



Research article

Medical assertion classification in Chinese EMRs using attention enhanced neural network

Zhichang Zhang*, Yu Zhang, Tong Zhou and Yali Pang

College of Computer Science and Engineering, Northwest Normal University, Lanzhou, 730070, China

* **Correspondence:** Email: zzc@nwnu.edu.cn; Tel: +86-130-3876-9329.

Abstract: Electronic medical records (EMRs), such as hospital discharge summaries, contain a wealth of information only expressed in natural language. Automated methods for extracting information from these records must be able to recognize medical concepts in text and their semantic context. A contextual property critical to reason on information from EMRs is the doctor's belief status or assertion of the patient's medical problem. Research on the medical assertion classification (MAC) can establish the foundation for various health data analyses and clinical applications. However, previous MAC studies are mainly based on traditional machine learning methods which mostly require manually constructed features and the original unlabeled data cannot be easily and effectively applied to classification or classification tasks. Furthermore, external medical knowledge such as various medical dictionary bases, which provides rich explain and definition information about medical entity, is rarely utilized in existing neural network models of medical information extraction. In this study, we propose a deep neural network architecture enhanced by medical knowledge attention layer through combining GRU neural network with CNN model to classify the assertion type of medical problem such as disease and symptom in Chinese EMRs. The attention layer in the model is applied to integrate entity representations learned from medical dictionary bases as query for encoding. Experimental results on own manually annotated corpus indicate our approach achieves better performance compared to existing methods.

Keywords: Chinese electronic medical record; medical entity assertion classification; deep learning; attention mechanism

1. Introduction

The electronic medical records (EMRs), sometimes called electronic health records (EHRs) or electronic patient records (EPRs), is one of the most important types of clinical data and often contains valuable and detailed patient information for many clinical applications. This paper studies the technology of structuring EMRs and medical information extraction, which are key foundations for health-related various applications.

As a kind of medical information extraction technology, the medical assertion classification (MAC) in EMRs, which is formally defined for the 2010 i2b2/VA Challenge, aims to recognize the relationship between medical entities (Disease and Symptom) and patients. Given a medical problem or entity mentioned in a clinical text, an assertion classifier must look at the context and choose the status of how the medical problem pertains to the patient by assigning one of seven labels: present, absent, conditional, possible, family, occasional, or history. The assertion is reflected in two aspects: whether the entity occurs to the patient, and how the entity occurs to the patient. As the basis of medical information processing, assertion classification in EMRs is of great importance to many EMRs mining tasks. When there are many researches about MAC for English EMRs, few studies have been done on Chinese texts.

Based on the above, we study the MAC methods for Chinese EMRs in this paper. According to the corresponding task of the 2010 i2b2/VA Challenge, we divide the assertion categories of medical entity into seven categories: present (当前的), possible (可能的), conditional (有条件的), family (非患者的), occasional (偶有的), absent (否认的) and history (既往的). The definitions and Chinese sentence examples of different kind of assertion category are shown in the Table 1.

Table 1. Definition of assertion categories.

Category	Definition	Example sentence
Present	Symptoms or illness that must be present in the patient	患者意识昏迷 8 小时 (The patient was unconscious for 8 hours.)
Absent	A denial of illness or symptoms	患者无眩晕无头痛 (No vertigo , no headache)
Conditional	A condition or illness that occurs only under certain conditions	饮酒后易休克 (Shock after drinking)
Possible	Possible disease or symptom	术后可能有红肿现象 (Postoperative redness may occur) 肿瘤待查 (Tumor waiting for investigation)
Family	A condition or condition that is not the patient's own	直系亲属患有癫痫病史 (History of epilepsy in immediate relatives)
Occasional	An illness or symptom that does not currently occur frequently	偶有头晕症状 (Occasionally dizziness)
History	Past illnesses or symptoms	2 年前因痛风入我院治疗 (Gout was admitted to our hospital two years ago for treatment)

Although there have been lot of methods proposed for recognizing entity assertion category from EMRs, most of them used traditional machine methods such as Support Vector Machine (SVM) and Conditional Random Field (CRF), which are mainly based on feature engineering. However, feature engineering is relatively time-consuming and costly, and resulting feature sets are both

domain and model-specific. While deep neural network approaches on various medical information extraction tasks have achieved better performance compared to traditional machine learning models, research on entity assertion classification of EMRs using deep neural network model is still few.

