Wen Hao-Liang
d09525007
d09525007@ntu.edu.tw

Homework 1
Applied Deep Learning

2022-03-22

1 Model description

A multi-layer gated recurrent unit (GRU) RNN is applied in intend classification and slot tagging:

$$h_t = \mathbf{GRU}(\mathbf{x}_{t-1}, h_{t-1}), \tag{1}$$

where $\mathbf{x}$ is the embedding of the t-th token. Adam algorithm is adopted to update the weights, which associated to minimize the negative log likelihood loss function (NLL loss). The dropout is set to fix at 0.2 for all tasks and the max number of words use for training is set as the maximum in the training set.

2 Cross validation

The training of each problem includes the cross validation with 5 folds. After 5 folds of training, the best parameters for the model will be determined by the average accuracy.

3 Data prepossessing

Firstly, the word exist in the data are collect to be encoded into token, and, the pretrained model *glove* is applied to get the embedded data of each token.

4 Result

4.1 Intent classification

To find the best parameters for intent classification, different number of layers, batch size and hidden size of model are trained at fixed dropout. According to Fig. 1, the accuracy of four and five layers are very unstable and bad. As a result, two and three layers of the model are chosen to compare their accuracy under different parameters setting. According to Fig. 2, both two and three layers of model have the better accuracy at batch size 256. However, according to Fig. 3, two layers of hidden size 1024 at batch size 256 has the best accuracy. Also, one can find that the issue of overfitting can be neglected from Fig. 4. The model get the public score 0.92800 and private score 0.92400 on Kaggle.

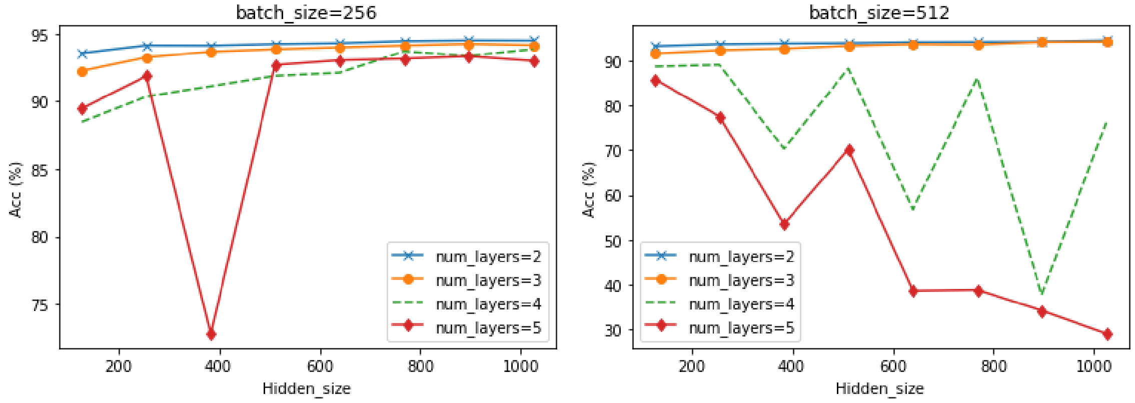


Figure 1: Plot of accuracy versus hidden size for different number of layers.

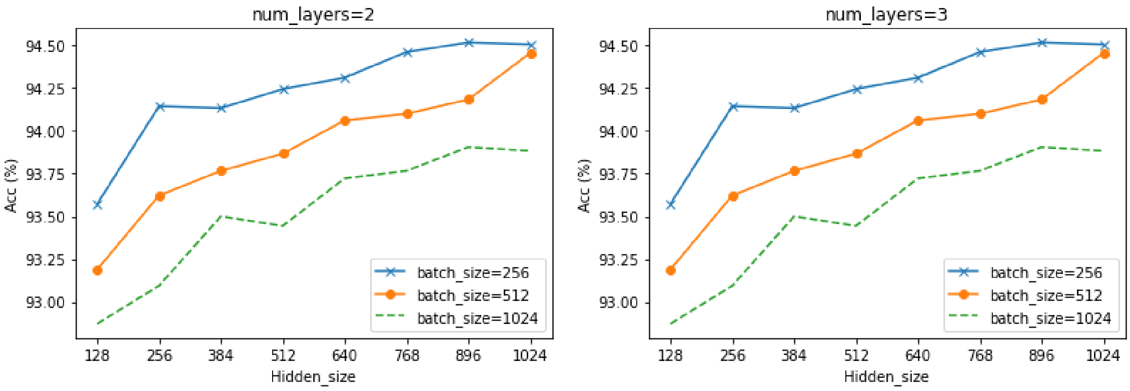


Figure 2: Plot of accuracy versus hidden size for two and three layers for different parameters.

