OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

SURVEY ON CHAT BOT SYSTEM FOR CANCER PATIENT

Pooja Nehul[1], Bhakti Lohar[2], Uttkarsha Jagtap[3], Shreya Rajurkar[4], Prof. Gauri Virkar[5]

UG Student, Department of Computer Engineering, JSPM's BSIOTR, Wagholi Pune[1,2,3,4]

Assistant Professor, Department of Computer Engineering, JSPM's BSIOTR, Wagholi Pune[5]

pujanehul@gmail.com, bhaktilohar10@gmail.com, jagtaputtkarsha98@gmail.com, shreyarajurkar@gmail.com,

gaurivirkar05@gmail.com

ABSTRACT: Normally Users are not aware about all the treatment or symptoms regarding the particular disease for small problem user have to go personally to the hospital for check-up which is more time consuming. Also handling the telephonic calls for the complaints is quite hectic. Such a problem can be solved by using medical Chat Bot by giving proper guidance regarding healthy living. The medical chat-bots functioning depends on Natural language processing that helps users to submit their problem about the health. The User can ask any personal query related to health care through the chat-Bot without physically available to the hospital by Using Google API for voice-text and text voice conversion. Query is sent to Chat Bot and gets related answer and display answer on android app.

Keywords: *Chat-Bot, Natural Language Processing, Health-Care, Random-Forest*

--- .˙.˙.˙.---

I INTRODUCTION

Input data is provided as relevant questions with their answers. This data can be converted into structured form that a Chat-bots can learn from. The bot can train itself using Natural Language Processing (NLP), Natural Language Understanding (NLU) and other Machine Learning skills. Natural Language Processing is "ability of machines to understand human language the way it is written or spoken". The objective of NLP is to make computer/machines as intelligent as human beings in understanding language.

The main purpose of the scheme is to build the language gap between the user and health providers by giving immediate replies to the Questions asked by the user. Today's people are more likely addicted to internet but they are not concern about their personal health. They avoid going in hospital for small problem which may become a major

disease in future. Establishing question answer forums is becoming a simple way to answer those queries rather than browsing through the list of potentially relevant document from the web. Many of the existing systems have some limitation such as There is no instant response given to the patients they have to wait for experts acknowledgement for a long time. Some of the processes may charge amount to perform live chat or telephony communication with doctors online. [7] This system allows computer to communication between human to computer by using natural language processing (NLP).

There are three analyses which understand natural language i.e. identification of main linguistic relations is completed to parse subject into object of the sentences. After that description of the texts is done. The semantic interpretation uses knowledge of word meaning Chatbot is an Entity which imitates human discussion in its particular

accepted set-up together with a text or vocal language with techniques such as Natural Language Processing (NLP). The aim of this system is to replicate a person's discussion. The development of chatbot application canbe done with making a user interface to send input and receive response. It is a system that interacts with user by keeping the track of the state of interaction and recollecting the preceding commands to give functionality. The medical chat-bots can be developed by using artificial algorithms that scrutinize user's queries and recognize it and give reply to related query. A big disease can start from small problems such as headache which feels normal but it may beginning of big disease such as brain tumor most of the disease can be identified by common symptoms so the disease can be predicted if the patient body is analysed periodically. [6]

II RELATED WORK

Table: Literature Survey In Tabular Form

Sr. No.	Paper Name	Year	Methodology	Conclusion
1.	Application of synonous text-based dialogue system in mental health interventions	2017	Web-based chat technologies are compared.	The review provides an evaluation of individual synchronous Web-based chat technologies as a mode of psychological intervention and support
2.	A Self-Diagnosis Medical Chatbot Using Artificial Intelligence	2018	Natural language processing and pattern matching algorithm	The usage of Chatbot is user friendly and can be used by any person who knows how to type in their own language in mobile app or desktop version. A medical chatbot provides personalized diagnoses based on symptoms.
3.	A Novel approach for medical assistant using trained chatbot	2017	Natural language processing and pattern matching algorithm	It is possible to predict any possible problem even before they start to cause any damage to the body.
4.	MedChatBot: An UMLS based Chatbot for Medical Students	2012	AIML pattern technique for Pattern matching.	The chatbot for medical students by deploying the widely available medical knowledge source UMLS and the open source Chatterbean. MedChatBot could be incorporated in a tutoring system to support natural dialog. Although we have described the implementation in the context of the medical domain, the techniques can easily be applied to other domains too
5.	Conditional entropy based Retrivel model in Patient-Carer Conersational cases	2017	Natural Language Processing.	The results show that even in small training datasets, this method outperforms up to date methods for automated communication.

