Implementation of Artificial Intelligence Based Chatbot System with Long Term Memory

Shruti Katkade, Urjita Kerkar, Pravin Bhilare, Manish Gupta and Payel Thakur

IMPLEMENTATION OF ARTIFICIAL INTELLIGENCE
BASED CHATBOT SYSTEM WITH LONG
TERM MEMORY

Shruti Katkade, Urjita Kerkar, Pravin Bhilare, Manish Gupta, and Prof. Payel Thakur

Department of Computer Engineering, PCE, Navi Mumbai, India - 410 206

Abstract: This paper mainly explores a specific deep learning method to build a conversational agent. Nowadays the popularity of chatbot systems is on rise as they attempt to get into daily life and achieve some commercial success. Previous approaches used simple keywords & pattern matching methodologies, answering in a static manner irrespective of previous conversions. As an improvement to this technology would be a system that will work with sequence to sequence framework. Our proposed model makes use of this framework. Given the previous sentence or sentences and the next sentence in a conversation, the model converses by predicting the next sentence. The distinctive feature of our model is that it can be trained end-to-end hence requires much fewer hand-crafted rules. This straightforward model can generate simple conversations given a large conversational training dataset.

Keywords: Deep Learning, Recurrent neural network, end to end memory, LSTM model, seq-to-seq model.

I. INTRODUCTION

Before deep learning hit the scene a few years ago all chatbots used hard coded rules. The programmer had to think of all of the possible responses to a question and handcraft rules accordingly. Recently there has been a growing interest in developing chatbots using deep learning that use end to end systems as it is less cumbersome to build than the previous methodologies in which different components are trained to carry out a specific task and the result of each component is collectively used to predict a response. End to end systems are preferred when you have a really large dataset available to train your model. Many researchers have explored the use of deep learning recurrent neural network architecture to develop efficient chatbots.[8]

The most popular model used nowadays is the seq-to-seq recurrent neural network model used for sentence generation. Seq-to-Seq was initially proposed for machine translation, but it can be applied in generating the conversation as well. The seq2seq model has two encoder units and a decoder unit. The encoder unit takes the input sequence from the system and trains on it then it passes the last state of its recurrent layer as an initial state to the first recurrent layer of the decoder part. The decoder takes the last state of encoders last recurrent layer and uses it as an initial state to its first recurrent layer, the sequence we want to get is the input of decoder.This type of seq2seq model has shown impressive performance in various domains like speech recognition, machine translation, question answers etc. Conversation generation using deep learning includes two types of models: Retrieval based model and generative model. Retrieval based model has a repository of responses and uses some kind of heuristics to pick an appropriate response based on input and content hence it cannot handle unseen cases. On the other hand generative models create a response from scratch. But they are more prone to grammatical errors which can be handled if these models are trained precisely. This research is on generative models based method LSTM for conversation generation trained end to end. [8]

II. LITERATURE SURVEY

Sr No	Paper	Advantage	Disadvantage
1	Chatbot : Artificially Intelligent Conversational Agent,Ayesha To design a chatbot system for a particular	To design a chatbot system for a particular college using the divide and conquerstrate gy.It makes use of the	Chatbot system is developed only for android users.

	college using the divide and conquer Chatbot system is developed only for android Shaikh , Geetanjali Phalke ,2017 [1]	Dialogue Acts Classifiers,POS-taggerData Mining,Patten matching, Knowledge Database,cloud services to store responses.	
2	A Neural-network based Chatbot, Milla T Mutiwokuziv, Melody W Chanda,2017 [3]	To explore the capabilities of deep neural networks by designing a chatbot using recurrent neural networks(RNN).Implementing sequence to sequence long short-term memory cell neural network (LSTM) in conjunction with Google word2vec.	Textual input provided To the system.Voice recognition and voice Synthesis can be provided as input and output to the system.
3	Long Short-Term Memory Networks for Automatic Generation of Conversations, Tomohiro Fujita,2017 [2]	Generates conversational sentences via recurrent neural network and its coupled memory unit, long short-term memory (LSTM) which gives efficient answers.	Current system is particularly vulnerable to the case Of Notation Change and the sequential outputs are frequently neither Linguistic nor contextual valid.
4	A Neural Conversational Model, Oriol Vinyals,2015 [5]	To present a simple approach for conversational modelling using the recently proposed sequence to sequence framework.Chatbot is trained end-to-end and thus requires much fewer hand-crafted rules.	The model does not capture a consistent personality
5	Example-based Chat-oriented Dialogue System with Personalized Long-term Memory, Jeesoo Bang, Hyungjong Noh,2015 [4]	To design an Example-based chat-Oriented dialogue system with Personalization framework using long-term memory. To reduce human Labor, annotation cost and build rapport with the user. EBDM and personalization is used.	Detecting important topic keywords automatically will Enable system to extent to other language Easily. Complex Features such as parsing information can be used for sentence matching.