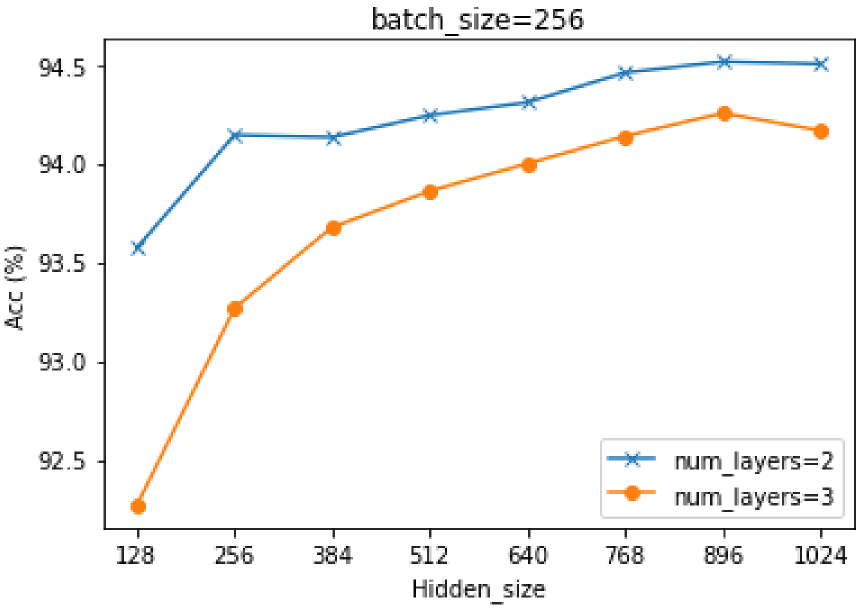


Figure 3: Plot of accuracy of two and three layers at batch size of 256.

