A Chatbot Developed by Deep Learning

with Variant of Seq2Seq Models

A Project Proposal for MSCI 641

Shuo Tian, 20800262
Tongdan Su, 20754736

May 27, 2019

1. Back information

1.1 Background

A chatbot is a program designed to simulate how a human would behave as a conversational partner. A simple chatbot may just scan for keywords within the input, then pull a reply with the most matching keywords, or the most similar wording pattern, from a database. While within deep learning, a chatbot could be much smarter to understand and reply to human's input in a real-time conversation.

In general, there are 2 different methods with deep learning to construct a chatbot: Retrieval-Based Models and Generative Models. Retrieval-Based Models are defined as models with a well-defined repository of responses. By simply matching the keywords or classifying within machine learning methods, the models would be able to choose the best answer from the repository, because these well-defined responses are usually manually inputted by a real human, they are more possible to a be both syntax-free and semantics-free. However, due to the limited repository capacity, Retrieval-Based Models can only work on a certain topic or situation prepared by the programmers.

By contrast, Generative Models are used to produce a new sentence aimed at real-time input. They are able to deal with a new and complex situation, while larger training data is needed. Besides, they are more easily to make the errors, generating a sentence contains all necessary elements but makes no sense, especially for the long sentences.

1.2 Statement of Design

This project will design and develop a chatbot based on the Generative Models. By collecting billions of comments from Reddit, training the data in a sentence to sentence model from TensorFlow, This

project would be able to reply the input by generating a new comment close to the natural human-speaking way in a real-time conversation.

2. Objective

1) Flexibility

To help the chatbot work in a more flexible way, as being able to talk more topics with users, this project is designed to build based on a Generative Model, which would produce the new sentences in the real-time conversation.

2) Smartness

To make the chatbot output more close to a human speaker, the chatbot model would be trained within deep learning, based on a sequence to sequence model.

3. Planed Approaches:

Chatbots that use deep learning are almost all using some variant of a sequence to sequence (Seq2Seq) model. This project will also apply to this model. Since 2014 the paper “Sequence to Sequence Learning with Neural Networks” has been published, the Seq2Seq model has seen numerous improvements. This project will also compare different models (e.g. Sequence to Sequence with Attention, Attention with Intention, and Copying Mechanisms et. al.) to achieve better performance**.**

4. Description of Datasets:

For sequence to sequence models, we need a large number of conversation logs. From a high level, this encoder-decoder network needs to be able to understand the type of responses (decoder outputs) that are expected for every query (encoder inputs). Some common

datasets are the Cornell Movie Dialog Corpus, the Ubuntu corpus, and Microsoft's Social Media Conversation Corpus.

While most people train chatbots to answer company specific information or to provide some sort of service, we are more interested in a bit more of a fun application. With this particular post, conversation logs from Reddit were used to train a Seq2Seq model and this dataset is around 10 GB, which will satisfy our expectations. Dataset can be download from the followed website: http://files.pushshift.io/reddit/comments/

5. **Evaluation:**

Although natural language tasks have developed many automatic evaluation metrics, for example, BLEU, METEOR and ROUGE, the results still have their specific weaknesses comparing with human judgments on dialogue system. Therefore, this project will focus on human judgments to evaluate the responses.

References

[1] Chia-Wei Liu, Ryan Lowe, Iulian V. Serban, Michael Noseworthy, Laurent Charlin, Joelle Pineau. " How NOT To Evaluate Your Dialogue System: An Empirical Study of Unsupervised Evaluation Metrics for Dialogue Response Generation ". arXiv:1603.08023.

[2] Jiatao Gu, Zhengdong Lu, Hang Li, Victor O.K. Li. " Incorporating Copying Mechanism in Sequence-to-Sequence Learning". arXiv:1603.06393.

[3] Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio. “Neural Machine Translation by Jointly Learning to Align and Translate". arXiv:1409.0473.

[4] Kaisheng Yao, Geoffrey Zweig, Baolin Peng. “Attention with Intention for a Neural Network Conversation Model”. arXiv:1510.08565.

[5] Jiwei Li, Will Monroe, Alan Ritter, Michel Galley, Jianfeng Gao, Dan Jurafsky. " Deep Reinforcement Learning for Dialogue Generation".arXiv:1606.0154.