III. PROPOSED WORK

An LSTM(Long-term Short-Term Memory) network is a Recurrent Neural Network(RNN). In this network LSTM cells replace neural network layers. Our model has memory and can remember information for a long period of time. It is trained end to end, with more accuracy as Bidirectional LSTM implementation has been used, which makes it stand out.[8]

3.1. System Architecture

The system architecture is given in Figure 2. Each block is described in this Section. (LSTM) in conjunction with Google word2vec.
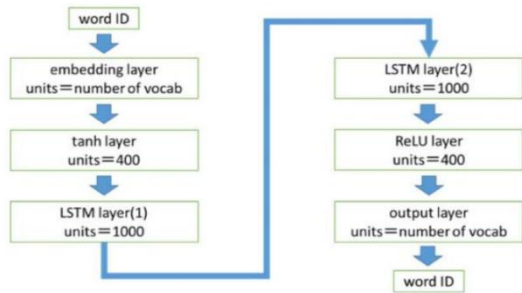
3.2. Existing System

Figure.1. LSTM model for conversation generation [2]
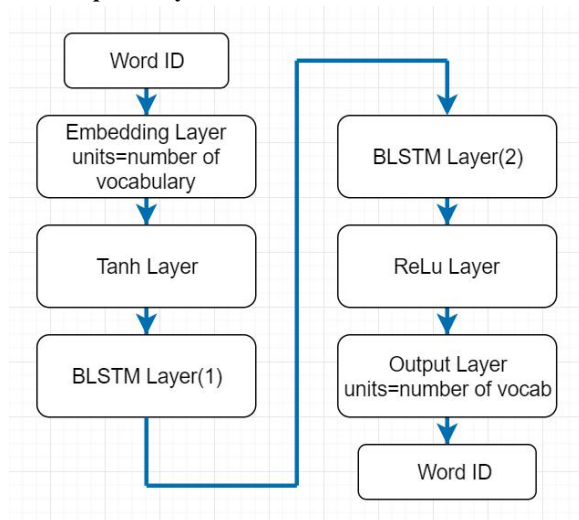
3.3. Proposed System

Figure. 2 Proposed system architecture[8]

A. Block 1: The input to the system is a sequence of sentences provided by the user which is later split into words. Each of the words is indexed to form a word ID for the specific word. Given a word ID to the system, it retrieves the word corresponding to the current id to find its next word ID. It then concatenates the two words for generating a candidate system snippet. Model's internal state is kept updating until the candidate snippet output is terminated by symbol<EOS>.

B. Block 2: The embedding layers task is to perform the word embedding. Each word is represented as real valued vectors ,of hundreds of dimensions ,in a predefined vector space. The vector is a distributed expression of each word

.
C. Block 3and 6: The tanh and ReLu are activation functions. To improve the representation keeping the expression size unaltered, the tanh and ReLu layer is Introduced.

D. Block 4 and 5: The Bidirectional LSTM network is more powerful than LSTM. For any point in the sequence BLSTM stores entire sequential information about points before and after that point. BLSTM networks are trained using back propagation algorithms over time. In BLSTM networks, the forward and backward passes are carried out in the same way as a regular network but the only difference here is that the hidden states are required to be unfolded for all time. [7]

E. Block 7: The last block is the output layer. The generated response is again separated word by word and is provided with a word ID.

