BERT-BiLSTM Model for Entity Recognition in Clinical Text

Zhengzhong Zhu[1], Liqing Wang[1]

[1]School of Inoformation Science and Engineering, Yunnan University, Yunnan 650091, P.R. China

Abstract

LivingNER2022, held by IberLEF 2022, proposes three subtasks to address the issue of the automatic system for semantic analysis of species mentioned in non-English inconsistent documents, such as name changes (obsolete species names), homonymy with commonly used words and so on, through a challenge on NER of species mentions and entity linking providing an exhaustively annotated large corpus of Spanish clinical case reports. Our team only participated in subtask A: Entity Recognition. This study introduces the system submitted to the LivingNER2022 by the zzz team. Our team used a BERT-BiLSTM model-based entity recognition for multidomain text features extraction. The BERT-BiLSTM architecture has been proved to better capture the global dependencies of the input text.

Keywords

LivingNER2022, BERT, BiLSTM,

1. Introduction

In recent years, natural language automatic knowledge extraction technology has been widely used in various fields, but it is rarely involved in the medical field. The collection of plain text clinical case report documents contains many corpora [1], especially for non-English content. High-quality guides and corpora that provide fine-grained text binding semantic annotations are needed to take advantage of biological references in biomedical texts. Annotation of species or organisms is essential for scientific disciplines such as medicine, biology, ecology/biodiversity, nutrition and agriculture. The LivingNER2022 shared task will address these issues by challenging NER's species mention and entity linking, providing a large corpus of fully annotated Spanish clinical case reports. Task 1 is LivingNER2022 specific NER track (species mention entity recognition). The purpose of this task is to give a set of plain text clinical case report documents and participants must return the accurate character offset of all mentioned species, including HUMAN and SPECIES. The rest of this article is arranged as follows. Section 2 introduces the model architectures used by zzz team. The experimental settings are obtained in Section 3. Section 4 presents the local performance of three models for the LivingNER2022 first task and the prediction scores returned by the LivingNER2022 to our team. Section 5 introduces the discussion of this paper, and Section 6 presents the conclusion of this paper and some opinions for future work.

IberLEF 2022, September 2022, A Coruña, Spain.
2251821381@qq.com (Z. Zhu)
https://github.com/2251821381 (Z. Zhu)
0000-0002-6138-7611 (L. Wang)

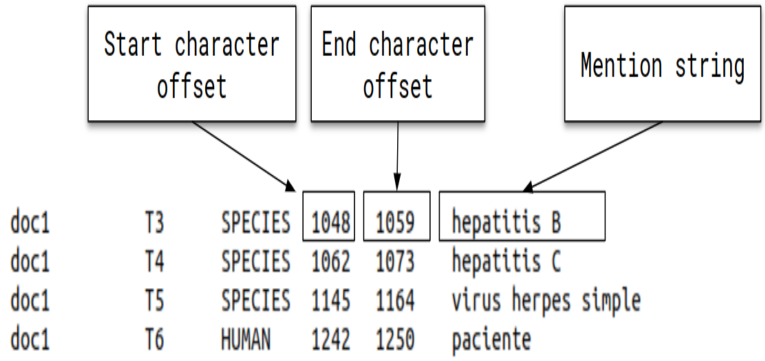

Figure 1: Example annotations for LivingNER – Species NER.

2. System Description

Our team used a BERT-based multilingual system and BiLSTM to predict entities. In this section, we detail the architecture of the system and how to handle inputs and outputs.

2.1. Model architecture

The BERT-BiLSTM model[2] consists of BERT module, BiLSTM. The overall model is shown in Figure 2.

The full name of BERT[3] is a bidirectional encoder representation from transformers[4], which is a pretrained language representation model. It emphasizes that instead of using the traditional one-way language model or shallow splicing of two one-way language models for pretraining. A new masked language model (MLM) is used to generate deep two-way language representation. The MLM model has two main advantages: First, it is used to pretrain bidirectional transformers to generate deep bidirectional language representation; Second, after pretraining, only an additional output layer is needed to add for fine-tuning to achieve the performance of state-of-the-art in a variety of downstream tasks. This process does not need to make task-specific structural changes to BERT. In addition, BERT can calculate the relationship between words, and extract the important features in the text by using the calculated relationship adjustment weight. Pretraining by using the structure of self attention[5] mechanism captures the real context information and can learn the relationship between continuous text segments relationship between sentences. Due to the fact that the semantic relationship between sentences is very important in the named entity recognition task, BERT model splices sentences L and M, and predicts whether M is behind L in the original text. So we use the BERT model to solve the polysemy problem and improve the named entity recognition task to achieve a remarkable effect.

