

# MedicalNER: Named Entity Recognition for Medical Text via Character-level BiLSTM-based Models

Internship Candidate: Shuyue Jia

October 16, 2020

## Abstract

With the development of Medical Artificial Intelligence (AI) System, Natural Language Processing (NLP) has played an essential role to process medical texts and build intelligent machines. Named Entity Recognition (NER), one of the most basic NLP tasks, is primarily studied since it is the cornerstone of the following NLP downstream tasks, e.g., Relation Extraction. In this work, a character-level Bidirectional Long-short Term Memory (BiLSTM)-based models were introduced to tackle the challenge of medical texts. The input character embedding vectors were randomly initialized and then updated during training. The character-level BiLSTM extracted features from medical order-matter sequential data. The followed Conditional Random Field (CRF) predicted the final entity tag. Results have shown that the presented method took advantage of the recurrent architecture and achieved competitive performances for medical texts. Promising results paved the road towards building robust and powerful medical AI engines.

**Keywords**— Named Entity Recognition (NER), Bidirectional Long-short Term Memory (BiLSTM), Convolutional Neural Network (CNN)

## 1 Introduction

Currently, researchers in the medical field have paid more attention towards building a robust Medical Artificial Intelligence (AI) System, taking IBM Watson as an example. Enterprise-level Medical AI products often suit up with a NLP engine to process medical texts and assist doctors. Multiple modules, however, made up with such a engine. Among those, Named Entity Recognition (NER), one of the most basic and cornerstone for medical NLP, plays a critical rule as its performance can directly influence the followed tasks, e.g., Relation Extraction (RE). Thus, the model needs to be elaborately designed and engineers try to achieve a trade-off between performance and model complexity, which can profoundly affect the deployment and operations (DevOps) in the industry. In this work, we presented character-level Bidirectional Long-short Term Memory (BiLSTM) to address the challenge of NER in the medical field.

## 2 Method

### 2.1 Character Embedding Layer

As for the character-level BiLSTM model, the input to the BiLSTM is an embedded vector of a character. Before training the model, the word embeddings were randomly initialized with the dimensionality  $K$ . During training, a Deep Neural Network (DNN) hidden layer was applied to update the word embeddings which were the distributed representations for input characters. The character embeddings were set while evaluating.

### 2.2 BiLSTM Model

In recent years, Recurrent Neural Network (RNN) [1] was broadly implemented to order-matter time-series signals [2]. One drawback of RNN was that it suffered a lot from gradient vanishing problem [3]. When there were a longer sequence, it cannot back-propagate errors and the long-term gradient was lost. To curb the problem, in this work, we employed BiLSTM model, which can learn the long-term dependencies and relief gradient vanishing problem. The input of BiLSTM was a sequence of medical texts with the dimensionality  $N \times D$ .  $N$  is the number of characters in the sequence, and  $D$  is the dimension of embedding for each character.

### 2.3 CRF Model

A softmax layer can output a probability distribution of entity tags. However, it does not consider the relationship between different tags. Thus, we applied the Conditional Random Field (CRF) model which can predict the possibility of each tag and the possibility between the tag and the followed tag. The CRF Model was optimized by Viterbi algorithm [4].

### 2.4 Model Architecture

The main model was a fully-connected DNN layer, BiLSTM, and CRF model [5]. The fully-connected DNN layer updated character embeddings, the BiLSTM extracted features from medical sequences, and the CRF model output predicted entity tag. The loss function we used was Negative Log-likelihood (NLL) Loss. The model parameters were learned through back-propagation algorithm.

## 3 Basics inside NER Task

### 3.1 Data Annotation

We used *BIOES tagging Scheme*: B: beginning, I: intermediate / inside, O: other, E: end, S: singleton / single.

### 3.2 Evaluation Metrics

Tag/entity level, Character level, and label-level: Accuracy, Precision, Recall, F1-score  
Tag-level: measure an entire entity (B-entity, (I-entity,) E-entity) is right or

not. Character-level: measure each character’s prediction is right or not. label-level: measure the performance from B / I / O / E / S.

## References

- [1] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.
- [2] Yimin Hou, Shuyue Jia, Shu Zhang, Xiangmin Lun, Yan Shi, Yang Li, Hanrui Yang, Rui Zeng, and Jinglei Lv. Deep feature mining via attention-based bilstm-gcn for human motor imagery recognition. *arXiv preprint arXiv:2005.00777*, 2020.
- [3] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks. In *International conference on machine learning*, pages 1310–1318, 2013.
- [4] Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. Neural architectures for named entity recognition. *arXiv preprint arXiv:1603.01360*, 2016.
- [5] Jing Li, Aixin Sun, Jianglei Han, and Chenliang Li. A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*, 2020.