Computer Science
Defense

Contextual Embedding Representation for Dialogue Systems

Zihao Wang
Emory University

Abstract: Context is a crucial element for conversational agents to conduct natural and engaging conversations with human users. By being aware of the context, a conversational agent can capture, understand, and utilize relevant information, such as named entity mentions, topics of interest, user intents, and emotional semantics. However, incorporating contextual information into dialogue systems is a challenging task due to the various forms it can take, the need to decide which information is most relevant, and how to organize and integrate it. ¡br¿¡br¿ To address these challenges, this thesis proposes exploring and experimenting with different contextual information in the embedding space across different models and tasks. Furthermore, the thesis develops models that overcome the limitations of state-of-the-art language models in terms of the maximum number of tokens they can encode and their incapacity to fuse arbitrary forms of contextual information. Additionally, diarization methods are explored to resolve speaker ID errors in the transcriptions, which is crucial for training dialogue data. ¡br¿¡br¿
The proposed models address the challenges of context integration into retrieval-based and generation-based dialogue systems. In retrieval-based systems, a response is selected and returned by ranking all responses from different components. A contextualized conversational ranking model is proposed and evaluated on the MSDialog benchmark conversational corpus, where three types of contextual information are leveraged and incorporated into the ranking model: previous conversation utterances from both speakers, semantically similar response candidates, and domain information associated with each candidate response. The performance of the contextual response ranking model exceeded state-of-the-art models in previous research, showing the potential to incorporate various forms of context into modeling. ¡br¿¡br¿
In generation-based systems, a generative model generates a response to be returned to the conversing party. A generative model is built on top of the Blenderbot model, overcoming its limitations to integrate two types of contextual information: previous conversation utterances from both conversing parties and heuristically identified stacked questions that tackle repetition and provide topical diversity in dialogue generations. The models are trained on an interview dataset and evaluated on an annotated test set by professional interviewers and students in real conversations. The average satisfaction score from professional interviewers and students is 3.5 out of 5, showing promising future applications. ¡br¿¡br¿
Additionally, to better understand topics of interest, topical clustering and diversity are investigated by grouping topics and analyzing the topic flow in the interview conversations. Frequent occurrences of some clusters of topics give a clear presentation of what scopes of topics an interview would touch on while maintaining a great selection of unique topics for individuals. Based on this observation, another generative model architecture integrating topical information is proposed that generates the next topic of interest in the conversation flow in parallel to generating utterances. This work is ongoing, with the expectation of improving the performance of the previous generative model.
¡br¿¡br¿ Day/time: March 30th, 11:00 am - 12:30 pm
Room: Math CS E306 ¡br¿¡br¿ Zoom Option: https://us02web.zoom.us/j/9910064905?pwd=aThaYVd1eFBkRWp
Meeting ID: 991 006 4905
Passcode: 290751

Thursday, March 30, 2023, 11:00 am
Mathematics and Science Center: MSC E306

Computer Science
Emory University