38

A Survey on NLP Chatbots

Prathamesh Tambe

S.Y MSC (C.S), Department of Computer Science, Maharashtra Institute of Technology

tambeprathamesh16@gmail.com

Payal Kawade

S.Y MSC (C.S), Department of Computer Science, Maharashtra Institute of Technology
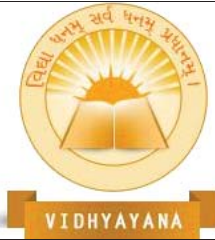
Payalkawade74@gmail.com

Charan Shekar

S.Y MSC (C.S), Department of Computer Science, Maharashtra Institute of Technology

Abstract-

Natural Language Processing is a branch of Artificial Intelligence that helps machines to understand and interpret human language. An NLP-based chatbot is a software program based on artificial intelligence system that interacts with a client through written or auditory means. These programs are frequently developed to assist clients on websites. Many domains like E-commerce, healthcare or customer services have utilized chatbots to improve their communication with customers. The system uses keywords to identify the context of the conversation. NLP enables chatbots to interpret, evaluate, and prioritize questions based on their complexity, allowing bots to reply to consumer enquiries faster than a person. Faster answers aid in the development of consumer trust and, as a result, greater business.

Index Terms- Natural Language Processing (NLP), Natural Language Toolkit (NLTK), Natural Language Understanding (NLU), Chatbot.

Volume 8, Special Issue 7, May 2023
4th National Student Research Conference on
"Innovative Ideas and Invention in Computer Science & IT with its Sustainability"

Page No. 520

I.   INTRODUCTION

Human beings have become the most advanced species on the planet and made so much technological progress because of our ability to communicate and share knowledge. The medium we use to communicate is called Language. Our interaction with computers, however, was limited to the Graphical User Interfaces provided by the operating systems. In recent times, we have started using language even to communicate with computers. We type questions into our Search Engines, we talk to our smartphone assistants, this has all been possible due to continuous research in the field of Natural Language Processing. An NLP system needs to perform at least one of two tasks i.e., Natural Language Understanding, understanding the context of the human language and Natural Language Generation, generating a response.
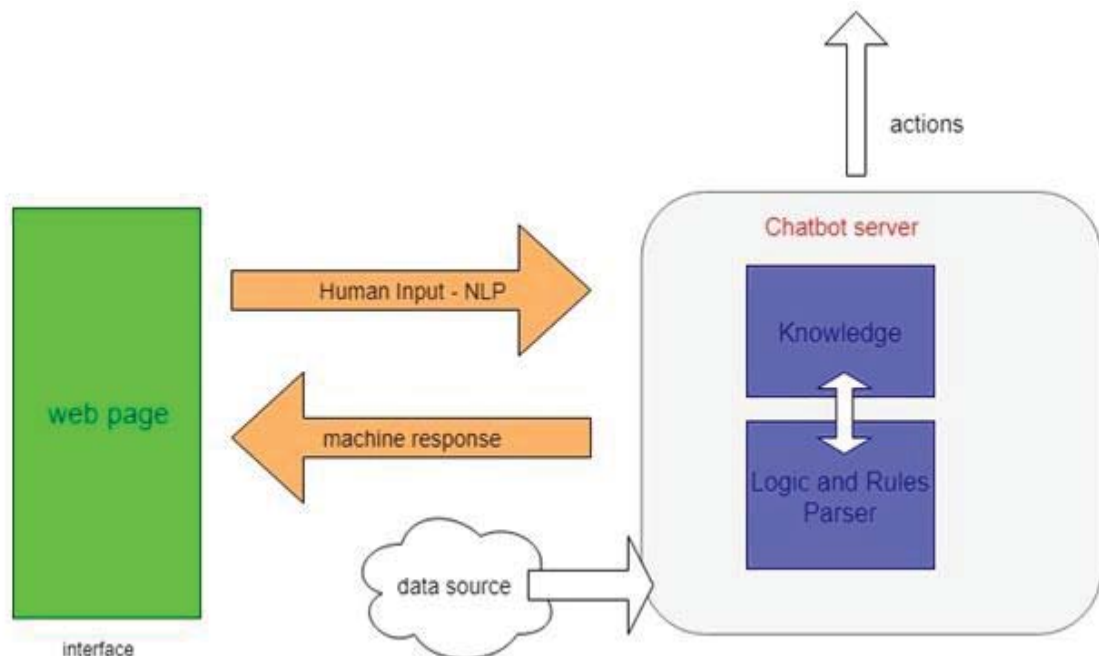
A chatbot is a more complex system that must perform both understanding user input and then generating an appropriate response. This provides us a simplified medium to interact with computers. In countries where the work force is limited, chatbots provide a simpler alternative to solve customer queries.

