Overview

Paper: <https://vtechworks.lib.vt.edu/bitstream/handle/10919/107136/Wang_P_T_2021.pdf?sequence=1&isAllowed=y>

This chatbot project had 3 stages(versions) in total.

1st stage is a GPT2-based chatbot (Deprecated).

2nd stage is to upgrade to use T5 model (one victim model + 4 perpetrator models for stages)

3rd stage is to add RL to 2nd stage. (one victim model + one perpetrator model + A classifier model for RL)

Training idea, pre-train on Convai2 dataset then fine-tune on PJ dataset

Google drive folders:

T5

<https://drive.google.com/drive/folders/1mcuHPN0XaXOVo9J7RCDR_S-pBs6kjXmy?usp=sharing>

BERT\_classifier

<https://drive.google.com/drive/folders/1YfuI307DFEeKHiqwUeRWWUBtFduPcULC?usp=sharing>

Data processing

Code and data are in BERT\_classifier

Raw PJ dataset is in BERT\_classifier/PJ-Dataset-html

**Step0. Html to csv.**

BERT\_classifier/PJ-Dataset-html/test/html2csv.ipynb

Step1. **First round clean**. Use CleanCSV.ipynb to do simple clean job + combine consecutive utterances by the same person into one. Save output in BERT\_classifier/PJdataset\_clean

Step2. **Second round clean using MoNoise.**

* Use CSV2TXT.ipynb to convert CSV files into txt file (saved inBERT\_classifier/PJdataset\_clean/TXT **)**
* Pass txt files into MoNoise for normalizing slangs (can only be used in Linux with command) and output to BERT\_classifier/PJdataset\_clean/pj\_out
* Then use TXT2CSV.ipynb to convert the output of MoNoise back to csv files, saved in BERT\_classifier/PJdataset\_clean/PJ\_clean\_final100/

Step3. **Get final training set(T5model with RL)**

For Convai2 data set BERT\_classifier/ConvAI2/ToTrain.ipynb

For PJ data setGetTrainPJ.ipynb

All training data saved inT5/data

Step extra  **Stage Classifer**

* Training data : chatLogs.tsv
* Albert\_classifier.ipynb

Step\*. **Get training set for 2nd version T5 model (by stages)**

* Get\_trainPJ\_mid.ipynb is to generate a mid file with a sentence and its former and after sentence with connection score, saved in PJ100\_mid
* Segment2train.ipynb take the mid file and generate training data for T5 model for each perpetrate stage

Training and testing

train&test.ipynb contains training and testing.

Training part contains all model, please refer to comments.

Testing part is to observe output of selected model, usually do it after selecting best epoch model..

Interact.ipynb is to select best epoch of each model, and has the **final demo of chatbot**.

Evaluation

Part1 BERT\_classifier/EVA\_set

Multi\_EVA.ipynb evaluate BLEU, ROUGE-L, BERTScore

Perplexity calculation.ipynb evaluate perplexity score

Part2 BERT\_classifier/EVA

EVA.ipynb is to evaluate online\_dialog\_eval

Part3 Human evaluation

All data and code inside BERT\_classifier/EVA\_set/humanEVA