Question Answering based University Chatbot using Sequence to Sequence Model

Naing Naing Khin
Natural Language Processing Lab
University of Computer Studies,Yangon
Yangon, Myanmar
naingkhin@ucsy.edu.mm

Khin Mar Soe
Natural Language Processing Lab
University of Computer Studies,Yangon
Yangon, Myanmar
khinmarsoe@ucsy.edu.mm

***Abstract*— Educational chatbots have great potential to help students, teachers and education staff. They provide useful information in educational sectors for inquirers. Neural chatbots are more scalable and popular than earlier ruled-based chatbots. Recurrent Neural Network based Sequence to Sequence (Seq2Seq) model can be used to create chatbots. Seq2Seq is adapted for good conversational model for sequences especially in question answering systems. In this paper, we explore the ways of communication through neural network chatbot by using the Sequence to Sequence model with Attention Mechanism based on RNN encoder decoder model. This chatbot is intended to be used in university education sector for frequently asked questions about the university and its related information. It is the first Myanmar Language University Chatbot using neural network model and gets 0.41 BLEU score.**

Keywords*— *chatbot, RNN, Seq2Seq learning, Attention Mechanism, Natural Language Processing(NLP), question answering system

I. INTRODUCTION

Chatbots are intelligent systems that can communicate with human using natural language while providing answers. Chatbots can be used well in the field of marketing, medicine, banking, education and transportation today. Other than sharing information, chatbots can offer more benefits in serving the education purposes for educational support. Chatbot can assist in educational sectors where people can obtain ideas or expand knowledge by using the Natural Language Processing and Machine Learning methods. NLP allows technology to understand what we are saying and how to response to it. Without NLP, AI that requires language inputs is relatively useless. By combing the NLP and Machine learning techniques, we can create many powerful tools for language processing. There are many chatbots using ruled-based and machine learning based on different methods and algorithms. Some chatbots like SimSimi, Mitsuku, A.L.I.C.E were famous and nowadays the machine learning chatbots like Siri, Alexa, Cortana are useful in human computer interaction. The development of a chatbot will be difficult if it is developed using pattern matching or rule-based approach. Neural network models generate the promising results to be used in question answering system for (FAQ) systems [1]. Sequence to sequence model is used to predict the sequences of text, image and speech recognition. This model becomes as a solution to any sequence to sequence problems where the inputs and outputs have different sizes and categories. It is based on Recurrent Neural Network(RNN) model that reads a word from an input sentence and then predicts the output words, which is constructed to be a sentence [5]. The aim of this research is to find out the answers for given questions through the sequence to sequence model. We have made some experiments on how an automatic communication can be built between humans and machine based on Seq2Seq neural network model.

This paper is composed of six sections. First section (1) is Introduction and the second section(2) is Related Work. The Methodology is in section(3) and Proposed System is in section(4). Implementation and results are displayed in section (5) and concludes the paper in section (6) respectively.

II. RELATED WORK

I. Sutskever, O. Vinyals, Q. V Le [4] presented an end-to-end approach to sequence learning that makes minimal assumptions on the sequence structure by using multilayered LSTM. They found that it is important to find a problem encoding that has the great number of short term dependencies and observed that LSTM can correctly translate very long sentences. Y. W. Chandra, S. Suyanto [5] focused on developing Indonesian chatbot based on sequence to sequence model. Evaluation was made on a small dataset from Telkom University's Admission department. The paper shows the comparison that the sequence to sequence model with Attention Mechanisim produces a quite higher BLEU score than the model without Attention Mechanism. K. Palasundram, N. M. Sharef, N. A. Nasharuddin, K. A. Kasmiran, A. Azman [8] provided sequence to sequence model performance for Education Chatbot. They investigated how a curated dataset can be developed sequence based on question-answer model. They proposed different question categories in natural language for Malay language in educational setting. K. Palasundram, N. M. Sharef, K. A. Kasmiran, A. Azman [9] made the enhancements of the sequence-to-sequence based natural answer generation models. This paper fills the several factors such as structural modifications, augmented learning, beam search and complementary mechanism. This also highlight unexplored areas in Seq2Seq Modeling and proposes potential future work for natural answer generation. A. Sojasingarayar [10] implemented encoder-decoder attention mechanism architecture using Recurrent Neural Network with Long Short Term Memory cells. These conversation agents are predominately used by businesses government organizations and non-profit organizations. The performance of the training was analyzed with the help of automatic evaluation metrics and by comparing output responses for a set of source utterances. This paper describes different combination of hyper parameters other than the configurations.

