Pragmatic Sentiment & Dialogue in Natural Language Processing

USC Viterbi School of Engineering

SHINE Summer High School Intensive in Next-Generation Engineering

Anushka Rane | anushkarane2018@gmail.com
Diamond Bar High School 2024
USC Viterbi Department of Computer Science | SHINE 2023

Overview

Large Language Models used for tasks involving Natural Language Processing have been limited by their lack of a living human experience– impacting their ability to understand “commonsense” or the underlying meanings and contexts in various situations.

The main research questions I explored are:

1. How can we use pragmatic sentiment and dialogue to challenge Large Language Models (LLMs) like BERT and GPT-3?
2. What makes a pragmatic sentiment analysis dataset most effective?

Objective & Impact

Professor Ren’s research at INK Lab focuses on advancing commonsense reasoning in Artificial Intelligence through fine tuning models utilizing datasets and algorithms with specified NLP tasks and situations. Professor Ren is also passionate about making NLP models more accessible and increasing understanding and transparency of the inner-workings of Artificial Intelligence. My mentor Hirona Arai’s PhD work centers around computational pragmatics with conversational questions and their intended meanings.

Methodology

When comparing pragmatic-tasked datasets to determine what makes a data set most effective for training NLP models, I trained BERT to 4 different datasets and compared the results.

In the BERT for Sentiment Analysis Tutorial [1], we modified the training loop code to give more data, made graphs using Matplotlib to visualize training data, and tested out 6 different hyperparameters (learning rate and epsilon value) to understand their impact on training results. Using the Hugging Face and Kaggle libraries, I found three other pragmatic sentiment analysis/detection datasets to fine tune BERT with.

1. Airplane Tweet Complaint Detection [1]
2. SMS Spam Detection
3. IMDB +/- Review Classification
4. Spam Detection

Once finalizing the three datasets, I cleaned split them where necessary by removing proper nouns, removing specific words, including certain special characters, and tokenizing and including emojis (for D3, expanded BERT Tokens). After training BERT with the cleaned datasets, I concatenated the datasets’ accuracies onto a graph (fig. 1). Lastly, I determined patterns in the datasets after comparing them to understand what makes a dataset most effective in training a model to perform pragmatic classification. (fig. 2)

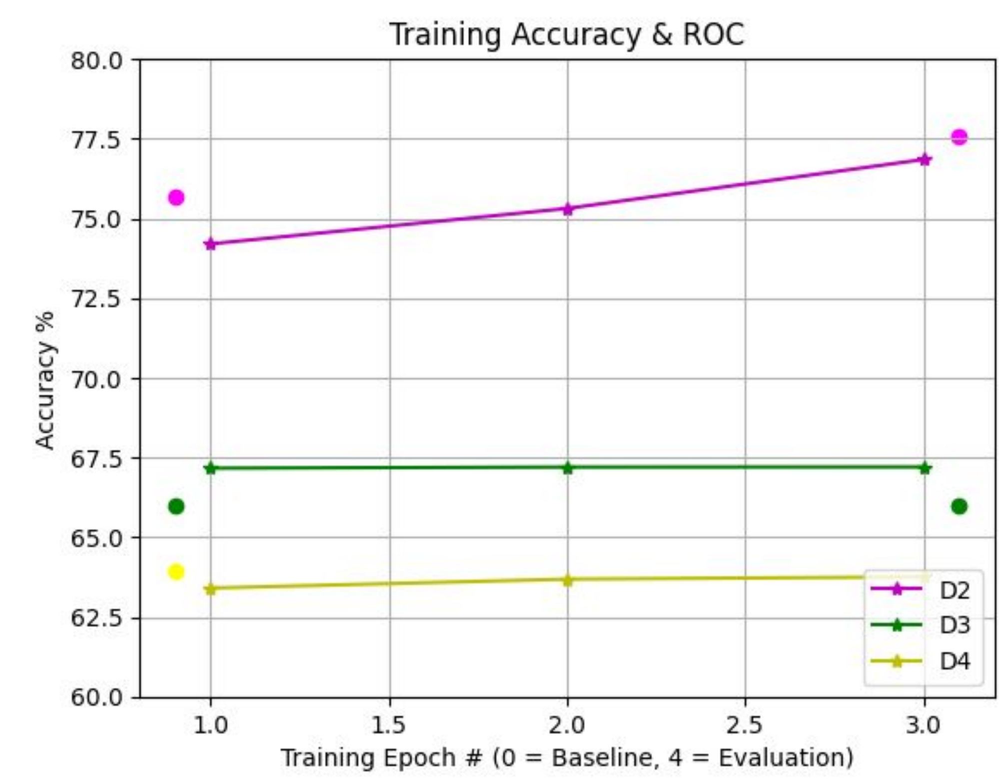
When questioning how to use pragmatic sentiment and dialogue to challenge LLMs, I noticed patterns in the outputs of GPT-3 to various pragmatic prompts.

