A Personalized Medical Assistant Chatbot: MediBot

Gajendra Prasad K. C.
Assistant Professor
Department of Computer Science & Engineering
Sapthagiri College of Engineering, Bangalore, India

Satvik Ranjan
UG Student
Department of Computer Science & Engineering
Sapthagiri College of Engineering, Bangalore, India

Tathagat Ankit
UG Student
Department of Computer Science & Engineering
Sapthagiri College of Engineering, Bangalore, India

Vivek Kumar
UG Student
Department of Computer Science & Engineering
Sapthagiri College of Engineering, Bangalore, India

Abstract

In recent years Machine Learning has a major impact in the field of medical science due to its ability to learn and analyze from the examples provided. There are lots of treatments that are available for various common diseases that can be detected by individuals and doctors just on the basis of symptoms. But it is not possible for any human to know about all diseases. So, the problem is that there isn't any place where anyone can have the details of the diseases based on the symptoms. What if there is a place where we can find our health problem just by entering symptoms. The proposed idea is to create a chatbot that can predict disease using Apriori algorithm just by providing symptoms. Also, by using Recurrent Neural Network algorithm this system can be simply used as a tool of communication. Hence it can be beneficial for people as early detection of the disease can help in better treatment and recovery. Moreover, this system will help people to keep track of their heath regularly and properly without going anywhere.
Keywords: Apriori Algorithm, Chabot, Disease Prediction, ML, NLP, RNN

I. INTRODUCTION

In today's fast paced life people often don't take proper care of their health. People try to avoid going to doctor for diagnosis either because they don't like it or they don't get time. Now let's imagine you have a friend who is a doctor then your life would be so easy. You can randomly have a conversation with your friend about your health and your friend responds by asking more about the symptoms which you are having and will then diagnose you for a disease which you are suffering from. But not everyone can have a friend who is a doctor and as a result people often end up ignoring their health conditions. Well not anymore. In today's era where technologies like Apple's Siri, Microsoft Cortana is at the forefront of the highly personalized virtual assistants. They can do things which people imagined few decades ago. Similar to these technologies we can build a Chatbot where everyone can have a doctor friend with whom he can talk in general as well as talk about health issues.

Chatbots are programs that mimic human conversation using Machine Learning algorithms. It is designed to be the ultimate virtual assistant helping one to complete tasks ranging from answering questions, getting driving directions, to playing one favorite tune etc. Chatbots has become more popular in business groups right now as they reduce customer service cost and handles multiple users at a time. They are being used in almost every domain from virtual assistant like Siri on mobile to customer support in tech industry to e-commerce websites. Chatbots are currently the one of the best trending technologies available. It is certainly one of the most advanced and time saving technology also. But yet to accomplish many tasks there is a need to make chatbots efficient in medical field as well. To address this problem this project provides a platform where human can interact with a chatbot which are highly trained on datasets using Machine Learning algorithms. Machine Learning algorithms take a more natural approach for computation rather than taking a logical approach. The output is depended on the dataset they are trained on. Even without knowing the proper logic behind an algorithm one can make use of those algorithms and can be benefited from it. It is like learning from the experiences. But in Machine learning the biggest challenge is the choice of algorithm that is to be used for the dataset. It is said that all Machine Learning algorithms are wrong but some are useful. There are number of models available for machine learning. For the proposed chatbot the current best industry algorithms are used.

II. LITERATURE REVIEW

A survey on the techniques that can be used to design Chatbots is made [5]. The different design techniques of nine papers that can be used to design a chatbot are compared. It also covers the significant improvements in Chatbots in the last decade. In addition, it discusses about the similarity and differences in the techniques that are used in the selected study and then with the Loebner Prize winning chatbots techniques. The proposed work presents the methods of teaching chatbots to process natural language text. It discusses about processing natural language using Recurrent Neural Network (RNN). The sequence to sequence long short-term