III. METHODOLOGY

A. Natural Language Processing and Chatbots

Natural Language Processing is a type of artificial intelligence technology that intends to translate, recognize, and understand human requests in the form of different languages. Question Answering System can handle

978-1-7281-9896-5/20/$31.00 ©2020 IEEE

documents related their respective sections to retrieve more precise answers using NLP techniques [2], [11]. We need to have an interface compatible with the ways of humans' receipt and pass information with communication. Now, NLP based chatbots can be used the user query written in their natural language and answer them quickly. NLP and machine learning techniques build chatbots for effective conversations without human interference. In today educational systems, the chatbot is used for teaching, learning and exploring the required information for the specific topic.

B. Sequence to Sequence Model

Sequence to sequence model is one of the current trend neural network models for sequential learning like dialogue systems, image captioning, speech recognition systems and so on. It consists of two Recurrent Neural Network(RNN), encoder and decoder. In question-answering problem, the input sequence is a collection of all words form the question. The encoder takes the input sequence and converts them into a fixed size feature vectors. During the encoding, the following steps are happen: (a) the string "ဘယ် အချိန် ကျောင်း ဆင်း တာလဲ", When does the class finish?(in English) is first tokenized into a list of tokens. "ဘယ်", "အချိန်", "ကျောင်း", "ဆင်း", "တာလဲ". (b) Each token is related with a vector which is a subset of input vocabulary. For every timestep(each input), the hidden state(vector) is updated according to the input. (c) The encoder network will do the embedding and produce the intermediate states(fixed length vectors). These vectors aims to encapsulate the information for all input elements in order to help the decoder make accurate predictions.

The decoder works differently during training and testing unlike the encoder part of the model. The decoder aims to generate the answer based on the hidden states of the encoder. The decoder takes its previous hidden state and computes the next hidden state by using hidden vectors from previous layer, the previous layer output and the original hidden vector. Each recurrent unit accepts a hidden state from the previous unit and produces the output with respective weight. Finally, the loss is calculated on the predicted outputs from each timestep of the model.

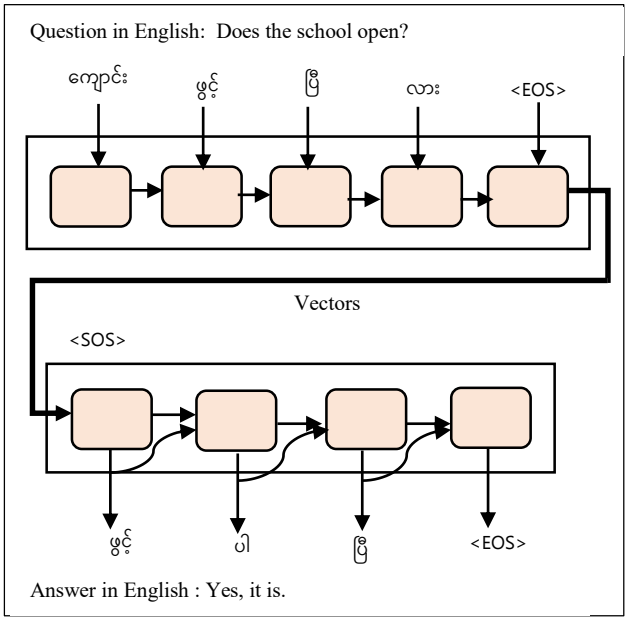

Fig. 1. Architecture of the Sequence to Sequence Model

C. Attention Mechanism

Attention mechanism is developed to increase the performance of RNN model which encodes the input sequence into one fixed size vector. The basic seq2seq does not handle well in decoding large sequences. When the length of the sequence gets larger, it can lose the amount of information. We add the Attention mechanism in this work, the model tries to predict an output word, it only uses parts of an input where the most relevant information is concentrated instead of an entire sentence. Attention was introduced as a solution to solve some issues among seq2seq models to have better work. Instead of using a fixed context, a distinct context vector is used for generating word [13], [14], [18]. Two-layered RNNs (with LSTM cell) is used as encoder and decoder. This model calculates both previous and next words because it consists of both forward and backward directions. Therefore, attention in deep learning can be interpreted as a vector of important weights to predict or infer one element, such as a pixel in an image or a word in a sentence. We estimate using the attention vector how strongly it is correlated with other elements and takes the sum of their values weighted by the attention vector as the approximation of the target element.

