

Medical Chatbot based on NLP

Manan Malhotra, Student, MIT School of Engineering, malhotramanan99@gmail.com

Harsh Tiwari, Student, MIT School of Engineering, ankitaharsh056@gmail.com

Manpreet Singh, Student, MIT School of Engineering, mr.singh29898@gmail.com

Abstract A chatbot is considered to be an intelligent robot that can have a conversation with a person in real time using natural language. Many companies now use online tools to manage customer inquiries as a result of the rise in Internet usage, and many of them are turning to chatbots to improve customer support and increase their productivity. Healthcare is such a field that needs a solution like this considering the pandemic. The presented solution make use of NLP techniques such as TF-IDF and cosine similarity to extract the keywords to match with similar queries from database and then provide appropriate reply based on that.

Keywords —Chatbot, cosine Similarity, Healthcare, keywords, natural language, NLP, TF-IDF.

INTRODUCTION

“Health is Wealth” is said for a reason and we all have acknowledged it recently. After living through the pandemic we have seen how important health is and why should we be concerned about it. Nowadays, health is very important as people are busy with their work at home or office. People have been avoiding going to the hospital due to the fear of pandemic.

During a recent research it was observed that the population of India would be at its peak by 2050 because of the exponentially increasing birth rate and decreasing death rate because of the advancements in the medical field. Accordingly, in a nation where ratio of healthcare workers is already small, this could become a bottleneck in the future with respect to medical services. According to the research approximately 60% of the individual visiting specialist have issues such as cold, cough, fever, etc. These can be self managed with the assistance of a chatbot.

Nowadays, most of the business models have adapted chatbots are using them be it travel industry, educational

institutes, e-commerce websites, etc. These have been demonstrating to be very effective as they have reduced the load on the customer support executives as majority of the queries that clients have are policy related or status of their requests which can be resolved with the help of chatbots. There are several advantages of deploying chatbots one of them is 24X7 support to the clients. They can get their queries answered even in the middle of the night.

With increase in the workload it is not convenient for everybody to visit doctors to get their queries answered every-time. And with advancement in technologies Artificial intelligence has given us the power to imitate the behaviour of human brain. With the increase in the computation powers and advancements it has become easier to deploy the chatbots and make them effective. Also comparing it to the customer support module of companies a company might require a 1000 executive but 1 chatbot could suffice the purpose of them. This makes chatbot the preference for the companies.

The proposed idea is to create a chatbot system using artificial intelligence that can answer user queries. Using the chatbot the user could get essential information about the diseases or the could also get details about disease using symptoms. The system works on question and answer module where user can input his queries and the chatbot would answer the users queries. The important keywords are separated from characteristic language and at that point if a match is found the critical answer would be given.

I. RELATED WORK

A. Literature Survey:

In this study, the main focus is on the emotion recognition, this is based on the deep learning, CNN. The emotion classification model is trained using the above methods. NLG and NLP are used for the purpose of understanding and user input and interaction. [5]

Here, a speech recognition chatbot is programmed to give answers to the asked questions. If the bot fails to understand the question then it is put to the third party expert system. This is designed as a friendly web based chatbot. The main focus of the study was not only text based but also voice based. In this, speech recognition demands a divided system with functionalities for gathering and processing input signals. The server used in this is SOAP based on the black box system.[6]

The intended target of the chatbot is to have a real time interaction with human being. In the given study the computer stores knowledge to describe a sentence and to make the decision to answer the question. The said knowledge us stored in the RDBMS.[7]

In this study, the implemented chatbot is based on the sequence comparison that understands the sequence in which sentences are and entered and also the response pattern of the sentences. The study describes about the input and output of the system. This has been explicitly created for entertainment purpose.[8]

Here, the study shows the implementation of the n-gram method to extract terms from sentences. In this n-gram is used to arrange and also reduce the input using Moro phonemes and phonemes as the final parameter. The final

