RESEARCH

A multitask bi-directional RNN model for named entity recognition on Chinese electronic medical records

Shanta Chowdhury[1], Xishuang Dong[1], Lijun Qian[1], Xiangfang Li[1*], Yi Guan[2], Jinfeng Yang[3] and Qiubin Yu[4]

[*]Correspondence: xili@pvamu.edu
[1]Center of Excellence in Research and Education for Big Military Data Intelligence (CREDIT), Department of Electrical and Computer Engineering, Prairie View A&M University, Texas A&M University System, Prairie View, Texas 77446, USA
Full list of author information is available at the end of the article

Abstract

Background: Electronic Medical Record (EMR) comprises patients' medical information gathered by medical stuff for providing better health care. Named Entity Recognition (NER) is a sub-field of information extraction aimed at identifying specific entity terms such as disease, test, symptom, genes etc. NER can be a relief for healthcare providers and medical specialists to extract useful information automatically and avoid unnecessary and unrelated information in EMR. However, limited resources of available EMR pose a great challenge for mining entity terms. Therefore, a multitask bi-directional RNN model is proposed here as a potential solution of data augmentation to enhance NER performance with limited data.

Methods: A multitask bi-directional RNN model is proposed for extracting entity terms from Chinese EMR. The proposed model can be divided into a shared layer and a task specific layer. Firstly, vector representation of each word is obtained as a concatenation of word embedding and character embedding. Then Bi-directional RNN is used to extract context information from sentence. After that, all these layers are shared by two different task layers, namely the parts-of-speech tagging task layer and the named entity recognition task layer. These two tasks layers are trained alternatively so that the knowledge learned from named entity recognition task can be enhanced by the knowledge gained from parts-of-speech tagging task.

Results: The performance of our proposed model has been evaluated in terms of micro average F-score, macro average F-score and accuracy. It is observed that the proposed model outperforms the baseline model in all cases. For instance, the micro average F-score and the macro average F-score are improved by 2.41% and 4.16%, respectively, and the overall accuracy is improved by 5.66%.

Conclusions: In this paper, a novel multitask bi-directional RNN model is proposed for improving the performance of named entity recognition in EMR. Evaluation results using real datasets demonstrate the effectiveness of the proposed model.

Keywords: recurrent neural network; multitask learning; word embedding; parts-of-speech tagging; named entity recognition; electronic medical records

Background

Electronic Medical Record (EMR) [1], a digital version of storing patients' medical history in textual format, has shaped our medical domain in such a promising way that can gather all information into a place for healthcare providers. It comprises both structured and unstructured data that consists of patients' health condition and information such as symptoms, medication, disease, progress notes, and discharge summaries. EMR facilitates medical specialists and providers to track digital information and monitor them for patients' regular check-up. It can also provide healthcare suggestions to patients even they live in a remote area. Moreover, when a patient switches to a new healthcare provider, the provider can easily obtain patients' medical history and current health condition by studying patient's EMR. Therefore, information extraction [2] from EMR is one of the most important tasks in medical domain. The intent of information extraction system is to identify and connect the related information and organize them in such a way that can help people to draw conclusions from it, and by avoiding the unnecessary and unrelated information.

To extract information like entity recognition from EMR is labor intensive and time consuming. Although there are many developed models for extraction of entity terms from textual documents, adopting these models for the purpose of medical entity recognition from EMR has been demonstrated as a challenging task, because most of the EMRs are hastily written and incompatible to preprocess [2]. Moreover, incomplete syntax, numerous abbreviation, units after numerical values make the recognition task even more complicated [3]. Standard Natural Language Processing (NLP) tools cannot perform efficiently when they are applied on EMR, since the entity terms of standard NLP is not designed for medical domain. Therefore, it is necessary to develop effective method to perform entity recognition from EMR.

In recent years, various deep learning based methods have been developed for Named Entity Recognition (NER) [4] from EMR. Convolutional Neural Network (CNN) model is used for NER by using data mining to enhance the performance [5]. Zao et al. [6] proposed multiple label CNN based disease NER architecture by capturing correlation between adjacent labels. Dong et al. [7] developed multiclass classification based CNN for mining medical entity types from Chinese EMR.

