

Contribution Document

1. The code we used is based on <https://github.com/huggingface/pytorch-openai-transformer-lm>.
2. The link to our repository is: https://github.com/wangxizhe2017/nlp_project.
3. We coded to merge the training datasets together, and merge the validation datasets together, then removed the duplicated stories among them, the code is located at https://github.com/wangxizhe2017/nlp_project/tree/master/data_cleaning.
4. We coded to apply pos tagging to both our training set and testing set, so we can extract the main components out and train on the simplified data. Also we can generate the antonym sentences based on the processed data. The code is located at: https://github.com/wangxizhe2017/nlp_project/tree/master/data_preprocessing
5. We coded to generate the antonym sentences based on WordNet, and use them as the second ending option in our training set, the code is located at: https://github.com/wangxizhe2017/nlp_project/tree/master/data_cleaning.
6. We put all those processed files in the 'data' folder. The files are located at https://github.com/wangxizhe2017/nlp_project/tree/master/data
7. We fine-tuned our model based on a different training set and test set, modified the batch size and running epoch.