# A Joint Model of Clinical Domain Classification and Slot Filling Based on RCNN and BiGRU-CRF

Yuming Li*†, Pin Ni*†, Junkun Peng*, Jiayi Zhu*, Zhenjin Dai*, Gangmin Li*‡ and Xuming Bai§
*Research Lab for Knowledge and Wisdom, Xi'an Jiaotong-Liverpool University, China
†Department of Computer Science, University of Liverpool, UK
§Department of Interventional Radiology, Second Affiliated Hospital of Soochow University, China
Email: ‡Gangmin.Li@xjtlu.edu.cn, §2005baixuming@163.com

***Abstract*—The task of the Intent Classification & Slot Filling serves as a key joint task in the voice assistant, which also plays the role of the pre-work in the construction of the medical consultation assistant system. How to distribute a doctor-patient conversation into a formatted electronic medical record to an accurate department (Intent Classification) to extract the key named entities or mentions (Slot Filling) through a specialized domain knowledge recognizer is one of the key steps of the entire system. In real cases, the medical vocabulary and clinical entities in different departments of the hospital often differ to some extent. Therefore, we propose a comprehensive model based on CMed-BERT, RCNN and BiGRU-CRF for a joint task of department identification and slot filling of the specific domain. Experimental results confirmed the competitiveness of our model.**

***Index Terms*—RCNN, BiGRU-CRF, CMed-BERT, NER**

## I. Introduction

We found that in real cases, the medical vocabulary and clinical entities involved in different medical fields often have certain differences. The common medical named entities involved in the electronic medical records of each specific department also have their characteristics, the general multi-tasking or ensemble learning models are difficult to achieve the ideal, balanced recognition accuracy and coverage in EMR of different diseases. Therefore, we are inspired by the joint task of the key steps in the voice assistant: Intent Classification & Slot Filling, which can translate into a joint task of Domain Classification and Named Entity Recognition (NER) in the medical field. The main steps are followed: firstly, classified the EMR to identify which department (domain) is it belongs and deploy a more proprietary domain named entity recognizer based on the previous classification results to improve the accuracy of key mentions recognition and extraction. This work as the first medical Intent Classification & Slot Filling task and corresponding joint model proposed in the medical field, we designed the RCNN & BiGRU-CRF joint task model, which is a comprehensive mechanism that first classifies the medical records using RCNN and identifies specific medical named entities according to the specific classified domain using BiGRU-CRF. The experiments are based on a Chinese EMR dataset which contains integration of 2 real-world EMR samples, SAHSU and CCKS 2019. After the 20 times comparison experiments with previous State-Of-The-Arts models in NER and classification, respectively, the effectiveness of our model can be confirmed.

## II. Literature Review

Meystre et al. [1] retrieved more than 200 articles, and found out most systems used only one or two specific clinical document types, mostly based on two different sets of approaches: pattern matching and machine learning. Li et al. [2] and Gao et al. [3] proved that Natural Language Processing is the best method to solve the task of electronic medical record processing.

Text classification is one of the basic tasks in the field of Natural Language Processing, its core is to classify text content for one or more categories, and has been applied in many areas of the actual [4], [5] Named Entity Recognition is one of the key tasks of domain-specific information extraction. It is a technique for locating named entities composed of words and assigning them to predefined categories (such as drugs, anatomical sites, diseases and diagnostics, surgery, etc. in this task). NER method can be divided into dictionary-based [6] and dictionary-based machine learning model [2], [7]. Pineda et al. [8] Used LSTM-RNN method to automatically classify unstructured medical description, and obtained accuracy and F1 Score higher than decision tree [9] and random forest [10].

It can be seen that both the Classification task and the Named Entity Recognition task have been developing continuously and rapidly [11], however, there is still a gap in the comprehensive Named Entity Recognition in the combination of classification.