Most recently, Recurrent Neural Network (RNN) such as Long Short-Term Memory (LSTM) is taking prominent place in NER due to its ability of dependency building in neighboring words. A hybrid LSTM-CNN is proposed in [8]. The authors used CNN to extract the features and fed them to LSTM model for recognizing entity types from CoNLL2003 dataset. Wang et al. [9] studied bi-directional LSTM architecture and concluded that this model is very effective for predicting sequential data. Moreover, the performance of the model is not based on language dependency. Simon et al. [10] and Vinayak et al. [11] used bi-directional RNN model on their Swedish EMR and Hindi dataset, respectively. In each case, the model shows better performance comparing to the state-of-the-art model. Similarly, the approach of using bi-directional RNN with LSTM cell has proven to perform well in extracting named entity recognition task [12].

In general, large corpus dataset is required to train deep learning models. However, there are limited number of corpus in many existing datasets that hinders the development of NER. Moreover, building labeled Chinese EMR data faces many challenges [13], and most organizations do not want to share their data publicly as the data contains private information of patients. In order to address this challenge,

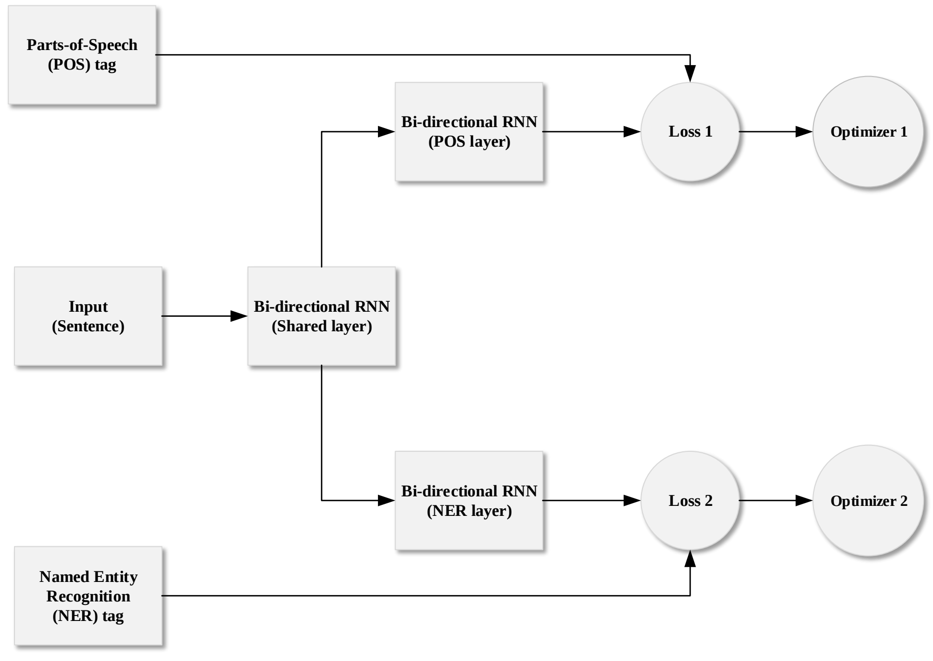

Parts-of-Speech (POS) tag
Bi-directional RNN (POS layer)
Loss 1
Optimizer 1
Input (Sentence)
Bi-directional RNN (Shared layer)
Bi-directional RNN (NER layer)
Loss 2
Optimizer 2
Named Entity Recognition (NER) tag

Figure 1 Framework of the proposed multitask bi-directional RNN model for NER.

a multitask bi-directional RNN model is proposed in this work for extracting entity terms from Chinese EMR. It is motivated by the observation that the performance of multitask learning model is much better comparing to individual learning approach when there is limited corpus dataset [14]. The framework of the proposed multitask bi-directional RNN model for NER is given in Figure 1.

Methods

In this work, a multitask bi-directional RNN model is proposed for extracting entity terms from Chinese EMR. The proposed model can be divided into two parts: shared layer and task specific layer, see Figure 1. Specifically, vector representation of each word is a concatenation of word embedding and character embedding in the proposed model, see Figure 2. Bi-directional RNN is used to extract context information from sentence. Then all these layers are shared by two different task layers, namely the parts-of-speech tagging task layer and the named entity recognition task layer. These two tasks layers are trained alternatively so that the knowledge learned from named entity recognition task can be enhanced by the knowledge gained from parts-of-speech tagging task.