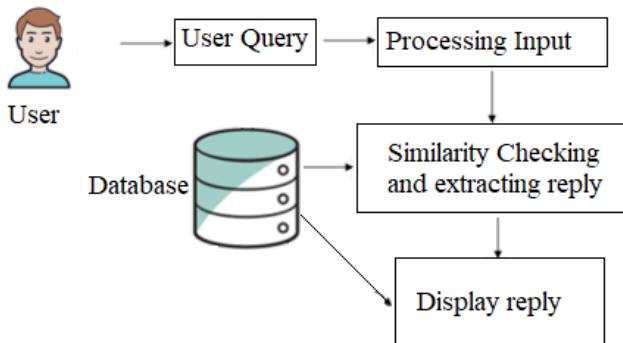
statement is redirected to expert system. The main principle is pattern matching and the dataset is clustered using the k means clustering algorithm.[9]

In this study the chatbot is designed for the healthcare domain in android. The google api is used here. In this the user sends a voice or text message then user gets the appropriate response from the chatbot. In this the SVM algorithm is used to classification of dataset and the proper algorithm is used for the purpose of discarding redundant words.[10]

This is study that has created a document similarity checking model which is based on the tf-idf algorithm. The number of documents that are present online are checked using n-gram and then the cosine similarity is calculated.[11]

The chatbot in this study is developed for customer service, which works as a health service. This model uses a contrast of N-gram, Tf-IDF and cosine similarity. A database is created to store the questions and answers. This method shows very promising results. This is study shows a contrast between unigram, bigram and trigram.[12]

In the given study the chatbot is created for the healthcare domain with the help of AI. This shows a web based solution with the purpose of giving patients response to



their queries by making use of pattern matching techniques.[13]

In this, a bot is implemented using the raspberry pi which works with the mobile application which can be connected using the bluetooth. This bot is capable to move. This particular bot is based on the AI and makes use of Support Vector Machine algorithm in the background. This is a form of scripted chatbot that has been combined with prediction algorithm for disease prediction.[14]

The study, smart doctor is also a chatbot that has been developed. The approach used for development in this is using decision tree algorithm from artificial intelligence. A prototype had been developed for 100 diseases. The principle is a question and answer approach. Although the method gave satisfactory result but this method cannot be considered for chatbot because of the computational complexity.[15]

In this study, model used for the development is based on the SVM algorithm from artificial intelligence. In this study various other methods such as KNN and Naives Bayes were also explored. The results obtained from the research are promising and could prove beneficial. This study makes use of google api for text to voice and voice to text. Medicine api is also being used for giving medical diagnosis.[16]

B. Research Gaps

The chatbots deployed by most of the websites such as telecom and marketing for customer service are scripted, i.e., they are based on rigid syntax. This is the main reason due to which while interacting with the chatbot mostly we end up with “Sorry I didn’t quite get what you said.” Which indicates there is no response set to question asked by the user. A lot of research is being carried out in making the conventional chatbots to be more communicative and interactive. There are multiple ways in which this can be achieved; selection of the appropriate method is done considering the domain in which the chatbot is going to be deployed. Making communication interactive majorly requires NLP and ML techniques in the system. Also considering the seq-seq model for chatbots it requires a lot of computation power and time, and yet the results are satisfactory only for 5 word sentences. All the above-mentioned issues need to be considered while implementing a chatbot system.

II. PROPOSED SYSTEM

In our proposed system the user can chat with the bot regarding the queries through text. This system is designed as an expert system that can answer the users health related queries. User can also ask about their health and information for that particular disease. The system can be used by doctors also to help their patient better understand about the disease. The data of the chatbot stored in the database. Bot will resolve users queries related to health and specialists doctors to consult apart from this the chatbot will also be capable of having general conversation with the user. Bot will also be able to suggest analgesics for minor health issues.

III. SYSTEM ARCHITECTURE

The system implemented in the study is chatbot based on the Tf-idf and cosine similarity. In which the user queries are taken as documents then the Tf-idf is calculated for that and then compared with the tf-idf scores of the documents from the dataset and then document with the highest score is given as the reply. Cosine Similarity is used to calculate the similarity between two documents. This backend has been developed in the python. Whereas the GUI is being developed using HTML and CSS. Also to integrate these Flask framework is being used.

IV. ALGORITHMS

We are using two algorithms to implement chatbot system.

1. TF-IDF(Term frequency-inverse document frequency)
2. Cosine Similarity algorithm

1. TF-IDF

Term frequency(tf):

Usually, while designing a model with the intention of understanding text, it is observed that all the stop words are

removed. A different approach is to check the related importance of words using Tf-idf.