We first started with annotating 200 examples of questions said in various tones, and whether their intended meanings (provided by ChatGPT) were reasonable. We then provided annotations for intent {0: seeking information, 1:requesting something, 2: offering something, 3:seeking some other/further information, 4: rhetorical commentary, 5:other}, power dynamic {0: linear, 1: non-linear}, better intended meanings, and other possible situations. Then, I provided ChatGPT 150+ prompts of different difficulty levels that tasked it with:

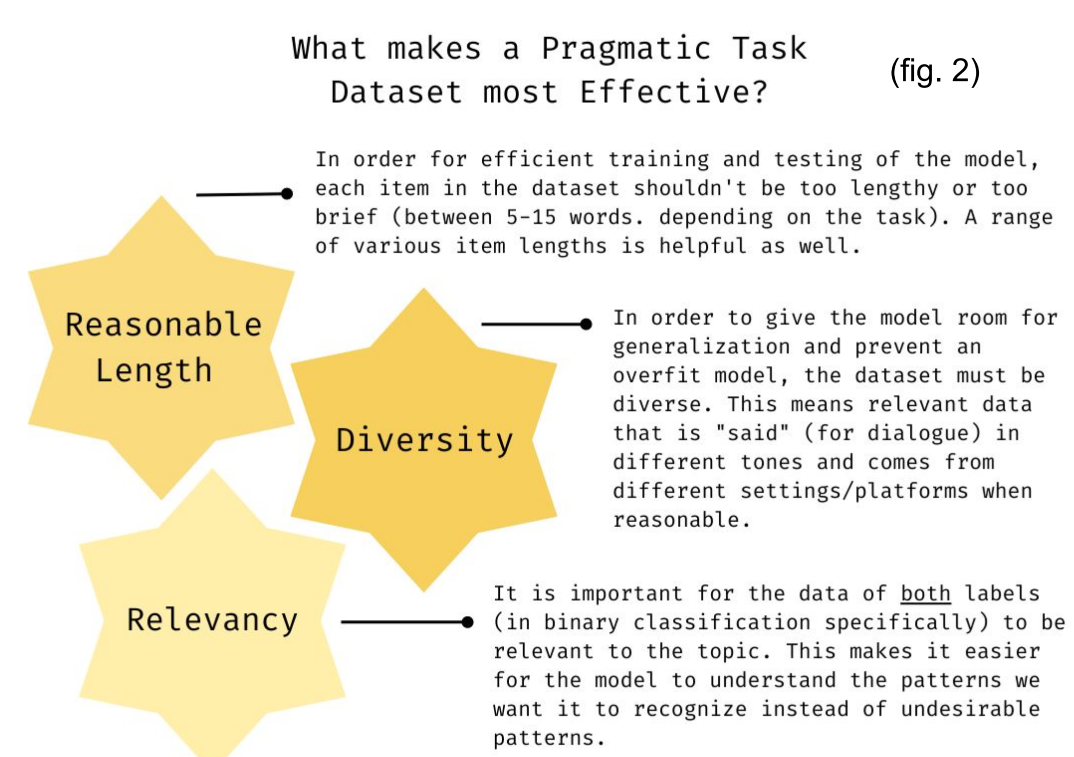
- Providing (3) contexts (including roles) for questions given tone
- Providing tone explanation and (3) possible responses
- Including power dynamics in context outputs
- Forming a coherent discourse with multiple contexts
- Further questions about conversationality of outputs

Findings

(fig. 1)



(fig. 2)



