AI Chatbot for Regional Languages

Ketan Makwana
Computer Engineering
A.P.Shah Institute of technology
Thane,India
mketan16@gmail.com

Gaurav Hulmukh
Computer Engineering
A.P.Shah Institute of technology
Thane,India
ghulmukh10@gmaill.com

Ganesh Kamath
Computer Engineering
A.P.Shah Institute of technology
Thane,India
ganeshkamath1997@gmail.com

Abstract—**An AI chatbot aims to make a conversation between both human and machine. The machine has been embedded knowledge to identify the sentences and making a decision itself as response to answer a question. The response principle is matching the input sentence from user. From input sentence, it will be find to get the similarity of sentences and train itself.The knowledge of chatbot are stored in the database. The chatbot will be made using the python language using Natural Language Processing. The front-end of the chatbot will be developed using Dialogflow in which the input provided by the user will be in the form of text or voice and the reply for the input will be received in immediate time, Also it aims for paying the Light-bills or any other transactions. The user has to provide the input for the first time(bill number,credit card number)and it will remind the user about the deadline of the payment. In this paper chatbot is use tensorflow,tflearn,NLTK for data training,and also use Textblob for Tokenization,parsing and etc.**

Keywords- Chatbot,machine learning,Natural Language Processing,Dialogflow,NLTK,Textblob,tensorflow.

I. INTRODUCTION

What is Chatbot?

[1] Ai chatbot is a application that was designed to converse with users using speech and natural language, and plays a crucial role in human-machine interaction. As shown in Fig. 1, chatbot is generally composed of automatic speech recognition (ASR),natural language understanding (NLU), dialog management (DM), natural language generation (NLG), and text-to-speech synthesis (TTS) . NLG in chatbot components generates natural language sentences that correspond to structured semantic representations. The semantic expression includes the intention and the information an chatbot intends to convey. Thus, effective NLG dramatically improves the usability of human-machine interaction. In recent years, various deep neural network (DNN) technologies have been applied to natural language processing tasks. Specifically, the recurrent neural networkc (RNN) has demonstrated unequivocal advantages in tasks such as machine translation, image captioning , video captioning , video action recognition video retrieval , information extraction , and NLG Among these applications, NLG is a particularly challenging task in an chatbot..Currently, conventional DNN-based NLG usually employs a sequence decoding method that generates the words of a sentence in sequential order, moving forward or backward, using input generated by the word in the preceding step

A chatbot is an artificial intelligence (AI) software that can simulate a conversation (or a chat) with a user in natural language through messaging applications, websites, mobile application or through the telephone.

Why are chatbots important?

A chatbot is often described as one of the most advanced and promising expressions of interaction between humans and machines. However, from a technological point of view, a chatbot only represents the natural evolution of a Question Answering system leveraging Natural Language Processing (NLP). Formulating responses to questions in natural language is one of the most typical Examples of Natural Language Processing applied in various enterprises end-use applications.

The purpose of the chatbot was to develop friendly interaction with the human. It can be considered as one of the best assistant of the human as the machine has the ability to perform the task as human do and many companies like facebook,amazon,Twitter are investing a lot of money for developing an effective chatbot.

II. LITERATURE REVIEW

A. Natural Language Toolkit

The Natural Language Toolkit (NLTK) is data package. In NLTK has a pre-trained Punkt tokenizer for Dataset. Removing **Noise** that is everything that isn't in a standard number or letter. Removing **Stop words ,** Sometimes, some extremely common words which would appear to be of little value in helping select documents matching a user need are excluded from the vocabulary entirely. These words are called *stop words*

Stemming*: Stemming is the process of reducing inflected (or sometimes derived) words to their stem, base or root form —generally a written word form. Example if chatbot were to stem the following words: "takes", "taking", "taken", "and Stemtization", the result would be a single word "tak".*

B. Natural Language processing

Natural Language processing is the potential of the machines to understand, analyze and generate human speech. The goal of NLP is is to make interactions between computer and human feel like two humans are actually communicating with each other. With NLP ,the computer has the ability to understand the natural language spoken by the human,understand the meaning of it,and if needed, respond to it by generating natural language to communicate back to the person. NLP is a field of computer science that has been around for a while,but it has gained popularity in recent years as advances in technology has made it easier to develop computers with NLP abilities. In 1950,Alan Turing has published his famous paper titled Computer Machinery And Intelligence. The paper proposed a test to determine if a machine can trick a human in identifying that whether the person or human judge communicating is communicating with the person or with the robot. This test became popular as the Turing test and passing it would signal the goal of artificial intelligence. The most essential part of the Turing Test is communication, the machine must communicate with human effectively,which is called natural language processing.

