Chatbot using NLP and Deep Learning

Ravi Khevaria

[1]*Department of Information Technology, Maharaja Agrasen Institute of Technology, GGSIP University, Delhi, India*

Abstract: Today the use of Machine Learning, Artificial Intelligence, Natural Language Processing, and Deep Learning is on the rise. In today's world using chatbot for customer care services, banking services, credit card companies, and start-ups is common. Using these conversational agents not only eases the work of the companies but also deals with the queries fastly. By using neural network we can train the chatbot to reply as close to as human does. So, by using RNN we can build and train a model accordingly. I have taken the dataset of movie dialogues and used it to build our seq2seq model. By using this dataset and training our model we can build a chatbot that can reply according to the input given by the user. There are many chatbots like code-based or interfaced but they are not good at holding the conversation just like humans. By using the seq2seq model Based on encoder and decoder architecture, we can build a chatbot that can emulate like human beings.

Index Terms- Chatbot, NLP, RNN, LSTM.

I. INTRODUCTION

Chatbots or Virtual Assistants make the interaction between computers and humans simple. A chatbot uses Artificial Intelligence and can interact with the human in a more realistic way in natural language via various chat rooms, websites, mobile apps, and messaging applications. When any response is generated to user queries in a more realistic human-like language, it is generated via Natural Language Processing. Chatbots provides various services one of which is improving customer experience. In Chatbots various crucial role like training, optimizing and configuring the chatbot system is provided by human intervention. There are two different tasks that form the basis of a chatbot: 1. Analysis of User Request 2. Generating and returning the response of the User Request Analysis. The ability of the chatbot to identify the user intent in order to extract relevant entities and generating an adequate response to the User Request is the core function. If any failure occurs to correctly acknowledge the User Request, it will fail to provide the Desired Output. The Response to the User by the Chatbot can be predefined generic text, text obtained from a database having different Outputs, Data stored in various Systems.

II. RESEARCH AND IDEA

The Inputs Given by the User is processed by using Recurring Neural Network (RNN). RNN is specially used to handle the sequences. The internal state of the RNN is responsible for handling the dependency between successive inputs. This model is well suitable for different NLP tasks. RNN along with Tensorflow and Keras can be used for training the final model. For the Chatbot, I have used the seq2seq model which consists of two different components, encoder, and a decoder. The encoder encapsulates the information of the user's input text into a fixed representation. The decoder than take the representation and generate the variable-length response which is best suitable for the input. which can be used for our sequence to sequence models. The network of the encoder and decoder should be able to understand the type of responses that are predicted for every query. We will use various datasets like the Ubuntu corpus, Microsoft's Social Media Conversation Corpus and the Cornell Movie Dialog Corpus. Here we will use the Cornell Corpus Dataset.