This can be calculated by the total number of times the word appears in the document by the total number of words in the document. All the documents have their own frequency.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{i,j}}$$

Inverse Data Frequency (IDF):

Calculating the term frequency is not enough because words like the, a, an, and, etc. appear many times but considering the meaning these are not very important so inverse document frequency is used which basically shows how unique a word is. This can be calculated by taking log of the number of documents divided by the number of documents that contain the word.

$$idf(w) = \log\left(\frac{N}{df_t}\right)$$

2. Cosine Similarity:

Cosine similarity is used to find the similarity between the two vectors by measuring the cosine angle between them. This technique is also used in the domain of data mining to measure the cohesion in the clusters

Cosine similarity = $\frac{AB}{|A||B|}$.

VIII. CONCLUSION

While there have been several studies in the domain of chatbots and a lot of various models have been implemented but there is none which is cost efficient and provides satisfactory results. The Tf-idf based chatbot is cost efficient and can be run on local machines. The main advantage of using retrieval based model is that it can also be trained using smaller datasets and still give satisfactory results. This model also gives us an advantage of reducing the training time as we do not have to worry about the grammar of the chatbot as it functions on a retrieval based model. This projects success is promising and has more scope for advanced chatbot development in the field of healthcare.

REFERENCES

[1] Mohammed Javed, P. Nagabhushan, B.B. Chaudhari, "A Direct Approach for Word and Character Segmentation in Run-Length Compressed Documents with an Application to Word Spotting", 13th International Conference on Document Analysis and Recognition (ICDAR), 2015.

[2] Jerome r. Bellagarda, "Parts-Of-Speech tagging by Latent Analogy", IEEE Journal of Selected Topics in Signal Processing, Vol. 4, No. 6, 2010.

[3] Kyo-Joong, DongKun Lee, ByungSoo Ko, Ho-Jin, Choi, "A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation", IEEE 18th International Conference on Mobile Data Management, 2017.

[4] Ming-Hsiang Su, Chung-Hsien Wu, Kun-Yi Huang, Qian-Bei Hong, Hsin-Min Wang, "A Chatbot Using LSTM-based Multi-Layer Embedding for Elderly Care".

[5] K. Oh, D. Lee, B. Ko and H. Choi, "A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation," 2017 18th IEEE International Conference on Mobile Data Management (MDM), Daejeon, 2017, pp. 371-375. doi: 10.1109/MDM.2017.64

[6] Du Preez, S.J. & Lall, Manoj & Sinha, S. (2009). An intelligent web-based voice chat bot. 386 - 391.10.1109/EURCON.2009.5167660

[7] Bayu Setiaji, Ferry Wahyu Wibowo, "Chatbot Using a Knowledge in Database: Human-to-Machine Conversation Modeling", Intelligent Systems Modelling and Simulation (ISMS) 2016 7th International Conference on, pp. 72-77, 2016.

[8] Dahiya, Menal. (2017). A Tool of Conversation: Chatbot. INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING. 5. 158-161.2017.

[9] C.P. Shabariram, V. Srinath, C.S. Indhuja, Vidhya (2017). Ratatta: Chatbot Application Using Expert System, International Journal of Advanced Research in Computer Science and Software Engineering, 2017

[10] Mrs Rashmi Dharwadkar1, Dr.Mrs. Neeta A. Deshpande, A Medical ChatBot, International Journal of Computer Trends and Technology (IJCTT) – Volume 60 Issue 1-June 2018

[11] Farheen Naaz, Farheen Siddiqui, modified n-gram based model for identifying and filtering near-duplicate documents detection, International Journal of Advanced Computational Engineering and Networking, ISSN: 2320-2106, Volume-5, Issue-10, Oct.-2017

[12] N-gram Accuracy Analysis in the Method of Chatbot Response, International Journal of Engineering & Technology. (2018)

[13] Intelligent Healthbot for Transforming Healthcare, ISBN: 978-93-5351-521-8 (2019)

[14] Ashwini Shangrapawar, Artificial Intelligence Based Healthcare Chatbot System, IRJET e-ISSN:2395-0056(2020)

[15] Zeeshan Bhatti, Artificial Intelligence based Smart Doctor using Decision Tree Algorithm, JICT (2017)

[16] Tamizharasi B., Jenila Livingston L.M.* and S. Rajkumar, Building a Medical Chatbot using Support Vector Machine Learning Algorithm, (NCSET) 2020