H. N. Io, C. B. Lee, "Chatbots and Conversational Agents: A Bibliometric Analysis"

This paper used bibliometric analysis to examine past research on chatbot and CA. The results are helpful in identifying the research agenda for the future. The publication information and citation report in the literature databases shows a potential opportunity on chatbot research with the emergence of deep learning technology. There are also research opportunities on applications of chatbot and CA in areas other than education, psychology, and linguistics. The results of the analyses using CiteSpace and Bibliometrix indicate that past research on chatbot and CA is highly fragmented and there is a large space of research opportunity to explore. This study proposed several suggestions for future research, which could help to understand chatbot and CA not only from the technical point of view, but also from the business and users' point of view. The results of this study would hopefully help to encourage future research on chatbot and CA.

Honghao WEI, Yiwei Zhao, "Building Chatbot with Emotions"

In this paper, we test three mechanisms for generating responses with specific sentiment. Due to time constraints, we have not finished training phase for Deep Reinforcement Learning model. We compare the performance of vanilla Seq2Seq model and both static and dynamic Emotion Chatting Machine. Automatica evaluation shows that the emotion embedding, internal memory and external memory in ECM help to generated reasonable responses on both content and emotion level. In the future work, we would continue the training phase of Deep Reinforcement Learning and report the results as soon as we can. In addition, we notice the emotion tag for dataset has influential effects on the performance. We would collect more samples on live journal and train better multi-category classifiers with different neural network architecture for sake of better emotion tag results.

Sameera A. Abdul-Kader, Dr. John Woods, "Survey on Chatbot Design Techniques in Speech Conversation Systems"

In this paper, the literature review has covered a number of selected papers that have focused specifically on Chatbot design techniques in the last decade. A survey of nine selected studies that affect Chatbot design has been presented, and the contribution of each study has been identified. In addition, a comparison has been made between Chatbot design techniques in the selected studies and then with the Loebner Prize winning Chatbot techniques. From the survey above, it can be said that the development and improvement of Chatbot design is not grow at a predictable rate due to the variety of methods and approaches used to design a Chatbot. The techniques of Chatbot design are still a matter for debate and no common approach has yet been identified. Researchers have so far worked in isolated environments with reluctance to divulge any improved techniques they have found, consequently, slowing down the improvements to Chatbots. Moreover, the Chatbots designed for dialogue systems in the selected studies are, in general, limited to particular applications. General-purpose Chatbots need improvements by designing more comprehensive knowledge bases.

S. J. du Preez, "An Intelligent Web-Based Voice Chat Bot"

This paper presents the design and development of an intelligent voice recognition chat bot. The paper presents a technology demonstrator to verify a proposed framework required to support such a bot (a web service). While a black box approach is used, by controlling the communication structure, to and from the web-service, the web-service allows all types of clients to communicate to the server from any platform. The service provided is accessible through a generated interface which allows for seamless XML processing; whereby the extensibility improves the lifespan of such a service. By introducing an artificial brain, the web-based bot generates customized user responses, aligned to the desired character. Questions asked to the bot, which is not understood is further processed using a third-party expert system (an online intelligent research assistant), and the

response is archived, improving the artificial brain capabilities for future generation of responses.

Bayu Setiaji, "Chatbot Using A Knowledge in Database"

The development of chatbot application in various programming language had been done with making a user interface to send input and receive response. Designing and building tables as representation of knowledge in the database had been started from entity-relationship diagram resulting 11 entities and its cardinalities. Making use of structured query language (SQL) for pattern matching had been done within stored program. The stored program consists of 4 stored procedures and 21 stored functions employed as pattern matching and supporting processes. Bigram method can be used not only for Indonesian language words, but also other languages with some boundaries.

There are a few typical chat bots systems that closely resembles with our expected system. These systems are as follows.

1.   Alexa By Amazon

Alexa is a virtual digital assistant developed by Amazon for its Amazon Echo and Echo Dot line of computing devices. Alexa's capabilities mimic those of other intelligent assistants such as Apple Siri, Microsoft Cortana, Google Assistant and Samsung Bixby.