Therefore, this paper proposes a novel model for MAC of Chinese EMRs. We build a deep network (called GRU-Softmax) as baseline, which combines Gated Recurrent Unit (GRU) neural network (a type of Recurrent Neural Networks, RNNs) and softmax, to classify named entity assertion from Chinese EMR. Compared with RNN, the advantage of GRU-Softmax lie in that GRU neural networks have strong expressive ability to capture long context without time-intensive feature engineering.

Furthermore, in order to obtain character level characteristics in EMRs text, we train Chinese character-level embedding representation using Convolutional Neural Network (CNN), and combine them with word-level embedding vector acquired from large-scale background training corpus. Then the combined vectors are sent to GRU architecture to train entity assertion classification model. In addition, to enhance the representation and distinguish ability of characters and their contexts, we integrate the medical knowledge attention (MKA) learned from entity names and their definition or descriptions in medical dictionary bases (MDBs) in the model.

On the whole, the contributions of this work can be summarized as follow: (1) we introduce a deep neural network that combines GRU neural networks and Sfoamax to classify medical assertion at first time; (2) We compared the influence of character-level representation extracted by CNN on the model; (3) we use medical knowledge attention (MKA) to integrate entity representation from external knowledge (medical dictionary bases, MDBs).

The remainder of this paper is composed as follows. In section 2 we summarize the related work about MAC. In section 3 we present our attention-based CNN-GRU-Softmax network model for MAC in Chinese EMRs. In section 4 we show the experimental results and give some analysis. Finally, we summarize our work and outline some ideas for future research.

2. Related work

Research of entity assertion classification is to study the relation classification between entity and patient on the basis of entity known. Chapman et al. [1] proposed a classification model named as NegEx based on regular expression rules, which classifies disease entity as "existing" or "nonexistent", and can obtain 85.3% of F value on more than 1000 disease entities. Based on NegEx method and combining regular expression rules and trigger words, Harkema et al. [2] proposed the ConText method to classify the disease entities into one of six categories. On six different types of medical records, 76% to 93% of F values could be obtained, indicating that the distribution of modified disease entities varied greatly in different styles of texts.

Based on the evaluation data of the 2010 i2b2 Challenge, researchers proposed many classification methods based on rules, SVM, CRF, etc. The most effective concept extraction systems used support vector machines (SVMs) [3–11], either with contextual information and dictionaries that indicate negation, uncertainty, and family history [6,10], or with the output of rule-based systems [3,6,8]. Roberts et al. [4] and Chang et al. [11] utilized both medical dictionary and rules. Chang et al. complemented SVM with logistic regression, multi-logistic regression, and boosting, which they combined using voting mechanism. The highest classification effect in the evaluation was obtained by the Bi-level classifier proposed by de Bruijn et al. [5], who used the

cTAKES knowledge base created an ensemble whose final output was determined by a multi-class SVM, and the evaluation result F could reach 93.6%. Clark et al. [12] used a CRF model to determine negation and uncertainty with their scope, and added sets of rules to separate documents into different zones, to identify and scope cue phrases, and determine phrase status. They combined the results from the found cues and the phrase status module with a maximum entropy classifier that also used concept and contextual features.

3. The assertion classifier

In this paper, we propose a neural network architecture combining GRU-CNN-Softmax network with Medical Knowledge Attention that will learn the shared semantics between medical record texts and the mentioned entities in the medical dictionary bases (MDBs). The architecture of our proposed model is shown in Figure 1. After querying pretrained character embedding tables, the input sentence will be transformed respectively to the corresponding sequences of pretrained character embeddings and random generated character embedding matrixes for every character. Then a CNN is used to form the character level representation and a GRU is used to encode the sentence representation after concatenating the pretrained character embeddings and character-level representation of the sentence. Afterwards, we treat the entity information from MKBs as a query guidance and integrate them with the original sentence representation using a multi-modal fusion gate and a filtering gate. At last, a Softmax layer is used to classify.