memory cell neural network (LSTM) is used to train the model. In addition, it also talks about the challenges of implementing a Recurrent Neural Network based chatbot [1]. Disease diagnosis system using several machine learning algorithms is proposed. A detailed comparison of four Machine Learning algorithms to predict disease based on symptoms provided, is also presented in this paper. The purpose was to find the best Machine Learning algorithm to diagnose diseases early to help doctors and patients, as early prediction of disease can improve treatment efficiency [12]. The work proposed here discusses the need of a chatbot on e-commerce site to enhance user experience. The chatbot gives information about the products on an e-commerce sites which is useful for a customer to buy exactly what they want. The chatbot is designed using Artificial Intelligence Markup Language (AIML). It highlights the multi-lingual properties of the chatbot. The Graphical User Interface (GUI) is developed in Bangla Language. The proposed work also discusses the limitations of AIML Knowledge Based System to process natural language [11]. A problem solving chatbot using Deep Neural Network is proposed [4]. The chatbot is trained to remember different Data Structures such as array, queue, stack and tree. It can take inputs in text or voice format and respond form a knowledge base. The output is based on the learning provided during the training of the chatbot. The main purpose is to show that chatbot can also be used for problem solving.

III. PROPOSED SYSTEM

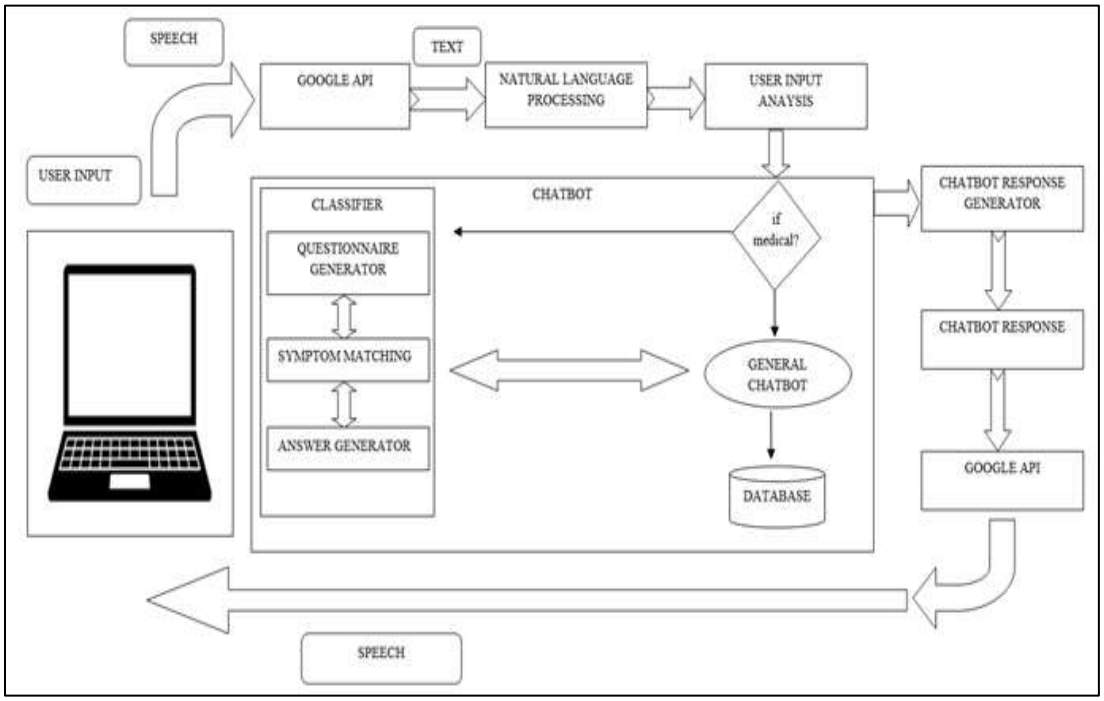

Fig. 3(a): System Architecture

In the proposed system we have tried to enhance the ability of a chatbot from being used just as a tool of communication to predict disease based on symptoms provided as input. To achieve this, the system combines various machine learning algorithms. The system design is as follows:

A. User Interact with Chabot

The user starts a conversation with the chatbot by providing input through speech. The input is recorded through a microphone. It is just like starting a normal conversation with any human being. The conversation with the chatbot is completely voice based.

