

RESEARCH

Open Access



Entity recognition in Chinese clinical text using attention-based CNN-LSTM-CRF

Buzhou Tang¹, Xiaolong Wang¹, Jun Yan² and Qingcai Chen^{1*}

From The Sixth IEEE International Conference on Healthcare Informatics (ICHI 2018)
New York, NY, USA. 4-7 June 2018

Abstract

Background: Clinical entity recognition as a fundamental task of clinical text processing has been attracted a great deal of attention during the last decade. However, most studies focus on clinical text in English rather than other languages. Recently, a few researchers have begun to study entity recognition in Chinese clinical text.

Methods: In this paper, a novel deep neural network, called attention-based CNN-LSTM-CRF, is proposed to recognize entities in Chinese clinical text. Attention-based CNN-LSTM-CRF is an extension of LSTM-CRF by introducing a CNN (convolutional neural network) layer after the input layer to capture local context information of words of interest and an attention layer before the CRF layer to select relevant words in the same sentence.

Results: In order to evaluate the proposed method, we compare it with other two currently popular methods, CRF (conditional random field) and LSTM-CRF, on two benchmark datasets. One of the datasets is publically available and only contains contiguous clinical entities, and the other one is constructed by us and contains contiguous and discontinuous clinical entities. Experimental results show that attention-based CNN-LSTM-CRF outperforms CRF and LSTM-CRF.

Conclusions: CNN and attention mechanism are individually beneficial to LSTM-CRF-based Chinese clinical entity recognition system, no matter whether contiguous clinical entities are considered. The contribution of attention mechanism is greater than CNN.

Keywords: Chinese clinical entity recognition, Neural network, Convolutional neural network, Long-short term memory, Conditional random field

Introduction

With rapid development of electronic medical information systems, more and more electronic medical records (EMRs) are available for medical research and application. In EMRs, plenty of useful information is embedded in clinical text. The first step to use clinical text is clinical entity recognition that finds which words form clinical entities and which type each entity belongs to.

In the last decades, a large number of methods have been proposed for clinical entity recognition. The methods includes early rule-based methods, machine learning methods based on manually-crafted features in

past a few years and recently deep neural networks. The most popular machine learning method used for clinical entity recognition is conditional random field (CRF) [1], and the most popular deep neural network is LSTM-CRF [2]. However, most studies focus on entity recognition in English clinical text rather than other languages. It is necessary to investigate the latest methods for entity recognition in other languages, for example Chinese.

To promote development of entity recognition in Chinese clinical text, the organizers of China conference on knowledge graph and semantic computing (CCKS) launched a challenge was launched in 2017 [3]. The challenge organizer provided a dataset (called CCKS2017_CNER) with only contiguous clinical entities following the guideline of i2b2 (Informatics for

* Correspondence: qingcai.chen@gmail.com

¹Key Laboratory of Network Oriented Intelligent Computation, Harbin Institute of Technology, (Shenzhen), Shenzhen 518055, China
Full list of author information is available at the end of the article



Integrating Biology and the Bedside) challenge for English clinical text in 2010 [4]. Nearly all systems proposed for CCKS2017 challenge adopted CRF or LSTM-CRF. In addition, discontinuous clinical entities composed of discontinuous words, accounting for around 10% in English clinical text, also widely exist in Chinese clinical text. No study have ever considered discontinuous entities in Chinese clinical text.

In this study, we propose a novel deep neural network, called attention-based CNN-LSTM-CRF, for entity recognition considering both contiguous and discontinuous entities in Chinese clinical text. Attention-based CNN-LSTM-CRF is an extension of LSTM-CRF by adding two layers. A dataset (called ICRC-CNER) containing both Chinese contiguous and discontinuous entities is constructed by us (the intelligence computing research center (ICRC) of Harbin institute of technology, Shenzhen) and used to evaluate attention-based CNN-LSTM-CRF. Experiments conducted on CCKS2017_CNER and ICRC_CNER show that our proposed method outperforms CRF and LSTM-CRF. It should be stated that this paper is an extension of our previous paper [5].

Related work

Clinical entity representation is very important for recognition. As there exist contiguous and discontinuous entities in clinical text, we could not adopt named entity representation in the newswire domain directly for clinical entities. In order to represent contiguous and discontinuous clinical entities in a unified schema, Tang et al. [6, 7] extended the schemas, such as “BIO” and “BIOES” by introducing new labels for contiguous word fragment shared by discontinuous clinical entities or not, that are “BIOHD” and “BIOHD1234”. Wu et al. [8] proposed a schema, called “Multi-label” to give each word multiple labels, each one of which corresponds the label of the token in one clinical entities.