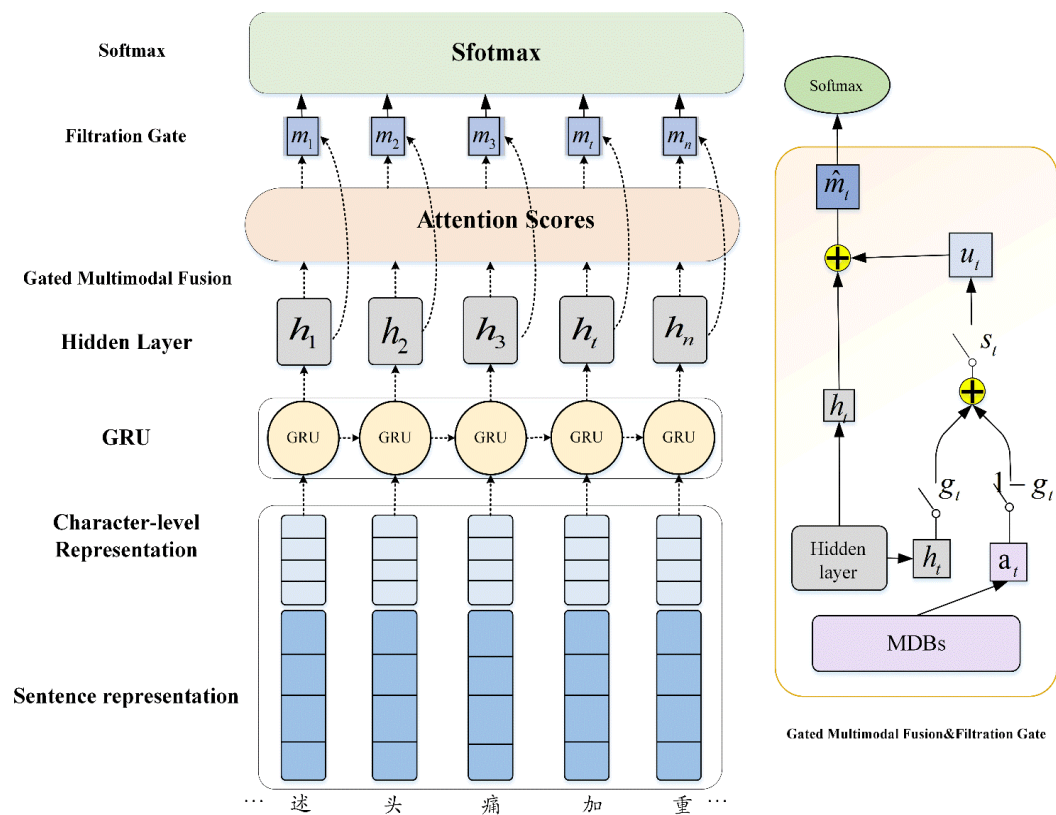


Figure 1. The framework of our model. The right part is the GMF and Filtering Gate.

3.1. CNN

As described in Figure 2, we firstly train Chinese character embeddings from a large unlabeled Chinese EMR corpus, then CNN is used to generate sentence character-level representation from the character embedding matrix sequence to alleviate rare character problems and capture helpful morphological information like special characters in EMRs. Since the length of sentences is not consistent, a placeholder (padding) is added to the left and right side of character embeddings matrix to make the length of every sentence character-level representation vector matrix sequence equal.

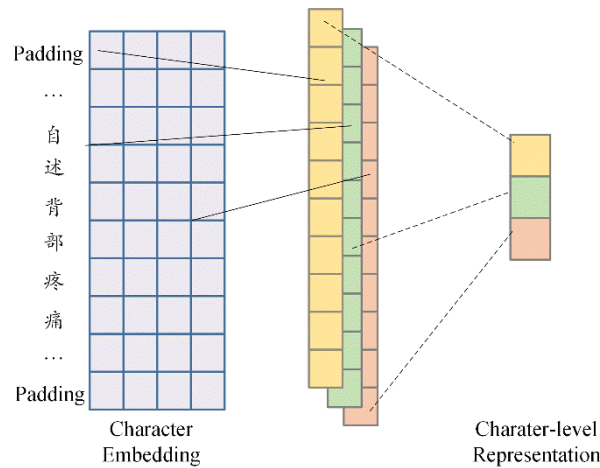


Figure 2. Character-level representation of a sentence by CNN.