## III. Methodology

The EMR text will be converted into a specific vector-matrix based on context by using our previously proposed C-BERT language model as an introduction of prior knowledge. And the RCNN model [12] as a type of classifier which mainly consists of 3 parts: Recurrent CNN layer, max-pooling layer and output layer, which is used to determining the domain it should belong. After being distributed to the corresponding domain according to the RCNN classification results, and embedding by the corresponding domain exclusive language model "C-BERTs", which pre-trained by corresponding specific domain knowledge, and the output matrix will be inputted into the corresponding exclusive trained downstream NER model to

978-1-7281-0858-2/19/$31.00 ©2019 IEEE

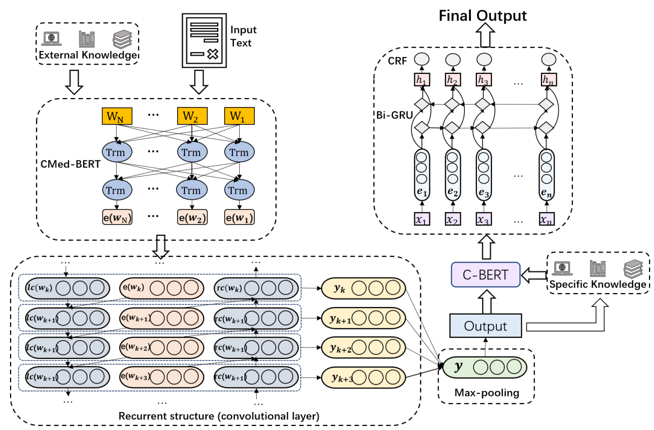


Fig. 1. Architecture of Clinical Domain Classification and Slot Filling Model

recognize the specific named entity or mentions. In the NER section, BiGRU is used to learn the content of the context text and capture the entities, and mapping is conducted through CRF layer to finally output the category of entities (Fig. 1).

### A. Intent Classification

*1) CMed-BERT layer:* Devlin et al proposed BERT [13] in 2018, the first unsupervised, in-depth and bi-directional language representation model for Natural Language Processing pre-training. Different from simple context information accumulation by ELMo [14] and OpenAI GPT [15], BERT jointly modulates context in all layers to pre-train depth bi-directional representation, which can achieve better semantic disambiguation effect. In the proposed model, the input text is assumed to be $[W_1, W_2, ..., W_n]$ where n is the length of the text, $W \in R^V$ and V is the vocabulary size. We uses the classic Chinese literatures of biomedical field as the training corpus of original BERT, and defines the trained domain knowledge language model as CMed-BERT. The descriptions of the training corpora for CMed-BERT are listed in Table 1.

TABLE I
CORPORA DESCRIPTION

| Corpus Source | Description | Size(word) |
|---|---|---|
| Medical Books | We use 13 specific books as training corpus, including *Psychiatry*, *Clinical Drug Therapy*, ect. | 4384503 |
| SAHSU | The electronic medical records of the Second Affiliated Hospital of Soochow University, including 5090 electronic medical record. | 2002202 |
| Online Resources | All the data were collected from 4 professional Chinese health websites including ”39 Health (39.net)”, ”XunYiWenYao (XYWY.com)”, ”Feihua Health (fh21.com.cn)”, ”NetEase Health (jiankang.163.com)”. | 29092216 |

*2) Recurrent Convolutional Neural Networks (RCNN) Layer :* RCNN model mainly consists of three parts, which are Recurrent CNN layer, max-pooling layer and output layer. In the Recurrent CNN layer, for each word, RCNN will recursively calculate its left context vector and right context vector, and then splicing these two parts of vectors with the word vector of the current term as the vector representation of the word. Word embedding for each word is the output of CMed-BERT layer $((e(w_1), e(w_2), ..., e(w_N)))$, the right-side context vector $rc(w_k)$ and the left-side context vector $lc(w_k)$ are defined as equation (1) and (2):

$$rc(w_k) = f\left(W^{(r)} rc(w_{k+1}) + W^{(rs)} e(w_{k+1})\right) \quad (1)$$

$$lc(w_k) = f\left(W^{(l)} lc(w_{k-1}) + W^{(ls)} e(w_{k-1})\right) \quad (2)$$

In which $W^{(r)}, W^{(rs)}, W^{(l)}, W^{(ls)}$ are all weight matrixes, used for calculating the right context, right semantic, left context and left semantic respectively. The three vectors are spliced together as a vector representation of the current word, then pass in a full connection layer with tanh activation function to calculate the potential semantic vector of word $k$:

$$y_k = \tanh\left(W\left[lc(w_k); e(w_k); rc(w_k)\right] + b\right) \quad (3)$$

In the max-pooling layer, RCNN gets the most important elements $\max_{k=1}^{n} y_k$ of all potential semantic vectors.The resulting vector $\boldsymbol{y}$ is the vector representation of the whole text. In the output layer, RCNN passes the resulting text vector into a full connection layer with softmax to get the probability distribution of the current text in each category $c$, $p(c|D, \theta)$, in which D is the input text, $\theta$ is a parameter of the network.

