Recall/%	F1 score/%
SVM[37]	67.98	57.35	62.21
CNN-Max[5]	63.23	58.92	61.00
LSTM-Max[20]	57.54	55.40	56.45
LSTM-Att[20]	65.23	56.77	60.71
CRNN-Max[21]	67.27	62.89	65.01
CRNN-Att[21]	66.69	58.57	62.05
CNN-Multi[19]	73.05	66.58	69.67
"CNN+SegAtt"+Tmp	74.83	69.65	72.15
"CNN+SegAtt"+Tmp+Wgt	76.58	69.98	73.13

5. Conclusion and Future Work

In this paper, a new CNN-based "CNN+SegAtt"+ Tmp+Wgt medical relationship classification model is proposed. Segmented attention mechanism and Tmp typed entity representation help to extract more accurate and richer features at the semantic layer, and Wgt ensures the learning ability of the model in each category according to the original distribution of each category, while significantly improving the model's fitting ability to small category samples. The performance of the proposed model is better than all existing models on the 2010 i2b2/VA dataset.

Future work includes (1) evaluating our model on more pre-trained language models and will extend to other tasks such as named entity recognition and QA. (2) The use of additional information to train more effective models for further addressing the long-tail problem.

6. Acknowledgment

This work was supported by the national key research and development project (No.2019YFC0121502).

7. References

[1] Wasserman, Richard C. "Electronic medical records (EMRs), epidemiology, and epistemology: reflections on EMRs and future pediatric clinical research." Academic pediatrics 11.4 (2011): 280-287.

[2] Peng H, Gao T, Han X, et al. "Learning from context or names? an empirical study on neural relation extraction." arXiv preprint arXiv:2010.01923 (2020).

[3] Zhong, Zexuan, and Danqi Chen. "A frustratingly easy approach for joint entity and relation extraction. 466." arXiv preprint arXiv:2010.12812 467 (2020).

[4] Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli and Christopher D. Manning. "Position-aware attention and supervised data improve slot filling." Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017: 35-45.

[5] Sunil Kumar Sahu, Ashish Anand, Krishnadev Oruganty and Mahanandeeshwar Gattu. "Relation extraction from clinical texts using domain invariant convolutional neural network." Proceedings of the 15th Workshop on Biomedical Natural Language Processing. 2016.

[6] Riedel, Sebastian, Limin Yao, and Andrew McCallum. "Modeling relations and their mentions without labeled text." Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Berlin, Heidelberg, 2010.

[7] Han Xu, Pengfei Yu, Zhiyuan Liu, Maosong Sun and Peng Li. "Hierarchical relation extraction with coarse-to-fine grained attention." Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018:2236-2245.

[8] Zhang, Ningyu, et al. "Long-tail relation extraction via knowledge graph embeddings and graph convolution networks." arXiv preprint arXiv:1903.01306 (2019).

[9] Xiaodan Zhu, Colin Cherry, Svetlana Kiritchenko, Joel Martin and Berry de Bruijn. "Detecting concept relations in clinical text: Insights from a state-of-the-art model." Journal of biomedical informatics 46.2 (2013): 275-285.

[10] D’Souza, Jennifer, and Vincent Ng. "Ensemble-based medical relation classification." Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. 2014: 1682-1693.

[11] Kim, Jisung, Yoonsuck Choe, and Klaus Mueller. "Extracting clinical relations in electronic health records using enriched parse trees." Procedia Computer Science 53 (2015): 274-283.

[12] Xu, Yan, et al. "Classifying relations via long short term memory networks along shortest dependency paths." Proceedings of the 2015 conference on empirical methods in natural language processing. 2015: 1785-1794.

[13] Patrick Verga, David Belanger, Emma Strubell, Benjamin Roth and Andrew McCallum. "Multilingual relation extraction using compositional universal schema." Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016: 886-896.

[14] ChunYang Liu, WenBo Sun, WenHan Chao and WanXiang Che. "Convolution neural network for relation extraction." International Conference on Advanced Data Mining and Applications. Springer, Berlin, Heidelberg, 2013: 231-242.

[15] Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou and Jun Zhao. "Relation classification via convolutional deep neural network. " Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. 2014: 2335-2344.