We have an input question x of length n and try to output a target answer y of length m: $x = [x_1, x_2, \dots, x_n]$ and $y = [y_1, y_2, \dots, y_m]$. The encoder uses bidirectional Recurrent Neural Network with forward hidden state $\overrightarrow{h_t}$ and a backward one $\overleftarrow{h_t}$. The concatenation of these two states represents the encoder state. The motivation is to include both the preceding and following words in the annotation of one word. The decoder network has hidden state $s_t = f(s_{t-1}, y_{t-1}, c_t)$ for the output word at position t, $t = 1, \dots, m$ where the context vector c_t is computed as a weight sum of activation states in forward and backward directions.

$$c_t = \sum_{i=1}^{n} \alpha_{t,i}\, h_i \qquad (1)$$

where n denotes the length of the input sentence, h_i is the hidden state at time step i and $\alpha_{t,i}$ denotes how much attention is given by the input for the generation of output word.

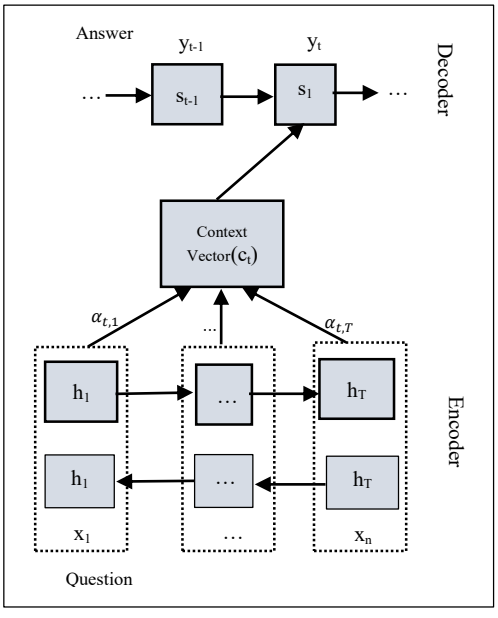

Fig. 2. Illustration of Attention Mechanism

IV. Proposed System

A. Data Collection

Data is the most important thing when we develop a system in Natural Language Processing applications and it is also time consuming task especially for low-resourced languages. Most NLP models need large amounts of training data and language-specific processing. However, such amount of data is not available for most languages to build a language model. Myanmar Language is also a low-resourced language and we face the problem of data scarcity in creating training data. We have collected a corpus that contains 5000 question answer pairs (in a total of about 10K sentences) to train the conversation model. The data is still closed domain type which comes from the admission team of University of Computer Studies, Yangon, Myanmar and from its Facebook page messenger. We also collect manually the data from the students, their parents and people who want to inquire the information about the university.

B. Data Preprocessing

Data preprocessing is the basic important step for Natural Languages before training the model. Myanmar sentences are usually written in two kinds of Myanmar fonts: non-Unicode and Unicode. So, for the proposed model training, it needs to be unique and converted to Unicode font using a Myanmar language font conversion tool (http://burglish.my-mm.org/latest/trunk/web/fontconv.htm). Then, the collected data is segmented with the Myanmar word segmenter of UCSY NLP Lab. There are stop words in Myanmar language like "နှင့်, ဤ, ထို, ဟို, ယင်း, ၎င်း, သည်, က, မှ, ၌, ၍", etc. These stop words are removed. Normalization is the task of transforming a text into a canonical form and it also maps the identical words: for example, such as "အဆင်ပြေသလား", "အဆင်ပြေရဲ့လား", "အဆင်ပြေမလား", "Are you ok?" in English to "အဆင်ပြေလား" in Myanmar Language.

