

CHATBOT USING DEEP LEARNING

¹Neeraj Singh Kadayan, ²Aditya Dave, ³Vinit.K,

¹Student, ²Student, ³Assitant Professor,

Department of Computer Science,

SRM Institute of Science and Technology, Chennai, India

Abstract-

Chatbots are very famous now a days and there are a lot of applications of chatbot. Chatbots are interactive conversational application that introduce new way for the people to interact with the computer system. The purpose of this study is to create a chatbot that can be used on various platforms like twitter, twitch, discord. The technology we have used is Neural Machine Translation (NMT). For a better result and shorter training time compared to deep natural language processing, we try for the Bidirectional Recurrent Neural Network (BRNN). We have used the reddit comments as dataset for our training our model. We are developing a chatbot that will talk to people and is humorous in nature and doesn't sound like a bot.

Index terms- Artificial Intelligence, Chatbot, Deep Learning, Communication, Neural Networks.

I. INTRODUCTION

Now a days digital interaction with the devices is a bit limited, with respect to what technology and features being offered. No matter how simple a device technology is, there is always some innovation and new to learn. The bots we develop solve this situation by texting with the users and thus solving their problems. Chatbots are the most simple technology that we have to interact with users who are not bots but actual human being.. Since the chatbots behave like real human beings, Artificial Intelligence (AI) technology is used to develop them. Deep learning is one of the techniquethat duplicates the human behaviour. It finds similar replies through the dataset and educates itself to reply to the question being asked by users.

Deep learning is one of the most promising technology that can solve the problems like vision and the problems encountered in Natural Language Processing(NLP). We have certain chatbots that are already developed around us, like Alexa, Cortana, Siri and Charles the AIwhich have problems interacting with users and therefore increases the difficulty to handle users. Also these chatbots cannot keep up with the conversation going on with the users and lose track at some point of the conversation. Additional problem to this is, these bots are designed to deal with specific tasks designed by the developer, thereby limiting their domain. Conclusively leading them to a situation where they are unable to keep up with the current affairs and topics such as polity, times now etc.

II. MODULE AND ALGORITHM DESCRIPTION

The algorithm used is BRNN, seq2seq, Neural Machine Translation.

2.1 Recurrent Neural Network

The recurrent neural network (RNN) takes the data present in the present and also in the past while making the input sequence. The recurrent neural network has two nodes that are the encoder and the decoder. The RNN also has one hidden state present. The RNN only works in the forward direction. That is why we are going to use the Bidirectional neural networks which are alot more efficient as they have two hidden states. The bidirectional neural network connects one output to two hidden layers of opposite direction. With this type of generative deep learning the output layer is capable of getting information from back(past) and forward(future) states simultaneously. The bidirectional recurrent neural networks were invented by Schuster and Paliwal in 1997 in order to increase the input information available to the network.

The neural machine translation model contains two Bidirectional recurrent neural networks. One works as encoder and other one works as decoder.

In recurrent neural network, for every single word v_n , there is a dense vector called the recurrent state l_n , that combines v_n with l_{n-1} , where l_{n-1} is the processed recurrent state. Eventually, as given in the equation (1),

$$l_0 = f(l_{n-1}, v_n), l_0 = 0 \quad (1)$$

where, $l_n \in R_{d_h}$, d_h - number of dimensions of recurrent state, f - Non-linear transformation. The recurrent state h_n , is the compact summary of the words till n^{th} position.