3.2. GRU

The Gate Recurrent Unit (GRU) is a branch of the Recurrent Neural Network (RNN). Like LSTM, it is proposed to solve such problems as the gradient in long-term memory and reverse propagation. We choose to use GRU [13] in our model since it performs similarly to LSTM [14] but is computationally cheaper.

The GRU model is defined by the following equations:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (1)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (2)$$

$$\hat{h}_t^o = \tanh(W_h x_t + U_h (h_{t-1} * r_t) + b_h) \quad (3)$$

$$\hat{h}_t^o = \tanh(W_h x_t + U_h (h_{t-1} * r_t) + b_h) \quad (4)$$

In particular, z_t and r_t are vectors corresponding to the update and reset gates respectively, where $*$ denotes elementwise multiplication. The activations of both gates are elementwise logistic sigmoid functions $\sigma(\cdot)$, constraining the values of z_t and r_t ranging from 0 to 1. h_t represents the

output state vector for the current time frame t , while \tilde{h}_t^0 is the candidate state obtained with a hyperbolic tangent. The network is fed by the current input vector x_t (sentence representation of previous layer), and the parameters of the model are W_z, W_r, W_h (the feed-forward connections), U_z, U_r, U_h (the recurrent weights), and the bias vectors b_z, b_r, b_h . The Gate Recurrent Unit (GRU) is shown in Figure 3.

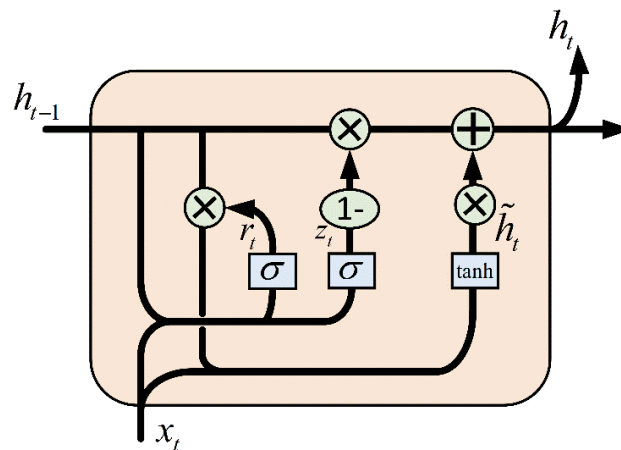


Figure 3. The architecture of gate recurrent unit (GRU).

3.3. Attention mechanism

Concerning rich entity mention and definition information containing in MDBs, the medical knowledge attention is applied to integrate entity representations learned from external knowledge bases as query vector for encoding. We use a medical dictionary to encode entity information (entity mention and definition) into attention scores as entity embeddings.

$$a_t = f(eW_A h_t) \quad (5)$$

Where e is the embedding for entity, and W_A is a bi-linear parameter matrix. We simply choose the quadratic function $f(x) = x^2$, which is positive definite and easily differentiable.

3.4. Fusion method

Based on the output of GRU and attention scoring, we design a gated multimodal fusion (GMF) method to fuse the features from output of hidden layer h_t and attention scoring a_t . When predicting the entity tag of a character, the GMF trades off how much new information of the network is considering from the query vector with the EMR text containing the character. The GMF is defined as:

$$h_{a_t} = \tanh(W_{a_t} a_t + b_{a_t}) \quad (6)$$

$$h_{h_t} = \tanh(W_{h_t} h_t + b_{h_t}) \quad (7)$$

$$g_t = \sigma(W_{g_t} (h_{a_t} \oplus h_{h_t})) \quad (8)$$

$$m_t = g_t h_{a_t} + (1 - g_t) h_{h_t} \quad (9)$$

where $W_{a_t}, W_{h_t}, W_{g_t}$ are parameters, h_{h_t} and h_{a_t} are the new sentence vector and new query vector respectively, after transformation by single layer perceptron. \oplus is the concatenating operation, σ is the logistic sigmoid activation, g_t is the gate applied to the new query vector h_{h_t} , and m_t is the multi-modal fused feature from the new medical knowledge feature and the new textual feature.