RNN [15] is an artificial neural network which can capture previous word information of a sequence in its memory. It computes each word of input sequence (x_1, x_2, $\cdots$, x_n) and transforms it into a vector form (y_t) by using the following equations:

$$h_t = H(U_{xh}x_t + U_{hh}h_{t-1} + b_h). \tag{1}$$

$$y_t = U_{hy}h_t + b_y. \tag{2}$$

where U_{xh}, U_{hh}, U_{hy} denote the weight matrices of input-hidden, hidden-hidden and hidden-output processes, respectively. h_t is the vector of hidden states that capture the information from current input x_t and the previous hidden state h_{t-1}.

Here the bi-directional RNN is used to exploit both past and future context, where forward hidden states compute forward hidden sequence while backward hidden states compute backward hidden sequence. The output y_t is generated by integrating the two hidden states. In this work, we use a special form of bi-directional RNN, the bi-directional RNN with LSTM cell [16]. LSTM is a special kind of RNN where hidden states are replaced by memory cells to capture long term dependent contextual phrase. The computation of LSTM is quite similar to RNN except for the hidden units, and it is given below:

$$i_t = \sigma(U_{xi}x_t + U_{hi}h_{t-1} + U_{ci}c_{t-1} + b_i). \tag{3}$$

$$g_t = \sigma(U_{xg}x_t + U_{hg}h_{t-1} + U_{ci}c_{t-1} + b_g). \tag{4}$$

$$c_t = g_t c_{t-1} + i_t \tanh(U_{xc}x_t + U_{hc}h_{t-1} + b_c). \tag{5}$$

$$y_t = \sigma(U_{xy}x_t + U_{hy}h_{t-1} + U_{cy}c_t + b_y). \tag{6}$$

$$h_t = y_t \tanh(c_t). \tag{7}$$

where i, g, c, o and σ are the input gate, forget gate, cell activation vector, output gate, and logistic sigmoid function of LSTM cell, respectively. These gates and activation functions soothe LSTM to avoid the limitation of vanishing gradients by storing long term dependencies terms of a sequence.

The shared layer contains two consecutive parts, illustrated by Figure 2 and Figure 3. In Figure 2, each word is represented by a vector developed by Mikolov [17]. The vector is built as a concatenation of word embeddings [18] and character embeddings. Bi-directional RNN with LSTM cell is used to extract features at the character level and represent the features as character embeddings. Word embedding is achieved by word to vector representation. Character embeddings and word embeddings are then combined to represent each word in a vector representation. In Figure 3, another bi-directional RNN with LSTM cell is used to extract context information from text sequence. Then the outputs (contextual word representations) are shared by two different bi-directional RNN with LSTM cell for two different tasks: parts-of-speech tagging and named entity recognition. These two task layers are trained alternatively so that knowledge from parts-of-tagging task can be used to improve the performance of named entity recognition task [19]. The detailed settings of the proposed model is shown in Table 1.

Results

Dataset Details

The EMR dataset used in our experiment was collected from the departments of the Second Affiliated Hospital of Harbin Medical University, and the personal information of the patients have been discarded. An annotated/labeled corpus consisting