Alexa responds to voice control by returning information on products, music, news, weather, sports and more. The back-end engine for Amazon's Alexa runs on Amazon Web Services in the cloud, enabling Alexa to learn an individual or family's preferences and expand its functionality over time. In addition to the Echo products, Alexa is also supported by Amazon's Fire HD tablet and Fire TV set-top box products.

2.   Siri By Apple

Siri's artificial intelligence-like (AI) natural language processing engine and backend services. Perhaps a useful simplification is to suggest that Siri has three layers: voice processing, grammar analysis-context learning engine and services. Importantly, Siri not only understands spoken words, it understands context. Understanding context requires deciphering natural language.

How Siri Works Once the Siri microphone button is touched, whatever is said is recorded, compressed and sent to Apple's data centers where Apple hosts Siri's Nuance speech-to- text and AI-like natural language processing engine. Siri then figures out what has been said. Depending upon the request, Siri creates a voice response and either sends it back to the iPhone.

3.   Natasha By Hike

Natasha works somewhat similar to Siri or google now. When you send a message or give an input. They recognize certain keywords from your input and match them with their library and based on those keywords they produce some combination of words which could answer your question.

For example if you say "what is your name." it recognize 'what' and 'name' as keyword and produce a combination of possible matching words. in this case it'll be My name is Natasha or I am Siri based on your questions and previous interactions.

4.   eBey By Google

eBay Shop Bot is a personalized shopping assistant that helps people find the best deals from eBay's one billion listings. Our vision is to make shopping with eBay as easy as talking to a friend, whether you are looking for something specific or just browsing for inspiration.

Shoppers have complex needs, which are often not fully met by traditional search engines. While offline retailers can offer human experts to help, there aren't enough experts around to help us with everything we want and need.

The science of AI provides contextual understanding, predictive modelling, and machine learning abilities. Combining AI with eBay's breadth of inventory and unique selection will enable us to create a radically better and more personal shopping experience for virtually anyone that owns a mobile phone.

III PROPOSED SYSTEM

Following are the main modules of the proposed system:

1. User / Patient Login to System: User registers on Chat-bot application. Then ask queries regarding to the health care and medical details.

2. Chat BOT Responding System:

• NLP Processing and Sentiment Analysis for Complaint: When patient complaint is submitted to the system, NLP is applied and sense of the complaint is detected. The sense of the words is found using part of speech tagging and word-net dictionary. By using the sentiment analysis negation level of a complaint is detected. And user complaints are prioritized accordingly.

• Search Questions in knowledge database: Once the negation level of the complaint is detected, furthermore, the exact question in the complaint is detected using WorldNet.

3. all data as well as medicine and food alert display.

Figure 1: System Architecture

4. Ask some Questions: You can ask some questions regarding cancer. And its related to voice- text and text-voice conversation. Using Google API for inter conversion of text-voice and vice versa.

Medicine Alert and Food Alert: Medicine dosage details you can ask medical dosage related queries to this app in voice and system gets output for medicine API and speak out and display

