

Environment:

Tensorflow2.2

Python:3.7

gensim

jieba

scipy

numpy

scikit-learn

Dataset:

There are 500 pieces of news information in cnews.train.text file.

How to run it:

1. Firstly, run python train_word2vec.py.
2. Then, run python text_train.py to train model using training data.

```
Training and evaluating...
Epoch: 1
Epoch: 2
step: 100,train loss: 0.335, train accuracy: 0.875, val loss: 0.076, val accuracy: 0.963,training speed: 0.084sec/batch *
Epoch: 3
step: 200,train loss: 0.164, train accuracy: 0.984, val loss: 0.023, val accuracy: 1.000,training speed: 0.164sec/batch *
Epoch: 4
step: 300,train loss: 0.171, train accuracy: 0.953, val loss: 0.009, val accuracy: 1.000,training speed: 0.266sec/batch
Epoch: 5
Epoch: 6
step: 400,train loss: 0.085, train accuracy: 0.984, val loss: 0.007, val accuracy: 1.000,training speed: 0.022sec/batch
Epoch: 7
step: 500,train loss: 0.052, train accuracy: 1.000, val loss: 0.005, val accuracy: 1.000,training speed: 0.106sec/batch
Epoch: 8
step: 600,train loss: 0.027, train accuracy: 1.000, val loss: 0.003, val accuracy: 1.000,training speed: 0.198sec/batch
Epoch: 9
step: 700,train loss: 0.034, train accuracy: 1.000, val loss: 0.003, val accuracy: 1.000,training speed: 0.268sec/batch
Epoch: 10
```

3. Finally, run python text_test.py to test the test data.

```
Testing...
Test Loss: 0.085, Test Acc: 98.13%
Precision, Recall and F1-Score...
E:\Applications\WPy64-3741\python-3.7.4.amd64\lib\site-packages\sklearn\metrics\_classification.py:1272: UndefinedMetricWarning: Recall and F-score are i
defined and being set to 0.0 in labels with no true samples. Use 'zero_division' parameter to control this behavior.
  _warn_prf(average, modifier, msg_start, len(result))
      precision    recall  f1-score   support

      0         1.00      0.96      0.98         50
      1         1.00      1.00      1.00         57
      4         0.00      0.00      0.00          0
      9         0.00      0.00      0.00          0

   accuracy          0.98         107
  macro avg          0.50      0.49      0.49         107
 weighted avg          1.00      0.98      0.99         107

Confusion Matrix...
[[48  0  1  1]
 [ 0 57  0  0]
 [ 0  0  0  0]
 [ 0  0  0  0]]
Time usage:2.089 seconds...
```