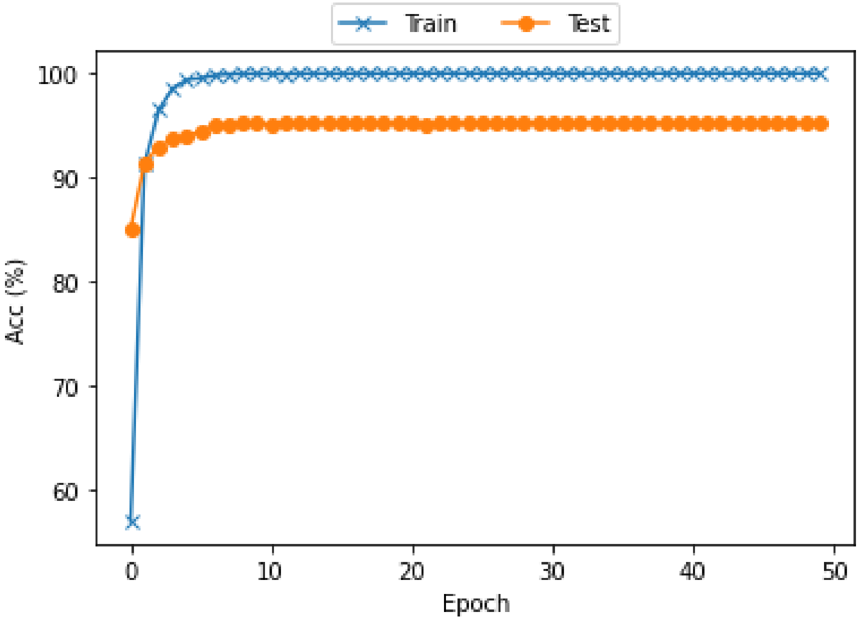
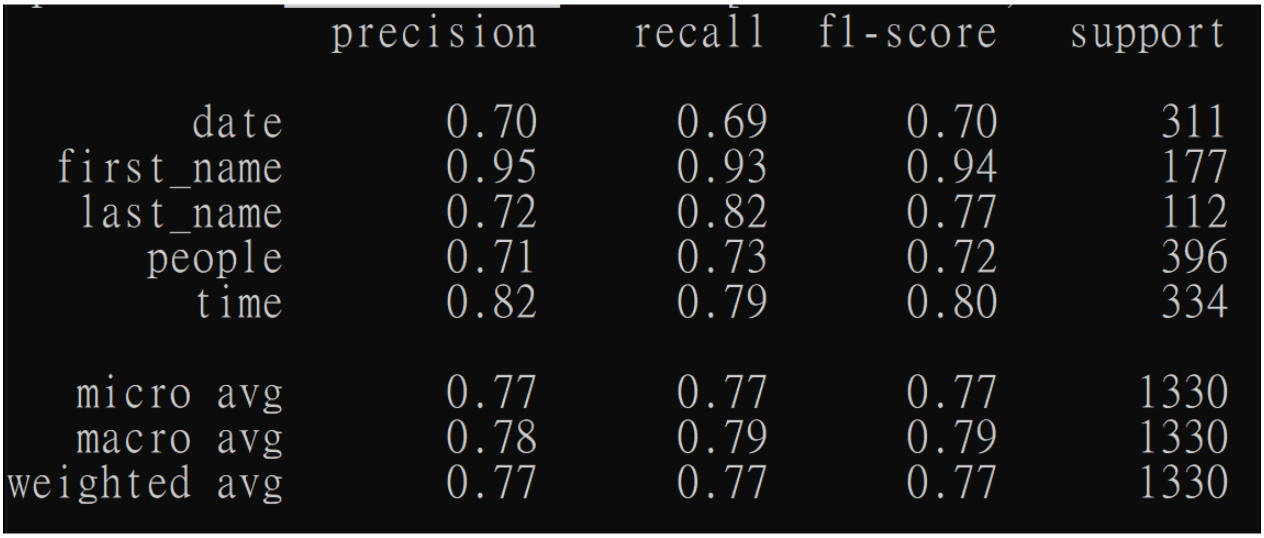


Figure 4: Learning curve of best model.

4.2 Slot tagging

According to Fig. 6 (a)-(c), one can find out that batch size of 512 is the best choice for training. And, according to Fig. 6 (d), two layers of hidden size equals to 384 is the best. Also, one can find that the issue of overfitting can be neglected from Fig. 7. According to the classification report using seqeval, the classsification of time-realted label is the hardest task. The average accuracy in the report is similar to the joint accuracy calculate in the code. The best model from the training get the public score 0.75495 and private score 0.75723 on Kaggle.