3. Requirement Analysis

The details about the dataset along with hardware and software requirements is given in this section.

3.1 Dataset and Parameters

1. Movie Dialogue Dataset:

Table 3.1 Dataset Used for General Conversation [6]

Dataset	Items
movie_characters	9034
movie_conversations	9030
movie_lines	666256
movie_titles	616

2. Question Answer Dataset:

Table 3.2 Dataset Used for First Women in India and Sports[9][10]

PARAMETERS	ITEMS
ArticleTitle	1720
Question	8500
Answer	8500
DifficultyfromQuestioner	3245
DifficultyfromAnswerer	1200
ArticleFile	1500

We manually created the dataset for experimental purposes by extracting and integrating questions from female centric datasets, sports specific datasets. Our dataset contains around 10,000 pairs of questions and answers. Each question maps to a single answer.

Structure of dataset is:

<id>+<question>

<id+1>+<answer>

The dataset was divided in ratio 3:2 for training and validation.

Dataset format:

LINE	TEXT
1+	Who was the first lady president of india?
2+	Pratibha Patil[2007-2012]
3+	Which football player has been voted as the best playmaker in the world for 2016?
4+	Lionel Messi

3.2 Hardware

SR NO.	NAME	SPECIFICATION
1	C.P.U with minimum clock speed:	Intel celeron 700 Mhz processor / Quad core Intel core i3 or greater processor
2	Storage:	4GB of hard-drive space
3	GPU	NVIDIA TESLA K80
4	R.A.M	4B system memory

3.3 Software

SR NO.	NAME	SPECIFICATION
1	Compatible Operating System:	Compatible with Ubuntu 14.04 or up.
2	Language:	Python, HTML, JavaScript ,CSS.
3	Drivers:	Microsoft drivers ,Sound Card drivers
4	A.P.I.:	Responsive voice.JS, Spacy, Flask.

FUTURE SCOPE

A voice enabled chatbot can be created as our proposed chatbot only takes textual input. Models accuracy can be improved by fine tuning the hyperparameters used. This chatbot web application can be built for android platform and can also be embedded with websites.

V. REFERENCE

[1]. "Chatbot: Artificially Intelligent Conversational Agent", Ayesha Shaikh, Geetanjali Phalke, 2nd International Conference on Communication and Electronics System(ICCES), Milla Mutiwokuziva, Melody Chanda, Prudence Kadebu, Addlight Conference

Location-Coimbatore, India, 2017.

[2]. "Long Short-Term Memory Networks for Automatic Generation of Conversations",Tomohiro Fujita ,18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Tomohiro Fujita, WenjunBai, Changqin Quan, 26-28 June 2017, Location-Kanazawa, Japan,2017.

[3]. "A Neural-network based Chat Bot" , Milla Mutiwokuziva, Melody Chanda, 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS), Bayu Setiaji, Ferry Wahyu Wibowo, 25-27 Jan 2016, Bangkok, Thailand,2017.

[4]. "Example-based Chat-oriented Dialogue System with Personalized Long term Memory", Jeesoo Bang, Hyungjong Noh, International Conference on Big Data and Smart Computing (BIGCOMP), 9-11 Feb 2015 Conference Location- Jeju, South Korea ,2015.

[5]."A Neural Conversational Model",Oriol Vinyals,Quoc V. Le, 2015

[6].https://www.cs.cornell.edu/~cristian/Cornell_Movie-Di alog s_ Corpus.html

[7].https://medium.com/@BhashkarKunal/conversational-ai-chat bot-using-deep-learning-how-bi- Directional -lstm-machine-read ing-38dc5cf5a5a3

[8] "AI Based Chatbot System With Long Term Memory" Shruti Katkade , Urjita Kerkar , Pravin Bhilare , Manish Gupta , Payel Thakur Assistant Professor, IJESC, December 2019

[9] First Women in India in various fields

[10]ICC World Cup Trivia Questions & Answers | Cricket