Learning BLSTM-CRF with Multi-channel Attribute Embedding for Medical Information Extraction

Jie Liu(✉), Shaowei Chen, Zhicheng He, and Huipeng Chen

College of Computer and Control Engineering, Nankai University, Tianjin, China
jliu@nankai.edu.cn, {chenshaowei,hezhicheng,chenhp}@mail.nankai.edu.cn

Abstract. In Recent years, medical text mining has been an active research field because of its significant application potential, and information extraction (IE) is an essential step in it. This paper focuses on the medical IE, whose aim is to extract the pivotal contents from the medical texts such as drugs, treatments and so on. In existing works, introducing side information into neural network based Conditional Random Fields (CRFs) models have been verified to be effective and widely used in IE. However, they always neglect the traditional attributes of data, which are important for the IE performance, such as lexical and morphological information. Therefore, starting from the raw data, a novel attribute embedding based MC-BLSTM-CRF model is proposed in this paper. We first exploit a bidirectional LSTM (BLSTM) layer to capture the context semantic information. Meanwhile, a multi-channel convolutional neural network (MC-CNN) layer is constructed to learn the relations between multiple attributes automatically and flexibly. And on top of these two layers, we introduce a CRF layer to predict the output labels. We evaluate our model on a Chinese medical dataset and obtain the state-of-the-art performance with 80.71% F1 score.

Keywords: Medical information extraction · Multi-channel Convolutional neural network

1 Introduction

Recently, online medical and health services have been rapidly developing, and a great deal of medical doctor-patient question and answer data has been accumulated on the Internet. Due to the great value and application potential of this information, text mining about online medical text data has been an active research field in recent years. A fundamental work of these studies is information extraction (IE), whose aim is to extract pivotal contents from the medical texts such as diseases, symptoms, medicines, treatments and checks. And these contents can be further used for other text mining works including information retrieve [20], Pharmacovigilance (PV) [2], and drug-drug interactions [19] tasks.

© Springer Nature Switzerland AG 2018
M. Zhang et al. (Eds.): NLPCC 2018, LNAI 11108, pp. 196–208, 2018.
https://doi.org/10.1007/978-3-319-99495-6_17

IE is an important research in natural language processing (NLP), which focuses on extracting knowledge from unstructured text [6]. Hand-crafted regular expressions, classifiers, sequence models and some other methods are always used in IE. For decades, Conditional Random Fields (CRFs) [13] have been widely considered as effective models. After that, to construct an end-to-end model which can automatically learn semantic relations between words without any hand-crafted features, neural networks have been introduced into CRF methods. Furthermore, the neural network based CRF models have achieved great success in IE tasks including name entity recognition (NER) [3,8], opinion extraction [10], and text chunking [12].

As a domain-specific task, medical IE is a challenging work because the online medical text data always has plenty of professional terminologies and noise. Thus, adopting simple neural network based models is not enough. In order to capture more information, many approaches have been proposed [1,11] to introduce side information and prior knowledge. However, depending on the LSTM-CRF structure, the existing methods always neglect the classical attributes of text such as syntax and morphology. Furthermore, these attribute are not difficult to obtain and can greatly improve the performance.

In this paper, we focus on the medical IE and aim to utilize the classical attributes of data to improve the performance. To achieve this, we propose a novel bidirectional LSTM-CRF model with multi-channel convolution neural network (MC-BLSTM-CRF), and introduce multiple attributes which covers aspects of lexical, morphological, and domain-specific information. These attributes play a strong guiding role in information extraction and are easy to obtain. A multi-channel convolution neural network (MC-CNN) is built to learn the hidden representations of multiple attributes automatically and flexibly. To evaluate our model, we construct a Chinese medical dataset which composed of doctor-patient question and answer data, and achieve the state-of-the-art result on it. Experimental results demonstrate that our approach can discover more meaningful contents than baseline methods. The main contributions of our work can be summarized as follows:

- We propose a MC-BLSTM-CRF model with multi-channel attribute embedding for medical IE task, which can model relations between tokens, multiple attributes and labels effectively.
- The proposed method has excellent extensibility and can flexibly capture meaningful information which is neglected by existing models.
- The experimental results on a Chinese medical IE dataset show that our model substantially outperforms other baseline methods.