B. Decision Making

As soon as the chatbot starts the conversation with the user, the user is asked to choose from the options provided. Two options are provided: General Conversation and Medical Assistant. Based on the response from the user the chatbot decides if it has to continue a general conversation or to help the user for self-diagnosis.

C. Response from chatbot

The response is generated based on the decision making in the previous stage. If the user needs help related to health the chatbot enters into a questionnaire. It starts asking about different symptoms and then finally predicts a disease. If the user does not need any help then the chatbot continues with the normal conversation. All the response from the chatbot is voice based.

IV. MODULE DESCRIPTION

A. Speech-To-Text

In this module the user voice input is being converted into text. This is to be done because text input is required for the next stage. We will be using a Google Speech Recognition API for Speech-to-text conversion. We can speak in the microphone of our laptops and the Google API will translate it into text. This Google API is very powerful and it can also translate several other languages into text. But for this proposed system only English is required. However, this API does not work offline and it requires a working internet connection for processing the speech.

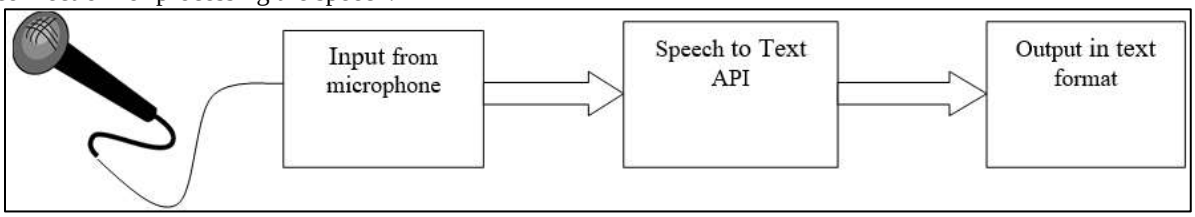

Fig. 4(a): Speech-to-Text Processing

B. Natural Language Processing

The computer system cannot understand the natural language of the humans. Hopefully we have a branch of computer science, called Natural Language Processing, which deals with the natural language which human speak. The goal of NLP is to take the unstructured output of the API and produce a structured representation of the text that contains spoken language understanding. The output of the Google API, which will be a text input, is given as input to our chatbot system. The chatbot uses Natural Language Processing to process the text input. After processing of the text input, the chatbot responds with a series of questions in order to understand the situation of the user better. A number of methods are available to extract meaning from spoken words. We have used Bag of Words. In Bag of Words sentence structure order and syntax is ignored only the count of number of occurrences of each word is taken into consideration. This is used to form a vector space model in which unnecessary words are removed. It is used to extract meaning information from the spoken words.

C. Response generation

The response generation is the most important phase of the chatbot system. This chatbot will respond based on the requirements of the user. The interaction with the chatbot is classified in mainly two types: General Response and Medical Assistant.

1) General Response

The general response will be created based on the training of the chatbot. There are number of ways for training the chatbot. In this system we have used the Sequence-to-Sequence (seq2seq) model to train the system. The model has been trained on the Cornell Movie Corpus dialogue dataset.

2) Medical Assistant

If the user wants any medical diagnosis of the disease based on the symptoms provided, then the chatbot uses the Apriori Algorithm to predict the disease. The model is trained on dataset available from the New York Presbyterian Hospital. This algorithm is best suited for frequently occurring item sets in the dataset.

D. Text-to-Speech

After the response generator analysis and text input it generates an output according to the user input. The output can either be general conversational speech or it can be a questionnaire of symptoms to predict disease. The output from the response generator has to be converted from text to speech in both the cases. For this purpose, the Google text-to-speech API is used to generate a voice output from the chatbot. The API is capable of responding in a number of languages but for this project only English is used.

