21/07/2022

Intel-PA Individual Research Direction Proposal

Abdul Rehman

The NLP comprehension core of the IntelPA has four components that relate to their own research fields.

1. **Conversation tree entailment:** Inferencing or contrasting the statements to the sentences on the pre-designed conversation tree (dialogue state tracker).
2. **General Chitchat:** Context-based text generation for the response to any conversation that is not part of the tree. Fine-tuning the pre-trained models like GPT2 and OPT is the current approach.
3. **Entity extraction:** Extracting time duration, names, locations, Boolean or numerical variables, etc as standardized variables from sentences. This feature will support the decision-making in the conversation tree and will provide the summarized information to the administrator.
4. **Paralingual recognition:** Integration of the facial and speech emotions and personality with the conversation tree will help the bot to make more nuanced judgments.

The challenge: Most goal-oriented conversation bots are either trigger-driven or use reinforcement learning. Whereas the general chitchat bots can talk about anything in any context without any intended goal. Even the biggest general NLP models often generate incomprehensible out-of-context replies in a normal conversation. Mixing the two types of bots is the main task here.

Proposed approach: My initial thoughts are to design a general-chitchat bot with the goal-entailment as the loss index. In other words, the text can be generated using the chitchat model when the conversation is distant from the current dialogue node. However, if that distance (loss index) keeps increasing with each reply then a fallback response will intercept the conversation to bring it back to the current dialogue state on the tree.
Another more sophisticated way to bring back the conversation to the dialogue node will be the convergence of the loss index based on the entailment scores of multiple model-generated sentences. The chitchat model can generate 10 possible replies for a question, then our dialogue-tree model will pick the reply that has the highest entailment score (minimize the distance). Within a few reply turns, the conversation will hopefully converge near the current node topic. This method, however, doesn't work in real-time with the computing power available to us.

The core NLP problems are more of engineering challenges instead of research challenges. There is a possible research space here for goal-oriented chatbots, but all the current techniques require training data. Therefore, that leaves us with the only option of engineering the conversation tree until we have collected enough samples to try creating new machine learning models.

My PhD research proposal is focused on paralingual speech recognition because that is an area that is less explored and has many interesting challenges to solve. I will be working on this part in the background (70%) until it's ready for integration with the Intel-PA project when the IntelPA streaming server is functional.