DUTIR at the CCKS-2019 Task1: Improving Chinese Clinical Named Entity Recognition using Stroke ELMo and Transfer Learning

Nan Li, Ling Luo, Zeyuan Ding, Yawen Song, Zhihao Yang* and Hongfei Lin

College of Computer Science and Technology, Dalian University of Technology, Dalian, China 116024
*Corresponding Author: yangzh@dlut.edu.cn

Abstract. As a fundamental task in medical information extraction, clinical named entity recognition (CNER) has received much attention. To accelerate the development of Chinese CNER, the 2019 China conference on knowledge graph and semantic computing (CCKS) challenge organized a task for Chinese CNER. This paper describes our neural network approach for the task. In the approach, a novel stroke ELMo is proposed to learn contextualized character embeddings from the language model. Then the character-based BiLSTM-CRF model is used to recognize clinical entities. In order to make full use of the existing related data, transfer learning is used to further improve the performance. On the official test set, our best submission achieves the F-scores of 85.16% and 92.92% under the "strict" and "relaxed" criteria, respectively.

Keywords: Entity Recognition, Chinese Clinical Text, Stroke ELMo, Transfer Learning.

1 Introduction

In recent years, the medical information processing has become a popular research focus as the generation of larger amount of electronic medical records and the potential requirements for medical information services and medical decision supports. As a fundamental task for medical information extraction, clinical named entity recognition (CNER) has received much attention and has been organized as a shared-task in many challenges [1-3]. To promote the performance of CNER on the Chinese clinical text, the 2019 China conference on knowledge graph and semantic computing (CCKS 2019) organized the CNER task to identify and extract the related medical clinical entities (i.e., anatomy, disease, imaging examination, laboratory examination, drug and operation) from Chinese clinical text.

Named entity recognition (NER) is a fundamental task for information extraction, and many NER methods have been proposed. For the English NER task, several similar neural network architectures [4-6] have been proposed. Among others, the model of bidirectional Long Short-Term Memory with a conditional random field layer (BiLSTM-CRF) exhibits promising results [5]. Different from English, there is no clear word boundary information (e.g., the space in English text) in Chinese text. Therefore, Chinese NER methods can be roughly divided into two categories (i.e., word-based method and character-based method) according to the way of text segmentation. Most previous works show that character-based methods outperform word-based methods for Chinese NER since the latter suffers the potential issue of error propagation of word segmentation.

For the Chinese NER task in medical domain, most Chinese CNER works focus on exploring the additional features (i.e., domain dictionary feature and Chinese characters feature) to improve the performance of the character-based NER models. For example, Ji et al. [7] used drug dictionary and post-processing rules to improve the performance of the BiLSTM-CRF model for Chinese CNER. Wang et al. [8] proposed five different ways of adding dictionary features to improve the model performance. Yang and Huang [9] developed a CRF model based on rich features (i.e., char embedding, POS, radical, PinYin, dictionary and rule features), which achieves the F-score of 89. 26% and ranked the first in the CCKS 2018 CNER task. However, these methods depend on effective feature engineering, i.e., the design of effective features using various NLP tools and knowledge resources, which is still a labor-intensive and skill-dependent task.

In this paper, we describe our CNER approach based on the neural network for the CCKS 2019 CNER task. Compared with other methods, the key advantage of our method is that need less feature engineering. In the approach, first a novel stroke ELMo is proposed to learn contextualized character embeddings from Chinese stroke information. The stroke ELMo is learned from a language model, which is pre-trained on a large Chinese medical text. Then the character-based BiLSTM-CRF model is used to recognize clinical entities. In addition, transfer learning with the CCKS 2018 corpus is used to further improve the performance. The experimental results show that adding our stroke ELMo and transfer learning can improve the performance of the model. Our method achieves the state-of-the-art result on the CCKS 2019 CNER dataset. This paper presented a Chinese clinical named entity recognition method, which combines Chinese stroke ELMo and Transfer Learning. On the CCKS 2019 official test set, our best submission achieves the F-scores of 85.16% and 92.92% under the “strict” and “relaxed” criteria, respectively.

