General Conversation AI

(Mandate-1)

IMT2019019 - Chandrahas Sai Kotagiri
IMT2019045 - Kopparapu Sai Krishna

Problem:

General Conversation AI (not Domain-specific) which converses by predicting the next sentence depending on the previous sentence/Conversation. We would like to expand this to the COVID-19 guide and fact-checker which provides queried info regarding Covid (Does Not include diagnosis).

Introduction:

There are many ChatBots(Domain Specific) on the internet which makes the term chatbot sound like a cliche in Natural Language processing but this model we present stands out as it is not restricted to any domain(like customer care etc) but a bot which wants to have a casual conversation. We use neural networks to do more than just a mere classification i.e. mapping a complicated structure to another. In this case, these structures are query and response. We use sequence to sequence framework (using recurrent networks) to achieve this as this framework hardly needs any feature engineering and domain specificity.

Challenges:

- For a General conversation Bot, it matters a lot on which data we are training, as the model is expected to answer different combinations of phrases, especially combinations used in person-to-person conversation. To deal with this we've decided to train the model on a dataset extracted from 617 movie scripts which have 220,579 conversational exchanges between 10,292 pairs of movie characters which would suffice for training

- Incase of COVID helper we need to make sure that the data we are providing in return to the query must be highly reliable and authenticated since it is a sensitive issue to deal with and false information might cause unnecessary panic among the users. To deal with this we are using WHO's Frequently Asked Query Database which is a trustworthy source and precise.

Datasets:

We are using two datasets one for General Conversation and the other for Covid Helper

1) **Cornell Movie-Dialogs Corpus:** This dataset contains conversations that are extracted from 617 raw movie scripts. In this dataset, there are two text files movie_conversations.txt and movie_lines.txt which we use for training our model. Link to the dataset - https://www.cs.cornell.edu/~cristian/Cornell_Movie-Dialogs_Corpus.html

2) **WHO FAQs about COVID-19:** This dataset contains frequently asked questions regarding coronavirus and the answers provided by WHO to these questions. You can find the dataset in the below link with the file name WHO_FAQ.xlxs https://github.com/wearetriple/ai-faqbot-who

Evaluation metric:

Basically in a conversation, we are replacing one person with a machine, so there won't be any automatic evaluation metric. The evaluation metric should involve humans. One thing we can do is create a group of people and ask them to create a set of 100 questions and then we take them and generate the output of the chatbot. Now we can ask them to give 1 or 0 points to each output of the chatbot and calculate the efficiency. But this takes a long time. So a better way is to compare the performance of our chatbot with a popular rule-based chatbot (CleverBot) by giving the same set of questions to both the chatbots and asking a person to say which chatbot gave the better answer (he can also give tie if both of them answered equally well). In this process, we record the scores given by the person and compute the relative performance of our chatbot.

Eg: Let's say we have a set of 50 questions and the person preferred our chatbot in 20 questions and CleverBot is preferred in 17 questions and in 13 questions he can't differentiate which is better. This implies our chatbot is slightly better. But, to get good results after following this evaluation metric we need to test on a large number of questions.

Link to CleverBot - https://www.cleverbot.com/
Resource from where we took this evaluation metric-https://arxiv.org/pdf/1506.05869.pdf