IV ALGORITHMS USED

1. Natural Language Processing (NLP):-

For implementation of chatbot Natural Language Processing is used. The Figure 2 shows the several stages of NLP which are briefed further.
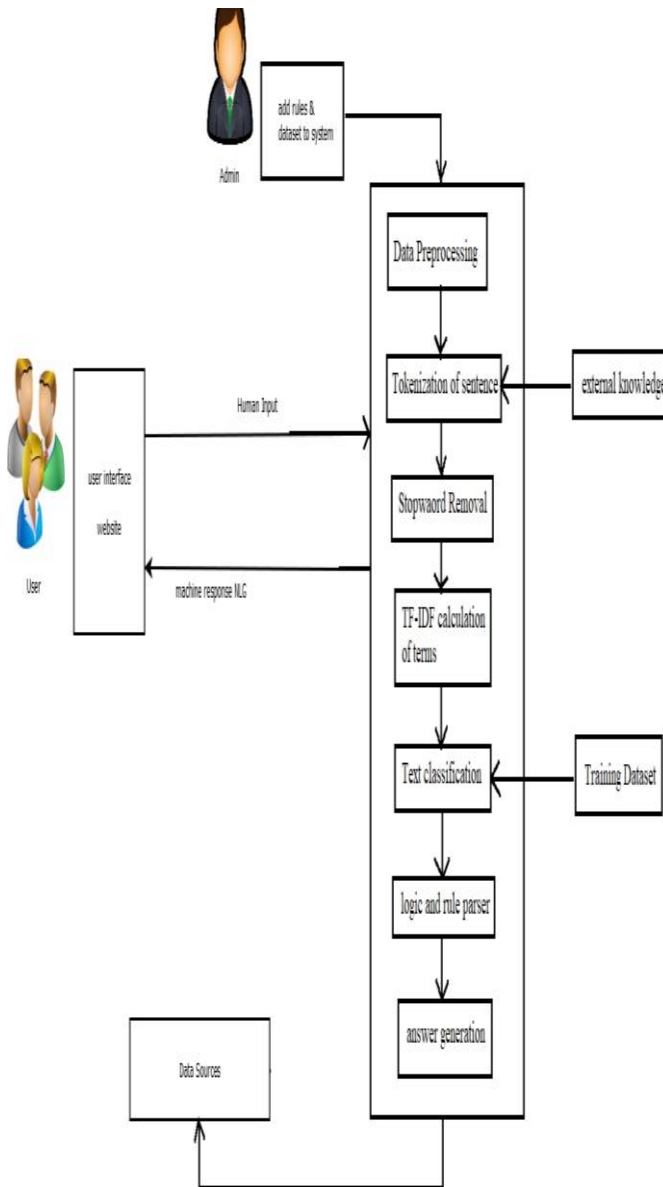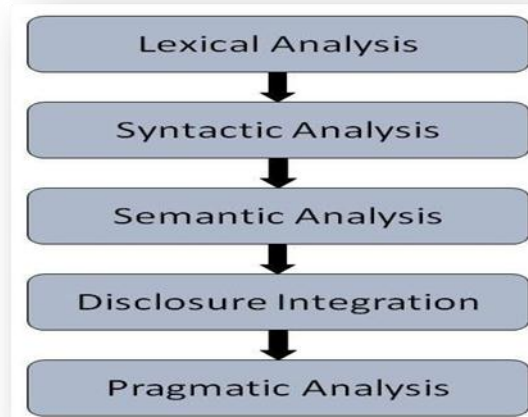

Figure 2: Stages of Natural Language Processing

i) Lexical Analysis: It involves identifying and analysing the structure of words. Lexicon of a language means the collection of words and phrases in a language. Lexical analysis is dividing the whole chunk of txt into paragraphs, sentences, and words.

ii) Syntactic Analysis (Parsing):- It involves analysis of words in the sentence for grammar and arranging words in a manner that shows the relationship among the words. The sentence such as "The school goes to boy" is rejected by English syntactic analyser.

iii) Semantic Analysis: It draws the exact meaning or the dictionary meaning from the text. The text is checked for meaningfulness. It is done by mapping syntactic structures

and objects in the task domain. The semantic analyser disregards sentence such as "hot ice-cream".

iv) Discourse Integration: - The meaning of any sentence depends upon the meaning of the sentence just before it. In addition, it also brings about the meaning of immediately succeeding sentence.

v) Pragmatic Analysis: - During this, what was said is re-interpreted on what it actually meant. It involves deriving those aspects of language which require real world knowledge.

NLP Algorithm:-

Input: Query q, answers(A0….An)

Output: relevant answer,

Procedure: parse uses input sentence q

If(q==chat)

{

Add chat to chat table

Grammer parsing

Generate relevent reply

}

Else if(q==statement)

{

Add statement to database

Apply grammer parsing

Generate relevent reply

}

Else if(q==Question)

{

Generate tokenization of question (T0, ….Tn)

Remove stopword using corpus

Tf = fTn in(s)/ l(s)

IDF = l(s) in D / l(s) Є Tn

Find most important term

Match term TA in dataset

If(! match){

No answer found

Else{

Generate answer reply to user

}

2. Data Set

In proposed system, two datasets will be used:

i) **Text Query:**

The proposed work consists of a chat bot which will take input as text queries with limitations.

ii) **Voice Query:**

The proposed work consists of a chat bot which will take input as voice queries with limitations.