In the past several years, as a number of manually annotated corpora have been publically available for clinical entity recognition in challenges such as the Center for Informatics for Integrating Biology & the Beside (i2b2) [4, 9–11], ShARe/CLEF eHealth Evaluation Lab (SHEL) [12, 13], SemEval (Semantic Evaluation) [14–17], etc., lots of machine learning methods, such as support vector machine (SVM), hidden markov model (HMM), conditional random field (CRF), structured support vector machine (SSVM) and deep neural networks, have been applied to clinical named entity recognition. Among these methods, CRF is the most frequently used method whole performance relies on manually-crafted features, whereas deep neural networks, especially LSTM-CRF, which have ability to avoid

feature engineering, are recently introduced for clinical entity recognition. Common features, such as N-grams and part-of-speech, and domain-specific features, such as section information and domain dictionaries, are usually adopted in CRF. For LSTM-CRF, there are a few variants such as [18, 19], which extend the basic LSTM-CRF by introducing character-level word embeddings or attention mechanism.

Methods

The overview architecture of attention-based CNN-LSTM-CRF is shown in Fig. 1. It consists of the following five layers: 1) Input layer, which takes the representation of each Chinese character in a sentence; 2) CNN layer, which represents the local context of a Chinese character of interest within a sliding window (e.g. $[-1, 1]$ in Fig. 1); 3) LSTM layer, which uses a forward LSTM and a backward LSTM to model a sentence to capture global context information of a sentence; 4) Attention layer, which determines relativity strength of other Chinese characters to a Chinese character of interest; 5) CRF layer, which predicts a label sequence for an input sentence by considering relations between neighbor labels. The five layer is presented in detail in the following sections.

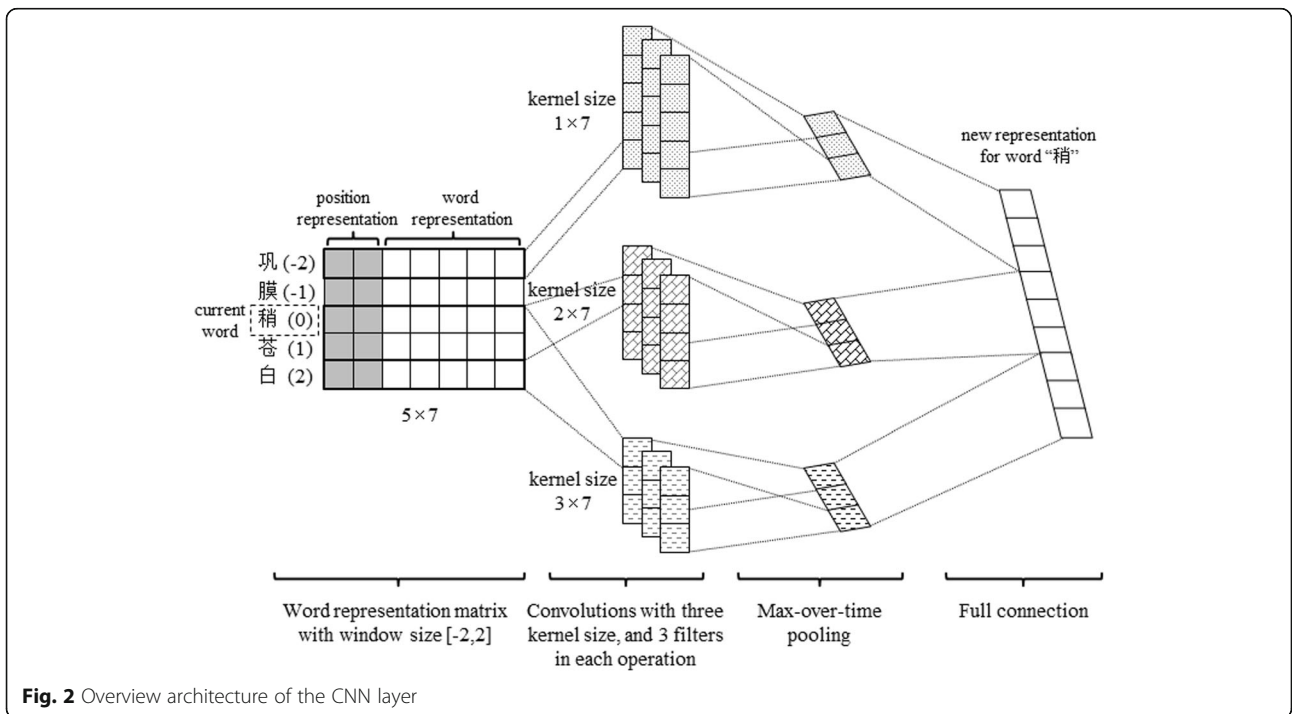
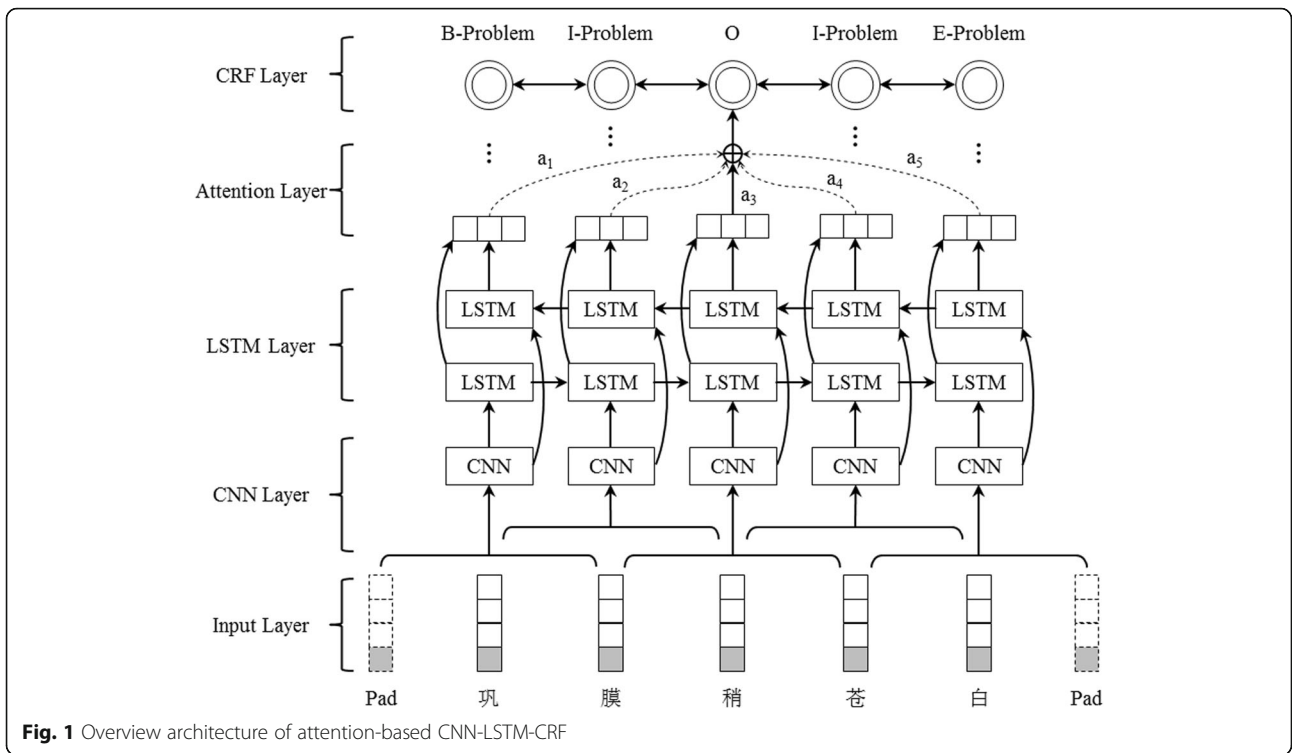
Input layer

