COMP 4901K Project Report

JIANG Wei, SUN Yushi and ZHANG Chi

1. Overview

In this project, we aim to predict the tag of each token in a given sentence from the scholarly literature about COVID-19 with a rather than high precision rate.

To achieve such a goal, our group adopts the bidirectional and unidirectional versions of LSTM and GRU models with ensemble technique whose accuracy rate on the validation set is over 92.01%. In the following sections of this report, we will introduce the details of our model and give an evaluation of it.

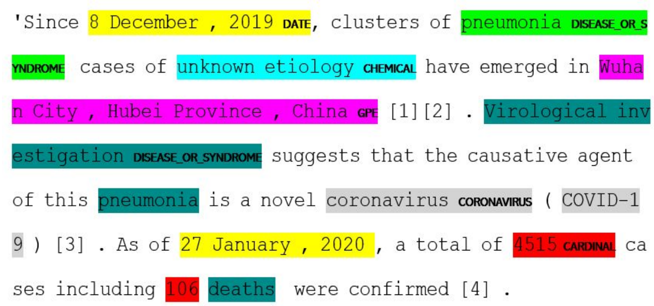

Figure 1: Visualization of the expected result

2. Methodology

In this project, we adopt RNN as the basic structure to complete the name entity recognition task. Two types of models are brought into consideration: LSTM [1] and GRU [2] with their respective bi-directional and unidirectional versions. Model ensemble technique is used to further improve the performance of our models. This section will introduce the structures of our base models and the model ensemble technique we used in this project.

Note: The code with pre-trained models on Kaggle: https://www.kaggle.com/ysunbp/4901k-project

2.1. Structures for base models

For the base models, we employ a double-RNN structure. As shown in Figure 2, the input text index data is first processed into word embeddings by the default embedding layer, with the embedding dimension of 60, provided by Keras. Then, the word embeddings are passed through a dropout layer to avoid overfit and improve the generalization of our models. After that, four different types of double-RNNs are used, which are Bi-LSTM+LSTM, Bi-LSTM+GRU, Bi-GRU+LSTM and Bi-GRU+GRU, with the hidden size of 120. Moreover, Batch Normalization Layer is added after the double-RNN structure to improve the generalization and the trainability of our model. Finally, a Timedistributed Dense Layer is used to generate the final prediction of name entities.

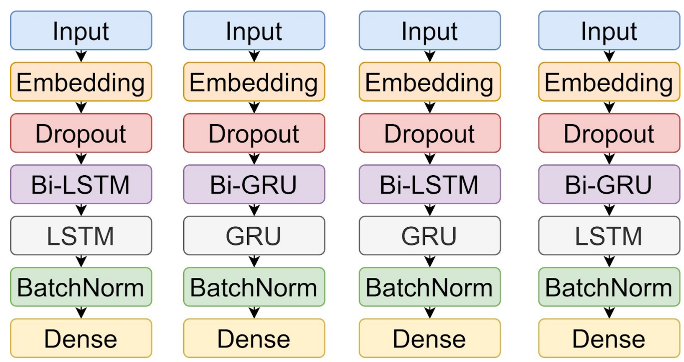

Figure 2: The four base models

2.2. Model Ensemble technique

A weighted ensemble mechanism is employed in this part. Generally, we have four double-RNN base models and we want to generate the final prediction by weighing the output vectors from these models. In this way, the diversities of different models are captured and a better performance can be achieved. Initially, we used average weight for each model and a validation accuracy of over 90% was achieved. Then, we assigned different weights (0.25 for