To understand the customers input, the chatbot must transform unstructured human language into organized data that machines can decipher. When a client delivers a text to the chatbot, it must utilize algorithms to extract context and meaning from each line in order to gather data.

Once the task of understanding the user input is done, the chatbot now has to generate appropriate response. The task of natural language generation is complex because it includes constructing an entire sentence from the context of the response. Many Deep Learning algorithms that work on sequential data like RNNs can be employed for this task. However, an RNN does not store context. An LSTM model that has a cell state that can store the context and generate more accurate response may be efficient in applications where context of the response is important. The response thus can be a predefined output or a text sentence generated from the knowledgebase using these deep learning models.

The progress in the field of NLP has led to more accurate chatbots which are being applied in a variety of sectors in industry as well as our everyday life.

Volume 8, Special Issue 7, May 2023
4th National Student Research Conference on
"Innovative Ideas and Invention in Computer Science & IT with its Sustainability"

Page No. 521

The figure 1 shows generalized chatbot architecture.

The communications platform used is a web page. The webpage will take the customers input query and provide the reply to their inquiry. When the web page receives the message, it sends it to chatbot server. The chatbot searches a predefined database for the right response and sends it back to the web server.
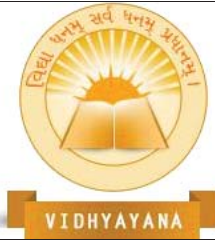
II. IDENTIFY, RESEARCH AND COLLECT IDEA

1.NLP Applications

Natural Language Processing may be used in a variety of applications, including machine translation, email spam detection, information extraction, summarization, and question answering.

2. Spam Filtering

The process of text classification uses various machine learning techniques, such as Rule Learning, to categorize text and filter out spam. One such technique, the Multi-variate Bernoulli model, considers the presence or absence of specific words in a document, regardless of their order or frequency.

Volume 8, Special Issue 7, May 2023
4th National Student Research Conference on
"Innovative Ideas and Invention in Computer Science & IT with its Sustainability"

Page No. 522

3. Information Retrieval

The retrieval of data is focused with locating terms of relevance in textual material. Extracting things such as names, places, events, dates, times, and prices is an effective means of summarizing relevant information to a customers' requirements in many applications. NLP employs a number of approaches to extract valuable information from primary sources, such as Parts of Speech (POS) tagging and stop-words and stemming.
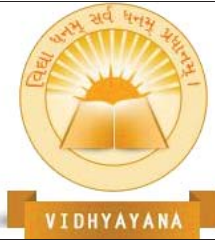
4.Summarization

In our digital era, data overload is a real phenomenon, and our range and access to information and knowledge already beyond our capacity to grasp it. This tendency is not losing momentum, therefore the capacity to summarize data while maintaining meaning is essential. This is essential because it not only allows us to detect and comprehend the relevant information in a huge number of facts.

III. WRITE DOWN YOUR STUDIES AND FINDINGS

In [1], an approach to replace standardized queries with a natural language based chatbot to allow users to easily get the database output has been presented. The system is based on automatic synthesis of commands from the natural language input given by the user. The chatbot is designed using Xatkit chatbot model on the Xatkit runtime engine.

Web based voice chat models [2], have also been very popular as they increase the ease of use of a particular service. A specific web service framework to implement the chatbot is created with a blackbox model so that the web server can respond to any type of user requests. The web framework is assisted by an nlp based intelligent model that takes in the client input in voice format, extracts the text and converts the text into a web request which is given to the server. The web framework allows seamless processing of XML.

Chatbots have also been appleid to learning a certain language. [3] explores a Computer Simulation in Eductaional Communication (CSIEC) based model which acts as a texting partner focusing on the teaching of the English Language. A naive approach of semantic and syntactical analysis of the textual input is used and the inference is generated by logical reasoning. The chatbot stores a context of the users wordbase, personality and common-sense knowledge together to form a natural lamguage for the chatbot.

Volume 8, Special Issue 7, May 2023
4th National Student Research Conference on
"Innovative Ideas and Invention in Computer Science & IT with its Sustainability"

Page No. 523

Healthcare is a field where the application of chatbot can do wonders. An NLP based healthcare chatbot [4] can analyse the keywords in the text input to get symptoms and provide basic diagnosis. It can also work as a reference to keep a user healthy. A knowldegebase of keywords needs to first be created for the chatbot to be able to extract symptoms. The chatbot uses NLTK to analyse input and an NLP based engine to provide appropriate response.