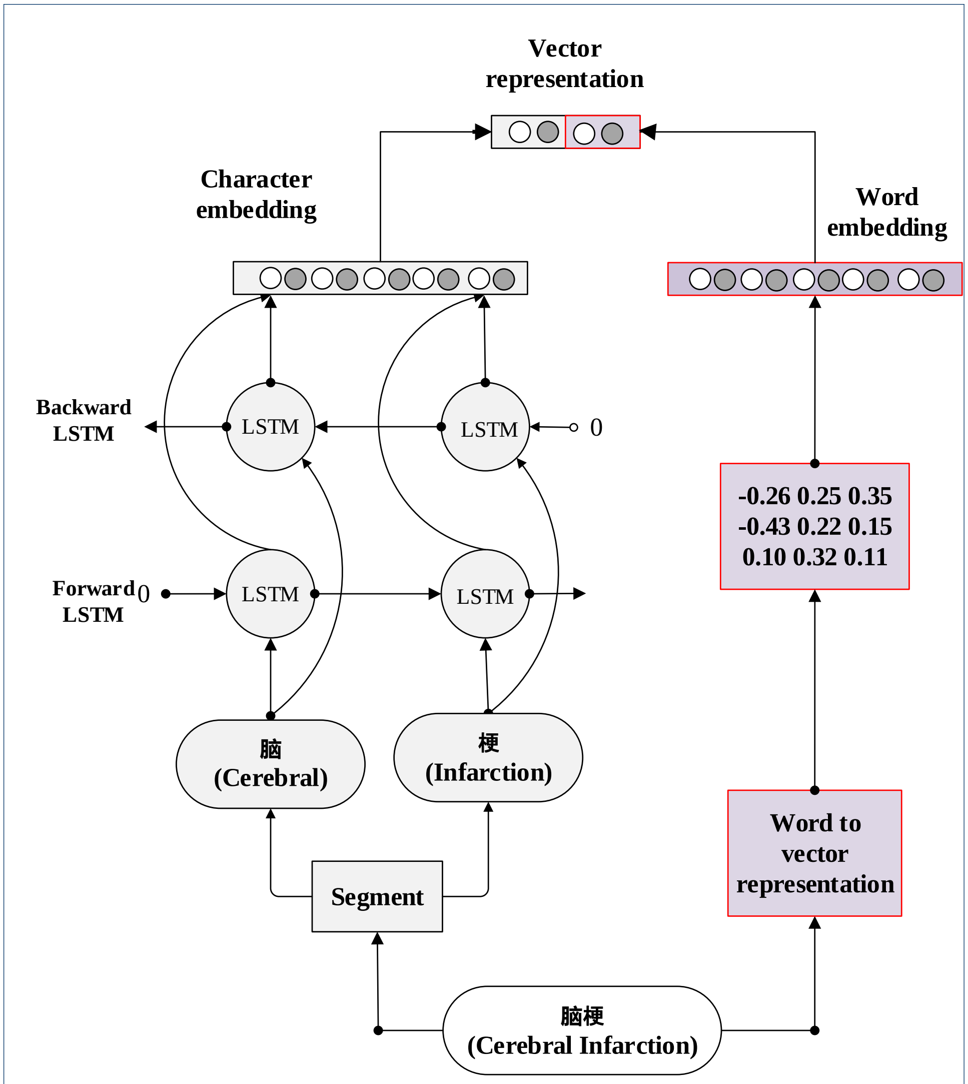


Figure 2 Vector Representation as concatenation of word embeddings and character embeddings. Vector representation of each word is presented as concatenation of word embeddings and character embeddings. The flow of word embedding is highlighted by red shaded box and character embedding is highlighted by white shaded box.

Table 1 The proposed network architecture.

Name	Description
Input	Sentences in EMR
Word Embedding	Mikolov model
Character Embedding Layer	150 LSTM cells for each hidden layer, one forward hidden layer and one backward hidden layer, Dropout = 0.5
Parts-of-speech tag (POS) layer	150 LSTM cells for each hidden layer, one forward hidden layer and one backward hidden layer, Dropout = 0.5
Named Entity recognition (NER) Layer	150 LSTM cells for each hidden layer, one forward hidden layer and one backward hidden layer, Dropout = 0.5
Output	Softmax