2 Method

In this section, our approach for CNER is described. Fig. 1 shows the processing flow of our method. Firstly, some preprocessing steps including sentence splitting, and stroke generation are performed. Secondly, a traditional character embedding is learned with large amounts of unlabeled data by the cw2vec[1] tool. Moreover, the Chinese stroke ELMo is introduced into the model. Then with the embeddings as input, the BiLSTM-CRF model is trained by the annotated training set. In order to make full use of the existing related data, we used the CCKS 2018 corpus for transfer

[1] https://github.com/bamtercelboo/cw2vec

learning. Finally, we trained five models with different random seeds, and the results from these models are combined by majority voting. In addition, some post-processing steps are employed. The detailed description of our method is presented in the following sections.

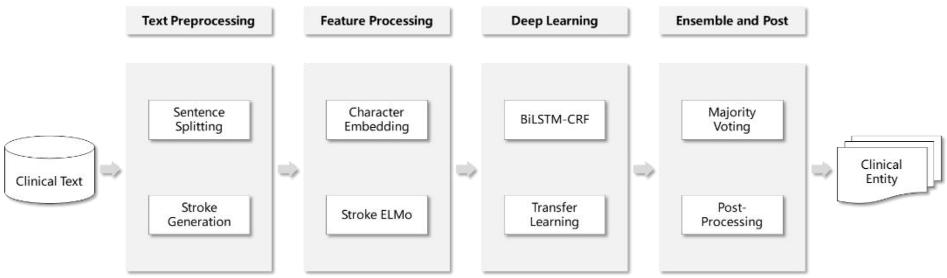

Fig. 1. The processing flow of our method

2.1 Features

Traditional Character Embedding

The character embedding is used as the basic features of our method since the character-based methods outperform word-based methods for Chinese NER [10, 11]. To achieve a high-quality character embeddings, we downloaded the medical abstracts from Chinese CNKI[2] and collected the Chinese electronic medical record texts provided by CCKS, a total of 1,568,458 texts, as unlabeled data. Then these texts were used to train 200-dimensional character embedding by the cw2vec tool as pre-trained traditional character embeddings.

Stroke ELMo

Above-mentioned traditional character embedding only learns a global vector representation for a character. However, a character can have completely different senses or meanings in the contexts. For example, in the sentence “取病理提示胃恶性肿瘤，给予奥沙利铂+**多**西他赛方案化疗，近期患者精神可，大便次数**多**，小便正常，无腹痛、腹胀等不适。”, the character senses of “多” are different depending on its context. Reasonably, the character “多” based on its two different word senses should want two different vector representations. Recently, Embedding from Language Model (ELMo) [12] is one such method that provides deep contextual embeddings. The original ELMo is proposed for English text and it produces word embeddings for each context where the word is used, thus allowing different representations for varying senses of the same word. Che et al. [13] applied ELMo to multiple languages, including Chinese. They firstly spilt the text into the words by Chinese word segmentation tool, and then used the ELMo model to obtain the contextualized word embedding. However, due to the word segmentation error, the Chinese word ELMo did not significantly improve the model performances in their

[2] http://www.cnki.net/

tasks. Different from their method, we proposed a stroke ELMo model to learn the contextualized Chinese character embedding. Fig. 2 shows the architecture of the stroke ELMo.

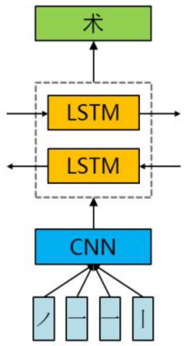

Fig. 2. The architecture of stroke ELMo model

Inspired by the English character vector, we construct the stroke ELMo model using the Chinese stroke sequence information. Our intuition is that the semantic related Chinese characters usually have the similar internal structure. For example, the characters of “森”, “林”, “木” are composed of different numbers of “木”. “呕” and “吐” are related with the character of “口”. Therefore, the basic stroke sequence of the Chinese character is used as the language model input. Then the language model is pre-trained on the large-scale corpus to learn the internal structure relationship. Concretely, first the stroke sequence of every character in corpus vocabulary is obtained with HanDian[3] website. Then the stroke embeddings are fed into a convolutional layer. And the max pooling layer is used to extract features from the convolution layer. Next, these feature vectors are fed into a BiLSTM-based bidirectional language model (biLM). After the pre-training of the biLM with above-mentioned unlabeled data, stroke ELMo extracts the intermediate layer representations from the biLM and performs a linear combination to obtain the 512-dimensional contextualized character embeddings.