As we all know, Chinese text processing is different from English text processing as there is no separator between words. Therefore, word segmentation is usually a first step for Chinese text processing. However, there is no publicly available Chinese word segmentation tool in the clinical domain, and Chinese word segmentation tools developed in other domains have been proved detrimental to Chinese clinical entity recognition [20]. Therefore, in this study, Chinese clinical sentences are segmented into single Chinese characters as shown in Fig. 1 (“巩膜稍苍白” – “slight pallor of the sclera” was segmented into “巩”, “膜”, “稍”, “苍”, “白”).

Formally, given a Chinese clinical sentence $s = w_0w_1\dots w_n$, where w_t ($1 \leq t \leq n$) is the t -th Chinese character, we follow the previous study [21] to represent w_t by $x_t = [c_{wt}; r_{wt}]$, where c_{wt} and r_{wt} are embeddings of w_t and its radical respectively, and ‘;’ is the concatenation operation.

CNN layer

Convolutional neural network (CNN), as shown in Fig. 2, is employed to extract local context information of a Chinese character of interest in the following four steps:



- 1) Input matrix. The context of w_t within a window of $[-m, m]$, $w_{t-m} \dots w_{t+m}$, is represented by $Q = [[x_{t-m}; p_{-m}], \dots, [x_{t+m}; p_m]]$, where p_i ($-m \leq i \leq m$) is position embeddings for the distance of w_i relative to w_t .
- 2) Convolution operation. Convolution kernels of different size M are employed for feature extraction. Suppose that there are L filters (feature maps) for each size, let $W^{(u,v)}$ ($1 \leq u \leq M, 1 \leq v \leq L$) denotes the v -th filter of size u . Then, the following convolution operation is applied on Q :

$$F_i^{(u,v)} = \sigma \left(W^{(u,v)} \otimes Q_{[i:i+k-1]} + b^{(u,v)} \right) \quad (1 \leq i \leq m-k+1) \tag{1}$$

where $F_i^{(u,v)}$ is the i -th feature extracted from context matrix Q by filter $W^{(u,v)}$, σ is the element-wise sigmoid function, \otimes is the element-wise product, and $b^{(u,v)}$ is a bias vector. All features extracted by filter $W^{(u,v)}$ can be represented as $F^{(u,v)} = (F_1^{(u,v)}, F_2^{(u,v)}, \dots, F_{m-k+1}^{(u,v)})$.