of 500 discharge summaries has been manually created. The EMR data are written

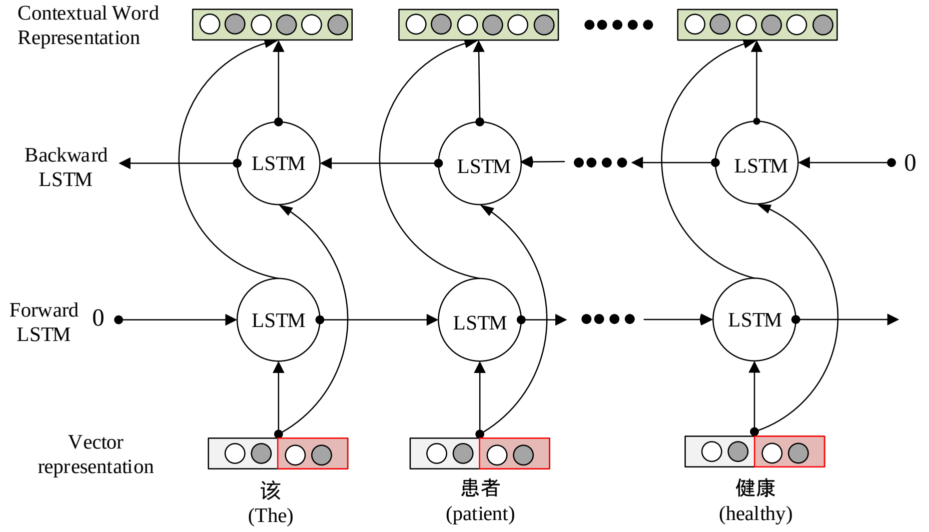


Figure 3 Contextual word representation from vector representation. To extract relevant context information from sentence, bi-directional RNN with LSTM cell is used to extract information from a vector associated with word embedding (red shaded box) and character embedding (white shaded box) to form contextual word representation (green shaded box).

in Chinese with 27,110 sentences. The annotation was made by two Chinese physicians (A1 and A2) independently [7] [13]. It is categorized into five entity types: disease, symptom, treatment, test, and disease group. An annotation example is shown in Figure 4. The character n-grams are conducted by word segmentation and named entity recognition on Chinese sentences. In the domain of natural language processing (NLP) on Chinese, the first step is to segment the sentence into words containing n-gram characters since for Chinese the minimum semantic units are words, not individual characters. It can be accomplished by NLP tools like Stanford Word Segmenter [20, 21]. Then for recognizing medical concepts from EMR, we define the named entity classes and use different labels to indicate these classes. For example, B/I/O labels denote the beginning word, inside word, and outside word of the named entities. Moreover, for named entity recognition on EMR, we attach the medical information to these three labels in order to denote different categories of named entities. For example, B_disease and B_treatment are denoting beginning words of disease and treatment named entities, respectively. The descriptions of entity types are given in Table 2.

Table 2 Name of the entity types and their descriptions.

Entity Types	Description
Disease	phrases related to disease concept
Symptom	phrases of symptom concept
Disease group	phrases of the cruelty of diseases
Treatment	phrases of protocol and surgery name
Test	phrases represent different tests name prescribed for patient

The categorized entity types are labeled in BIO format: B, starting of the medical entity type; I, inside of the medical entity type; O, apart from the entity type. The categorization of entities in BIO format is given in Table 3.

Table 3 BIO format of entity types.

	Categories						Total
NER type	Disease	Symptom	Disease group	Treatment	Test	Other	6
BIO format	B_dis I_dis	B_com I_com	B_dit I_dit	B_tre I_tre	B_tes I_tes	other	11

Experimental settings

In this experiment, our proposed model is employed to extract medical information from EMR dataset. The key hyper parameters are: Number of hidden neurons for each hidden layer: 150, Minibatch size: 20, Number of epoch: 100, Optimizer: Adam optimizer, Learning rate: 0.01, Learning rate decay: 0.9. They are determined by trial and error.

Evaluation metric

Different metrics in terms of micro-average F score (MicroF), macro-average F score (MacroF) [22] and accuracy have been used to evaluate the performance of our proposed model. Accuracy is calculated by dividing the number of predicted entities that is exactly matched with dataset entities over the total number of entities in the dataset. MicroF is calculated by MicroP and MicroR values whereas MacroF is affected by the average F values of each class:

$$F = \frac{2PR}{P+R}. \tag{8}$$

where P indicates precision measurement that defines the capability of a model to represent only related entities [23] and R (recall) computes the aptness to refer all corresponding entities:

$$P = \frac{TP}{TP+FP}. \tag{9}$$

$$R = \frac{TP}{TP+FN}. \tag{10}$$

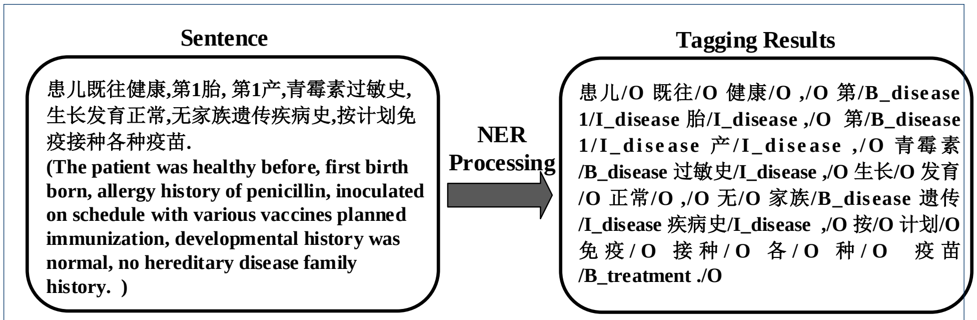
whereas TP (True Positive) counts total number of entity matched with the entity in the labels. FP (False Positive) measures the number of recognized label does not match the annotated corpus dataset. FN (False Negative) counts the number of entity term that does not match the predicted label entity. Then,