The remainder of this paper is organized as follows. In Sect. 2, we discuss the related work. Section 3 introduces the details of the MC-BLSTM-CRF model. Section 4 discusses the experiments setting and results. Finally, we conclude our work in Sect. 5.

2 Related Work

Several works have been proposed to medical IE, and these works can be divided into two categories: traditional methods and neural network based methods.

Traditional IE methods such as Hidden Markov Models (HMM) and CRF [13,15,17,18] have achieved great performance and been widely used in various tasks including medical IE. For instance, Bodnari et al. (2013)[1] developed a supervised CRF model based on features and external knowledge for medical NER. Jochim and Deleris (2017) [11] proposed a constrained CRF approach to consider dependence relations and probability statements in the medical domain. However, these models rely heavily on feature engineering.

To capture features automatically, neural network based models have been proposed and been frequently adopted for IE tasks. Collobert et al. (2011) [4] used a CNN to capture relations between tokens over word embeddings with a CRF on top. Huang et al. (2015) [9] introduced a BLSTM-CRF model for word encoding and joint label decoding based on rich hand-crafted spelling features. Both Lample et al. (2016) [14] and Ma and Hovy (2016) [16] proposed a BLSTM-CRF model with character-level encoding. For medical IE tasks, Chalapathy et al. (2016) [2] and Zeng et al. (2017) [21] both adopted a BLSTM-CRF model to provide end-to-end recognition without hand-craft features. Dong et al.(2017) [5] presented a transfer learning based on BLSTM to employ domain knowledge for enhancing the performance of NER on Chinese medical records.

Although neural network based models are useful, side information and prior knowledge are also important to domain-specific IE. Many studies have proposed various approaches to introduce side information and prior knowledge [1,5], but they usually need to design complex features or train model on external corpora which make these methods more complicated and less scalable. Moreover, they always neglect the traditional attributes which is meaningful for IE.

Comparing with existing approaches, a MC-BLSTM-CRF model with multi-channel attribute embedding proposed by us can learn the relations between traditional attributes of raw data by devising a multi-channel CNN structure.

3 Methodology

In this section, we describe the components of our attribute embedding based MC-BLSTM-CRF model in details. Given a word sequence $X = \{x_1, x_2, \cdots, x_n\}$, its corresponding multi-attribute sequence $M = \{m_1, m_2, \cdots, m_n\}$ and label sequence $Y = \{y_1, y_2, \cdots, y_n\}$, the goal is extracting important information by modeling the conditional probability $P(Y \mid X)$. To achieve this, we need to consider the semantic relations between words, context relations between attributes and the transfer relations between labels. Thus, we propose an attribute embedding based MC-BLSTM-CRF model which introduces a multi-channel CNN based into a neural network based CRF method. Figure 1 illustrates the overall framework of the attribute embedding based MC-BLSTM-CRF.

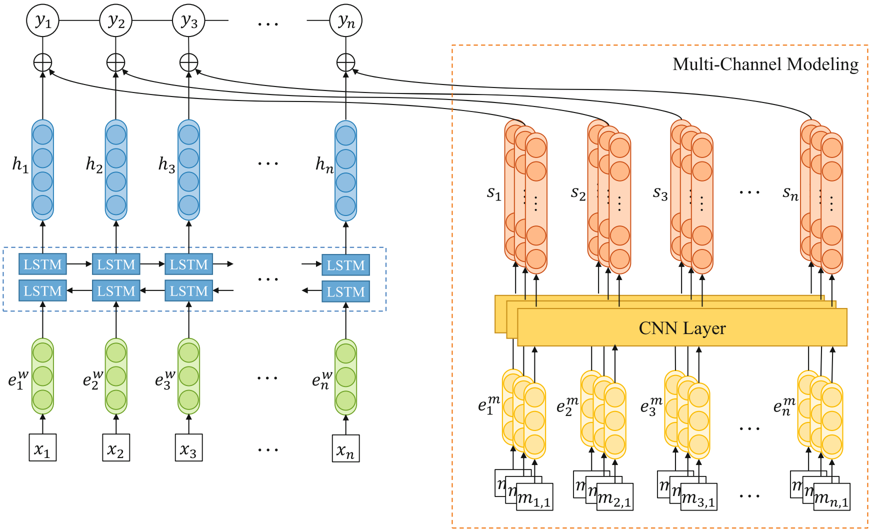


Fig. 1. The overall framework of MC-BLSTM-CRF.

