

Chatbot Personal Assistant Using Natural Language Processing (NLP)

¹Asha Gaud, ²Bhumi Mota, ³Dhananjay Kumbhar, ⁴Veer Kumar, ⁵Prof. Shashank Gothankar
^{1,2,3,4} Student, and Telecommunications Department, MCT Rajiv Gandhi Institute of Technology, Mumbai, India

⁵Professor, Electronics and Telecommunications Department, MCT Rajiv Gandhi Institute of Technology, Mumbai, India

Abstract—The modern era of technology has a significant impact on society. With the creation of digital personal assistants, chatbots have become a popular entity in conversational services. Chatbots are software programs that use natural language understanding and processing. Chatbots are not just restricted to helping the user to complete his tasks such as booking a movie ticket or finding the nearest restaurant, but they also provide a source of entertainment, play a major role in home automation projects, and give business strategy tips and help in other ways. In this paper, we will provide an insight into what a chatbot is and the types of chatbots. We also propose a classification based on the current market trends, ease of usability, and requirements.

Index Terms-Chatbot, NLP, Artificial Intelligence, Machine Learning, Natural Language Processing, Ryan.

INTRODUCTION

A chatbot is a computer program that uses Natural Language Processing, a subset of AI to communicate or interact with users. It is a the next massive aspect of conversational services. They are virtual individuals that can discuss with any human being effectively. The usage of chatbot systems as a medium of conversation is very common for most of the companies, organizations and educational institutions these days. The main motive is to simulate human conversation through text or voice using Natural Language Processing (NLP). It can understand one or more human languages by using NLP. The structure of the bot integrates computational algorithms and language model to build casual chat which covers enormous NLP techniques. Chatbots are being

used is several industries these days for information delivery, performing tasks, weather prediction, flight reservations, answering educational queries, or product purchases. The bot is customized for the queries of the college and it also gives an overview of the college. A few examples for talking computer systems today are Apple's Siri, Google Voice Assistance by Google and other big companies. This concept first came into picture in 1960s. Conversational bots are lately relying on applying deep learning strategies. Chatbots have become popular lately as good-sized use of message offerings and advancement in NLU. NLP People communicate with each other through text and speech. Human way of exchanging information is called as Natural Language. Every day we share a large quality and quantity information in several languages. However, computers cannot decipher this information as its in natural language and computers understand and communicate in 0s and 1s. The information delivered is valuable and can offer profitable bits of knowledge. Hence, we would like the computers to be able to understand it and respond intelligently and rationally.

Natural Language Processing or NLP refers to the branch of AI that provides the machines the power to read, understand and derive meaning from human languages. NLP has its roots to early 1950s. NLP combines the sector of linguistics and computing to decipher language structure and guidelines to form models which may comprehend, break down and separate significant details from text and speech. Natural Language Processing (NLP) is a subfield of computer science,

linguistics, and AI that is concerned with interactions between computers and humans, especially how to program computers to process and analyze large amounts of natural data. We aim to obtain a computer capable of "understanding" the contents of message or document, including the contextual nuances of the language. It can extract information from the documents and categorize them. Most of the NLP techniques rely on machine learning techniques to derive the corresponding meaning from human language.

However, chatbots have a big potential for implementation in personalized corporate accounts on every social media platforms. Chatbots are designed to be the ultimate virtual assistant, helping you to accomplish various tasks ranging from answering questions, getting driving directions, turning up the thermostat in your smart home, or play your favorite tunes and even draft a question and answer survey. Chatbots are being made to ease the pain that the industry is facing today. The purpose of chatbots is to support and help to scale business teams in their relationship with customers and consumers at the same time.

In literature [3], author has implemented the chatbot who helps student. The project is about interaction between users and chatbot which can be accessed from anywhere anytime. The chatbot can be easily attached with any university or college website with few simple language conversions. Chatbot provides various information related to university or college and also students-related information. In literature [4], author has implement the Affective computing explores the development of systems and devices that can perceive, translate, process, and reproduce human emotion. It is an interdisciplinary field which includes computer science, psychology and cognitive science. An inspiration for the research is the ability to simulate empathy when communicating with computers or in the future robots. This paper explored the potential of facial expression tracking. The developed chatbot summarizes emotional state of the user during the survey through percent-ages of the tracked facial expressions throughout the conversation with the chatbot. Facial expression tracking for happy, neutral, and hurt had 66.7%. Moreover, the developed program was tested to track expressions simultaneously per second. It can track 17 expressions

with stationary subject and 14 expressions with non-stationary subject in a span of 30 seconds.