3.5. Filtering gated

When decoding the combination of the multimodal fusion feature m_t at position t , the impact and necessity of the external medical knowledge feature for different assertion is different. Because the multimodal fusion feature contains external knowledge feature more or less and it may introduce some noise. We therefore use a filtering gate to combine different features from different signal that better represent the useful information. The filtering gate is a scalar in the range of $[0, 1]$ and its value depends on how much the multimodal fusion feature is helpful to label the tag of the assertion.

s_t and the input feature to the decoder \hat{m}_t are defined as follows:

$$s_t = \sigma(W_{s_t, h_t} h_t \oplus (W_{m_t, s_t} m_t + b_{m_t, s_t})) \quad (10)$$

$$u_t = s_t (\tanh(W_{m_t} m_t + b_{m_t})) \quad (11)$$

$$\hat{m}_t = W_{\hat{m}_t} (h_t \oplus u_t) \quad (12)$$

where $W_{m_t, s_t}, W_{s_t, h_t}, W_{m_t}, W_{\hat{m}_t}$ are parameters, h_t is the hidden state of bidirectional LSTM at time t , u_t is the reserved multimodal features after the filtering gate filter out noise, and \oplus is the concatenating operation. The architecture of gated multimodal fusion and filtering gate are shown in Figure 1.

After we get the representation \hat{m}_t of sentence, we use softmax function to normalize and output entity assertion probability.

4. Experiments

In this section, we evaluate our method on a manually annotated dataset. Following Nadeau et al., we use Precision, Recall, and F1 to evaluate the performance of the models [18].

4.1. Datasets

We use our own manually annotated corpus as evaluation dataset, which consists of 800 de-identified EMR texts from different clinical departments of a grade-A hospital of second class in Gansu Province. The annotated entity number of every entity assertion category in the dataset is shown in the Table 2.

Table 2. Number statistics of different entity assertion categories in the evaluation dataset.

Category	Training	Test	Total
Present (当前的)	2025	1013	3038
Absent (否认的)	1877	921	2799
Conditional (有条件的)	204	102	306
Possible (可能的)	844	420	1264
Family (非患者本人的)	235	117	352
Occasional (偶有的)	249	147	396
History (既往的)	342	171	513
Total	5778	2889	8667

4.2. Embedding and setting

We use Google's Word2Vec to train Chinese character embeddings on our 30 thousand unlabeled Chinese EMR texts which is from a grade-A hospital of second class in Gansu Province. Random generated character embeddings are initialized with uniform samples from $[-\sqrt{\frac{3}{dim}}, \sqrt{\frac{3}{dim}}]$, where we set $dim = 30$.

Table 3 gives the chosen hyper-parameters for all experiments. We tune the hyper-parameters on the development set by random search. We try to share as many hyper-parameters as possible in experiments.

Table 3. Parameter Setting.

Parameter	Value
Character-level representation size	50
Pretrained character Embedding Size	100
Learning Size	0.014
Decay Rate	0.05
Dropout	0.5
Batch Size	10
CNN Window Size	3
CNN Number of filters	50

4.3. Experiment results

In this part, we describe all of models in the following experimental comparison.

GRU+Softmax: We combine gated recurrent unit (GRU) neural network and Sftomaxto classify assertion of clinical named entity. In this model, the GRU neural network is used to help encoding character embedding vector and then the Softmax layer is used to decode and classify. To compare the impact of different methods on experimental performance, we will use this model as the baseline.

CNN+GRU+Softmax: This model is similar to the CNN-LSTM-CRF which was proposed by Ma and Hovy (2016)[15] and is a truly end-to-end system.

CGAtS(CNN+GRU+Attention+Softmax): This model is the CNN-GRU-Softmax architecture enhanced by medical knowledge attention (MKA). In this model the output of hidden layer h and the attention score a are used to encode text representation as follows:

$$c = \sum_{i=1}^L a_i * h_i \quad (13)$$

where L is the window size of text characters.