- 3) Max-pooling operation. After the convolution operation, a max-over-time pooling operation is employed on each filter to select the most significant feature as follows:

$$F_{max}^{(u,v)} = \max \left\{ F_1^{(u,v)}, F_2^{(u,v)}, \dots, F_{m-k+1}^{(u,v)} \right\} \tag{2}$$

Until now, the features corresponding to convolution kernels of size u are $F^{(u)} = (F_{max}^{(u,1)}, F_{max}^{(u,2)}, \dots, F_{max}^{(u,L)})$.

- 4) Full connection. Finally, all features outputted after max-pooling are concatenated together to represent the local context of w_b , that is $g_t = (F_t^{(1)}; F_t^{(2)}; \dots; F_t^{(M)})$.

After the CNN layer, the sentence representation becomes $g = (g_1, g_2, \dots, g_n)$.

LSTM layer

Taking $g = (g_1, g_2, \dots, g_n)$ outputted by the CNN layer as input, the LSTM layer produces a new representation sequence $h = (h_1, h_2, \dots, h_n)$, where $h_t = [h_{ft}; h_{bt}]$ ($1 \leq t \leq n$) concatenates the outputs of both forward LSTM h_{ft} and backward LSTM h_{bt} at step t . An LSTM unit is composed of one memory cell and three gates (input gate, forget gate and output gate), denoted by c_t, o_t, i_t and f_t respectively for the LSTM unit at step t . Taking g_t, h_{t-1}, c_{t-1} as

input at step t , the LSTM unit can produce h_t and c_t as follows:

$$\begin{aligned} i_t &= \sigma(W_{gi}g_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \\ f_t &= \sigma(W_{gf}g_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \\ c_t &= f_t \otimes c_{t-1} + i_t \otimes \tanh(W_{gc}g_t + W_{hc}h_{t-1} + b_c) \\ o_t &= \sigma(W_{go}g_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \\ h_t &= o_t \otimes \tanh(c_t) \end{aligned} \tag{3}$$

where σ is the element-wise sigmoid function, \otimes is the element-wise product, W_i, W_f, W_c and W_o (with subscripts: g, h and c) are the weight matrices, b_i, b_f, b_c and b_o are bias vectors.

Attention layer

An attention network, as shown in Fig. 3, is employed to determine relativity strength of other Chinese characters to the Chinese character of interest, under the assumption that the label of w_t is not determined by h_t only. For example, in a fragment “皮肤粗糙、苍白” (“hard and pale skin”), “皮肤粗糙” (“hard skin”) is a contiguous problem, and “皮肤...苍白” (“pale skin”) is a discontinuous problem with two words “皮肤” (“skin”) and “苍白” (“pale”). The word “皮肤” is not a clinical entity only when it appears with word “苍白”, which means that the label of word “皮肤” also depends on the word “苍白”.

Taking the representation sequence h outputted by the LSTM layer as input, the attention layer produces a new representation sequence $z = (z_1, z_2, \dots, z_n)$, where z_t at step t can be calculated as follows:

$$z_t = \tanh(h \cdot a_t^T) \tag{4}$$

where \tanh is the activation function, h is the representation matrix outputted by LSTM layer, a_t is the weight vector for each word in the sentence calculated as follows:

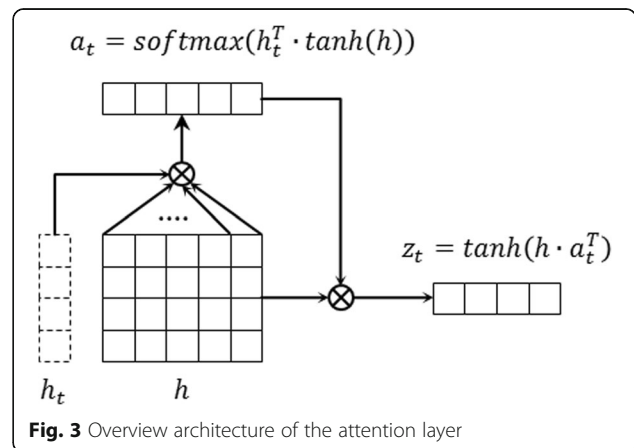


Fig. 3 Overview architecture of the attention layer

$$a_t = \text{softmax}(h_t^T \cdot \tanh(h)) \tag{5}$$

where *softmax* is the normalization function, h_t is the representation of h at step t . Finally, the new representation sequence z is applied for the label prediction in the next CRF layer.