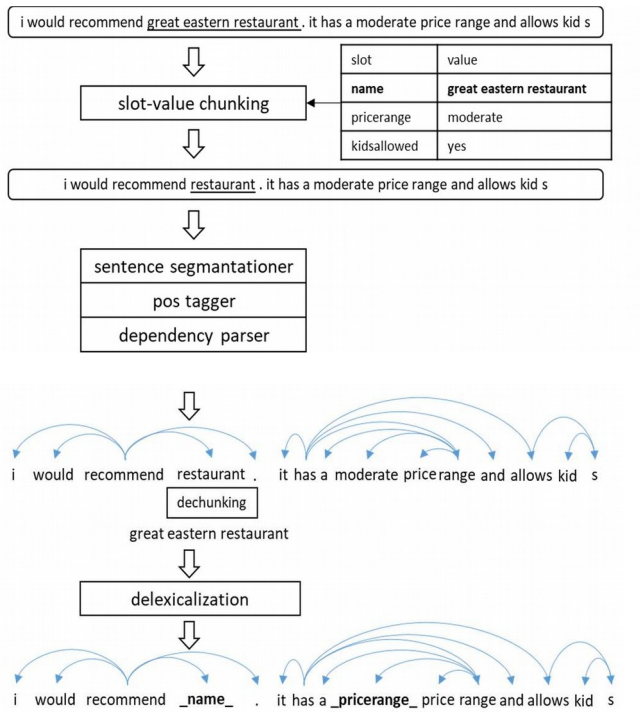

Figure 1

C. Tensor Processing Unit(TPU)

With machine learning gaining its relevance and importance everyday, the conventional microprocessors have proven to be unable to effectively handle it, be it training or neural network processing. GPUs, with their highly parallel architecture designed for fast graphic processing proved to be way more useful than CPUs for the purpose, but were somewhat lacking. Therefore, in order to combat this situation, Google developed an AI accelerator integrated circuit which would be used by its TensorFlow AI framework. This device has been named TPU (Tensor Processing Unit). The chip has been designed for Tensorflow Framework.

D. Tensorflow and Tensorflow Framework

In conversations, context is king! They are going to build a chatbot framework using Tensorflow and add some context handling to show how this can be approached. Tensorflow is an open source software library for high performance numerical computation. Its flexible architecture allows easy deployment of computation across a variety of platforms (CPUs, GPUs,TPUs), and from desktops to clusters of servers to mobile and edge devices. Originally developed by researchers and engineers from the Google Brain team within Google AI organization, it comes with strong support for machine learning and deep learning and the flexible numerical computation core is used across many other scientific domains.

TensorFlow is an open source library developed by Google for its internal use. Its main usage is in machine learning and dataflow programming. TensorFlow computations are expressed as stateful dataflow graphs. The name TensorFlow derives from the operations that such neural networks perform on multidimensional data arrays. These arrays are referred to as tensors. TensorFlow is available for Linux distributions, Windows, and MacOS.

E. TFlearn[4]

TFlearn is a modular and transparent deep learning library built on top of Tensorflow. It was designed to provide a higher-level API to TensorFlow in order to facilitate and speed-up experimentations, while remaining fully transparent and compatible with it. TFLearn features include Easy-to-use and understand high-level API for implementing deep neural networks,with tutorial and examples. Fast prototyping through highly modular built-in neural network layers, regularizes, optimizers, metrics...Full transparency over Tensorflow. All functions are built over tensors and can be used independently of TFLearn. Powerful helper functions to train any TensorFlow graph, with support of multiple inputs, outputs and optimizers. Easy and beautiful graph visualization, with details about weights, gradients, activations and more Effortless device placement for using multiple CPU/GPU.

The high-level API currently supports most of recent deep learning models, such as Convolutions, LSTM, BiRNN, BatchNorm, PReLU, Residual networks, Generative networks. In the future, TFLearn is also intended to stay up-to-date with latest deep learning techniques.

A chatbot framework needs a structure in which conversational intents are defined. One clean way to do this is with a JSON file. Each conversational intent contains:A Tag (a unique name),Patterns (sentence patterns for our neural network text classifier),Responses (one will be used as a response),later on well add some basic contextual elements. With their intents JSON file loaded,they can now begin to organize our documents, words and classification classes. They

create a list of documents (sentences), each sentence is a list of stemmed words and each document is associated with an intent (a class).The stem tak will match take, taking, takers, etc. They could clean the words list and remove useless entries but this will suffice for now. Unfortunately this data structure wont work with Tensorflow, we need to transform it further: from documents of words into tensors of numbers.

F. Dialogflow[5]

Dialogflow (formerly Api.ai, Speaktoit) is a Google-owned developer of human computer interaction technologies based on natural language conversations. The company is best known for creating the Assistant (by Speaktoit), a virtual buddy for Android, iOS, and Windows Phone smart phones that performs tasks and answers users question in a natural language.