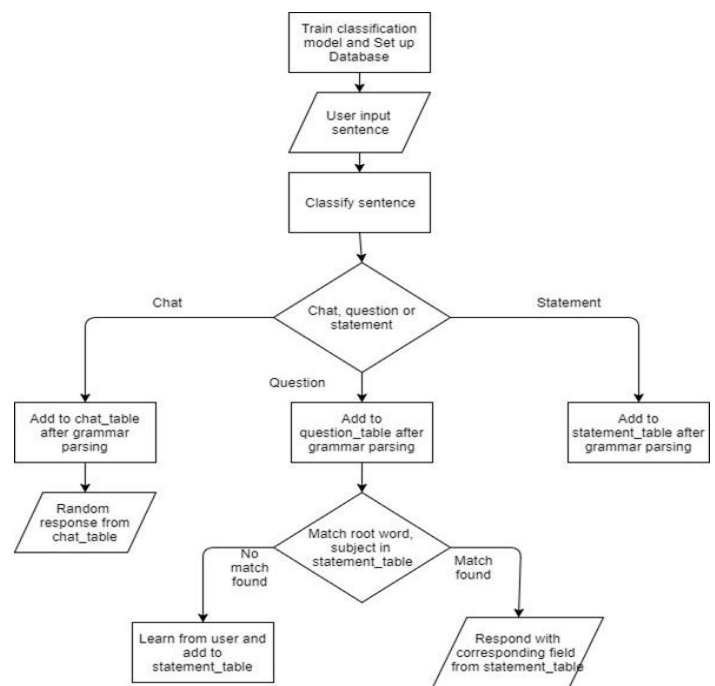Figure 3 gives brief overview of working implementation of this proposed system.

Figure 3: Implementation Flow Chart

3. Random Forest Algorithm

For classification of disease (cancer) we are using **Random-forest algorithm.** A random forest is an ensemble learning algorithm. The basic concept of the algorithm is that building a number of small decision-tree having less features and which is a computationally very cheap process. After building many small and weak decision trees parallel, then these trees can be combined to form a single and strong tree by averaging or taking the majority vote.

The algorithm is started by merging a combination of trees which each tree will vote and emerged for a class1. Suppose that there are N number of data and M number of input variables in a data set where the real data used in this system which comprised of data and input variables. Let k is

the number of sampling groups in the system, ai and bi be number of data and number of variables in group i where i is given as 1, 2, 3, 4… and k. Each sampling group is defined as follows.

$$E.g$$
$$Age > 55 \land BPH = yes \rightarrow class = diabetes\_risk\_group$$
$$[support=20\%, confidence=93\%]$$

i) ai is the data where ai is not greater than N variables which are selected randomly from N input variable.

ii) bi are the variables where bi is not greater M data which are selected randomly from M input variable.

iii) A tree increases to leaf nodes and predicts class.

After repetition of steps from 1 to 3 for the given k times, this   tree will take a structure like a forest. Then the classification of many models will be selected by considering a majority vote of all trees in the forest.

V CONCLUSION AND FUTURE WORK

In the proposed system we are developing a site and an android application which will provide answers to the analysed queries of the user. We are using large dataset which will ensures the better performance compared as earlier. Thus we build up a system which is useful for medical institute or hospitals to help the users to freely ask medical dosage related queries by voice. The system gets output for medicine API and speaks out and display all medicine alert and food alert. We are using NLP because we want to make a computer to communicate with users in their terms. User can get related answer displayed on android app and refer this answer for analysis.

REFERENCES

[1] "A Graph Based Chatbot for Cancer Patients", Belfin R VDepartment of Computer Science and Engineering.

[2] H. N. Io, C. B. Lee, "Chatbots and Conversational Agents: A Bibliometric Analysis"

[3] Honghao WEI, Yiwei Zhao, "Building Chatbot with Emotions"

[4] Sameera A. Abdul-Kader, Dr. John Woods, ``Survey on Chatbot Design Techniques in Speech Conversation Systems"

[5] S. J. du Preez, "An Intelligent Web-Based Voice Chat Bot"

[6] B. A. Shawar, and E. Atwell, "Different measurements metrics to evaluate a chatbot system," Proceedings of the Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies, 2007.

[7] J. Weizenbaum, "ELIZA — A Computer Program For the Study of Natural Language Communication Between Man And Machine," Commun. ACM, vol. 9, no. 1, pp. 36–45, 1966.

[8] B. A. Shawar and E. Atwell, "A comparison between ALICE and Elizabeth chatbot systems," Raport instytutowy, University of Leeds, 2002.

[9] S. Khalaf, "On Their Tenth Anniversary, Mobile Apps Start Eating Their Own," Flurry Analytics, 2016.

[10] C. Yin and Z. Jinran, "Seniors prove WeChat is not just for young," CHINA CHANNEL, 2017.

[11] S. Deveau, "Li Ka-Shing Buys Canada's Reliance Home for $2.1 Billion," Bloomberg, 2017.