CRF layer

The CRF layer takes sequence $z = (z_1, z_2, \dots, z_n)$ as input, and predicts the most possible label sequence $y = (y_1, y_2, \dots, y_n)$. Give a training set D , all parameters of CRF layer (denoted as θ) are estimated by maximizing the following log-likelihood:

$$L(\theta) = \sum_{(s,y) \in D} \log p(y|z, \theta) \tag{6}$$

where y is the corresponding label sequence of sentence s , p is the conditional probability of y when given s and θ . Assuming that $S_\theta(z, y)$ is the score of label sequence y for sentence, then the conditional probability p can be calculated as the normalization of $S_\theta(z, y)$. In order to take full advantage of dependencies between neighbor labels, the model incorporates a transition matrix T with an emission matrix E to calculate the score of label sequence $S_\theta(z, y)$, as follows:

$$S_\theta(z, y) = \sum_{t=1}^n (E_{y_t, t} + T_{y_{t-1}, y_t}) \tag{7}$$

where $E_{y_t, t}$ is the probability that word z_t with label y_t , and T_{y_{t-1}, y_t} is the probability that word z_{t-1} with label y_{t-1} followed by z_t with label y_t . We can maximize the log-likelihood (6) over all training set D by the dynamic programming, and find the best label sequence for any input sentence by maximizing score (7) using Viterbi algorithm.

Dataset

We evaluate the attention-based CNN-LSTM-CRF on two datasets: CCKS2017_CNER and ICRC_CNER. CCKS2017_CNER contains 400 Chinese clinical records with five categories of clinical entities, 300 records are treated as a training set and the remainder 100 records are treated as a test set. In this dataset, all clinical entities are contiguous, and the total number of them is 39,359. ICRC_CNER contains 1176 Chinese clinical records with the other five categories of clinical entities, 600 records are treated as a training set, 176 records are treated as a development set and the remainder 400 records are treated as a test set. In this dataset, both contiguous and dis contiguous clinical entities are manually annotated, and the total number of clinical

entities is 91,185. Table 1 list the statistics of the two datasets, where “#” denotes the number of “#”, and the numbers of contiguous entities and discontinuous entities in ICRC_CNER are given in separated rows (the numbers of contiguous entities in the upper rows, and the number of discontinuous entities in the lower rows).

Evaluation and experiments setup

We start from two baseline methods: CRF and LSTM-CRF, then investigate the effects of the CNN layer and attention layer respectively, and finally compare attention-based CNN-LSTM-CRF with other state-of-the-art systems on CCKS2017_CNER. Following previous studies [7, 17], clinical entities in CCKS2017_CNER are represented by “BIO”, and that in ICRC_CNER are represented by “BIOHD1234” and “Multi-label” respectively. The features utilized in CRF are the same as [21], including bag-of-words, part-of-speech, radical information, sentence information, section information, general NER, word representation, dictionary feature, etc. It should be stated that LSTM-CRF here is the same as that used in the best system of CCKS 2017 [21].

The performances of all systems are measured by micro-averaged precision, recall and F1-score under two criteria: “strict” and “relaxed”, where the “strict” criterion checks whether predicted entities exactly match with gold ones in boundary and category, while the “relaxed” criterion relaxes the condition in boundary, and only checks whether predicted entities overlap with gold ones. The “strict” measures are the primary measures.

The hyper-parameters used in LSTM-CRF and attention-based CNN-LSTM-CRF are: dimension of Chinese character embeddings-50, dimension of radical embedding-25, dimension of position embedding-20, size of convolution kernels in the CNN layer-1/2/3, number of filters of each size-32, size of LSTM unit-100, size of sliding window-[- 2,2], dropout probability-0.5 and training epochs-30. The Chinese character embeddings are pre-trained by the word2vec tool (<https://github.com/tensorflow/tensorflow/tree/r1.1/tensorflow/examples/tutorials/word2vec>) on a large unlabeled dataset provided by CCKS2017, and the radical embeddings are randomly initialized. The parameters of all deep neural network models are estimated using stochastic gradient descent (SGD) algorithm.

Results

Table 2 shows the performances of different methods on CCKS2017_CNER and ICRC_CNER, where the highest measures are in bold (the following sections also use the same way to denote the highest measures), and the

Table 1 Statistics of CCKS2017_CNER and ICRC_CNER for entity recognition in Chinese clinical text