2.2 BiLSTM-CRF Model

Similar to many NER tasks, we modeled the CNER as a sequence labeling problem. And the BIOES (i.e., Begin, Inside, Outside, End and Single) tagging scheme is used. We employed the BiLSTM-CRF model for the CNER task, whose architecture is illustrated in Fig. 3. Firstly, a sentence is represented as a sequence of embeddings. Next, the embeddings are given as input to a BiLSTM layer. In the BiLSTM layer, a forward LSTM computes a representation of the sequence from left to right, and another backward LSTM computes a representation of the same sequence in reverse. These two distinct networks use different parameters, and then the representation of a word is obtained by concatenating its left and right context

[3] http://www.zdic.net

representations. Then a tanh function on top of the BiLSTM layer is used to learn higher features. Finally, the CRF layer is added to predict the best label sequence path in all possible tag paths.

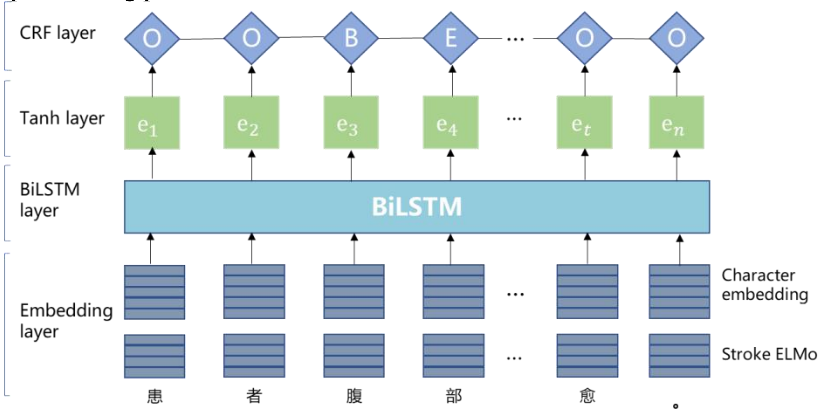

Fig. 3. The architecture of BiLSTM-CRF model

2.3 Transfer Learning

In this section we describe in details the transform learning used in our model. In order to make full use of the existing related data, transfer learning with the CCKS 2018 corpus is used to further improve the performance. The data of CCKS2018 is a CNER task, but the extract entity types are not same as the task of CCKS2019. CCKS2018 need to identify and extract the related medical clinical entities (i.e., anatomy, symptom, independent symptom, drug and operation) from Chinese clinical text. Although the labeling rules are different, they are relevant in the anatomy, medicines and operation. So we used CCKS 2018 CNER as the source domain and CCKS 2019 CNER as the target domain for transform learning. In the context of deep learning, transfer learning usually presupposes a neural architecture and sharing of some or all of its parameters across tasks. We explored three different architectures of transfer learning model for CNER, as shown in Fig. 4.