To model semantic relations, each word x_i is mapped to a word embedding vector $e_i^w \in R^{d_w}$ firstly, where d_w is the dimension of word embedding. And then, we exploit a LSTM layer to obtain the hidden representation h_i which contains context semantic information by encoding embedding vector e_i^w:

$$h_i = F(e_i^w, h_{i-1}), \tag{1}$$

where F is the encoding function, and $h_i \in R^{d_h}$, where d_h is the dimension of hidden vectors.

3.1 Bidirectional LSTM

LSTM [7] is a type of RNN which can capture long-distance semantic relation by maintaining a memory cell to store context information. The memory cell is constantly updated in the encoding process, and the proportions of information are determined by three multiplicative gates including input gate, forget gate and output gate. Although various LSTM architectures have been explored, we adopt the basic LSTM model similar to [16]. Formally, the encoding process at the t-th time step is implemented as follow:

$$i_t = \sigma\left(W_{hi}h_{t-1} + W_{ei}e_t^w + b_i\right), \tag{2}$$

$$f_t = \sigma\left(W_{hf}h_{t-1} + W_{ef}e_t^w + b_f\right), \tag{3}$$

$$\widetilde{c_t} = tanh\left(W_{hc}h_{t-1} + W_{ec}e_t^w + b_c\right), \tag{4}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \widetilde{c_t}, \tag{5}$$

$$o_t = \sigma\left(W_{ho}h_{t-1} + W_{eo}e_t^w + b_o\right), \tag{6}$$

$$h_t = o_t \odot tanh\left(c_t\right), \tag{7}$$

where c_t, i_t, f_t, and o_t represent the memory cell, input gate, forget gate and output gate respectively. e_t^w and h_t donate the word embedding vector and hidden state vector at time t. Both σ and $tanh$ are the activation functions, and $\odot$ represents the element-wise product. W_* and b_* are network parameters which donate the weight matrices and bias vectors.

Although LSTM can solve the long-distance dependency problem, it still lose some semantic information due to the sequential encoding way of LSTM. For example, h_t only contains the semantic information before time step t. Therefore, a bidirectional LSTM (BLSTM) is needed to model both the forward and backward context information in the following form:

$$\begin{aligned}\overrightarrow{h_t} &= F\left(e_t^w, \overrightarrow{h_{t-1}}\right),\\ \overleftarrow{h_t} &= F\left(e_t^w, \overleftarrow{h_{t+1}}\right),\end{aligned} \tag{8}$$

and the two hidden states are concatenated to obtain the final output as follow:

$$h_t = \left[\overrightarrow{h_t}, \overleftarrow{h_t}\right]. \tag{9}$$

3.2 Multi-channel CNN

Due to professional terminologies existing in the medical text, adopting the simple neural network based CRF models to medical IE is not enough. Many studies tried to introduce side information and domain-specific knowledge into the neural network based models. However, they usually neglect the traditional attributes of text, which is important for medical IE, such as syntactic attribute. Therefore, we propose to integrate multiple attributes of words including syntactic, morphological, and semantic information into existing IE models.

To achieve this, we need to consider how to model context relations among multiple attributes. In previous IE studies, CNN [16] has been mainly used to encode character-level representation, and the benefit of CNN has been proved. Inspired by these studies, to model relations among attributes flexibly and effectively, a multi-channel CNN structure is construct in parallel with capturing semantic by LSTM. A CNN structure for an attribute is defined as a channel. Figure 2 shows the CNN structure of one channel.

Formally, given a multi-attribute sequence $m_i = \{m_{i,1}, m_{i,2}, \cdots, m_{i,k}\}$ of word x_i, where k denotes the number of channels and attribute categories. Firstly, we map each category of attributes to an embedding vector $e_{i,l}^m \in R^{d_m^l}$, where $e_{i,l}^m$ and d_m^l represent the l-th attribute embedding of word x_i and its dimension respectively.