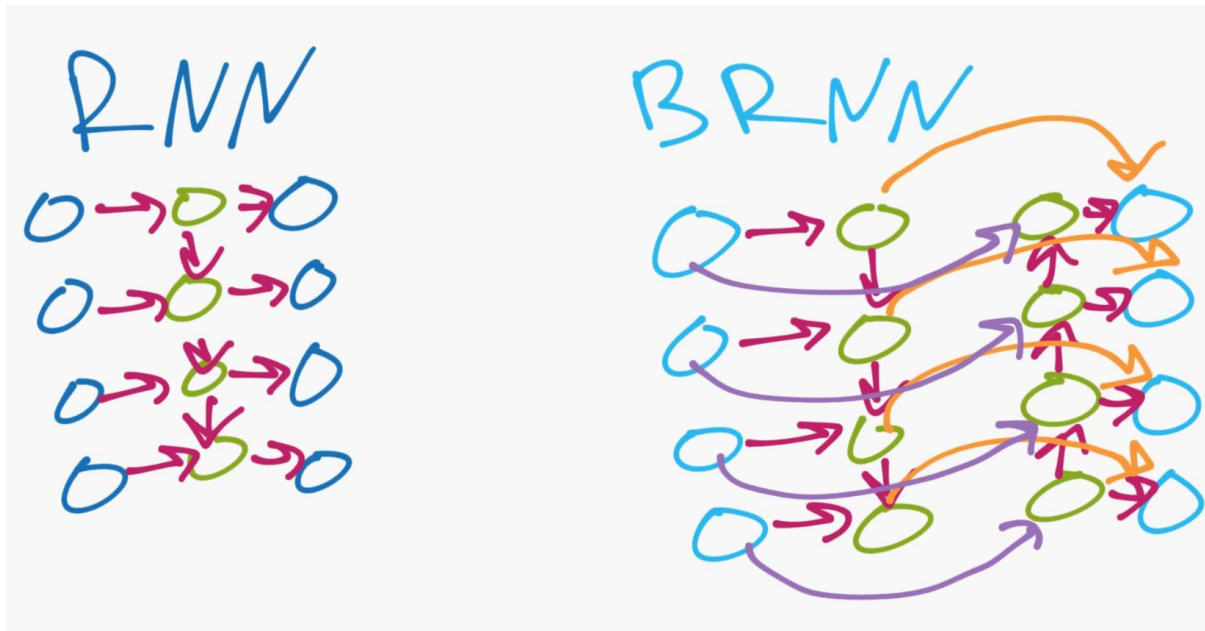
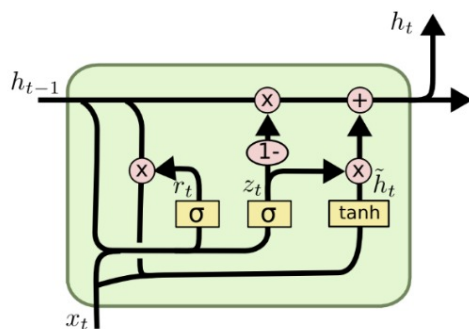


Fig.1 Bidirectional Recurrent Neural Networks

2.2 Gated Recurrent Cell

GRU or Gated Recurrent Unit is a variation of Recurrent Neural Networks (RNN) which uses gating mechanism. It was introduced by Cho, et al. in 2014. It solves the problem of vanishing gradient and can retain information from the distant past. GRU uses two gates, an update gate (Z_t) and a reset gate (R_t) according to the equation (1). These gates decide what information should be passed to the next step. The update gate determines the amount of information, from previous steps, that are to be passed along the next steps. GRU is pictured in Fig.2.



$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Fig.2 Gated Recurrent Unit

2.3 Sequence-to-Sequence Model (Seq2Seq)

In the present time, chatbot are a significant assistant to the application. Given the user input to the machine, it will analyze it and generate the response. In the current years, the models are made using the sequence-to-sequence (SEQ2SEQ) model. The advantage of sequence-to-sequence architecture is, it helps to process a long phrases of input and maintain a rich informational content between the user and chatbot.

A sequence-to-sequence(SEQ2SEQ) model has two recurrent neural networks : one is encoder and another one is decoder. Encoder used to process the input of the query from the user and decoder generates the relevant output of the query. Decode is having direct access of the word embedding with the encoder processed content to generate the most relevant output.

The objective function of seq2seq model is maximum-likelihood estimation (MLE). So this modules, may ignore the input and generate the generic response which may be frustrating for the user. For example, if user has given the input “How are you?”, so the response may be generic like “I am working”.

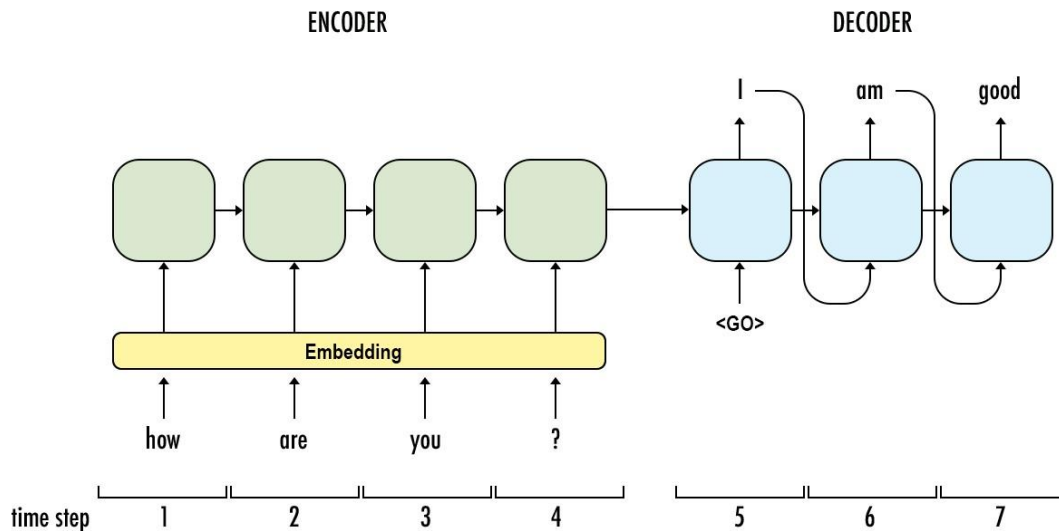


Fig.3 Sequence-to-Sequence Model Architecture (seq2seq)

III. METHODOLOGY

3.1 Selecting the dataset:

The data that we have selected for this is one month of reddit comments which is upto 45GB in size.