Dataset (CCKS2017_CNER)	#Record	#Clinical Entity					
		#Body	#Disease	#Symptom	#Test	#Treatment	#All
Training (300)	300	10,719	722	7831	9546	1048	29,866
Test (100)	100	3021	553	2311	3143	465	9493
Total (400)	400	13,740	1275	10,142	12,689	1513	39,359
Dataset (ICRC_CNER)	#Record	#Clinical Entity					
		#Medication	#Disease	#Symptom	#Test	#Treatment	#All
Training	600	1293	11,470	5270	17,024	3065	38,122
		0	7441	75	7	107	7630
Development	176	475	3594	1738	5276	938	12,021
		0	2421	37	3	41	2502
Test	400	999	7932	3353	11,326	2020	35,630
		3	5153	57	6	61	5280
Total	1176	2767	22,996	10,361	33,626	6023	75,773
		3	15,015	169	16	209	15,412

performances of each method using “BIOHD1234” and “Multi-label” on ICRC_CNER are listed in separated rows (the performance measures in the upper rows correspond to “BIOHD1234”, and the performance measures in the lower rows correspond to “Multi-label”). Our method achieves highest “strict” F1-scores of 90.61% on CCKS2017_CNER and 83.32% on ICRC_CNER, outperforming CRF and LSTM-CRF by 0.44 and 0.32% respectively. All methods using “Multi-label” shows better performance than that using “BIOHD1234”.

In order to investigate effects of the CNN layer and attention layer in our method respectively, we remove one or two of them from attention-based CNN-LSTM-CRF, and present the results in Table 3, where only precisions,

recalls and F1-scores under the “strict” criterion are listed, “w/o” denotes “without”, and our method without both CNN layer and attention layer just becomes LSTM-CRF. When the CNN layer is removed from our method, the F-score slightly increases on CCKS2017, but slightly decreases on ICRC_CNER. When the attention layer is removed, the F-scores on both two datasets decreases slightly. When both CNN and attention layers are removed, the F-scores on both two datasets decreases greatly. The experimental results indicates that both CNN and attention layers are individually beneficial to LSTM-CRF, the contribution of attention layer is greater than CNN layer, but they may hurt each other some times. It may be because contiguous entities only depend on neighbor Chinese characters which are

Table 2 Performances of different methods on the two datasets: CCKS2017_CNER and ICRC_CNER

Dataset	Method	Strict (%)			Relaxed (%)		
		Precision	Recall	F1-score	Precision	Recall	F1-score
CCKS2017_CNER	CRF	91.22	88.20	89.69	95.73	92.57	94.13
	LSTM-CRF	90.68	89.67	90.17	95.18	94.12	94.65
	Our Method	90.73	90.49	90.61	94.84	94.59	94.71
ICRC_CNER	CRF	81.84	78.86	80.32	93.75	90.34	92.01
		83.42	79.90	81.62	94.02	90.05	92.00
	LSTM-CRF	83.55	82.26	82.90	93.80	92.35	93.07
		82.71	83.30	83.00	92.77	93.42	93.09
	Our Method	82.96	82.60	82.78	93.30	92.90	93.10
		82.66	83.99	83.32	92.57	94.07	93.31

Table 2 shows the performances of different methods on CCKS2017_CNER and ICRC_CNER, where the highest measures are in bold (the following sections also use the same way to denote the highest measures)

Table 3 Effects of the CNN layer and attention layer in our method

Method	CCKS2017_CNER (%)			ICRC_CNER (%)		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Our method	90.73	90.49	90.61	82.66	83.99	83.32
w/o CNN	91.11	90.48	90.79	83.72	82.53	83.12
w/o attention	90.61	90.23	90.42	83.16	83.29	83.22
w/o both	90.68	89.67	90.17	82.71	83.30	83.00

Table 2 shows the performances of different methods on CCKS2017_CNER and ICRC_CNER, where the highest measures are in bold (the following sections also use the same way to denote the highest measures)

captured by the CNN layer and attention layer repeatedly, whereas discontinuous entities depend on skipping words which may benefit from the attention layer.

Furthermore, we also compare our method with the best system of the CCKS2017 challenge, which employed several individual methods, such as rule-based method, CRF, LSTM-CRF without additional features and LSTM-CRF with additional features (the same as the baseline method LSTM-CRF used in this paper), and further used a voting method to integrate all the results of these methods. The best individual method is LSTM-CRF with additional features, which is inferior to attention-based CNN-LSTM-CRF as mentioned above (shown in Table 2). Following the same way to integrate CRF, LSTM-CRF without additional features and our method together, we obtain a “strict” F1-score of 91.46%, higher than that of the best system of the CCKS2017 challenge (i.e., 91.02%) [21].