Bi-directional Long Short-Term Memory [6] was proposed in 1997. It is the most popular recurrent neural network at present. BiLSTM is the abbreviation of Bi-directional Long Short-Term Memory, a combination of forward LSTM and backward LSTM. LSTM model is composed of input gate, forgetting gate, cell state, and output gate added on the basis of RNN. In the

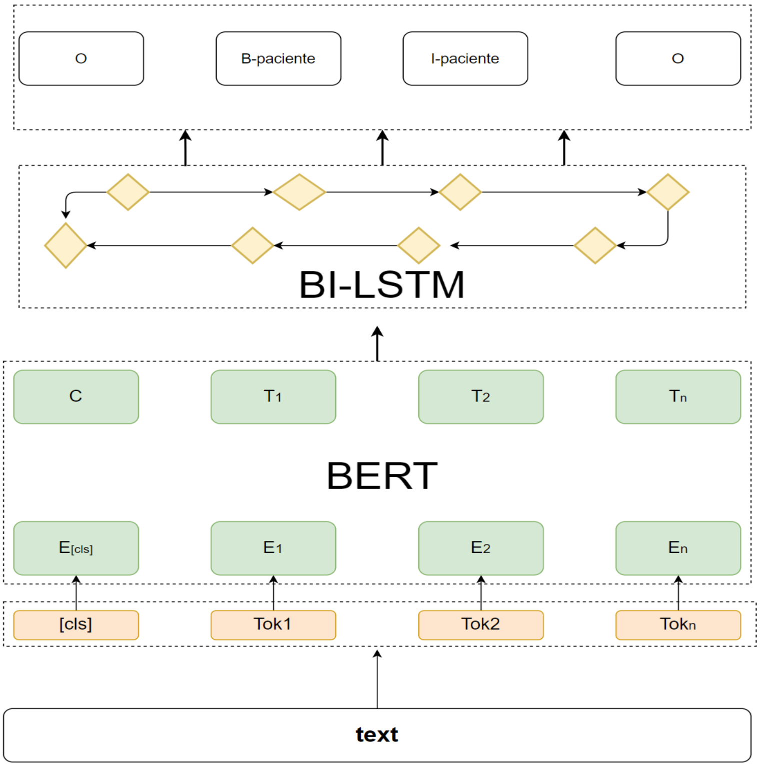

Figure 2: Architecture of the model

process of network training, information can be added or removed through the gate structure. Different neural networks can decide which relevant information to remember or forget through the gate structure of the unit state. It can be seen that it is very suitable for sequence annotation tasks with upper and lower relationships, so it is often used to model context information in NLP. The model first generates word vectors based on context information[7] through the pretraining of BERT model, then inputs the trained word vectors into the BiLSTM model for further training, and finally outputs entity information through a softmax layer.

2.2. Input handling

Our team considered training sets and test sets appear in the form of a document. The length of many sentences will exceed the BERT default sentence length of 512. If the document text is directly input into the BERT model for training, the sentence length will exceed the maximum length. Our solution is to split the sentence according to space and period first and then send the split sentences to the model for training. However, this will cause some problems. The goal of the task is to find out the entities, the span of the entities, and the number of entities in the document. After splitting the long sentences of the document, although the entities can be identified, it will cause the span of the entities and the number of entities in the document to be incorrect. Therefore, it is necessary to correct the output prediction file. In addition,our team use the standard BIO annotation format [8] for entities. The classifier could label each token of the text according to 5 categories: O, B-HUMAN, I-HUMAN, B-SPECIES, I-SPECIES, "B-X" means that the fragment of this element belongs to type X and this element is at the beginning of this fragment, "I-X" means that the fragment of this element belongs to type X and this element is in the middle of this fragment, "O" means that it does not belong to any type.

2.3. Output handling

The model's output contains a series of tags, and each tag is assigned to an entity tag. For each tag, if it is the beginning of an entity, we identify the character range it spans and add this range list to the result[9]. In the input processing stage, we input the whole document into the model in terms of sentences, so only a small part of the predicted entity span is correct. By recording the location of each entity and the length of the sentence where the entity is located, and by means of boundary detection, we try to make the location of the entity in the prediction file as correct as possible.

3. Experimental settings

The Bidirectional LSTM is constructed with 384 hidden units. During training, we use the backpropagation algorithm[10] and SGD optimizer with an initial learning rate of 2e-5. The training batch size was 1. All models are trained on a single GTX 3060Ti GPU. In the first step, we train the model until the F1-score does not improve significantly on the validation set. The loss function uses the cross entropy loss function[11].

$$L = -[y * \log(p) + (1 - y) * \log(1 - p)],$$

y is the real label value (positive class value is 1, negative class value is 0), and p is the predicted probability value, which represents the difference between the real sample label and the prediction probability. Second, in the prediction file correction stage, we use the boundary detection algorithm to compare the entity file predicted by the model with the test set and correct the Mark, off0 and off1 of the entities in the prediction file. The specific method is to find the off0 and off1 of the entities according to the filename of the predicted entities. Meantime we set the range of the parameter i to (-100, 100) until the value of off0 plus i and the value of off1 plus i are equal to the true span of the label.