This is because the reddit data has a good set of parent comments on which there are multiple replies. We selected reddit comments as our source of data as it has a hierarchical architecture and thus reduces our pre processing time. It consists of more than 1.4 billion comments.

This data is to be converted into comment and reply pairs, in order to define input-output for the chatbot.

The attributes that has to be taken into account are :

“comment_id”, “parent_id”, “score” and “subreddit”.

After making the data base from the raw data we need to buffer the data to reduce the redundancy and unwanted data. After buffering the data we had to convert the database into the format that was needed for training the model. We only selected the comments which had a score of more than “2”, score is the amount of upvotes that the user got on reddit forum. After this we selected neural machine translation as our training model.

3.2 Selecting the Model:

The model is based on seq2seq (sequence to sequence) algorithm. It is called as Neural Machine Translation(NMT). It has two BRNNs as encoder and decoder. The bidirectional recurrent neural network (BRNN) not only takes the data present in the present, the past, but also takes in consideration the data in future while making the input sequence. The bidirectional neural network connects one output to two hidden layers of opposite direction. With this type of generative deep learning the output layer is capable of getting information from back(past) and forward(future) states simultaneously. The bidirectional recurrent neural networks were invented by Schuster and Paliwal in 1997 in order to increase the input information available to the network.

The neural machine translation model contains two Bidirectional recurrent neural networks. One works as encoder and other one works as decoder.

3.2.1 Encoder

For each query $K_m = \{v_{p,1}, \dots, v_{p,D_p}\}$ in training model S , the query level sequentially reads the words of the query level and updates according to the equation (2) given below :

$$h_{p,q} = GRU_{enc}(h_{p,q-1}, W_{p,q}), q = 1, \dots, D_p \quad (2)$$

3.2.2 Decoder

The RNN decoder predicts the next query Q_m based on the previous queries $Q_{1:m-1}$, i.e., to find the probability given in equation (3):

$$P(Q_m|Q_{1:m-1}) = \prod_{n=1}^{N_m} P(w_n|w_{1:n-1}, Q_{1:m-1}) \quad (3)$$

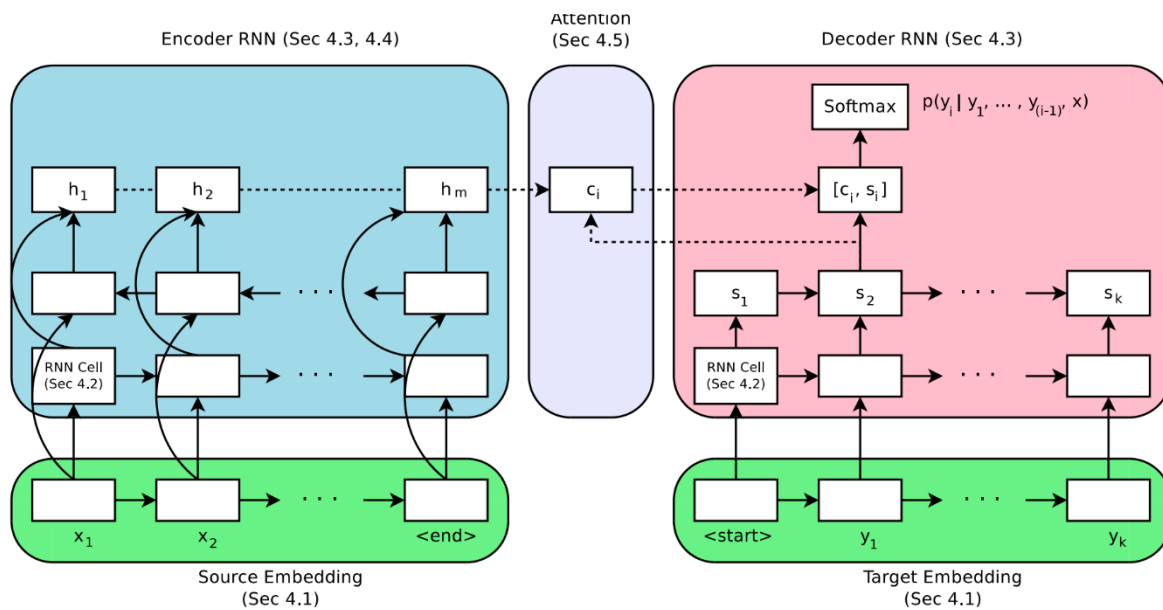


Fig.4 Neural

Machine Translator

3.3 Training the model:

We are using one month worth of training data created from the reddit comments. This data is enough for the model to atleast reach a acceptable BLEU score.