Define $e_l^m = \left\{e_{1,l}^m, e_{2,l}^m, \cdots, e_{n,l}^m\right\}$ to represent the l-th attribute embedding of a word sequence and extend its outer border to a padded sequence

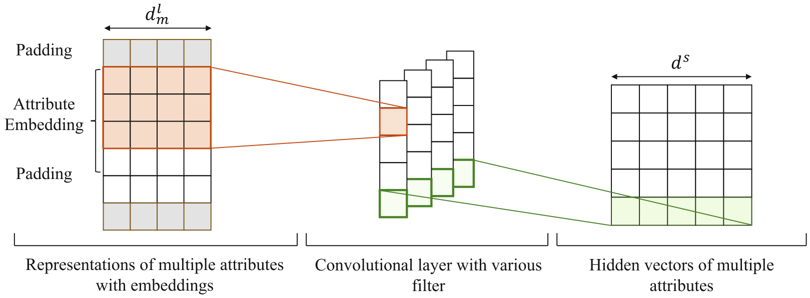

d_m^l
Padding
Attribute Embedding
Padding
d^s
Representations of multiple attributes with embeddings
Convolutional layer with various filter
Hidden vectors of multiple attributes

Fig. 2. The CNN structure of one channel.Different channels of CNN are used to capture context relations between different types of attributes.

$\{0_{\lfloor 1 \rfloor}, \cdots, 0_{\lfloor \frac{\beta}{2} \rfloor}, e_{1,l}^m, \cdots, e_{n,l}^m, 0_{\lfloor 1 \rfloor}, \cdots, 0_{\lfloor \frac{\beta}{2} \rfloor}\}$, where β is the size of CNN windows. The hidden representation of the l-th attribute can be obtained as follow:

$$s_{i,l} = W_l r_{i,l} + b_l, \tag{10}$$

$$r_{i,l} = \left\{ e_{i-\lfloor \frac{\beta}{2} \rfloor, l}^m, \cdots, e_{i,l}^m, \cdots, e_{i+\lfloor \frac{\beta}{2} \rfloor, l}^m \right\}, \tag{11}$$

where $r_{i,l}$ denotes the l-th attribute embedding of the current word, its left neighbors, and its right neighbors in convolution window. $s_{i,l}$ is the hidden representation of $e_{i,l}^m$. W_l and b_l are the network parameters which donate the weight matrices and bias vectors of l-th attribute. To control the proportions of various attributes in the final representations flexibly, we adopt a variety of weight matrices and bias vectors for different categories of attributes.

Finally, we obtain the complete representation of the i-th word by concatenating multi-attributes of each word and the hidden word vector as follow:

$$z_i = [h_i, s_{i,1}, s_{i,2}, \cdots, s_{i,k}]. \tag{12}$$

3.3 CRF

Sequence labeling for IE can be considered as a special classification problem. However, we cannot intuitively use a classifier because there are some dependencies across the output labels that can be overlooked by the classifier. For example, in NER tasks with BIO tagging scheme, I-LOC should not follow B-ORG.

Accordingly, we model the label sequence jointly with a CRF layer. Instead of predicting each label independently, CRF can model the relations between adjacent labels with a transition score and learn the interactions between a pair of token and label with a state score. Formally, given a hidden representation sequence $Z = \{z_1, z_2, \cdots, z_n\}$, and an output label sequence $Y = \{y_1, y_2, \cdots, y_n\}$, CRF is used to model the conditional probability $P(Y \mid X)$. The matrix of transition scores can be denoted by $A \in R^{k \times k}$ where k is the number of distinct labels, and the matrix of state scores can be denoted by $P \in R^{n \times k}$. Thus, we define the probability of a tag sequence as follows:

$$S(Z, y) = \sum_{i=1}^{n} A_{y_{i-1}, y_i} + \sum_{i=1}^{n} P_{i, y_i}, \tag{13}$$

$$p(y \mid Z) = \frac{\exp(S(Z, y))}{\sum_{\widetilde{y} \in Y_Z} \exp(S(Z, \widetilde{y}))}, \tag{14}$$

where Y_Z represents all possible label sequences for input Z. During training, we use the maximum conditional likelihood estimation for parameters learning:

$$\log(p(y \mid Z)) = S(Z, y) - \log \sum_{\widetilde{y} \in Y_Z} \exp(S(Z, \widetilde{y})). \tag{15}$$

While predicting, we search the output sequence which obtains the maximum conditional probability given by:

$$y^* = \text{argmax}_{\widetilde{y} \in Y_Z} S(Z, \widetilde{y}), \tag{16}$$

with the Viterbi algorithm.