V. ALGORITHMS USED

Two algorithms are used in this proposed system. They are:

A. Sequence-to-Sequence Model

The seq2seq model consists of two RNN: an encoder and a decoder. The encoder takes a sentence as input and processes one word one at a time. The decoder generates words one by one in each time step of the decoder's iteration. After one complete iteration, the output is generated.

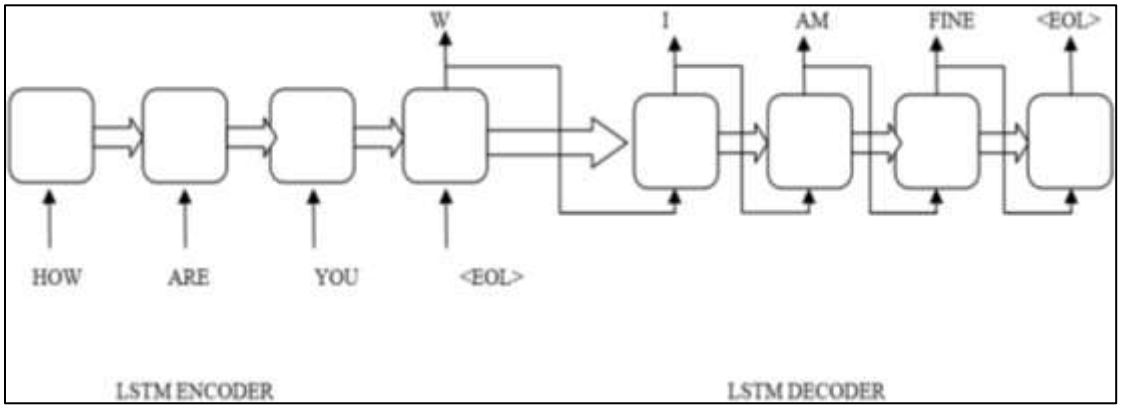

Fig. 5(A): Seq2Seq Model

B. Apriori

The Apriori algorithm is used for finding frequent item sets in a dataset for Boolean association rules. The apriori principle can reduce the number of items sets we need to examine. The algorithm uses bottom up approach where frequent subsets are extended one at a time, known as candidate generation and group of candidates are tested against the data. It states that if an item set is infrequent, then all its supersets must also be infrequent. This means that if {pale eyes} was found to be infrequent, we can expect {pale eyes, cold} to be equally or even more infrequent, so in consolidation the list of popular item sets, we need not consider {pale eyes, cold}, nor any other item set configuration that contains pale eyes.