In literature [5], Author has implemented the Machine learning based chatbots yields more practical results. Chatbot which gives responses based on the context of conversation tends to be more user friendly. The chatbot we are proposing demonstrates a method of de-veloping chatbot which can follow the context of the conversation. This method uses TensorFlow for developing the neural network model of the chatbot and uses the nlp techniques to maintain the context the conversation. This chatbots can be used in small industries or business for automating customer care as user queries will be handled by chatbots thus reducing need of human labour and expenditure.

In literature [6], we have seen that, the jollity chatbot is implemented in Rasa, an open-source conversational AI framework and it is easy to customize. The proposed method has added 12 intents with each more than 8 text examples constituting a total of 100 input samples in nlu.md and their response in domain.yml. The flow of interactions is given in stories.md. The jollity chatbot is deployed in Telegram using ngrok and the server URL details and the access token are given in the credentials.yml. The system is experimented with various evaluation measures like accuracy of the intents, accuracy of the stories and the confusion matrix to show that the proposed jollity chatbot system is more robust and can identify the user intents appropriately.

1.1 Natural Language Processing:

People communicate with each other through text and speech. Human way of exchanging information is called as Natural Language. Every day we share a large quality and quantity information in several languages. However, computers cannot decipher this information as its in natural language and computers understand and 566 International Journal of Research Publication and Reviews Vol(2) Issue (4) (2021) Page 565-574 communicate in 0s and 1s. The information delivered is valuable and can offer profitable bits of knowledge. Hence, we would like the computers to be able to understand it and respond intelligently and rationally.

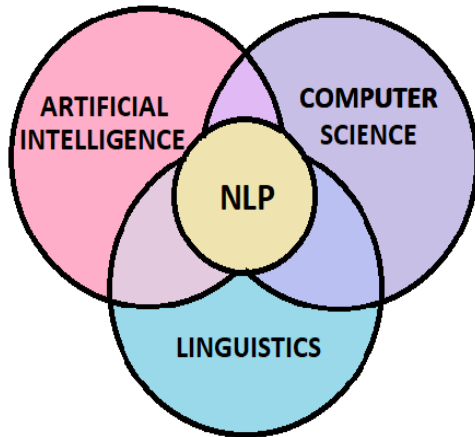


Fig 1. About NLP[1]

Natural Language Processing or NLP refers to the branch of AI that provides the machines the power to read, understand and derive meaning from human languages. NLP has its roots to early 1950s. NLP combines the sector of linguistics and computing to decipher language structure and guidelines to form models which may comprehend, break down and separate significant details from text and speech. Natural Language Processing (NLP) is a subfield of computer science, linguistics, and AI that is concerned with interactions between computers and humans, especially how to program computers to process and analyze large amounts of natural data. We aim to obtain a computer capable of "understanding" the contents of message or document, including the contextual nuances of the language. It can extract information from the documents and categorize them. Most of the NLP techniques rely on machine learning techniques to derive the corresponding meaning from human language.

2. RESEARCH METHODOLOGY

The methodology section outline the plan and method that how the study is conducted. This includes building the dataset, sample of the study, Data and Sources of Data, study's variables and analytical framework. The details are as follows;

2.1 Speech recognition

It is using speech recognition for registering user input using the microphone. Converting it into text form, Searching for its answers from the processed Corpus, and returning the output using text-to-speech. It will

continue taking user input and answering until the user says Bye/Goodbye.

2.2 Usage of Wake word

Personal assistant devices are one of the main use cases for Speech to Text technology for main stream users, not including accessibility tools like screen readers. "Wake words" engage devices to process what they hear, like "Hey Google" or "Alexa", often sending it to the cloud if a connection has been established. Here, in our project we have used the wake word 'Ryan' in order to invoke our chatbot, which would then be ready to take user queries and other interactions.