CGAtFuFiS(CNN+GRU+Softmax+All): This is our model. Unlike the previous one, we employed a gated multi-modal fusion (GMF) mechanism and a filtration gate.

The performance on each of seven categories obtained by all models are shown in Figure 4, and their overall performance on the evaluation dataset is shown in Table 4.

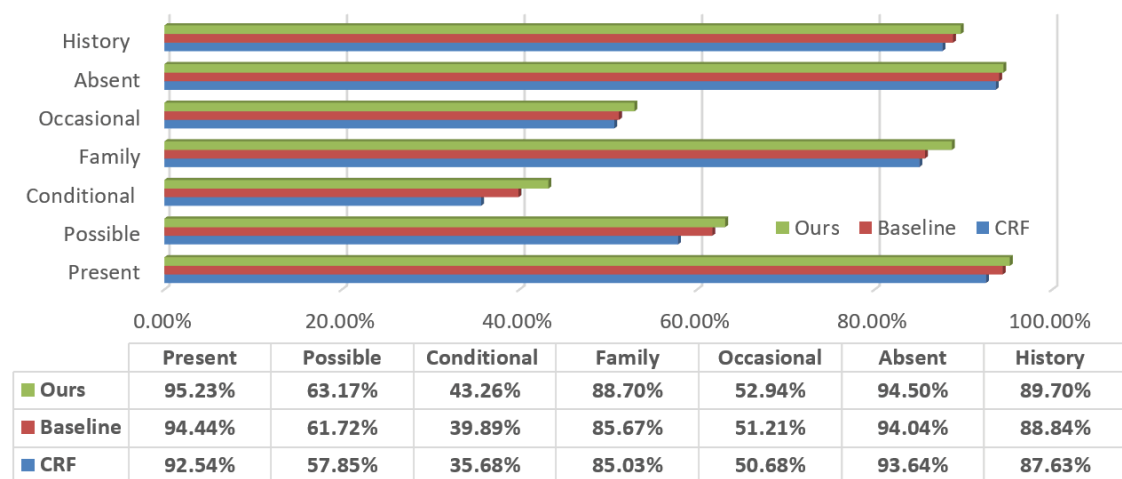


Figure 4. Experimental results of different assertion categories.

We compare our model with the baseline. Table 4 shows the overall assertion classification performance obtained by our method and others, from which we can see that our model CGAtFuFiS obtains the best $F1$ -score of 92.84%.

Table 4. Performance of different models on the total evaluation dataset.

Model	<i>Precision (%)</i>	<i>Recall (%)</i>	<i>F1 (%)</i>
Softmax	89.13	88.31	88.72
GRU+Softmax(Baseline)	90.21	90.77	90.49
CNN-GRU-Softmax	92.95	89.65	91.27
CGAtS	90.76	93.34	92.03
CGAtFuFiS	92.19	93.48	92.84

5. Discussion

The experimental results of different models on our manually annotated datasets are shown in Table 4 and 5. Compared with the baseline model, all other models have improved performance and the updated neural network model is better than the traditional machine learning methods on the MAC task.

The convolution layer in convolution neural network can well describe the local features of characters, and the most representative part of the local features can be further extracted through the pooling layer. Therefore, our experimental results show that CNN-GRU-Softmax model is superior to GRU-CRF model. The performance of the CGAtS model is better than CNN-GRU-Softmax. This result shows that, the rich information of entities and their corresponding semantic definition from MDBs is surely useful for MAC. CGAtFuFiS model is slightly better than CGAtS model and indicates that for the clinical NER task in Chinese EMRs it is helpful to fuse the features from EMR text context with the external knowledge dictionary utilizing gated multimodal fusion (GMF). Since supplement of external information in MKBs sometimes causes noise to the model, we therefore use a filtering gate to combine and weight different features. As shown by the experimental results, the filtering gate is helpful to improve the overall performance of our model.

Due to the sublanguage characteristic of Chinese EMRs, the expression of clinical named entity is very different from those in general text. Using the entity information contained in the MABs as the classification query vector can lead the decoder to focus on the entity itself. We combine text itself and MABs features together with a multi-modal fusion gate as the query vector, then set up a filtering gate to filter out useless feature information. The experimental results show that our model CGAtFuFiS, which integrates CNN, GRU, medical knowledge attention, gated multimodal fusion, filtering gate, and Softmax, achieves the best F1 score on the evaluation corpus.