Skills Learned

Training models like BERT in Google Colab

- Using Hugging Face’s Transformers Library to fine tune BERT Classifier
- Storing training data in graphs and externally using pickle

Deepened python knowledge and introduced to new applications of it

Dataset searching, cleaning, splitting, and preprocessing

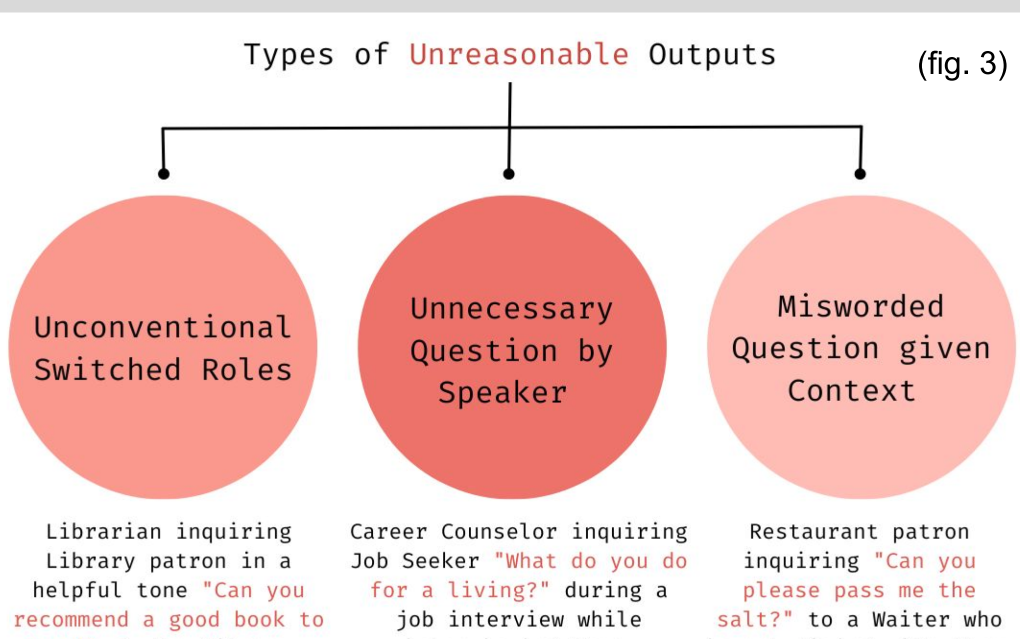
Understanding Binary Classification to read AUC and ROC graphs

Using Pandas in dataset & data analysis code

Converting emojis to text for BERT to use when training

Further Findings Conclusion

(fig. 3)



Given the manner and question, OpenAI’s ChatGPT forms a context in which the questions would be asked but occasionally doesn’t take into consideration but the speaker and listener already know or can observe. ChatGPT also struggles with writing language which would be acceptable in normal conversation by including overly-formal dialogue and “blunt” reference to contextual situation. ChatGPT’s reasonable outputs are often common scenarios that many have experienced or seen depicted, contrasting to its occasionally unreasonable made-up contexts. When giving an identification task, certain text information such as tone labels, setting, and speaker/listener roles proved to be helpful for the model. We can further challenge LLMs like GPT-3 with pragmatic reasoning by introducing power dynamics, sarcasm/humor, and euphemistic language into prompting and outputs.

Advice to Future SHINE Students + Relation to Coursework

The experiences and skills I have acquired from SHINE have not only deepened my interest in Computer Science, but also give me the confidence to be successful in my CS class next school year, college classes, and any other future projects I work on regarding NLP and Data Analysis. My advice to future SHINE students would be to strive to make the most of SHINE in all aspects. Participate in cohort events, make connections with peers, ask questions, and put your best effort into your research.

Acknowledgements

I would like to thank Professor Ren for giving me the opportunity to work in his lab, and my mentor Hirona Arai for all the support and guidance throughout the program. I am also grateful to the SHINE Faculty Ms. Lopez and Ms. Shim, my center mentor, for making this summer a great experience and learning opportunity.

Citations

[1] Qi, P. (2023, July 1). PragmaticQA: a dataset for pragmatic question answering in conversations. ACL Anthology.

[2] Elias, G., & Tran, C. (2023). Tutorial: Fine-tuning BERT for Sentiment Analysis. Skim AI.