2.3 Program Flow

For this chatbot, we have used If-Else control statements to parse through various functionalities of the chatbot. Here, at each stage of the control loop statements we have a possible user intent, and if the speech recognition input matches the pre-set user intents, the chatbot will respond accordingly to satisfy the user query.

2.4 Intent Classification and Entity Extraction

While working on this project, we understood that every user query is unique and no two user inputs may sound exactly the same even if they intend to ask the same question. If we go ahead with preset keywords to recognise intents, then we our chatbot won't be optimised to cater to all kinds of variations in user queries and for this reason we require an Intent classification and Entity extraction mechanism that can handle wide range of user queries for the same questions.

2.5 Dataset

For this project, we have built our own training data to suite the kind of user queries we intend to target. This training data is basically a Comma Separated Value(CSV) file which has patterns of intents and corresponding tags that trace bck to their entities in the main function of our program. Using large number of examples for training data, we have achieved good accuracy on identifying user intent accurately.

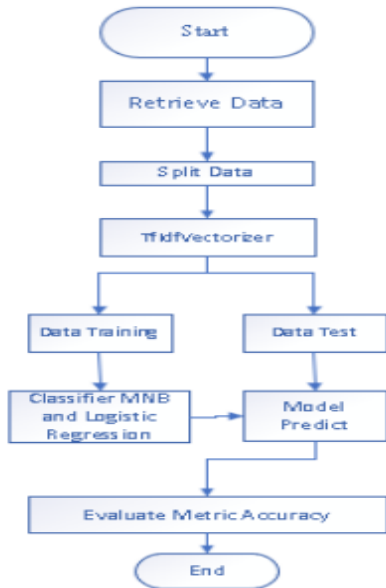


Fig 2. Project Flowchart

3. ALGORITHM

Multinomial Naive Bayes algorithm is a probabilistic learning method that is mostly used in Natural Language Processing (NLP). The algorithm is based on the Bayes theorem and predicts the tag of a text such as a piece of email or newspaper article. It calculates the probability of each tag for a given sample and then gives the tag with the highest probability as output. It is based on the following formula:

$$P(A|B) = P(A) * P(B|A)/P(B)$$

Where we are calculating the probability of class A when predictor B is already provided.

P(B) = prior probability of B

P(A) = prior probability of class A

P(B|A) = occurrence of predictor B given class A probability

This formula helps in calculating the probability of the tags in the text.

3.1 Advantages of Naïve Bayes Theorem

- It is easy to implement as you only have to calculate probability.
- You can use this algorithm on both continuous and discrete data.
- It is simple and can be used for predicting real-time applications.

- It is highly scalable and can easily handle large datasets.

4. RESULTS AND DISCUSSION

From the evaluation results of the experimental classification model to determine the intent classification on chatbot, the accuracy of the Multinomial Naive Bayes model is 75.43%, and the Logistic Regression model is 54.22%. This shows that there is an accuracy distance between the two classification model with logistic regression which is more accurate at 75.43. In this study, we still use training data with a reasonably limited amount, the amount of data is not the same in each class of intent, so it is possible that errors will occur when predicting class intent with less training data. The accuracy of the two methods will experience differences in accuracy distance when the training data for each intent class has the same amount of data in each class. We have found a decrease in the accuracy of the logistic regression model when the intent class had the same amount of data in each class.

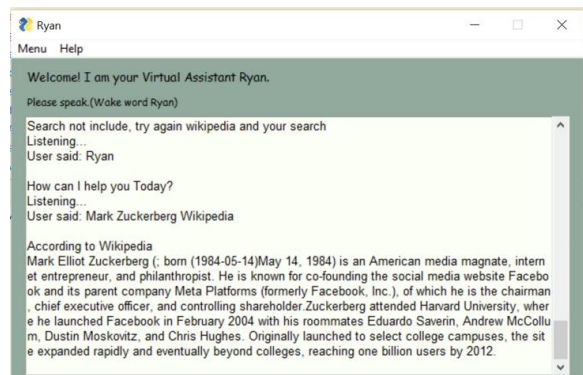


Fig 3. Invoking the Chatbot using Wakeword 'Ryan'