### B. Slot Filling

*1) C-BERT Layer:* CMed-BERT is another pre-train model which is determined by the output of the RCNN layer (maximum possible category $c$), training corporate is specific knowledge related to this category

*2) Bi-GRU Layer:* Gated Recurrent Unit (GRU) is the variant of Long Short Term Memory (LSTM) variant, and also the simplified LSTM by reducing the number of network parameter, but the model will not be affected by the performance, in the solution to the problem of long-term dependence on sequence at the same time can effectively shorten the training time. Bidirectional GRU can capture textual characteristics of the context simultaneously, and get the output sequence $[h_1, h_2, h_3 ... h_n]$

*3) Conditional Random Field:* Conditional Random Field (CRF) is a conditional probability distribution model of another set of output random variables given a set of input random variables. The CRF layer learns the conditional probability $p(y|h)$, where $h = [h_1, h_2, ..., h_n]$ is the representation sequence produced by the upper Bi-GRU layer, $y = y_1, y_2, , y_n$ For the tag sequence. Given the input $h$, the conditional probability of the tag sequence $y$ is calculated as equation 4:

$$p(y|h; \theta) = \frac{\prod_{i=1}^{N} \psi(h_i, y_i, y_{i-1})}{\sum_{y' \in Y(s)} \prod_{i=1}^{N} \psi(h_i, y'_i, y'_{i-1})} \quad (4)$$

TABLE II
THE SAMPLE OF COMPARISON RESULTS

| Task | Original Model | | | | Pre-trained Model | | | |
|---|---|---|---|---|---|---|---|---|
| | Model Name | Precision | Recall | F1 | Model Name | Precision | Recall | F1 |
| Department / Domain (Intent) Classification | BERT-CNN-GRU | 0.6296 | 0.5835 | 0.5791 | CMedBERT-CNN-GRU | 0.6465 | 0.5785 | 0.5735 |
| | BERT-Dropout-AVRNN | 0.6862 | 0.6585 | 0.6417 | CMedBERT-Dropout-AVRNN | 0.6787 | 0.6569 | 0.6556 |
| | BERT-BiGRU | 0.6091 | 0.5918 | 0.5741 | CMedBERT-BiGRU | 0.6044 | 0.5870 | 0.5678 |
| | BERT-Dropout-BiGRU | 0.6561 | 0.6506 | 0.6310 | CMedBERT-Dropout-BiGRU | 0.6522 | 0.6195 | 0.6213 |
| | BERT-AVRNN | 0.6610 | 0.6701 | 0.6487 | CMedBERT-AVRNN | 0.7040 | 0.6716 | 0.6596 |
| | BERT-AVCNN | 0.6706 | 0.6605 | 0.6487 | CMedBERT-AVCNN | 0.7858 | 0.7268 | 0.7170 |
| | BERT-RCNN | 0.7389 | 0.7114 | 0.7028 | ***CMedBERT-RCNN*** | ***0.7456*** | ***0.7531*** | ***0.7372*** |
| NER (Slot Filling) | BERT-CNN-LSTM | 0.6330 | 0.6910 | 0.6421 | C-BERT-CNN-LSTM | 0.6481 | 0.7248 | 0.6839 |
| | BERT-BiLSTM | 0.6852 | 0.7640 | 0.7214 | C-BERT-BiLSTM | 0.7095 | 0.7536 | 0.7303 |
| | BERT-BiLSTM-CRF | 0.6897 | 0.7716 | 0.7282 | C-BERT-BiLSTM-CRF | 0.7201 | 0.7874 | 0.7516 |
| | BERT-BiGRU | 0.7047 | 0.7437 | 0.7221 | C-BERT-BiGRU | 0.7125 | 0.7635 | 0.7367 |
| | BERT-BiGRU-CRF | 0.7214 | 0.7820 | 0.7499 | **C-BERT-BiGRU-CRF** | **0.7240** | **0.8072** | **0.7632** |