IV. PREPROCESSING

4.1 Word Embedding:

Word embedding can be defined as vectors representation of particular words as shown in Fig.1. These are a class of techniques which is used to represent individual words as a vector in a predefined vector space. Words havingsimilar meaning have similar representation or occupy adjacent space. Algorithms like Word2Vec, Glove, etc are used for learning word embedding. For example, words like good, well, nice, etc. will occupy adjacent space in a predefined vector space.

"a"	"abbreviations"	"zoology"	"zoom"
1	0	0	0
0	1	0	1
0	0	0	0
⋮	⋮	⋮	⋮
0	0	0	0
0	0	1	0
0	0	0	1

Fig 5 Basic Word Embedding.

4.2 SQLite:

We made the database out of the raw reddit data where there was unwanted information present. We only required the comment id, parent id, subreddit and scores. And tabulated this data in the form of a table and stored it locally in the system. We used SQLite for doing the same and viewing our database.

V. RESULT

In the end, after training the model, it was observed that the chatbot is more responsive and more intelligent than the traditional chatbots made using rule based method. Thereby our chatbot was ready to make its way into a working prototype in the industry format.

VI. CONCLUSION AND FUTURE WORK

The purpose of this paper is to study the applications of chatbots and to find if we can create a chatbot that can be used on multiple areas like twitch twitter facebook or even in commercial websites. We made the chatbot using neural machine translation model which made it possible for the chatbot to become intelligent and responsive compared to the traditional rule based chatbots. It is observed that while existing chatbots can perform these operations, they are not as efficient or robust as the new chatbots are made using deep learning algorithms.

The paper concludes with illustrating the benefits of the evolved deep learning techniques over conventional rule based chatbots made using request response technique. It shows that human like intelligence and better selection of responses can be achieved by using neural machine translation model for making chatbots.

Even though this model is far better than the traditional rule based models but it still has its own downsides. Like it requires a long time for us to buffer the data and create the dataset for training the model and also the model takes a lot of time for training. All these problems can be eliminated in the future by using a hybrid model that is based on both RNN and LSTM. Also with new DRL techniques that are emerging in Deep Learning, many new techniques for making chatbots are bound to come. With these new methods we may be able to increase the BLEU scores and also obtain good Turing test results.

VII. REFERENCES

- [1] Jincy Susan Thomas and Seena Thomas. Chatbot Using Gated End-to-End Memory Networks. In Proceeding of IRJET. March 2018
- [2] Hao Zhou, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu and Bing Liu. Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory. In Proceeding of Association for the Advancement of Artificial Intelligence 2018
- [3] Alessandro Sordani, YoshuaBengio, Hossein Vahabi, Christina Lioma, Jakob G. Simonsen and Jian-Yun Nie. A Hierarchical Encoder-Decoder for Generative Context-Aware Query Suggestion. In Proceeding of 24th ACM International Conference on Information and Knowledge Management(CIMK 2015), Melbourne, Australia
- [4] Iulian V. Serban, Alessandro Sordani, YoshuaBengio, Aaron Courville and Joelle Pineau. Building End-to-End Dialogue Systems Using Generative Hierarchical Neural Network Models. In Proceeding of 13th AAAI Conference on Artificial Intelligence(AAAI-16), pages 3776-3783
- [5] Chae Won Park and DaeRyongSeo. Sentiment Analysis of Twitter Corpus Related to Artificial Intelligence Assistants. In Proceeding of 5th International Conference on Industrial Engineering and Applications 2018, pages 495-498
- [6] Honghao WEI, Yiwei Zhao and JunjieKe. Building Chatbot with Emotions.
- [7] Siwei Lai, Kang Liu, Shizhu He and Jun Zhao. How to Generate a Good Word Embedding. In Proceeding of IEEE Intelligent Systems, 2016
- [8] BjarkeFelbo, AlenMislope, Anders Sogaard, IyadRahwan and Sune Lehmann. Using millions of emoji occurrences to learn any domain representation for detecting sentiment emotion and sarcasm. MIT (07 Oct 2017)
- [9] Ilya SutsKever, OriolVinyals and Quoc V. Le. Sequence to Sequence Learning with Neural Network. Google Scholar (14 Dec 2014)