Discussion

In order to investigate on which category of clinical entity how our method performs, we list the performance of our method on each category of clinical entity under “strict” criterion in Table 4. Our method performs well on some categories, such as “Test” and “Medication” on ICRC_CNER, “Symptom”, “Test” and “Body” on CCKS2017_CNER dataset. However, it also performs not

Table 4 Performances of our CNN-LSTM-Attention model on each category under “strict” criterion

Category	ICRC_CNER (%)			CCKS2017_CNER (%)		
	Pre.	Rec.	F1	Pre.	Rec.	F1
Disease	82.84	81.67	82.25	85.06	77.22	80.95
Symptom	77.06	76.01	76.53	94.92	96.28	95.60
Test	84.19	89.03	86.55	93.66	93.48	93.57
Treatment	77.53	79.58	78.54	77.63	79.08	83.98
Medication	87.88	89.72	88.79	/	/	/
Body	/	/	/	86.89	87.36	87.12

Table 2 shows the performances of different methods on CCKS2017_CNER and ICRC_CNER, where the highest measures are in bold (the following sections also use the same way to denote the highest measures)

very well on some categories, such as “Disease” and “Treatment” on both datasets, especially “Symptom” on ICRC_CNER dataset, which is much worse than that on CCKS2017_CNER, may because of a large number of discontinuous clinical entities in “Symptom” category on ICRC_CNER.

In previous studies, in English clinical text, recognizing discontinuous entities have been proved much more difficult than contiguous entities, and the “strict” F1-score difference on the two types of clinical entities exceeded 25% [21]. However, that difference in Chinese clinical text is around 15% as shown in Table 5. It means that discontinuous entities in Chinese clinical text is much easier than that in English clinical text. Among three method, our method achieves the highest “strict” F1-scores on both two types of clinical entities.

Although our method shows better overall performance than CRF and LSTM-CRF, it does not always achieve highest “strict” F1-score on all categories of clinical entities. Figure 4 shows the performances of different methods on each category of clinical entity. Our method achieves the highest “strict” F1-scores on all categories except “Medication” on ICRC_CNER and “Symptom” on CCKS2017_CNER. It may be caused by different guidelines. The limitations of this study are: 1) the proposed method is also applicable to entity recognition in English text, but we do not compare it on English datasets. The experiments will be conducted in the future. 2) there also some other extensions of LSTM-CRF on tasks in other domains, we do not compare them with our method in this study. Comparing our method with them and introducing their characteristics into our method to form new methods are other two cases of our future work.

Conclusions

In this study, we propose a novel deep neural network for entity recognition in Chinese clinical text, which extends LSTM-CRF by introducing a CNN layer and an attention layer. The CNN layer is used to capture local context information of the Chinese character of interest, and the attention layer is used to determine relativity

Table 5 Performances of methods on contiguous and discontinuous clinical entity under “strict” criterion on ICRC_CNER

Method	Contiguous entity (%)			Discontinuous entity (%)		
	Precision	Recall	F1-score	Precision	Recall	F1-score
CRF	83.52	84.35	83.93	82.73	58.26	68.37
LSTM-CRF	83.35	87.35	85.30	78.70	63.62	70.36
Our method	83.57	87.75	85.61	77.17	65.76	71.01

Table 2 shows the performances of different methods on CCKS2017_CNER and ICRC_CNER, where the highest measures are in bold (the following sections also use the same way to denote the highest measures)

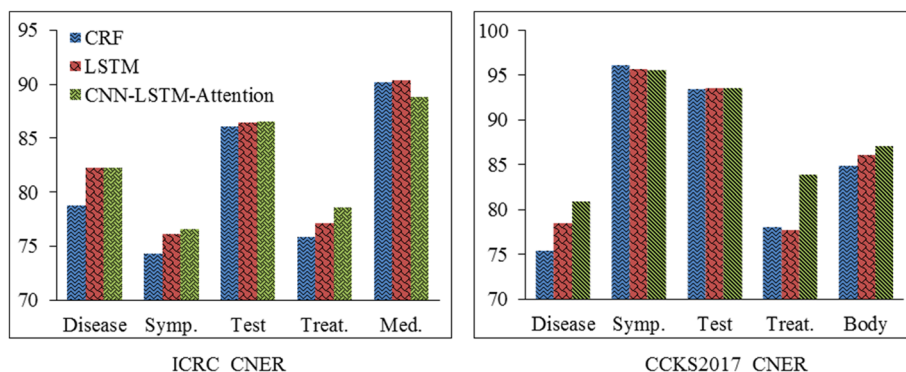


Fig. 4 "strict" F1-scores of different methods on each category of clinical text

strength of other Chinese characters to the Chinese character of interest. Experiments on two benchmark datasets shows the effectiveness of our proposed method.

Acknowledgements

Not applicable.

Funding

This paper is partially supported by the following grants: NSFCs (National Natural Science Foundations of China) (61573118, 61473101 and 61472428), Special Foundation for Technology Research Program of Guangdong Province (2015B010131010), Strategic Emerging Industry Development Special Funds of Shenzhen (JCYJ20160531192358466 and JCYJ20170307150528934) and Innovation Fund of Harbin Institute of Technology (HIT.NSRIF.2017052). Publication costs are funded by JCYJ20160531192358466 grant.

Availability of data and materials

The datasets that support the findings of this study are not available as there are many privacy information in the clinical notes, and no related act can be referred for the medical data publication in China.