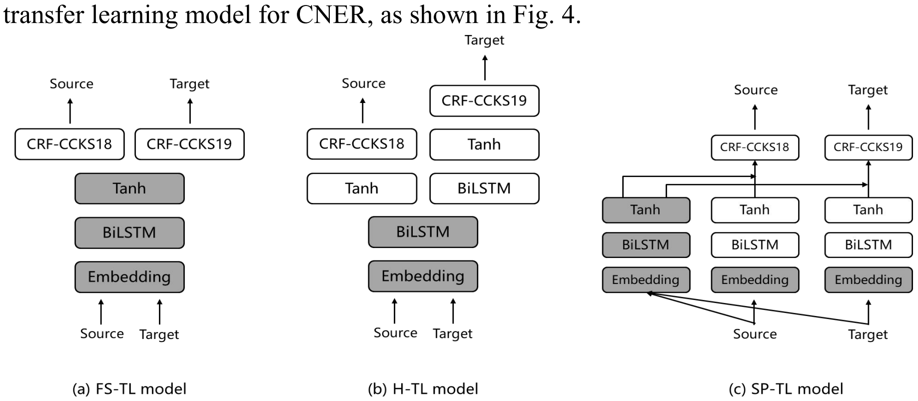

Fig. 4. The architectures of transfer learning models

Fully-shared transfer learning model (FS-TL). In the FS-TL model (as shown in Fig. 4(a)), all parameters of the general model except the output layer are shared. Each task has a task-specific CRF layer which makes the prediction based on the representation produced by the final tanh layer.

Hierarchical transfer learning model (H-TL). In the H-TL model (as shown in Fig. 4(b)), all parameters of the embedding layer and the first BiLSTM layer are shared. Each task has the task-specific tanh layer and CRF output layer, and CCKS 2019 task has one more task-specific BiLSTM layer.

Shared-private transfer learning model (SP-TL). In the SP-TL model (as shown in Fig. 4(c)), in addition to the shared embedding, BiLSTM and tanh layers, each task has the task-specific BiLSTM and tanh layers. Then representations of the shared part and the task-specific part are concatenated and passed to the task-specific CRF output layer. This architecture enables the model to selectively utilize the shared and task-specific information.

2.4 Ensemble and Post-Processing

To alleviate the influence produced by the random initialization of the model, we trained five models with different random seeds and used a majority voting approach to combine the predicted results. In addition, for performance optimization, we also employed several common post-processing steps including tagging consistency, deleting space and bracket balance.

3 Experiments

3.1 DataSet

In the CCKS-2019 CNER challenge, organizers provided a corpus including the training and test sets. The training set consists of 1000 medical records annotated with six categories of entity, including anatomy, disease, imaging examination, laboratory examination, drug and operation. And the test set consists of 379 medical records. In our experiments, we randomly selected the 20% of the training set as the development set to tune the hyper-parameters.

3.2 Evaluation

The evaluation metrics of this task include two criteria: 1) strict metrics which define a correct match as that the ground truth and extraction result share same mention, same boundaries and same entity type; 2) relaxed metrics which only consider the ground truth and the result have same entity type and overlap boundaries. All our evaluations were performed on the official test set using the evaluation tool of CCKS-2019 CNER challenge, which outputs micro-average precisions (Prec.), recalls (Rec.) and F-scores (F) via the strict metrics.

3.3 Experimental Results

In addition to the above-mentioned models, BERT [14] and Lattice LSTM [15] models are also tested for the CNER task. Table 1 shows the performances of various models on our development set. The result shows that our basic model (i.e., BiLSTM-CRF) achieves better performance than the BERT model. The main reason is that the BERT model is pre-trained on the text corpus of general domain. When our stroke ELMo is added into the basic model, the model achieves an improvement of 1.74% in F-score. It demonstrates that our stroke ELMo is effective for the CNER task.

In addition, several transfer learning models all achieve better performances than the models without transfer learning. Among the transfer learning models, FS-TL model achieves the highest F-score of 85.60%. It shows that CCKS 2018 corpus is helpful for the CCKS 2019 CNER task and our transfer learning method is effective. Finally, the ensemble of five FS-TL models with different random seeds achieves the highest F-score of 86.03%.

Table 1. Results of various models on our development set

Models	Anatomy	Disease	Image	Drug	Operation	Laboratory	Overall
BERT	80.11	77.84	84.56	92.77	79.49	87.00	81.59
Lattice LSTM	82.41	79.59	**89.42**	92.37	81.33	86.44	83.38
BiLSTM-CRF	82.06	78.83	86.47	92.20	80.82	86.54	82.81
BiLSTM-CRF+ELMo	83.45	81.34	87.08	95.09	82.71	86.99	84.55
BiLSTM-CRF+ELMo(E)	83.34	81.30	87.68	95.25	83.04	87.63	84.62
H-TL	84.17	81.17	85.30	95.83	83.62	87.09	84.88
SP-TL	84.45	81.70	86.73	95.97	83.79	86.31	85.18
FS-TL	85.24	81.13	86.54	**96.52**	83.79	**88.46**	85.60
FS-TL(E)	**85.52**	**82.46**	86.87	96.24	**83.84**	87.85	**86.03**

Note: (E) denotes the ensemble model.