$$MacroF = \frac{1}{T}\sum_{j=1}^{T} F_j. \tag{11}$$

$$MacroP = \frac{1}{T}\sum_{j=1}^{T} P_j. \tag{12}$$

$$MacroR = \frac{1}{T}\sum_{j=1}^{T} R_j. \tag{13}$$

where T denotes the total number of categorized entities and F_j, P_j, R_j are F, P, R values in the j^{th} category of entities [7].



患儿既往健康,第1胎, 第1产,青霉素过敏史,生长发育正常,无家族遗传疾病史,按计划免疫接种各种疫苗.
(The patient was healthy before, first birth born, allergy history of penicillin, inoculated on schedule with various vaccines planned immunization, developmental history was normal, no hereditary disease family history.)

患儿/O 既往/O 健康/O ,/O 第/B_disease 1/I_disease 胎/I_disease ,/O 第/B_disease 1/I_disease 产/I_disease ,/O 青霉素/B_disease 过敏史/I_disease ,/O 生长/O 发育/O 正常/O ,/O 无/O 家族/B_disease 遗传/I_disease 疾病史/I_disease ,/O 按/O 计划/O 免疫/O 接种/O 各/O 种/O 疫苗/B_treatment ./O

Figure 4 Tagging results on Chinese EMR [7].

Experimental results

Our experiments are implemented in different phases namely micro average, macro average and accuracy comparison. Precision, Recall and F-score are measured using our proposed multitask bi-directional RNN model and compared with the following classifiers: Naive Bayes (NB), Maximum Entropy (ME), Support Vector Machine (SVM), Conditional Random Field (CRF) [7], and deep learning models including Convolutional Neural Network (CNN) [7], single task bi-directional RNN (Bi-RNN) and transfer bi-directional RNN [24], where NER can be defined as a multiclass classification problem for these classifiers [7]. Among all the models, we have considered Bi-RNN model as baseline model.

Firstly, performances are compared based on micro values and summarized in Table 4. The results show that our proposed multitask bi-directional RNN model outperforms other models. For instance, the MicroF value of our proposed model is improved by 2.41% and 4.67% compared to the baseline model (Bi-RNN) and CNN, respectively.

Table 4 Comparison results of MicroP, MicroR and MicroF measure.

Model	MicroP	MicroR	MicroF
Naive Bayes	78.07	77.91	77.99
Maximum Entropy	88.81	88.81	88.81
Support Vector Machine	90.52	90.52	90.52
Conditional Random Field [7]	93.15	93.15	93.15
Convolutional Neural Network [7]	88.64	88.64	88.64
Bi-RNN model	90.90	90.90	90.90
Transfer learning Bi-RNN model [24]	92.25	92.25	92.25
Our proposed model	**93.31**	**93.31**	**93.31**

Table 5 Comparison results of NER on discharge summaries.

	Bi-RNN model			Our proposed model		
Entity type	Precision	Recall	F-measure	Precision	Recall	F-measure
Disease	82.82	78.02	80.34	84.11	84.70	84.40
Symptom	80.26	80.11	80.19	88.08	84.01	86.00
Disease group	37.50	100	54.54	43.75	82.35	57.14
Treatment	68.89	78.58	73.41	73.91	82.06	77.77
Test	82.99	86.43	84.68	89.23	87.99	88.61
Macro average	70.91	**84.67**	74.63	**75.82**	84.22	**78.79**

Since micro average only measures the effectiveness of model on a large number of entity, macro average is computed to evaluate the model's performance in the case of small number of entity terms [25]. The macro average F-score is improved