Finally, we construct our MC-BLSTM-CRF model with the above three layers. For each word, the hidden representation is obtained by a BLSTM layer, and the hidden vectors of multiple attributes are computed by a multi-channel CNN layer. On the top of the two layers, we integrate the hidden representations of word and multiple attributes, and feed the output vector into the CRF layer to jointly decode the best label sequence. We use the loss function of label predicting as the overall optimization objective.

4 Experiments

4.1 Datasets

To validate the effectiveness of our proposed MC-BLSTM-CRF model, we test the performance of our model on Chinese medical IE task. For this task, we construct a medical dataset with data crawled from an online medical platform, haodf.com. For our research, we design five types of labels for medical entities: disease (D), symptom (S), medicine (M), treatment (T), and check (C). Detailed statistics of the dataset are shown in Table 1. Considering that lengths of most entities are short, we adopt the BIO (Beginning, Inside, Outside) tagging scheme.

To ensure the reliability of our experimental results, we divide the dataset into training set, validation set and test set according to the proportion of $4:1:1$, and use the 5-fold cross-validation.

Table 1. Statistics of the dataset. #Sent and #Token represent the number of sentences and words. #Entities, A#Entities and Avg L denote the total number of entities in each category, the average number of entities in each sentence and the average length of entities.

Diseases	**#Sent**	**#Tokens**
Gastritis	580	39312
Lung Cancer	513	34810
Asthma	690	45445
Hypertension	574	38564
Diabetes	560	31514

(a) Overall statistics of dataset.

Category	**#Entities**	**A#Entities**	**Avg L**
Disease	989	1.604	5.258
Symptom	1314	1.482	3.925
Medicine	1608	1.589	4.425
Treatment	500	0.590	4.694
Check	657	0.905	4.722

(b) Statistics of entities.

4.2 Attributes

In order to use the traditional attributes to help our model extract more meaningful contents better, we extract a series of attributes of raw data which are simple and easily accessible. The multiple attributes extracted by us can be divided into five categories as following, which cover three aspects of lexical, morphological, and domain-specific information.

- **POS attributes:** We use Ansj, which is an open source toolkit, to extract POS attributes. This kind of attributes represent the syntax information.
- **English acronym attributes:** In Chinese medical text, some professional terminologies are represented as English acronyms. And these acronyms can always provide some important information. Accordingly, we use "yes" or "no" to mark whether a word contains English acronyms.
- **Digital attributes:** Similar to English acronym, digits also play a unique role in sentences. We use "yes" or "no" to express if a word contains digits.
- **Suffix attributes:** In English IE, the suffix of words is often used to improve recognition performance, and the existing studies have proved the effectiveness of this operation. For Chinese medical IE, suffix information is also important. Thus, we choose the last characters of words as suffix attributes.
- **Body attributes:** Through observation, we found that medical entities are often related to body parts. Therefore, we build a dictionary of body parts and use "yes" or "no" to characterize the body attributes (whether

4.3 Experiment Setting

In our experiments, all embeddings are randomly initialized, and the dimensions of word embeddings, POS embeddings and suffix embeddings are set to 100, 40 and 50 respectively. Meanwhile, embeddings of English character, digital and body attributes are represented by One-Hot encoding with two dimensions. All the weight matrices are randomly initialized by sampling from (0.1, 0.1), and all the biases are initialized to zero. The size of the hidden units for word LSTM is set to 200, while the numbers of multi-channel CNN kernels are 90, 50, 2, 2 and

2 corresponding to POS, suffix, English character, number and body attributes respectively. The batch size and max iterations are set to 50 and 100.