C. Work Flow of the System

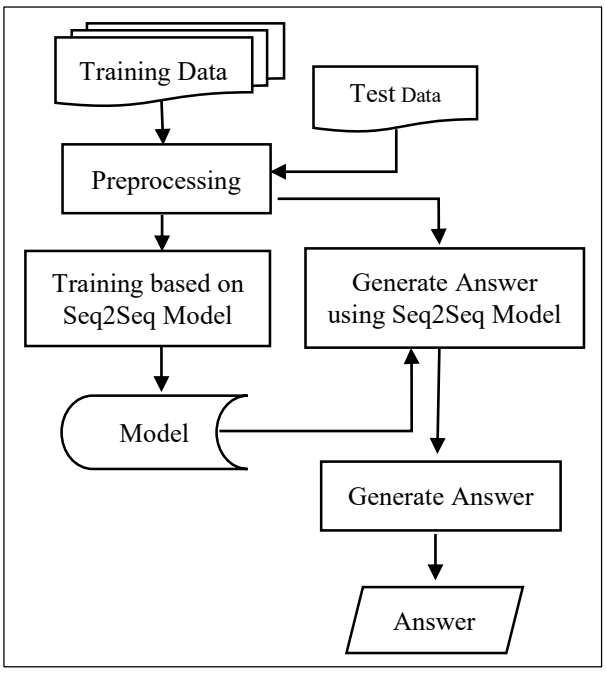

Fig. 3. Flow of the System

The proposed system involves two phases: training and testing. The data is preprocessed before training the conversation model. Next, the model is trained on sequence to sequence model and sequence to sequence with attention mechanism. The answer is generated and model evaluation is done by the Bilingual Language Understudy (BLEU) score method.

V. Implementation of the System

A. Experimental Results

There are many chabots using ruled-based and machine learning based in different languages. This is a University Chatbot using machine learning with Myanmar Language. We have proposed the system to get a support for university education sector. Our contribution is to create a University Related Information Corpus by Myanmar Language and to make enhancement the sequence to sequence model for automatic question answering Myanmar chatbot. It is closed domain type chatbot with Myanmar Language for Frequently Asked Questions about the university. The system uses Recurrent Neural Network based Encoder Decoder model via Attention mechanism. Attention was introduced to solve the problem of long sequences which is also a problem for most other NLP tasks as well. The aim of attention mechanism is to form direct-shortcut connections between the target and source by paying attention to the source content. We have found that the result with Attention mechanism is better than the baseline RNN encoder decoder method. The chatbot provides the useful information for inquirers who want to know the information according to their needs. The overall model accuracy is not very high and we still need some improvements for better model and also data preparation. We have used Python language with tensorflow library to process the implementation of the system. The sample output from the system is displayed in table 1.

TABLE I. Sample Output of the System

Question: (Human)	မင်္ဂလာ ပါ ရှင့်	Hello
Answer: (Bot)	ဟုတ်ကဲ့ မင်္ဂလာ ပါ	Yes, hello.
Question: (Human)	ရန်ကုန် ကွန်ပျူတာ တက္ကသိုလ် ရဲ့လိပ်စာ ကို သိ ချင် ပါတယ်	I want to know the address of University of Computer Studies, Yangon.
Answer: (Bot)	အမှတ် ၄ လမ်းမကြီး၊ ရွှေပြည်သာ မြို့၊ ရန်ကုန်တိုင်း ဖြစ်ပါတယ်	It is No.4 Main Street, Shwe Pyi Thar Township, Yangon Region.
Question: (Human)	အဆောင် လျှောက် ချင် လို ဘယ်လို လျှောက် ရ မ လဲ	How to apply for the accommodation?
Answer: (Bot)	ကျောင်းသား ရေးရာ ဌာန မှာ အသေးစိတ် သွားရောက် စုံစမ်း နိုင် ပါတယ်	Please inquire at Student Registration Department.
Question: (Human)	အမှတ် စာရင်း ထုတ် ဖို ဘာတွေ လိုအပ် သလဲ	What are needed for grading?

Answer: (Bot)	ဘွဲ့လက်မှတ် မိတ္တူ၊ နှစ် အလိုက် ခုံ နံပါတ်	Hard copy of graduation certificate, and roll number for each year
Question: (Human)	ရန်ကုန် ကွန်ပျူတာ တက္ကသိုလ် ကို ဘယ် ခုနှစ် က စတင် ဖွင့်လှစ် ခဲ့ သလဲ	When did the university open?
Answer: (Bot)	၁၉၉၈ ဇူလိုင်လ ၁ ရက်နေ ဖြစ်ပါတယ်	It is 1th July, 1998.