4. Result

The first task of livingNER2022 adopts the common evaluation indicators the micro-average F1-score, Precision and Recall. Table 1 displays the results (Precision, Recall , and F1-score) reported by BERT-LSTM, BERT, BERT-BiLSTM. It can be seen from Table 1 that the BRET-BiLSTM model performed best in the development set, with an F1-score of 0.88. So we chose the BERT-BiLSTM model for prediction. The Recall, Precision, F1-score of BERT-BiLSTM model on the test set returned by the LivingNER2022 to our team are 0.80,0.61 and 0.70, respectively.

Table 1
Comparison of Recall,Precision and F1-score of our three models on development set.

Model	Recall	Precision	F1
BERT	0.87	0.79	0.83
BERT-LSTM	0.83	0.90	0.87
BERT-BiLSTM	0.85	0.90	0.88

5. Discussion

We use a common BERT-BiLSTM model to complete this entity recognition task. The model performs excellently on task A, after training the model 10 epochs in total. Our final LivingNER2022 results of the Recall, Precision, and the F1-score are 0.80, 0.61 and 0.70, separately. The final results show that there is still a certain gap between our prediction results and the prediction results of other participants. Our final F1-socre is only 0.70. The possible reasons are two-fold. First, in the data preprocessing stage, we cut sentences with a length greater than 512, which will cause a big gap between the predicted entity span and the real span. Although the entity span is corrected in the output processing, there are still errors. Second, we use the traditional BERT and BiLSTM model to extract text features, making it difficult to achieve satisfactory performance in this task.

6. Conclusion

In this paper, we study and design a BERT-BiLSTM entity recognition method based on a new language model BERT. We use the pretraining model BERT as the basic representation, BiLSTM further extracts features. The final result of the system is not satisfactory. In future work, we will try to solve these problems with some latest methods, such as using model ensemble learning [12] and prompt learning[13] which may improve the result.

7. Acknowledgments

This work was supported by the Natural Science Foundations of China under Grants 61862064.

References

[1] S. L.-L. D. E. L. G. Antonio Miranda-Escalada, Eulàlia Farré-Maduell, M. Krallinger, Mention detection, normalization classification of species, pathogens, humans and food in clinical documents: Overview of livingner shared task and resources, Procesamiento del Lenguaje Natural (2022).

[2] R. Cai, B. Qin, Y. Chen, L. Zhang, R. Yang, S. Chen, W. Wang, Sentiment analysis about investors and consumers in energy market based on bert-bilstm, IEEE Access 8 (2020) 171408–171415.

[3] J. Kong, J. Wang, X. Zhang, Hierarchical bert with an adaptive fine-tuning strategy for document classification, Knowledge-Based Systems 238 (2022) 107872. URL: https://www.sciencedirect.com/science/article/pii/S0950705121010479. doi:https://doi.org/10.1016/j.knosys.2021.107872.

[4] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, et al., Huggingface's transformers: State-of-the-art natural language processing, arXiv preprint arXiv:1910.03771 (2019).

[5] K. Clark, U. Khandelwal, O. Levy, C. D. Manning, What does bert look at? an analysis of bert's attention, arXiv preprint arXiv:1906.04341 (2019).

[6] G.-A. Vlad, M.-A. Tanase, C. Onose, D.-C. Cercel, Sentence-level propaganda detection in news articles with transfer learning and bert-bilstm-capsule model, in: Proceedings of the second workshop on natural language processing for internet freedom: Censorship, Disinformation, and Propaganda, 2019, pp. 148–154.

[7] E. Alsentzer, J. R. Murphy, W. Boag, W.-H. Weng, D. Jin, T. Naumann, M. McDermott, Publicly available clinical bert embeddings, arXiv preprint arXiv:1904.03323 (2019).

[8] P. Stenetorp, S. Pyysalo, G. Topić, T. Ohta, S. Ananiadou, J. Tsujii, Proceedings of the demonstrations at the 13th conference of the european chapter of the association for computational linguistics, Association for Computational Linguistics Avignon, 2012.

[9] L. Pavanelli, E. T. R. Schneider, Y. B. Gumiel, T. C. Ferreira, L. F. A. de Oliveira, J. V. A. de Souza, G. P. M. Paiva, L. E. S. e Oliveira, C. M. C. Moro, E. C. Paraiso, et al., Pucrj-pucpr-ufmg at ehealth-kd challenge 2021: A multilingual bert-based system for joint entity recognition and relation extraction, in: IberLEF@ SEPLN, 2021.

[10] R. Rojas, The backpropagation algorithm, in: Neural networks, Springer, 1996, pp. 149–182.

[11] Z. Zhang, M. Sabuncu, Generalized cross entropy loss for training deep neural networks with noisy labels, Advances in neural information processing systems 31 (2018).

[12] Q. Duan, N. K. Ajami, X. Gao, S. Sorooshian, Multi-model ensemble hydrologic prediction using bayesian model averaging, Advances in water Resources 30 (2007) 1371–1386.

[13] N. Ding, Y. Chen, X. Han, G. Xu, P. Xie, H.-T. Zheng, Z. Liu, J. Li, H.-G. Kim, Prompt-learning for fine-grained entity typing, arXiv preprint arXiv:2108.10604 (2021).