[16] Mike Mintz, Steven Bills, Rion Snow and Dan Jurafsky. "Distant supervision for relation extraction without labeled data." Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. 2009: 1003-1011.

[17] Shengyu Liu, Buzhou Tang, Qingcai Chen, and Xiaolong Wang. "Drug-drug interaction extraction via convolutional neural networks. " Computational and mathematical methods in medicine 2016 (2016).

[18] Chanqin Quan, Lei Hua, Xiao Sun, and Wenjun Bai. "Multichannel convolutional neural network for biological relation extraction." BioMed research international 2016 (2016).

[19] He, Bin, Yi Guan, and Rui Dai. "Classifying medical relations in clinical text via convolutional neural networks." Artificial intelligence in medicine 93 (2019): 43-49.

[20] Sahu, Sunil Kumar, and Ashish Anand. "Drug-drug interaction extraction from biomedical texts using long short-term memory network." Journal of biomedical informatics 86 (2018): 15-24.

[21] Raj, Desh, Sunil Sahu, and Ashish Anand. "Learning local and global contexts using a convolutional recurrent network model for relation classification in biomedical text." Proceedings of the 21st conference on computational natural language learning *(CoNLL 2017)*. 2017.

[22] Asada, Masaki, Makoto Miwa, and Yutaka Sasaki. "Extracting drug-drug interactions with attention CNNs." BioNLP 2017. 2017: 9-18.

[23] Zheng, Wei, et al. "An attention-based effective neural model for drug-drug interactions extraction." BMC bioinformatics 18.1 (2017): 1-11.

[24] Yaocheng Gui, Qian Liu, Man Zhu and Zhiqiang Gao. "Exploring long tail data in distantly supervised relation extraction." Natural Language Understanding and Intelligent Applications. Springer, Cham, 2016. 514-522.

[25] Lei, Kai, et al. "Cooperative denoising for distantly supervised relation extraction." Proceedings of the 27th International Conference on Computational Linguistics. 2018: 426-436.

[26] Xu Han, Pengfei Yu, Zhiyuan Liu, Maosong Sun and Peng Li. "Hierarchical relation extraction with coarse-to-fine grained attention. " Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018: 2236-2245.

[27] Zhang, Ningyu, et al. "Long-tail relation extraction via knowledge graph embeddings and graph convolution networks." Proceedings of the 2019 Conference of the North, 2019.

[28] Defferrard, Michaël, Xavier Bresson, and Pierre Vandergheynst. "Convolutional neural networks on graphs with fast localized spectral filtering." Advances in neural information processing systems 29 (2016): 3844-3852.

[29] Hinton, Geoffrey E. "Learning distributed representations of concepts. " Proceedings of the eighth annual conference of the cognitive science society. Vol. 1. 1986.

[30] Tomas Mikolov, Kai Chen, Greg Corrado and Jeffrey Dean. "Efficient estimation of word representations in vector space." Proceedings of the International Conference on Learning Representations *(ICLR 2013)*. 2013.

[31] Daojian Zeng, Kang Liu, Yubo Chen and Jun Zhao. "Distant supervision for relation extraction via piecewise convolutional neural networks. " Proceedings of the 2015 conference on empirical methods in natural language processing. 2015. 2015: 1753-1762.

[32] Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling and Tom Kwiatkowski. "Matching the blanks: Distributional similarity for relation learning." Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019: 2895-2905.

[33] Wang, Ruize, et al. "K-adapter: Infusing knowledge into pre-trained models with adapters." arXiv preprint arXiv:2002.01808 (2020).

[34] Wenxuan Zhou, Kevin Huang, Tengyu Ma and Jing Huang. "Document-level relation extraction with adaptive thresholding and localized context pooling." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. No. 16. 2021.

[35] Sun, Weiyi, Anna Rumshisky, and Ozlem Uzuner. "Evaluating temporal relations in clinical text: 2012 i2b2 Challenge." Journal of the American Medical Informatics Association 20.5 (2013): 806-813.

[36] Özlem Uzuner, Brett R South, Shuying Shen and Scott L DuVall. "2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text." Journal of the American Medical Informatics Association 18.5 (2011): 552-556.

[37] Rink B, Harabagiu S, Roberts K. "Automatic extraction of relations between medical concepts in clinical texts." Journal of the American Medical Informatics Association, 2011, 18(5): 594-600.