

Hardware Environment

- Windows 10
- 1 GPU, Nvidia Geforce RTX 2080 Ti

Dependency Requirement

- Python 3.6
- Tensorflow 1.14.0
- sklearn, numpy, scipy

Dataset Construction Steps

- Download “clause_keywords.csv” and “w2c_200.txt” from the original ECPE dataset <https://github.com/NUSTM/ECPE>
- Put the above two files in the folder named “nega_data”
- Run the “preprocess.py” to get the manually labeled dataset, which will be stored in a file called “data.txt”
- Run the “gen_nega_samples.py” to generate the constructed dataset, which is stored in a file called “data_wneg.txt”
- If you prefer our training/testing split, please run “divide_fold.py” to get 20 files, which will be named as “foldx_train.txt” and “foldx_test.txt”, where “x” should be from 1 to 10.

Average Runtime Reminder:

- 2 hours per program for models containing LSTM and Self-Attention Module.
- 1.5 hours per program for models with Concatenation Module.

Run a program:

- Make sure you complete the dataset construction first
- Directly run “python programname.py”, where the “programname” is the python file you want to run.