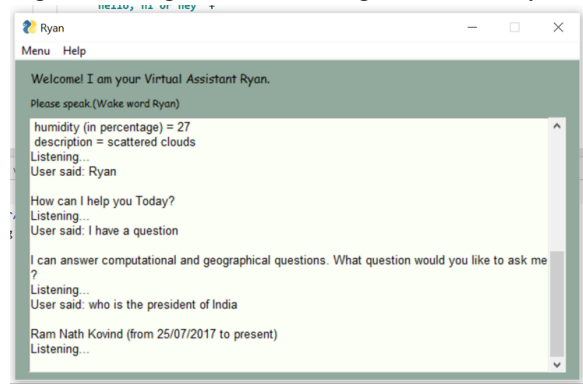


Fig 4. Output generated by Chatbot using Intent Classification

5. CONCLUSION

In this project the personal assistant chatbot we created is essentially a digital virtual assistant, and one of the major advantages of having such a digital assistant is that we get to perform hands-free operations that would otherwise require us to type commands or requests into a user interface. Second, the chatbot we created can be customised to suit user specific needs, we can add a particular user intent, by just adding a couple of lines of code and by training the machine learning model on a set of example statements pertaining to that specific user intent. Moreover, a conversational chatbot like Ryan can perform repetitive tasks with much more ease and efficiency than humans can, thereby, saving its users both time and effort to do a particular job. Despite, all these advantages, we as a team are aware of some of the limitations of our conversational chatbot, for instance, through several conversation trials with the chatbot, we as a team have realised that the speech recognition tool we are currently implementing struggles to understand regional English accents and may not yield the intended results with such users. This is a subject of research in the NLP field today and Baidu, a Chinese search engine company, is developing a 'deep speech' algorithm that is meant to recognise different accents and could be adopted by voice technology in the future. We must embrace the challenge of making chatbots understand language variations so the barriers of miscommunication are broken down. It will further a technology that will continue to go from strength to strength[8].

6. ACKNOWLEDGMENT

We extend our sincere thanks to honorable principal Dr. Sanjay Bokade, Principal of Rajiv Gandhi Institute of Technology, Mumbai, for allowing us to use college facilities like the Computer laboratory and the Central Library; these resources gave us the optimum environment to conduct project related work. Furthermore, we would like to thank our supervisors Prof. K.G Sarvarkar, Dept of EXTC and Prof. S.V Kulkarni, Dept of EXTC for guiding us through the course of this project.

REFERENCE

- [1] About NLP. (2017, January 12). [Photograph]. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2017/01/ultimate-guide-to-understand-implement-natural-language-processing-codes-in-python>
- [2] Aishwarya Gupta, Divya Hathwar , Anupama Vijayakumar, 2020, Introduction to AI Chatbots, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 07 (July 2020),
- [3] Drishti Malik Vibhor Sharma, Monika Goyal, "An intelligent behaviour shown by chatbot system", International Journal of New Technology and Research, 2017.
- [4] BalbirSinghBani,AjayPratapSingh, "College Enquiry Chatbot Using A.L.I.C.E", International Journal of New Technology and Research 2017.
- [5] Nitesh Thakur, Akshay Hiwrale, Sourabh Selote, Abhijeet Shinde, Prof. Nam- rata Mahakalkar, "Artificially Intelligent Chatbot", Department of Computer Science Engineering, Priyadarshini Institute of Engineering Technology, Nagpur.
- [6] Dahiya, "A tool of conversation: Chatbot", Dept. of Computer Science, Maharaja Surajmal Institute, Janakpuri, India, 2017.
- [7] Stöckl, Andreas. (2017). Classification of Chatbot Inputs. 10.13140/RG.2.2.28189.61920.
- [8] Williams, S., 2022. Why doesn't my voice assistant understand my accent? [The Knowledge. [online] The Knowledge. <https://www.b2bknowledge.com/2019/05/03/why-doesnt-my-voice-assistant-understand-my-accent/>>