Leveraging Decoder-Only Transformer Models for Dialogue Act and Emotion Tracking in Dialogue Systems

Ankur Agarwal

31 July 2023

Abstract

This work presents a novel dialogue system based on decoder-only transformer models. The system generates responses by effectively tracking dialogue acts and emotions from the conversation's history.

1 Introduction

The creation of proficient dialogue systems necessitates more than linguistic comprehension. It requires an understanding of dialogue acts and emotions, forming the intricate nuances of human conversations. Although Natural Language Processing (NLP) has made significant progress, the finer details of conversational dynamics still pose a challenge. In this work, a unique approach is introduced, employing decoder-only transformer models to comprehend and respond to the patterns in dialogue acts and emotions throughout a conversation. The system utilizes the DailyDialog dataset, a comprehensive resource covering a broad range of daily communication scenarios and emotional states.

2 Related Work

Historically, the quest for an ideal dialogue system has revolved around capturing conversational dynamics accurately. However, many extant models fall short in generating responses that resonate with the emotional and pragmatic context of the dialogue. Transformer models, despite their significant contributions to context-based tasks in NLP, occasionally struggle with the subtleties of dialogue acts and emotions. Decoder-only models, like GPT and OPT, have displayed immense potential in generating human-like text, but their capabilities

in consistently tracking dialogue acts and emotions could be further enhanced.

3 Approach

The presented system builds upon these insights, using a decoder-only transformer model like ChatGPT using the GPT model. The proposed system processes the concatenated dialogue history to generate a hidden state representation. This representation is then leveraged to generate a response, predict the dialogue act and emotion for each utterance in the dialogue history until the prediction of the upcoming one. This dual objective of generating a response and predicting dialogue history states influences the model, enhancing its comprehension and response-generation capabilities. The prediction of dialogue history also impacts the response generation in a significant way, ensuring the responses align more closely with the ongoing conversation's dynamics.

4 Experiment Setup

The DailyDialog dataset, comprising 11,318 transcribed dialogues, serves as the basis for the experiments. The dataset includes both dialogue act annotations and emotion annotations, offering a comprehensive range of communication scenarios for model training and testing. The dialogue act annotations encompass four categories, namely inform, question, directive, and commissive. The emotion annotations detail seven emotional states, including no emotion, anger, disgust, fear, happiness, sadness, and surprise.

For setting the conversational scheme, the usual concatenation-of-history approach is adopted with a "Response:" prompt at the end of the input. To have a healthy balance of learning responses for small and large history sizes, every sample dialogue is reused iteratively increasing its number of utterances in the conversation history and distinct samples of dialogues.

The prediction of the dialogue act and emotion is carried out using two independent multi-layer LSTM models which use a classifier for prediction from the last layer's hidden states. The usage of the two features is more beneficial at the output as compared to the input because it allows the decoder language model to generalize as opposed to passing them as inputs where the decoder model becomes dependent on its conditionality. The dialogue act/emotion sequence

prediction objective is used only for training purposes. It is expected the model is capable of predicting it if there needs to be a test case.

The drop in the loss first affects the LM for response generation until saturation. After the generation features stabilize, there is a conflict for generalization between the dialogue emotion/act features and the language generation features. The dual object reduces towards an equilibrium where all the losses are low and generalization improves. Due to constraints of time, the plots are not presented in the report.

5 Conclusion

This work proposes an innovative approach that leverages decoder-only transformer models to produce contextually relevant and emotionally resonant responses in dialogue systems. Preliminary results demonstrate an improved ability in the system's tracking of dialogue acts and emotions compared to its predecessors. This progress signifies a promising direction for creating more nuanced and human-like dialogue systems and illuminates the potential of blending advanced language understanding techniques with socio-linguistic considerations.