B. Testing and Evaluation

We have used 4700 question answer pairs for training phase and 300 for testing phase. Testing and Evaluation involves calculation or assessment for a system with appropriate methods. There are many evaluation methods for question answering systems such as ROUGE, BLEU, METEOR and so on. Now, the performance of the proposed system is evaluated with BLEU score method. The BLEU score which calculates the similarity between the output of the model and actual sentences. It is usually a number from 0 to 1. If the value is closer to 1, it can be recognized a better model and 0 is no match at all. It is formulated as:

$$BLEU = \frac{\sum_{n\_gram \in y^{\wedge}} Count_{clip}(n\_gram)}{\sum_{n\_gram \in y^{\wedge}} Count(n\_gram)} \quad (2)$$

where $Count_{clip}(n\_gram)$ is the largest number of n-grams in both output and actual sentences, $Count(n\_gram)$ is the number of n-grams found in the output, n is the length of contextual words(4gram in this system).

We have trained the models with the batch size of 64,128 and 256 respectively. Batch size is one of the most important hyper parameters to tune in modern deep learning systems. Sometimes, too large of a batch size will lead to poor generalization. The learning rate is 0.002, dropout rate is 0.9 and 128 neurons. Then, the BLEU scores on these models are as follow:

TABLE II. BLEU SCORE CALCULATION FOR RNN WITH ATTENTION MECHANISM

Batch Size	BLEU1	BLEU2	BLEU3	BLEU4
64	**0.41**	**0.34**	**0.23**	**0.14**
128	0.31	0.23	0.14	0.07
256	0.30	0.23	0.13	0.07

Table 2 describes the BLEU scores for RNN with attention mechanism by three batch size. It gets the highest of 0.41 BLEU with 64 batch size among them. In table 3, the baseline RNN model is tested with above batch sizes and have got 0.35 score. The higher score, the better result correlates to the human reference answers. However, the result of the system does not totally rely on these scores. Human judgment and manual analysis are also important for this system because human conversation style can vary on the same meaning of the sentence depend on its language nature. For example, "ကျောင်းက ဘယ်မှာ ရှိတာလဲ" and "ကျောင်း လိပ်စာ သိချင်လို့", "Where is the school?" and "I want to know the school address?" in English. They have the same meaning of question and we have to reply the same response to user. There are some challenges on these conditions.

TABLE III. BLEU SCORE CALCULATION FOR RNN

Batch Size	BLEU1	BLEU2	BLEU3	BLEU4
64	**0.35**	**0.21**	**0.12**	**0.06**
128	0.35	0.20	0.12	0.06
256	0.34	0.20	0.11	0.06

The loss function has an important job that improvements are a sign of a good model or not. Figure 4 shows the training loss to calculate the model error during the models training. The average loss value for RNN with attention mechanism is 0.72 and RNN model is 2.18 for this training.

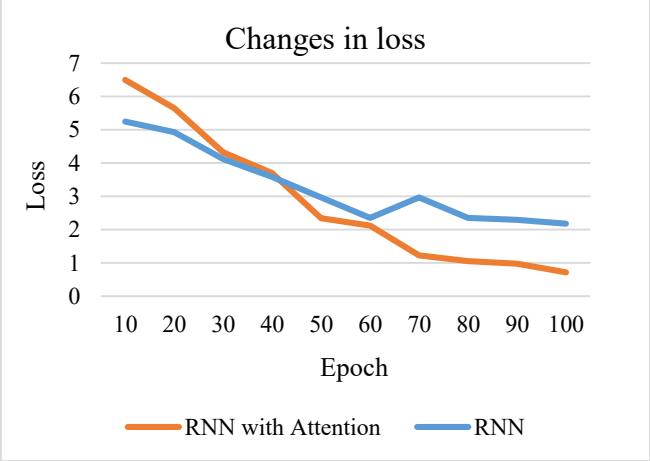

Fig. 4. Changes of loss between models

In figure 5, it displays the accuracy of the trained models. RNN with attention mechanism gets higher score than RNN. The score comparison of these two models are as follow:

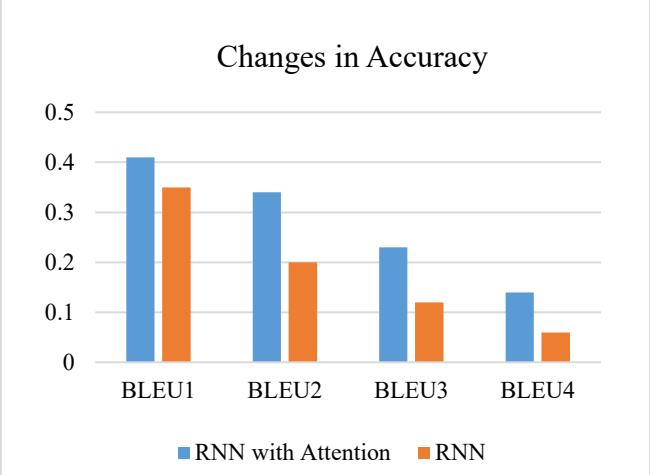

Fig. 5. Changes of accuracy between models

VI. Conclusion

Chatbots have been developed to assist the human requirements in our daily routine. In this paper, we have designed a Myanmar Language chatbot based on sequence to sequence model. The result shows that it can generate the required information on the user question about the university information. The chatbot helps the students, teachers and others who want to know information about the university. We have trained the conversation model on both CPU, GPU and implemented by Python language. Although there are some challenges, Seq2Seq model with attention mechanism provides some insights for the question answering system in the education chatbot. Further developments will lead to enhance the system for greater accuracy and solution.

References

[1] M. T. Mutiwokuziva, M. W. Chanda, P. Kadebu, A.Mukwazvure, T. Gotora, "A Neural-network based Chat Bot", 2017.

[2] T. Nguyen, M. Shcherbakov, "A Neural Network based Vietnamese Chatbot", Nov 2018.

[3] N. G. Aarthi, G. Keerthana, A. Pavithra, K. Pavithra, "CHATBOT FOR RETAL SHOP EVALUATION", March 2020.

[4] I. Sutskever, O. Vinyals, Q. V. Le, "Sequence to Sequence Learning with Neural Networks", 2014.

[5] Y. W. Chandra, S. Suyanto, "Indonesian Chatbot of University Admission Using a Question Answering System Based on Sequence-to-Sequence Model", 2019.

[6] T. H. Wen, D. Vandyke, N. Mrksic, M. Gasic, L. M. Rojas-Barahona, P. Su, S. Ultes, S. Young, "A Network-based End-to-End Trainable Task-oriented Dialogue System", April 2017.

[7] H. Mei, M. Bansal, M. R. Walter, "Coherent Dialogue with Attention-based Language Models", Nov 2016.

[8] K. Palasundram, N. M. Sharef, N. A. Nasharuddin, K. A. Kasmiran, A. Azman, "Sequence to Sequence Model Performance for Education Chatbot", 2019.

[9] K. Palasundram, N. M. Sharef, K.A. Kasmiran, A. A. Azman, "Enhancements to the Sequence-toSequence-Based Natural Answer Generation Models", March 2020.

[10] A. Sojasingarayar, "Seq2Seq AI Chatbot with Attention Mechanism", May, 2020.

[11] Y. Sharma, S. Gupta, "Deep Learning Approaches for Question Answering System", 2018.

[12] P. Muangkammuen, N. Intiruk, K. R. Saikaew, "Automated Thai-FAQ Chatbot using RNN-LSTM", 2018.

[13] M. R. Costa-jussa, A. Nuez, C. Segura, "Experimental Research on Encoder-Decoder Architectures with Attention for Chatbots" Vol. 22, No. 4, 2018.

[14] A. Ali, M. Z.Amin, "Conversational AI Chatbot Based on Encoder-Decoder Architectures with Attention Mechanism", Nov 2019.

[15] D. Hong, B. Wan, "Attention-based Recurrent Neural Networks for Question Answering".

[16] I. Ilievski, J. Feng, A Simple Loss Function for Improving the Convergence and Accuracy of Visual Question Answering Models".

[17] A. Galassi, M. Lippi, and P. Torroni, "Attention in Natural Language Processing", May 2020.

[18] Y. Nie, Y. Han, J. Huang, B. Jiao, A. Li, "Attention-based encoder-decoder model for answer selection in question answering", March 2017.

[19] A. Joshi, K, Popat, S. Gautam, "Automatic evaluation of Machine Translation (MT): BLEU, its shortcomings and other evaluation metrics", 2013.

[20] M. Dhyani, R. Kumar, "An intelligent Chatbot using deep learning with Bidirectional RNN and attention model", May 2020.