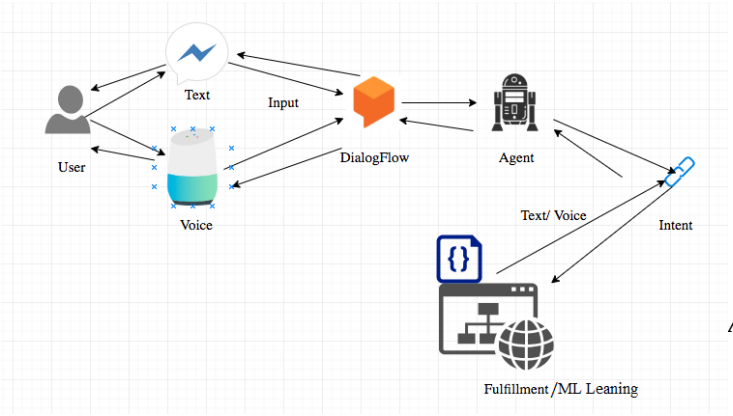

Figure 2

Text / Voice : The user interacts with an application like facebook messenger / Google home to start the interaction with the bot.

Dialogflow: Bot platform

Agent: A module within dialogflow which incorporates Natural Language Processing to understand what the user meant and to figure out what action has to be carried out. The agent transforms the user request into machine readable actionable data.

Intent: Support or the service that the user wants from the agent. Intent is configured by the developers. Intent determines the action by the code.

Fulfillment: This is the code. This part of the conversation lets you pass on the request from your bot to an external source and get response and pass it back to the user. This is achieved via Webhook. Setting up a webhook allows you to pass information from a matched intent into a web service and get a result from it.

G. TextBlob[3]

TextBlob is a Python library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more. TextBlob stands on the giant shoulders of NLTK and pattern, and plays nicely with both.

Feature:

- Noun phrase extraction
- Sentiment analysis
- Classification (Naive Bayes, Decision Tree)
- Language translation and detection powered by Google Translate
- Tokenization (splitting text into words and sentences)
- Word and phrase frequencies
- Parsing
- n-grams
- Word inflection (pluralization and singularization) and lemmatization
- Spelling correction
- Add new models or languages through extensions
- WordNet integration

III. Abbrevation and Objective

A. *Abbrevations*

1) NLP:Natural Language Processing
2) GPU:Graphics Processing Unit
3) CPU:Central Processing Unit
4) TPU:Tensor Processing Unit
5) LSTM:Long Short Term Memory
6) PReLU:Parametric Rectified Linear Unit
7) PReLU:Parametric Rectified Linear Unit

B. Objectives

1) To establish a friendly interaction with the human
2)To perform the tasks specified by the human such paying bills,maintaining remainder.
3)To communicate with the language the user is familiar.
4)For making appointments at a particular location in absence of human

IV. How to Chatbot working?

A. Chatbot will be working through 3 steps:

- Chatbot will transform conversational intent definitions to a Tensorflow model
- Next, chatbot will build a framework to process responses
- Lastly, Chatbot will show how basic context can be incorporated into our response processor

Chatbot will be using tflearn, a layer above tensorflow , and of course Python. As always Chatbot use ipython notebook as a tool to facilitate it work.

A chatbot framework needs a structure in which conversational intents are defined. One clean way to do this is with a JSON file.

Each conversational intent contains:

- a **tag** (a unique name)
- **patterns** (sentence patterns for our neural network text classifier)
- **responses** (one will be used as a response)

B. System Working

Front End: In this project the front end will be the Application which will be designed using Dialogflow. Human has to only one work that is speak or text. The rest of the processing will be done.

Back end: In the back end they will use chatbot will use python language as it is popular in machine learning,deep learning,IOT,etc. Their main concept will be Chatbot main concept will be NLP because using NLP,machine can analyze human speech,understand it and will give efficient result.

In chatbot first user have to manually enter all the details that is Bill no,credit card no,transaction id,which Will be stored permanently in database. Next time user have to just pronounce ‘Pay my light bill’ and the remaining tasks will be done by system.

Figure 3

ACKNOWLEDGMENT

We have great pleasure in presenting the report on AI Chatbot for Regional we take this opportunity to express our sincere thanks towards my guide Prof.Amol Kalugade Department of (Department of Computer Engineering), APSIT, Mumbai for Computer Engineering providing the technical guidelines and suggestions regarding line of work. We would like to express my gratitude towards his constant encouragement, support and guidance through the development of project.

We thank Prof.S.H.Malave Head of Department, (Computer Engineering), APSIT, Mumbai for his encouragement during progress meeting and providing guidelines to write this report.

We also thank the entire staff of APSIT, Mumbai for their invaluable help rendered during the course of this work. We wish to express our deep gratitude towards all my colleagues of APSIT, Mumbai for their encouragement.

REFERENCES

[1] Youngmin Park 1 , and Sangwoo Kang “Natural Language Generation using Dependency Tree Decoding for Spoken Dialog Systems”Gyeonggi-do 16082, Korea,2017

[2] Tensorflow:-https://www.tensorflow.org/

[3] Textblob:-https://textblob.readthedocs.io/en/dev/

[4] Tflearn:-http://tflearn.org/

[5] Dialogflow:-https://medium.com/swlh/how-to-build-a-chatbot-with-dialog-flow-chapter-4-external-api-for-fulfilment-3ab934fd7a00