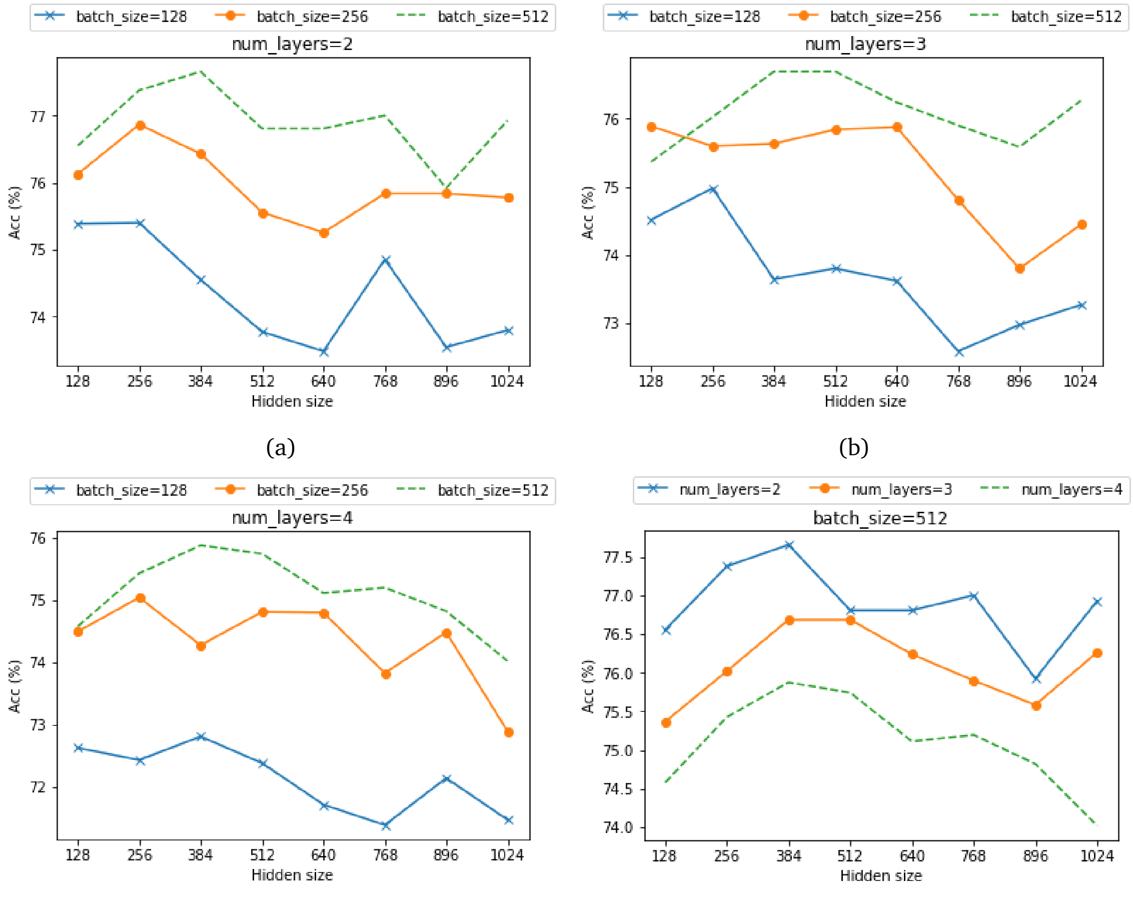


```
              precision    recall  f1-score   support

        date       0.70      0.69      0.70       311
  first_name       0.95      0.93      0.94       177
   last_name       0.72      0.82      0.77       112
      people       0.71      0.73      0.72       396
        time       0.82      0.79      0.80       334

   micro avg       0.77      0.77      0.77      1330
   macro avg       0.78      0.79      0.79      1330
weighted avg       0.77      0.77      0.77      1330
```

Figure 5: Classification$_reportusingseqevallibrary.$



(a)

(b)

(c)

(d)

Figure 6: Plot of accuracy for different parameters setting.

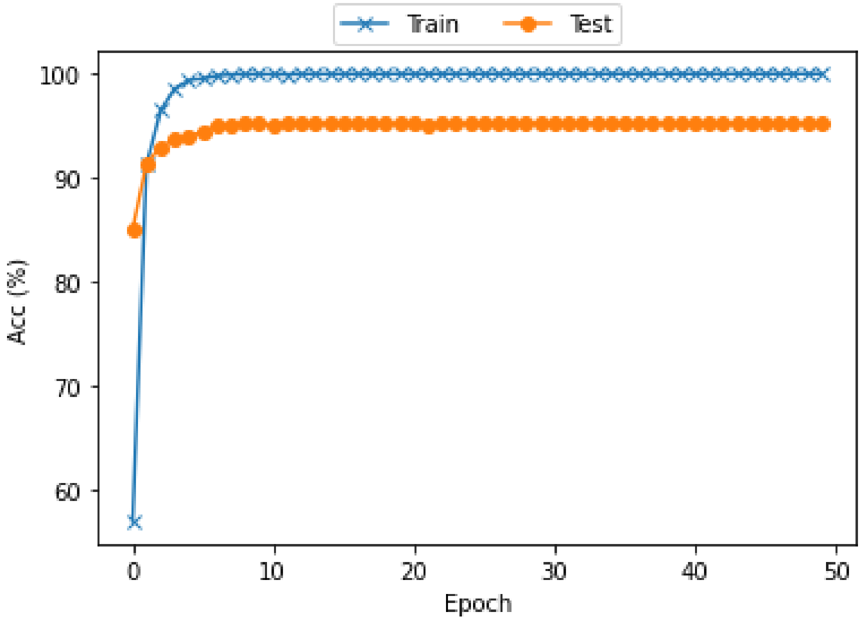


Figure 7: Learning curve of best model.