Univ.AI

TA BOT

Anah Veronica, Harshitha Kolukuluru, Janhavi Kashyap, Shivani GB

Motivation and About the Project

Conversational AI systems are considered to be the next generation of chatbots with reading comprehension capabilities that enhance user experience. The problem was to build such a system and apply it to the closed-domain setting of answering Univ.AI's frequently asked questions.

In our quest to build a good chatbot, we decided to try using different datasets for training our models and as many models as we could within a week for better understanding. We used two different datasets - the Stanford Question Answering Dataset or SQuAD2.0 and the Cornell Movie-Dialogs corpus as they differ in structure and the kind of data they contain.

Data and Labels

SQUaD - Question, Answer_Text
Cornell Movie Dialogues Corpus

Context - An article consisting of five paragraphs - Grading, Deadlines, Exercises/Homework/Projects, Course material and General - that answered the questions in the FAQ train set was written and annotated to contain the start index of each answer in the paragraph.

References

https://blog.keras.io/a-ten-minute-introduction-to-sequence-to-sequence-learning-in-keras.html

Model

1. **Base Model - SQUaD Data:** A simple RNN network with 2 RNN layers was created as a baseline model. This was used as the base to evaluate the performance of other models.
2. **Seq2Seq Model - SQUaD Data:** The sentences were converted into three numpy arrays: encoder_input_data, decoder_input_data, decoder_target_data. A basic LSTM-based model was created to make predictions
3. **BERT Model - FAQ (Modified) Data:** The goal is to find the span of text in the paragraph that answers the question. The context and the questions are fed as inputs to predict start_token and end_token
4. **Seq2Seq Model - Cornell Data:** We used a simple model with an embedding layer with pre-trained GloVe embeddings and two RNN layers as the baseline model.
5. **Transformer Model - Cornell Data :** Transformers are based on the idea of self-attention. Our model uses a multi-headed attention layer with scaled dot-product attention. Hyper-parameter tuning, mainly increasing the dropout rate while training, gave us improved results.
6. **GPT-2 Model - Cornell Data:** The GPT-2 is built with transformer decoder-only blocks and uses auto regression with masked self-attention. We trained GPT-2 on two separate variations of the FAQ data - 1) The original one with questions and answers and 2) FAQ data with custom context that we created for comparison.

Results

1. The base models, despite having better accuracies, had basically no predictive power.
2. Seq2Seq (SQUaD) predictions were influenced by the SQuAD training set despite fine-tuning with the FAQ data and didn't make a lot of sense.
3. BERT model was able to correctly answer only some of the test data questions. However, since the model was fine tuned on only the FAQ dataset, the model did not have enough data points to make accurate predictions.
4. Seq2Seq (Cornell Movie)
5. Transformer Model predictions make sense but very little (?)
6. OpenAI's GPT-2 is a great text generator but wasn't trained enough to work

Conclusion and Future Work

- We fine-tuned the GPT-2 model with the FAQ data which was just too small to make any conclusions about how it worked on QA tasks. Despite that, some of the outputs we got actually made a little sense. We'd like to train it on a bigger dataset with augmented data for longer.
- Text generation and question-answering are really interesting areas and there are a lot of ways we could've made our models better with techniques like ask-answer, distillation, etc.