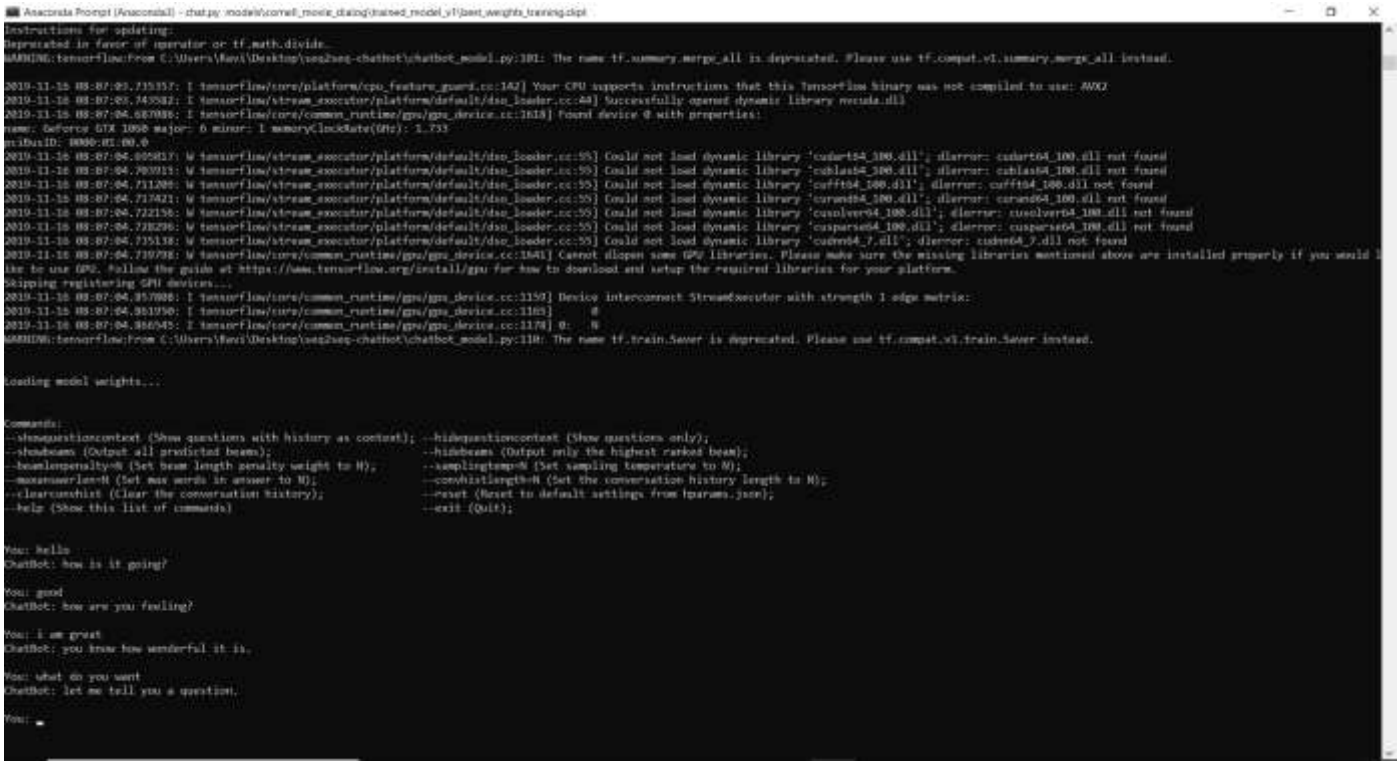
III. DATA PREPROCESSING

The Conversation data in the movie corpus contained Movie ID, Character ID, and Movie Line ID was separated by "+++++".

For Preprocessing, the data was cleaned by removing all the metadata. Now the separators were eliminated. All the data contained in an unsupported encoding format by UTF-8 standard and was hence removed. Also, various extraneous dialogues were also removed. After cleaning the source and target text were split for training and testing. The data is separated into Questions and Answers format and stored in the various dictionary. Now creating the last tokens, inverse dictionary and adding the end of the string. Now all the Questions and Answers are Translated into Integers and sorted. Now the decoding of the training set is done along with the test set. Now the Dictionaries are created to map the Questions words and the Answers words to a Unique Integer. Now the seq2seq model is built according to the dataset. Now creating the Decoder RNN. After Cleaning all the dataset and building the seq2seq model along with the decoder RNN training of the seq2seq Model is performed.

IV. RESULT

After training the dataset now the chatbot is tested by running various commands in the terminal. After training the model now the chatbot is ready to be tested. The output generated by the chatbot has moderate relevancy. Many of the outputs can be repetitive and generic. Also due to the lack of real-life quality data, the chatbot performed somehow above .average for imitating human interaction.



V. IMPROVEMENT

The performance of the chatbot can be optimized by using various real-life datasets. More realistic and high-quality conversational data could further be able to imitate a human being. For future training, personal chat history can also be incorporated to give the chatbot some personality. The replies of the Chatbot can look repetitive and lacks proper relevancy but this can be reduced by adding more diverse and healthy data.

VI. CONCLUSION

The Chatbot developed using the seq2seq model can be further improved with more robust and high-quality real-life datasets which can further increase the nature of the replies of the chatbot To further advance the project, Deep Reinforcement Learning can be applied that could improve the results significantly. The main techniques include Deep Neural Network, Recurrent Neural Network, seq2seq modeling with encoder and decoder. Future techniques to improve can include Long Short Term Memory Based RNN cells, Bi-directional LSTM, Neural Attention Model and Beam Searching. The training on Cornell Corpus produced results that can further be improved by more attention and speculation on training parameters. Training with other hyper-parameters and different datasets for further experimentation.

VII. ACKNOWLEDGEMENT

The success, as well as outcome of this project, required a lot of guidance and assistance. I am extremely privileged to have got this all along with the completion of my project. I have received supervision and assistance and I would not forget to thank Them. I heartily thank my guide, Assistant Professor Ms. Neha Singh for her guidance and suggestions during this project Work. I would also like to thank my parents and almighty who supported me throughout this journey.

REFERENCES

[1] Yin, Z., Chang, K. H., & Zhang, R. (2017, August). Deep probe: Information directed sequence understanding and chatbot design via recurrent neural networks. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 2131-2139). ACM.

[2] Cahn, J. (2017). CHATBOT: Architecture, design, & development. University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science.

[3] Wei, C., Yu, Z., & Fong, S. (2018, February). How to build a chatbot: Chatbot framework and its capabilities. In Proceedings of the 2018 10th International Conference on Machine Learning and Computing (pp. 369-373). ACM

[4] Nuruzzaman, M., & Hussain, O. K. (2019, May). Identifying facts for chatbot's question answering via sequence labeling using recurrent neural networks. In Proceedings of the ACM Turing Celebration Conference-China (p. 93). ACM

[5] Sidhu, J. S. (2019). Development of Chatbot using Deep NLP and Python.

[6] Deshmukh, V., & Nirmala, S. J. (2019, May). Open Domain Conversational Chatbot. In International Conference on Information, Communication and Computing Technology (pp. 266-278). Springer, Singapore...