Table 2 shows the result of our FS-TL (Ensemble) model with post-processing on the official test set. The best submission achieves the F-scores of 85.16% and 92.92% under the “strict” and “relaxed” criteria, respectively. We analyzed the results and found that the laboratory examination entity recognition has the poor performance. Therefore, we will focus on the laboratory examination entity recognition in our future work.

Table 2. Results of our best submission on the official test set

Criterias	Anatomy	Disease	Image	Drug	Operation	Laboratory	Overall
Strict (%)	85.99	82.81	88.01	94.49	86.79	75.65	85.16
Relaxed (%)	92.88	92.41	94.22	97.14	94.34	89.38	92.92

4 Conclusion

In this paper, we present a neural network approach to automatically recognize clinical entities from Chinese clinical texts. In this approach, a novel stroke ELMo is proposed to learn the contextualized Chinese character embedding. And we also explored the effect of transfer learning for the CNER task. The experimental results show that our stroke ELMo and transfer learning are effective to improve the performance of our model. At last, our best submission achieves the F-scores of 85.16% and 92.92% under the “strict” and “relaxed” criteria, respectively. In our future work, we will focus on the more effective extraction of laboratory examination entities.

Acknowledgement

This work was supported by the grant from the National Key Research and Development Program of China (No. 2016YFC0901902).

References

1. Uzuner Ö, Solti I, Cadag E: Extracting medication information from clinical text. Journal of the American Medical Informatics Association 2010, 17(5):514-518.
2. Sun W, Rumshisky A, Uzuner O: Evaluating temporal relations in clinical text: 2012 i2b2 challenge. Journal of the American Medical Informatics Association 2013, 20(5):806-813.
3. Bethard S, Savova G, Chen W-T, Derczynski L, Pustejovsky J, Verhagen M: Semeval-2016 task 12: Clinical tempeval. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016): 2016. 1052-1062.
4. Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P: Natural language processing (almost) from scratch. Journal of Machine Learning Research 2011, 12(Aug):2493-2537.
5. Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C: Neural architectures for named entity recognition. arXiv preprint arXiv:160301360 2016.
6. Ma X, Hovy E: End-to-end sequence labeling via bi-directional lstm-cnns-crf. arXiv preprint arXiv:160301354 2016.
7. Ji B, Li S, Yu J, et al. A BiLSTM-CRF Method to Chinese Electronic Medical Record Named Entity Recognition[C]. Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence, 2018:48.
8. Wang Q, Zhou Y, Ruan T, et al. Incorporating dictionaries into deep neural networks for the chinese clinical named entity recognition [J]. Journal of biomedical informatics. 2019,92:103133.
9. Yang X, Huang W. A Conditional Random Fields Approach to Clinical Name Entity Recognition[C]. CCKS2019: China Conference on Knowledge Graph and Semantic Computing.
10. He J, Wang H: Chinese named entity recognition and word segmentation based on character. In: *Proceedings of the Sixth SIGHAN Workshop on Chinese Language Processing: 2008.*

11. Li H, Hagiwara M, Li Q, Ji H: Comparison of the Impact of Word Segmentation on Name Tagging for Chinese and Japanese. In: *LREC: 2014*. 2532-2536.
12. Peters M, Neumann M, Iyyer M, et al., editors. Deep Contextualized Word Representations[C]. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), 2018:2227-2237.
13. Che W, Liu Y, Wang Y, et al. Towards better UD parsing: Deep contextualized word embeddings, ensemble, and treebank concatenation [J]. arXiv preprint arXiv:180703121. 2018.
14. Devlin J, Chang M-W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding [J]. arXiv preprint arXiv:181004805. 2018.
15. Zhang Y, Yang J. Chinese NER Using Lattice LSTM[C]. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018:1554-1564.