We chose Adam as the optimization method for training with learning rates 0.015. And we adopt a dropout rate of 0.5 to mitigate overfitting.

For evaluation, we calculate the precision, recall, and F1 score with full word matching based method. This method means that an entity is correct only if all words of this entity are correctly predicted.

4.4 Experiment Results

To verify the performances of MC-BLSTM-CRF, we compare it against a variety of representative baselines. We can divide the baselines into two categories: traditional methods and neural network based methods. Details of baselines are shown as follows:

- **CRF:** CRF is a classical method for IE. For tagging processing, it can capture the transfer relations between labels, and relations between tokens and labels. With this method, we need to manually design features.
- **BLSTM:** LSTM is a variant of RNN network. It can capture long-term distance semantic relations among tokens automatically without any manual features as input. And then, the hidden vectors will be fed into a softmax layer for tag prediction. To learn forward and backward semantic information, we adopt bidirectional LSTM (BLSTM).
- **CNN:** CNN can capture the semantic relations between tokens by convolution. Compared with LSTM, it has better parallelism and flexibility. But it can only capture the semantic context feature in a certain window around a given token. We also construct a softmax layer to predict labels.
- **BLSTM-CRF:** BLSTM-CRF is a model that uses CRF to replace the softmax layer for labeling.
- **CNN-CRF:** This method is an extension of the CNN method which replaces the softmax layer with a CRF layer to retain the label relations.

Table 2 illustrates the performance of our models and baseline models on the Chinese medical dataset, and lists the best result on valid set (Dev), test set (Test) and the test result corresponding to the best valid set result (Dev-Test). We can see that BLSTM-CRF and CNN-CRF models outperform CRF model with 0.5% and 2.3% on Test F1 score, which proves that the neural network can capture semantic features effective without heavy hand-crafted features.

Meanwhile, both the BLSTM-CRF and CNN-CRF models are superior to BLSTM and CNN architectures with 0.45% on Dev-Test F1 score respectively, because the CRF layer can learn the transfer relations among output labels besides the relations between states and labels.

Despite BLSTM and CNN can both capture effective semantic information of words, the performance of CNN is still lower than BLSTM with about 2.0% on Dev-Test F1 score. The reason is that CNN can only consider the context within a certain kernel window, while BLSTM can retain all of the important

Table 2. Performance comparison among MC-BLSTM-CRF and baselines.

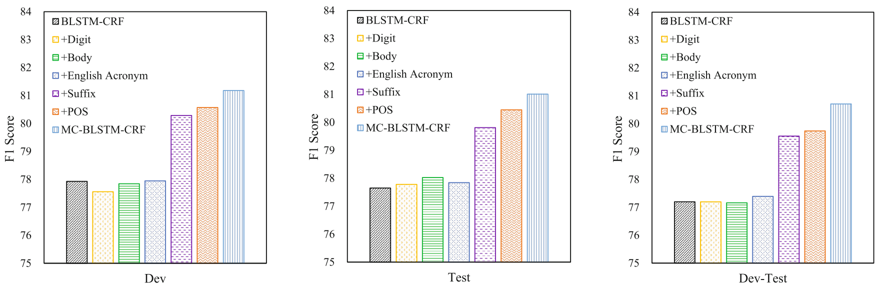
Model	Dev			Test			Dev-Test		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
CRF	-	-	-	81.49	70.11	75.36	-	-	-
CNN	79.64	71.87	75.55	78.74	71.54	74.93	78.80	70.86	74.61
CNN-CRF	80.08	72.67	76.18	79.68	72.33	75.81	79.07	71.50	75.06
BLSTM	79.28	74.86	76.99	79.35	75.10	77.12	79.34	74.40	76.75
BLSTM-CRF	81.28	74.87	77.93	80.25	75.22	77.65	80.17	74.49	77.20
Ours Model	**82.82**	**79.60**	**81.18**	**82.56**	**79.55**	**81.01**	**82.45**	**79.05**	**80.71**

information in sentences with the memory cell. Therefore, this proves the validity of using LSTM for word-level encoding in our model.