Table 6 Comparison results (%accuracy) on discharge summaries

Model	Entity type					
	Disease	Symptom	Disease group	Treatment	Test	Overall accuracy
Naive Bayes (NB)	44.82	51.72	N/A	59.00	65.96	58.91
Maximum Entropy (ME)	48.32	56.34	34.19	58.80	76.10	65.68
Support Vector Machine (SVM)	57.18	62.52	37.22	60.48	80.17	70.46
Conditional Random Field (CRF) [7]	77.33	77.83	**48.39**	**77.47**	**90.05**	**83.94**
Convolutional Neural Network(CNN) [7]	52.80	65.76	40.00	53.14	79.28	68.60
Bi-RNN model	73.83	79.35	28.00	67.99	82.63	77.85
Transfer learning Bi-RNN model [24]	74.30	82.60	44.00	68.20	86.79	80.75
Our proposed model	**76.86**	**87.22**	36.00	71.33	89.20	83.51

by 4.16% compared to the baseline model. Table 5 shows the comparison results of NER on discharge summaries. The F-measure ranged from 57.14% to 88.61% in different categorized entities when it is computed on our proposed model whereas the range is from 54.54% to 84.68% when it is computed from the baseline model.

The comparison results of accuracy on discharge summaries are given in Table 6. It is observed that the overall accuracy is improved by 5.66% compared to the baseline model. According to the evaluation results, our proposed model shows better performance on recognizing medical entity terms comparing with other models except CRF model. CRF uses the feature templates to extract features in order to build the NER model by introducing prior knowledge. On the other hand, the proposed model performs the NER task on Chinese EMRs without any prior knowledge.

It is observed that the best accuracy is enlisted as 89.20% in test terms and lowest performance is 36.00% in recognizing disease terms. The accuracy of recognizing disease terms is lowest comparing with other entities since there are very limited number of disease group (0.56%) [24] in sample which is not enough to train the model.

In addition, we examine how different features affect the model performance. We compare models built by word level features, character level features, and combined word level features and character level features. The comparison results are shown in Table 7. It is observed that combined features will improve the model performance.

Table 7 Comparison the results for character and word level feature

Embedding approaches	Character level	Word level	Character level+Word level
MicroF	77.25	93.22	**93.31**
MacroF	47.28	**81.23**	78.79
Accuracy	35.30	83.12	**83.51**

Discussion

In our proposed multitask model, we have been concentrating on improving the accuracy of named entity recognition task. Therefore, we have used different task layer (parts-of-speech tagging task) to enhance recognition performance which in turn improves the accuracy of named entity recognition task. More training time is needed for the proposed model since two task specific layers need to be trained, which involves two loss functions and two optimizers. We plan to use a joint loss function and joint optimizer to reduce the training time and improve the accuracy in our future research.

Conclusions

In this paper, a novel multitask bi-directional RNN model is proposed for improving the performance of named entity recognition in EMR. Two different task layers, namely parts of speech tagging task layer and named entity recognition task layer are used in order to improve the information extraction method from EMR dataset by sharing the word embedding and character embedding layer. The feature sharing layer has a great impact on improving the accuracy of extracting entity information. Evaluation results using real datasets demonstrate the effectiveness of the proposed model.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

SC, XD, LQ and XL come up with the proposed method; SC and XD complete the simulations; YG, JY, and QY provide the data sets and medical ground truth of the data. All authors proofread the manuscript.

Acknowledgements

This research work is supported in part by the Texas A&M Chancellor's Research Initiative (CRI), the U.S. National Science Foundation (NSF) award 1464387 and 1736196, and by the U.S. Office of the Assistant Secretary of Defense for Research and Engineering (OASD(R&E)) under agreement number FA8750-15-2-0119. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the U.S. National Science Foundation (NSF) or the U.S. Office of the Assistant Secretary of Defense for Research and Engineering (OASD(R&E)) or the U.S. Government.

Author details

[1]Center of Excellence in Research and Education for Big Military Data Intelligence (CREDIT), Department of Electrical and Computer Engineering, Prairie View A&M University, Texas A&M University System, Prairie View, Texas 77446, USA. [2]Schools of Computer Science and Technology, Harbin Institute of Technology, Harbin, China. [3]Schools of Software, Harbin University of Science and Technology, Harbin, China. [4]Second Affiliated Hospital of Harbin Medical University, Harbin, China.