In [5], a dedicated, simple and interactive chatbot that will suggest medicines and dosage according to age and symptoms is implemented. The chatbot takes in the user query in the form of a question. The NLP model will extract the keywords from a knowldegebase of medical text. The chatbot will suggest a medicine and dosage to the user in response using machine learning. The model can also predict disease based on the symptoms and even suggest doctors nearby.

[6] The chatbot is developed to check for symptoms and provide a basic solution or book an appointment to a nearby doctor thereby reducing the load on the telephone operator. The major challenge is to analyse user input and provide appropriate response.

In simpler terms, NLP technology improves the way we communicate with machines by enabling them to understand and generate language in a more natural and intuitive way. NLP has been applied to a variety of fields in the past few decades. A large number of these applications are very useful in daily life for example a machine that takes instructions by voice which is a more natural way of communication for humans. There has been a lot of research on this topic to develop more useful and practical systems. Natural Language Processing has shown great promise to develop computer interfaces that are simpler to use for people so that the sharing is seamless and natural rather than a set of defined rules which the computer understands. For programming, however, the importance of a formal high level programming language for interacting with a computer has always been taken for granted.

However, with the rise of social media, the natural language has also become less constricted with rules. A prime example is the use of twitter to express sentiments in 140 characters which has led to an increased ambiguity. [8] explores the use of Stanford CoreNLP toolkit to extract keywords from a tweet. It consists of two stages, in the first stage, the model parses a
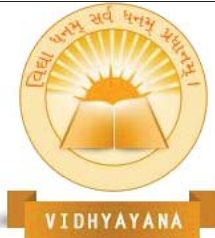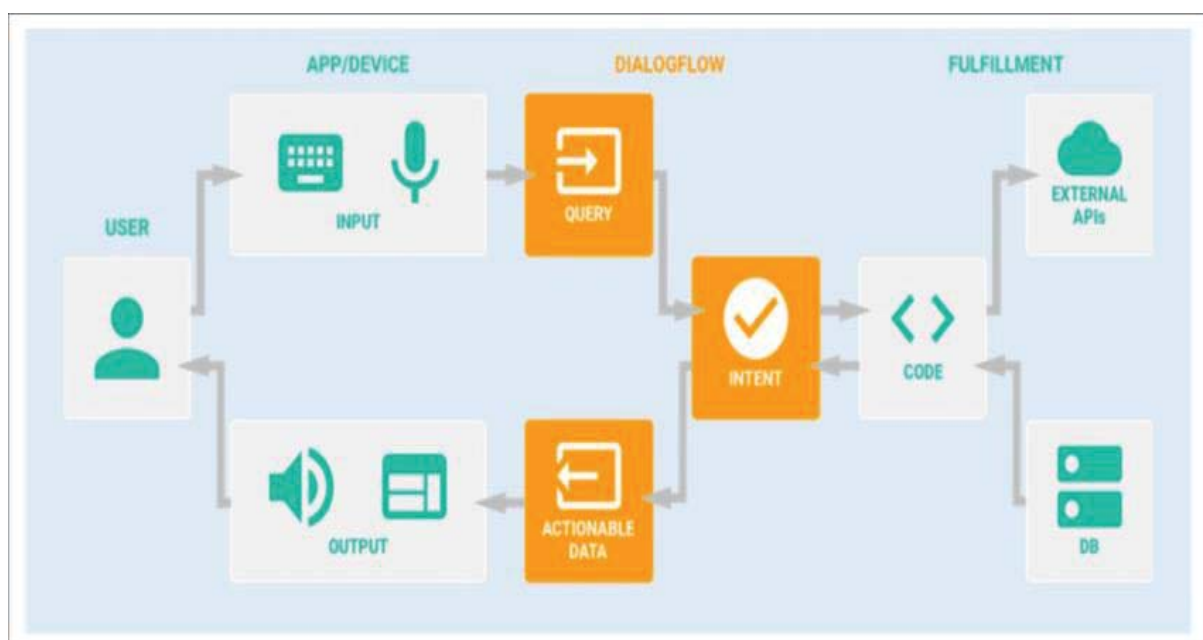
corpus of tweets from the profile of a telecommunication company and uses it to identify keywords and assign POS tags to the keywords. In the second stage, the approach uses Name-Entity recognition and lemmatization to get the root word. The performance of the entire model was validated using the Turing test.

[9] discusses the future of how NLP can be applied to various systems to make human machine interaction more efficient and reduce the problems of medium and interface to allow people to interact with machines in a natural and even a less rule constricted language.