Furthermore, compared with the baselines, we can find that the MC-BLSTM-CRF model proposed by us achieves higher scores with at least 3.2% on Dev, Test and Dev-Test F1 score. Therefore, it demonstrates that the attribute features captured by the multi-channel convolution layer can supplement extra information which is important to improve the performance.

4.5 Effectiveness Analysis

To demonstrate the effectiveness of multi-channel attribute embedding, we experiment with different attributes. Figure 3 shows the results. We can find that POS and suffix attributes play a important role in improving the performance of the model, while the English character, digital and body attributes have a slight promotion. We suppose that there are two reasons. Firstly, the POS and suffix attributes have more significant and diversified information. We extract 91 types of POS attributes and 11000 types of suffix attributes while other attributes are only "yes" or "no" value. And through a statistical analysis, we found that these two attributes have more indicative effect for information



(a) Dev (b) Test (b) Dev-Test

Fig. 3. Attribute analysis of MC-BLSTM-CRF.

extraction. For instance, drugs usually appear after a verb or punctuation mark and ends with “ 片 (tablet)”, “ 囊 (capsule)”, and so on. Secondly, in the process of encoding, the semantic representation of POS and suffix attributes are richer. The embeddings of POS and suffix attributes are vectors with 40 and 50 dimensions respectively while embeddings of other attributes are one-hot vectors with 2 dimension.

In general, the results prove that our method can use the traditional attributes to enhance the semantic information which is helpful to the performance. Furthermore, attribute fusion can bring a great promotion to the model.

4.6 Case Study

For better understand what information learned from the multi-channel CNN layer can be supplemented to the BLSTM layer, we randomly pick out three sentences from our dataset and compare the recognition results of MC-BLSTM-CRF model with LSTM-CRF.

According to Table 3, we can find that the recognition result of our model is obviously better than that of BLSTM-CRF, and the improvement is threefold: (1) the optimization of recognition boundary, (2) the recognition of medical terminologies, (3) the recognition of symptom descriptions. For example, in sentence 1, “ 下肢血管b超检 (Ultrasonic artery examination)” can be recognized by MC-BLSTM-CRF, while BLSTM-CRF can introduce the noise word “ 先行 (First)”. For sentence 2 and 3, MC-BLSTM-CRF can find more medical terminologies like “ 吉法酯片 (Gefarnate Tablets)” and symptom descriptions

Table 3. Case study.

Sentence	BLSTM-CRF	MFLSTM-CRF

建议先行下肢血管b超检查，看看有无动脉狭窄。
It is suggested that B Ultrasound should be checked first to see if there is any artery stenosis.
	先行下肢血管b超检查(C)
	First ultrasonic artery examination(C)
	动脉狭窄(D)
	Artery stenosis(D)
		下肢血管b超检查(C)
		Ultrasonic artery examination(C)
		动脉狭窄(D)
		Artery stenosis(D)

医生说是肠炎，给我开了腹可安，不管用，肚子经常咕噜响。
The doctor said it was enteritis and gave me Fukean Tablets. But it wasn’t useful and my stomach often grunted.
	肠炎(D)
	Enteritis(D)
		肠炎(D)
		Enteritis(D)
		腹可安(M)
		Fukean Tablets(M)
		肚子经常咕噜响(S)
		Stomach often grunted(S)

服用说明：雷贝拉唑钠胶囊每日1次1次1片，吉法酯片一天3次一次2片。
Instructions: Rabeprazole Sodium Enteric-coated Capsules, one time a day and one piece a time. Gefarnate Tablets, three times a day and two pieces a time.
	雷贝拉唑钠胶囊(M)
	Rabeprazole Sodium Enteric-coated Capsules (M)
		雷贝拉唑钠胶囊(M)
		Rabeprazole Sodium Enteric-coated Capsules (M)
		吉法酯片(M)
		Gefarnate Tablets(M)

like “ 肚子经常咕噜响 (Stomach often grunted)” consisting of multi-words than BLSTM-CRF.

5 Conclusion

In this paper, we proposed a attribute embedding based MC-BLSTM-CRF model for medical IE task. The main contribution of this model is to capture relations between attributes effectively and flexibly with a multi-channel CNN layer and use these attributes to improve recognition performance. Experimental results showed that our model outperforms the existing methods for IE. Meanwhile, the case study results showed our model's capability of learning domain-specific information which is helpful to improve recognition performance.

