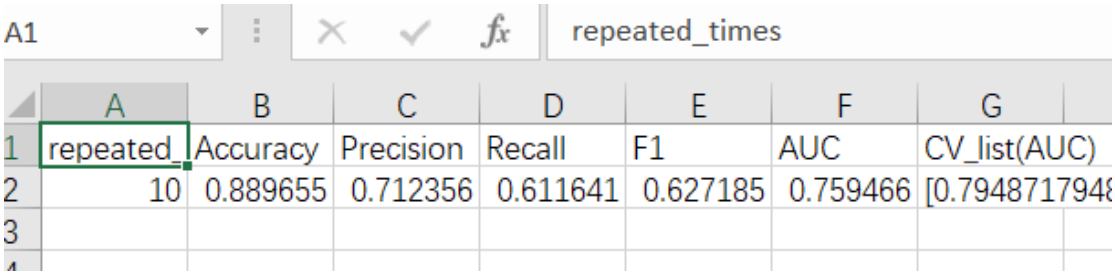
For detailed information about what each file represents and how to use it, please see the manul.pdf file.

You can follow these steps in this PDF to replicate the experimental results.

How to replicate the results reported:

1. The files that already exist in the results directory are the results of my previous experiment.They are also the data for Table 2 and Table 3 in the report. You can check the files with the word average in the results directory.



	A	B	C	D	E	F	G
1	repeated_	Accuracy	Precision	Recall	F1	AUC	CV_list(AUC)
2	10	0.889655	0.712356	0.611641	0.627185	0.759466	[0.7948717948
3							

This represents the average performance of ten ecperiments

2. Please modify the fifth line of the statistical_tests.py file and set the value of the project variable to one of pytorch, tensorflow, incubator-mxnet, keras, or caffe. Then run the script to obtain the results in Table 4 in the report.

3. You can also re-run the baseline.py and coursework.py files to get the new experimental results (but please note that the detailed information of the previous ten experiments will be overwritten. You can save the detailed results of the previous experiments in another directory on your computer).

	A	B	C	D	E
1	Accuracy	Precision	Recall	F1	AUC
2	0.931034	0.964286	0.666667	0.731481	0.794872
3	0.948276	0.973214	0.7	0.771953	0.815094
4	0.87931	0.637821	0.662264	0.648485	0.732075
5	0.87931	0.463636	0.472222	0.46789	0.490741
6	0.844828	0.569444	0.542017	0.548052	0.795518
7	0.896552	0.705357	0.573718	0.597222	0.778846
8	0.862069	0.928571	0.6	0.628205	0.775
9	0.913793	0.730769	0.771698	0.748918	0.890566
10	0.844828	0.4375	0.480392	0.457944	0.736695
11	0.896552	0.712963	0.647436	0.671698	0.785256

This represents the detailed information about the performance of each experiment

4. Although we set a lot of random seeds to make the experimental results as stable as possible, due to the multi-threaded processing used in the Word2Vec training process, the results of each experiment will have very small changes each time we run the script. However, using the statistical tests.py file to test, we find that the results are the same, and each time they are consistent with the data in Table 4 of the report. Our method can significantly improve the performance on all five datasets (except for the recall indicator of specific projects). The experimental results are reproducible.