About this supplement

This article has been published as part of *BMC Medical Informatics and Decision Making Volume 19 Supplement 3, 2019: Selected articles from the first International Workshop on Health Natural Language Processing (HealthNLP 2018)*. The full contents of the supplement are available online at <https://bmcmmedinformdecismak.biomedcentral.com/articles/supplements/volume-19-supplement-3>.

Authors' contributions

The work presented here was carried out in collaboration between all authors. BT and QC designed the methods and experiments, and contributed to the writing of manuscript. XW and JY provided guidance and reviewed the manuscript critically. All authors have approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Key Laboratory of Network Oriented Intelligent Computation, Harbin Institute of Technology, (Shenzhen), Shenzhen 518055, China. ²Yidu Cloud (Beijing) Technology Co., Ltd, Beijing 100191, China.

Published: 4 April 2019

References

- Lafferty J, McCallum A, Pereira FCN. Conditional random fields: probabilistic models for segmenting and labeling sequence data. 2001.
- Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv:150801991 2015.
- Li J, et al., editors. Knowledge Graph and Semantic Computing. Language, Knowledge, and Intelligence: Second China Conference, CCKS 2017, August 26–29, 2017, Revised Selected Papers. Vol. 784. Chengdu: Springer; 2018.
- Uzuner Ö, South BR, Shen S, Du Vall SL. i2b2/VA challenge on concepts, assertions, and relations in clinical text. *J Am Med Inform Assoc.* 2010;2011(18):552–6.
- Liu Z, et al. Chinese clinical entity recognition via attention-based CNN-LSTM-CRF. In: 2018 IEEE international conference on healthcare informatics workshop (ICHI-W). IEEE; 2018.
- Tang B, Wu Y, Jiang M, et al. Recognizing and Encoding Disorder Concepts in Clinical Text using Machine Learning and Vector Space Model; 2013; CLEF (Working Notes). p. 665.
- Tang B, et al. Recognizing disjoint clinical concepts in clinical text using machine learning-based methods. In: AMIA annual symposium proceedings. Vol. 2015. American Medical Informatics Association; 2015.
- Lin W, Ji D, Lu Y. Disorder recognition in clinical texts using multi-label structured SVM. *BMC Bioinform.* 2017;18(1):75.
- Uzuner Ö, Solti I, Cadag E. Extracting medication information from clinical text. *J Am Med Inform Assoc.* 2010;17:514–8.
- Sun W, Rumshisky A, Uzuner O. Evaluating temporal relations in clinical text: 2012 i2b2 challenge. *J Am Med Inform Assoc.* 2013;20:806–13.
- Stubbs A, Kotfila C, Uzuner O. Automated systems for the de-identification of longitudinal clinical narratives: overview of 2014 i2b2/UTHealth shared task track 1. *J Biomed Inform.* 2015;58:S11–9.
- UzZaman N, Llorens H, Derczynski L, et al. Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In: Second Joint Conference on Lexical and Computational Semantics (* SEM). Proc Seventh Int Workshop Semant Eval (SemEval 2013). 2013;2(2):1-9.
- Kelly L, et al. Overview of the share/clef ehealth evaluation lab 2014. In: International Conference of the Cross-Language Evaluation Forum for European Languages. Cham: Springer; 2014.
- Suominen H, Salanterä S, Velupillai S, Chapman WW, Savova G, Elhadad N, Pradhan S, South BR, Mowery DL, Jones GJ. Overview of the SHAre/CLEF

- eHealth evaluation lab 2013. In: International Conference of the Cross-Language Evaluation Forum for European Languages. Berlin, Heidelberg: Springer; 2013. p. 212–31.
15. Pradhan S, Elhadad N, Chapman W, Manandhar S, Savova G. Semeval-2014 task 7: analysis of clinical text. *SemEval*. 2014;199:54.
 16. Bethard S, Derczynski L, Savova G, Savova G, Pustejovsky J, Verhagen M. Semeval-2015 task 6: clinical tempeval. In: Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015); 2015. p. 806–14.
 17. Bethard S, Savova G, Chen W-T, Derczynski L, Pustejovsky J, Verhagen M. Semeval-2016 task 12: clinical tempeval. In: Proc SemEval; 2016. p. 1052–62.
 18. Liu Z, Yang M, Wang X, et al. Entity recognition from clinical texts via recurrent neural network. *BMC Med Inform Decis Mak*. 2017;17(2):67.
 19. Luo L, Yang Z, Yang P, et al. An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition. *Bioinformatics*. 2017;34(8):1381–8.
 20. Lei J, Tang B, Lu X, et al. Research and applications: a comprehensive study of named entity recognition in Chinese clinical text. *J Am Med Inform Assoc*. 2014;21(5):808.
 21. Hu J, et al. HITSZ_CNER: a hybrid system for entity recognition from Chinese clinical text. In: CEUR workshop proceedings; 2017. Vol. 1976.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