Acknowledgement. This research is supported by the National Natural Science Foundation of China under the grant No. U1633103 and 61502499, the Science and Technology Planning Project of Tianjin under the grant No. 17ZXRGGX00170, the Natural Science Foundation of Tianjin under the grant No. 18JCYBJC15800, and the Open Project Foundation of Information Technology Research Base of Civil Aviation Administration of China under the grant No. CAAC-ITRB-201601.

References

1. Bodnari, A., Deléger, L., Lavergne, T., Névéol, A., Zweigenbaum, P.: A supervised named-entity extraction system for medical text. In: Working Notes for CLEF 2013 Conference (2013)
2. Chalapathy, R., Borzeshi, E.Z., Piccardi, M.: An investigation of recurrent neural architectures for drug name recognition. In: Proceedings of the Seventh International Workshop on Health Text Mining and Information Analysis, pp. 1–5 (2016)
3. Chiu, J.P.C., Nichols, E.: Named entity recognition with bidirectional LSTM-CNNs. Computer Science (2015)
4. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.P.: Natural language processing (almost) from scratch. J. Mach. Learn. Res. **12**, 2493–2537 (2011)
5. Dong, X., Chowdhury, S., Qian, L., Guan, Y., Yang, J., Yu, Q.: Transfer bi-directional LSTM RNN for named entity recognition in Chinese electronic medical records. In: 19th IEEE International Conference on e-Health Networking, Applications and Services, pp. 1–4 (2017)
6. Hassan, H., Awadallah, A.H., Emam, O.: Unsupervised information extraction approach using graph mutual reinforcement. In: EMNLP, pp. 501–508 (2006)
7. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
8. Hu, Z., Ma, X., Liu, Z., Hovy, E.H., Xing, E.P.: Harnessing deep neural networks with logic rules. In: Proceedings of ACL (2016)
9. Huang, Z., Xu, W., Yu, K.: Bidirectional LSTM-CRF models for sequence tagging. Computer Science (2015)
10. Irsoy, O., Cardie, C.: Opinion mining with deep recurrent neural networks. In: Proceedings of EMNLP, pp. 720–728 (2014)

11. Jochim, C., Deleris, L.A.: Named entity recognition in the medical domain with constrained CRF models. In: Proceedings of ACL, pp. 839–849 (2017)
12. Kudoh, T., Matsumoto, Y.: Use of support vector learning for chunk identification. In: CoNLL, pp. 142–144 (2000)
13. Lafferty, J.D., McCallum, A., Pereira, F.C.N.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proceedings of the Eighteenth International Conference on Machine Learning, pp. 282–289 (2001)
14. Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., Dyer, C.: Neural architectures for named entity recognition. In: NAACL, pp. 260–270 (2016)
15. Luo, G., Huang, X., Lin, C., Nie, Z.: Joint entity recognition and disambiguation. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 879–888 (2015)
16. Ma, X., Hovy, E.H.: End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. In: Proceedings of ACL (2016)
17. Passos, A., Kumar, V., McCallum, A.: Lexicon infused phrase embeddings for named entity resolution. In: Proceedings of the Eighteenth Conference on Computational Natural Language Learning, pp. 78–86 (2014)
18. Ratinov, L., Roth, D.: Design challenges and misconceptions in named entity recognition. In: Proceedings of the Thirteenth Conference on Computational Natural Language Learning, pp. 147–155 (2009)
19. Segura-Bedmar, I., Martínez, P., de Pablo-Sánchez, C.: Using a shallow linguistic kernel for drug-drug interaction extraction. J. Biomed. Inform. **44**(5), 789–804 (2011)
20. Takaki, O., Murata, K., Izumi, N., Hasida, K.: A medical information retrieval based on retrievers' intentions. In: HEALTHINF 2011 - Proceedings of the International Conference on Health Informatics, pp. 596–603 (2011)
21. Zeng, D., Sun, C., Lin, L., Liu, B.: LSTM-CRF for drug-named entity recognition. Entropy **19**(6), 283 (2017)