6. Conclusion

In this work, we proposed a medical knowledge-attention enhanced neural clinical entity assertion classification model, which makes use of the external MABs in the way of attention mechanism. A gated multi-modal fusion module is introduced to decide how much MABs information is fused into the query vector at each time step. We further introduced a filtering gate module to adaptively adjust how much multi-modal information can be considered at each time step. The experimental results on the manually annotated Chinese EMR evaluation dataset show that our proposed approach improved the performance of MAC task obviously compared to other

baseline models.

In the future, we will explore a fine-grained clinical entity classification model for Chinese EMRs and method to extract entity semantic relation in Chinese EMRs.

Acknowledgments

We would like to thank the anonymous reviewers for their valuable comments. The research work is supported by the National Natural Science Foundation of China (NO. 61762081, No.61662067, No. 61662068) and the Key Research and Development Project of Gansu Province (No. 17YF1GA016).

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. W. W. Chapman, W. Bridewell and P. Hanbur, et al., A simple algorithm for identifying negated findings and diseases in discharge summaries, *J. Biomed. Inform.*, **34**(2001), 301–310.
2. H. Harkema, J. N. Dowling and T. Thornblade, et al., Con Text: an algorithm for determining negation, experiencer, and temporal status from clinical reports, *J. Biomed. Inform.*, **42**(2009), 839–851.
3. M. Jiang, Y. Chen and M. Liu, et al., Hybrid approaches to concept extraction and assertion classification—vanderbilt’s systems for 2010 I2B2 NLP Challenge, Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data, Boston, MA, USA: i2b2, (2010).
4. K. Roberts, B. Rink and S. Harabagiu, Extraction of medical concepts, assertions, and relations from discharge summaries for the fourth i2b2/VA shared task, Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data, Boston, MA, USA: i2b2, (2010).
5. B. de Bruijn, C. Cherry and S. Kiritchenko, et al., NRC at i2b2: one challenge, three practical tasks, nine statistical systems, hundreds of clinical records, millions of useful features, Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data, Boston, MA, USA: i2b2, (2010).
6. D. Demner-Fushman, E. Apostolova and R. Islamaj Dogan, et al., NLM’s system description for the fourth i2b2/VA challenge, Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data. Boston, MA, USA: i2b2, (2010).
7. C. Grouin, A. B. Abacha and D. Bernhard, et al., CARAMBA: concept, assertion, and relation annotation using machine-learning based approaches, Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data. Boston, MA, USA: i2b2, (2010).
8. G. Divita, O. Z. Treitler and Y. J. Kim, et al., Salt lake city VA’s challenge submissions. Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data, Boston, MA, USA: i2b2, (2010).

9. A. M. Cohen, K. Ambert and J. Yang, et al., OHSU/portland VAMC team participation in the 2010 i2b2/VA challenge tasks, Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data. Boston, MA, USA: i2b2, (2010).
10. P. Anick, P. Hong and N. Xue, et al., I2B2 2010 challenge: machine learning for information extraction from patient records. Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data. Boston, MA, USA: i2b2, (2010).
11. E. Chang, Y. Xu and K. Hong, et al., A hybrid approach to extract structured information from narrative clinical discharge summaries, Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data. Boston, MA, USA: i2b2, (2010).
12. C. Clark, J. Aberdeen and M. Coarr, et al., Determining assertion status for medical problems in clinical records, Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data. Boston, MA, USA: i2b2, (2010).
13. K. Cho, B. Van Merriënboer and D. Bahdanau, et al., On the Properties of Neural Machine Translation: Encoder-Decoder Approaches, Computer Science, (2014), arXiv preprint arXiv:1409.1259
14. D. Nadeau and S. Sekine, A survey of named entity recognition and classification, *Linguisticae Investigationes*, **30.1**(2007), 3–26.
15. X. Ma and E. Hovy, End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF, Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, (2016), arXiv preprint arXiv:1603.01354.



AIMS Press

©2019 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)