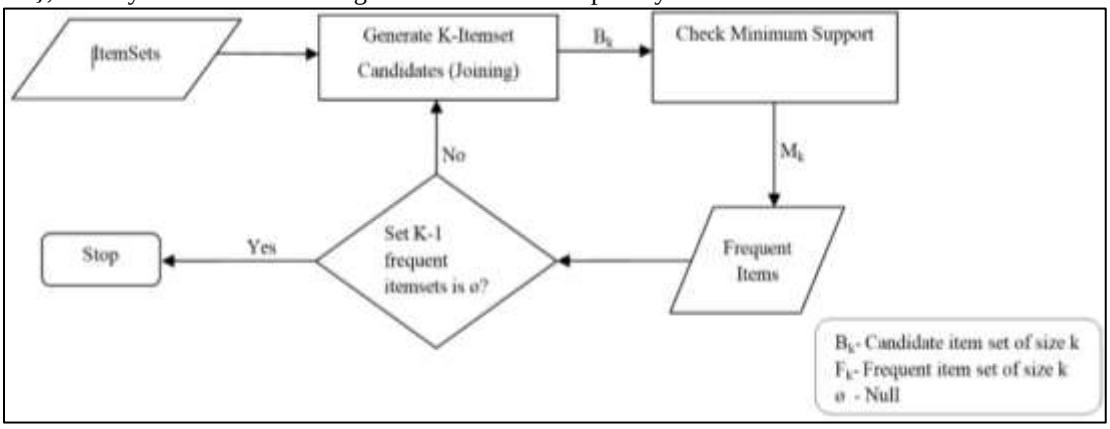

Fig. 5(b): Apriori Algorithm

VI. CONCLUSION

A chatbot design was implemented which can be used as a tool for general conversation as well as for self-diagnosis. The system can be very useful in the field of medical science for early and faster detection of disease. Also, the system helps an individual to keep track of their health more properly. However, the biggest challenge is the lack of correct and accurate medical dataset. Although there are a number of dataset available for developing a conversational chatbot, there is only one dataset available for disease prediction which is limited. There is one more big challenge that the seq2seq model requires a lot of time for training even though the hardware is capable of handling it. For future enhancements we can use an offline Application Programming Interface (API) for Speech-to-text conversion and Text-to-Speech conversation. Also, the symptom-diseases mapping can be done more accurately if a more reliable dataset is made available. Along with this, with the increasing use of smart wearable more accurate body measurements like heart rate, Body Mass Index (BMI) etc. can be given as input to the system and the disease prediction algorithm can be improved for more accurate prediction of diseases.

REFERENCES

[1] Milla T Mutiwokuziva, Melody W Chanda, Prudence Kadebu, Addlight Mukwazvure, Tatenda T Gotora, "A Neural-network based Chat Bot" in ICCES,2017
[2] Bhavika R. Ranoliya, Nidhi Raghuwanshi and Sanjay Singh, "Chatbot for University Related FAQs". IEEE trans. 978-1-5090-5013,2017
[3] Kai Yu,Zijian Zhao, Xueyang Wu, Hongtao Lin, Xuan Liu, "Rich Short Text Conversation Using Semantic Key Controlled Sequence Generation". IEEE trans. On ASL., to be published
[4] Ankil Shah, Bhargav Jain, Bhavin Agarwal, Saurabh Jain, Simon Shim, "Problem Solving Chatbot for Data Structure". IEEE trans.978-1-5386-4649-6,2018
[5] Sameera A. Abddul-Kader, Dr. John Woods, "Survey on Chatbot Design Techniques in Speech Conversation System", International Journal of Advanced Computer Science and Application, Volume 6-No.7,2015

[6] L.Sathish Kumar, A. Padmapriya, "Prediction for Common Disease using ID3 Algorithm in Mobile Phone and Television",International Journal of Computer Application, Volume 50-No.4,July,2012
[7] Balbir Singh Bani, Ajay Pratap Singh, "College Enquiry ChatBot Using A.L.I.C.E",International Journal of New Technology and Research, Volume 3-No.1,2017
[8] M. Dahiya, "A Tool of Conversation:Chatbot", International Journal of Computer Science and Engineering, Volume 5-No.5,20177
[9] Liciana Benotti, Maria Cecillia Martinez, Fernando Schapachnik, "A Tool for Introducing Computer Science with Automatic Formative Assessment". IEEE Transaction On Learning Technologies, Vol.11 ,No.2 ,2018
[10] Bingquan Liu, Zhen Xu, Chengjie Sun, Baoxun Wang, Xiaolong Wang, "Content-Oriented User Modeling for Personalized Ranking in Chatbots", IEEE Transaction On Audio, Speech And Language Processing, Vol.26, No.1,2018
[11] Md. Shahriare Satu, Tajim Md. Niamat Ullah Akhund, Mahammad Abu Yousuf. (2017,Feb.).Online Shopping Management System with Customer Multi-Language Supported Query handling AIML Chatbot. Available: https://www.researchgate.net/publication/313315555
[12] Allen Daniel Sunny, Sajal Kulshreshtra, Satyam Singh,Srinabh, Mr. Mohan Ba, Dr. Sarojadevi H., " Disease Diagnosis System By Exploring Machine Learning Algorithm", International Journal of Innovations n Engineering and technology(IJIET),Available: http://dx.doi.org/10.21172/ijiet.102.03