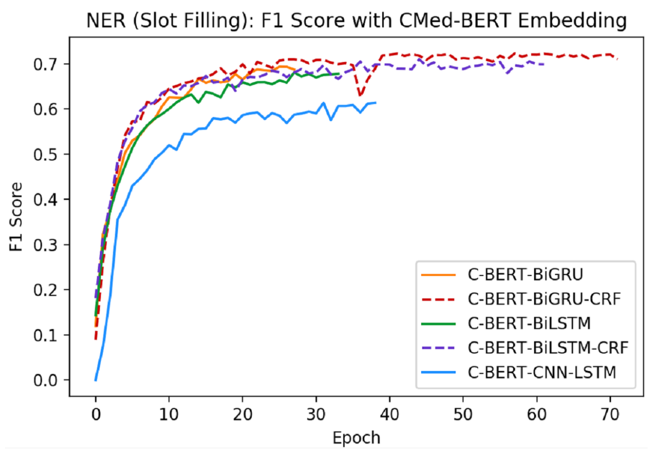


Fig. 2. NER (Slot Filling): F1 Score with C-BERT Embedding

## IV. RESULT

We have tested 7 classification model and 5 NER model, the results are showed in Tab. I and Fig. 2.

By comparing the previous State-Of-The-Arts models on the respective tasks of classification and NER, our joint model achieved better results in the corresponding sub-tasks than the other separate task-specific models (Tab. II). Experiments show that our proposed joint model has greater competitiveness than other specific mainstream models, respectively. It can also be seen from the table that for the same model, the effect of using domain-specific CMed-BERT pre-training model is better than that of original BERT, moreover, the effect of Named Entity Recognition task using C-BERT based on previous classification result is better than that without classification.

## V. ACKNOWLEDGEMENT

This work is partially supported by the AI University Research Centre (AI-URC) through XJTLU Key Programme Special Fund (KSF-P-02) and KSF-A-17. And it is also partially supported by Suzhou Science and Technology Programme Key Industrial Technology Innovation programme with project code SYG201840. We appreciate their support and guidance.

## REFERENCES

[1] Stephane M Meystre, F Jeffrey Friedlin, Brett R South, Shuying Shen, and Matthew H Samore. Automatic de-identification of textual documents in the electronic health record: a review of recent research. *BMC medical research methodology*, 10(1):70, 2010.

[2] Lishuang Li, Rongpeng Zhou, and Degen Huang. Two-phase biomedical named entity recognition using crfs. *Computational biology and chemistry*, 33(4):334–338, 2009.

[3] Hongyuan Gao, Erin J Aiello Bowles, David Carrell, and Diana SM Buist. Using natural language processing to extract mammographic findings. *Journal of biomedical informatics*, 54:77–84, 2015.

[4] Erik Cambria, Daniel Olsher, and Dheeraj Rajagopal. Senticnet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In *Twenty-eighth AAAI conference on artificial intelligence*, 2014.

[5] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 1480–1489, 2016.

[6] Carol Friedman, Philip O Alderson, John HM Austin, James J Cimino, and Stephen B Johnson. A general natural-language text processor for clinical radiology. *Journal of the American Medical Informatics Association*, 1(2):161–174, 1994.

[7] Yan Xu, Junichi Tsujii, and Eric I-Chao Chang. Named entity recognition of follow-up and time information in 20 000 radiology reports. *Journal of the American Medical Informatics Association*, 19(5):792–799, 2012.

[8] Arturo Lopez Pineda, Oliver J Bear, Guhan R Venkataraman, Ashley M Zehnder, Sandeep Ayyar, Rodney L Page, Carlos D Bustamante, and Manuel A Rivas. Fastag: automatic text classification of unstructured medical narratives. *bioRxiv*, page 429720, 2019.

[9] J. Ross Quinlan. Simplifying decision trees. *International journal of man-machine studies*, 27(3):221–234, 1987.

[10] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[11] Abeed Sarker and Graciela Gonzalez. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. *Journal of biomedical informatics*, 53:196–207, 2015.

[12] Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. Recurrent convolutional neural networks for text classification. In *Twenty-ninth AAAI conference on artificial intelligence*, 2015.

[13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[14] Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*, 2018.

[15] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. *URL https://s3-us-west-2. amazonaws. com/openai-assets/researchcovers/languageunsupervised/language understanding paper. pdf*, 2018.