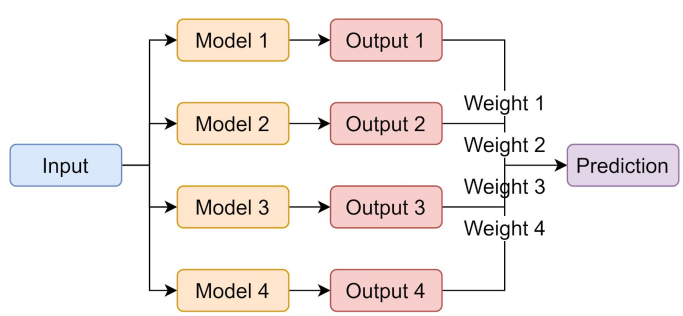

Figure 3: Model Ensemble

Bi-LSTM+LSTM and Bi-GRU+GRU, 0.5 for Bi-GRU+LSTM, 0 for Bi-LSTM+GRU) to these models based on their respective performance on validation set and a final performance of over 92% validation accuracy is achieved.

3. Evaluation

To improve accuracy of our model, we fine-tuned hyperparameters and tried different combinations of model structures.

3.1 Learning Rate

The learning rate affects the accuracy of the model to a great extent, so it needs to be carefully selected. With other hyperparameters being the same, we use Bi-LSTM+LSTM to tune the learning rate rather than using the ensembled system to save tuning time. Results are shown in Figure 4.

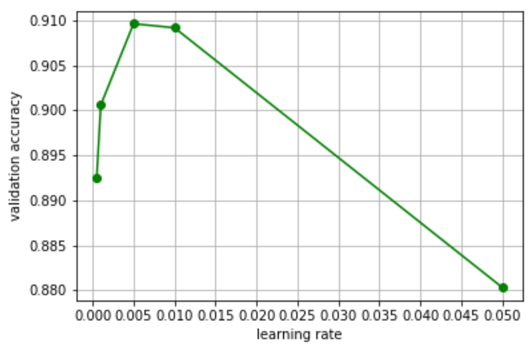

Figure 4

ID	Model	Validation Accuracy
1	Bi-LSTM+LSTM	0.9109664729
2	Bi-GRU+GRU	0.9126934536
3	Bi-LSTM+GRU	0.9078827852
4	Bi-GRU+LSTM	0.9146923544
5	Bi-LSTM+LSTM+LSTM	0.9087014377
6	Ensemble (1,2,3,4)	0.9078365009
7	Ensemble (1,2,4)	0.9200179351
8	Ensemble (1,2,4,4)	0.9201191819

Table 1

We use learning rates of 0.0005, 0.001, 0.005, 0.01, 0.05 to test the validation accuracy. The model can achieve pretty good validation accuracy with learning rates between 0.001 and 0.01 (over 90%) and we choose 0.005 here, which achieves the highest accuracy for Bi-LSTM+LSTM model. Since all models we ensembled have similar structures, this learning rate works fine for all individual models.

3.2 Model Structure

Both LSTM and GRU are recurrent neural networks with memory. GRU is like LSTM with additional forget gates. We experimented with different combinations of GRU and LSTM, and the first layer is chosen to be bidirectional to get information from both past and future states simultaneously. Results are shown in Table 1. Although we didn’t repeat the same experiment multiple times due to the long training time, Bi-GRU+LSTM generally achieves the highest validation accuracy among all combinations. One layer RNNs have a lower accuracy and three-layer RNNS have no help to improve the precision but only increase training time, so we choose to use two-layer RNNs.

3.3 Ensemble

We tried different ways of ensemble and the result is shown in Table 1. Among the four two-layer model structures, Bi-LSTM+GRU has the lowest accuracy while Bi-GRU+LSTM attained the highest score. Ensemble Modeling all the four models didn’t yield a good result, but after the worst-performing Bi-LSTM +GRU was removed, the overall model scored higher than any individual model. We also tried to give the best performing model, Bi-GRU+LSTM, twice weights than other models, and the verification accuracy is even higher, reaching 0.9201191819.

4. References

[1] F. Gers, “Learning to forget: continual prediction with LSTM,” 9th International Conference on Artificial Neural Networks: ICANN 99, 1999.

[2] K. Cho, B. V. Merrienboer, D. Bahdanau, and Y. Bengio, “On the Properties of Neural Machine Translation: Encoder–Decoder Approaches,” Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, 2014.