References

1. Gunter, T.D., Terry, N.P.: The emergence of national electronic health record architectures in the united states and australia: models, costs, and questions. Journal of medical Internet research **7**(1) (2005)
2. Ford, E., Carroll, J.A., Smith, H.E., Scott, D., Cassell, J.A.: Extracting information from the text of electronic medical records to improve case detection: a systematic review. Journal of the American Medical Informatics Association **23**(5), 1007–1015 (2016)
3. Tange, H.J., Hasman, A., de Vries Robbé, P.F., Schouten, H.C.: Medical narratives in electronic medical records. International journal of medical informatics **46**(1), 7–29 (1997)
4. Nadeau, D., Sekine, S.: A survey of named entity recognition and classification. Lingvisticae Investigationes **30**(1), 3–26 (2007)
5. Yao, C., Qu, Y., Jin, B., Guo, L., Li, C., Cui, W., Feng, L.: A convolutional neural network model for online medical guidance. IEEE Access **4**, 4094–4103 (2016)
6. Zhao, Z., Yang, Z., Luo, L., Zhang, Y., Wang, L., Lin, H., Wang, J.: Ml-cnn: A novel deep learning based disease named entity recognition architecture. In: 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 794–794 (2016)
7. Dong, X., Qian, L., Guan, Y., Huang, L., Yu, Q., Yang, J.: A multiclass classification method based on deep learning for named entity recognition in electronic medical records. In: Scientific Data Summit (NYSDS), 2016 New York, pp. 1–10 (2016)
8. Chiu, J.P., Nichols, E.: Named entity recognition with bidirectional lstm-cnns. arXiv preprint arXiv:1511.08308 (2015)
9. Wang, P., Qian, Y., Soong, F.K., He, L., Zhao, H.: A unified tagging solution: Bidirectional lstm recurrent neural network with word embedding. arXiv preprint arXiv:1511.00215 (2015)
10. Almgren, S., Pavlov, S., Mogren, O.: Named entity recognition in swedish health records with character-based deep bidirectional lstms. In: Proceedings of the Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM2016), pp. 30–39 (2016)
11. Athavale, V., Bharadwaj, S., Pamecha, M., Prabhu, A., Shrivastava, M.: Towards deep learning in hindi ner: An approach to tackle the labelled data scarcity. arXiv preprint arXiv:1610.09756 (2016)
12. Luong, M.-T., Manning, C.D.: Achieving open vocabulary neural machine translation with hybrid word-character models. arXiv preprint arXiv:1604.00788 (2016)
13. He, B., Dong, B., Guan, Y., Yang, J., Jiang, Z., Yu, Q., Cheng, J., Qu, C.: Building a comprehensive syntactic and semantic corpus of chinese clinical texts. Journal of biomedical informatics **69**, 203–217 (2017)
14. Zhang, Y., Yang, Q.: A survey on multi-task learning. arXiv preprint arXiv:1707.08114 (2017)
15. A Beginner's Guide to Recurrent Networks and LSTMs. https://deeplearning4j.org/lstm.html

16. Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/
17. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, pp. 3111–3119 (2013)
18. Habibi, M., Weber, L., Neves, M., Wiegandt, D.L., Leser, U.: Deep learning with word embeddings improves biomedical named entity recognition. Bioinformatics **33**(14), 37–48 (2017)
19. Sequence Tagging with Tensorflow. https://guillaumegenthial.github.io/sequence-tagging-with-tensorflow.html
20. Stanford Word Segmenter. https://nlp.stanford.edu/software/segmenter.html
21. Chang, P.-C., Galley, M., Manning, C.D.: Optimizing chinese word segmentation for machine translation performance. In: Proceedings of the Third Workshop on Statistical Machine Translation, pp. 224–232 (2008)
22. Yang, Y.: A study of thresholding strategies for text categorization. In: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 137–145 (2001)
23. Oliveira, J.L., Matos, S., Campos, D.: Biomedical named entity recognition: A survey of machine-learning tools. In: Sakurai, S. (ed.) Theory and Applications for Advanced Text Mining. InTech, Rijeka (2012). Chap. 8. doi:10.5772/51066. https://doi.org/10.5772/51066
24. 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: 2017 IEEE 19th International Conference one-Health Networking, Applications and Services (Healthcom), pp. 1–4 (2017)
25. Suominen, H., Zhou, L., Hanlen, L., Ferraro, G.: Benchmarking clinical speech recognition and information extraction: new data, methods, and evaluations. JMIR medical informatics **3**(2) (2015)