\Section {Proposed Methodology}

Modern chatbots need to interact with the user using a more natural way of communication which is the human language. People interact with each other using language and the same can be implemented to make interactions with machines more user friendly. The chatbot needs to have an interface with which it will interact with users. A virtual assistant uses a mobile phone as its interface, some voice assistants have dedicated hardware that takes in user voice input. Irrespective of the interface, the chatbot would take in voice as its input. The input command can be in any language. So, the chatbot needs to be integrated with a translator software which would convert one language text into the basic language of the chatbot. Fig 2. shows the entire process flow of a voice-based chat assistant.

Volume 8, Special Issue 7, May 2023
4th National Student Research Conference on
"Innovative Ideas and Invention in Computer Science & IT with its Sustainability"

Page No. 525

The voice input is converted into the text of the base language of the chatbot system. An NLP model is implemented for language conversion. Once the text is obtained, tokenization, lemmatization, POS tagging and other NLP techniques need to be implemented on the text to obtain the keywords of the command. The keywords will form the request query. Based on the query, the appropriate action needs to be taken. The intent of the keyword defines the context of the human command. The query processes the request and generates a response.

The response now needs to be given back to the user. If the response is a task, like setting up an alarm or playing music, the task can be performed. However, if the response is a chat, a text in the natural language needs to be generated. This is where Deep Neural Network come in. They predict the next word in the sentence from the previous work and context. An NLP model trained on a corpus of the language assisted by the deep neural network is used to generate the text in the base language. The text can then be given back to the translator model to be converted to the original language in which the user had given the comma.

IV. CONCLUSION

Chatbots have been implemented in a variety of sectors be it education, industry, customer service, healthcare and many more. As humans are changing the way they exchange information, the medium of communication with machines also needs to improve. A use of natural language to interact with machines instead of the present graphical user interfaces and strict constricted commands will provide more ease of use and make machines more user friendly. Chatbots range from text-based chat systems to smart voice-based assistants that are coupled with search engines and home automation systems truly define the modern way of interacting with machines. As people become more used to chatbots and voice assistants, NLP and AI models on which these systems are based also need to be more accurate and efficient. They need to properly understand the context of the user command even though the human language is ambiguous and perform the correct action to the command. Such systems are validated using Turing tests. Applying these chatbots to medical field can provide basic healthcare facilities at the tip of the user's fingers. Chatbots in customer service reduce the requirement of human labour and thus reduces the cost to company. NLP has constantly seen a lot of improvement supported by deep learning models to improve the accuracy of chatbot
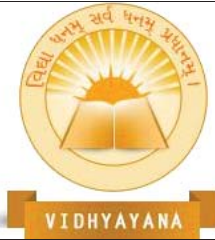
systems.

REFERENCES

[1] Eko Handoyo, and M. Arfan. "Ticketing Chatbot Service using Serverless NLP Technology." In 2018 5th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE), Semarang, Indonesia, 2018, pp. 325-330.

[2] Harsh Lal, Priyanshu Lal. "NLP chatbot for Discharge Summaries." In 2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCI), Jaipur, India, 2019, pp. 250-257.

[3] Lukas Tommy, Chandra Kiran, Leo Riska. "The Combination of Natural Language Processing and Entity Extraction for Academic Chatbot." In 2020 IEEE International Conference on Cyber and IT Service Management (CITSM), 2020, pp.1-6.

[4] Udayakumar Shanmugam, Sowjanya Mani, Sneha Shivakumar and Rajeshwari P. "Human-Computer text conversation through NLP in Tamil using Intent Recognition." In 2019 IEEE International Conference on Vision Towards Emerging Trends in Communication and Networking (VitECoN), Chennai, India, 2019, pp 1-5. 10.1109/IJCNN.2018.8489629.

[5] Sara Perez-Soler., Gwendal Daniel, Jordi Cabot. "Towards Automating the Synthesis of Chatbots for Conversational Model Query." In Springer BPMDS 2020/EMMSAD 2020, LNBIP, 2020, pp. 257-265, Switzerland.

[6] S. du Preez, M. Lall, and S. Sinha, "An intelligent web-based voice chatbot," in EUR OCON 2009, EUROCON'09 IEEE, 2019, pp. 386–391.

[7] Jiyou Jia, "A Computer assisted English learning Chatbot based on textual Knowledge and Reasoning (CSIEC)" in Elsevier KnowledgeBased Systems, Volume 22, issue 4, May 2009.AUTHORS

Volume 8, Special Issue 7, May 2023
4th National Student Research Conference on
"Innovative Ideas and Invention in Computer Science & IT with its Sustainability"

Page No. 527