A Smart Bot as an Interactive Medical Assistant with Voice-based System using Natural Language Processing

Aashita Tiwari, Ahitagnee Paul, P. Suganya, A.P(O.G)

Abstract-In this fast-moving world, people are ignorant about their health issues and avoid routine check-ups. It is very difficult for users to spend longer time on-line and explore health information. To solve the problem, voice-based application is provided to the user where user can interact with system and get inference of diseases and their remedies by giving the symptoms as input.For processing the given input, the data is normalized by using noun phrase extraction and medical term identifier. For getting more precise result, the system generates relevant questions to the user and accordingly provide remedy for problem. Question is generated by mapping the user input array with question generation matrix. The project is based on a digital medical aid through a smart bot using machine learning and optical character recognition techniques. The user can just talk to the bot and get to know what the possible causes and effects of the particular symptoms are, determine the illness and take appropriate actions. Basically, the purpose of this bot is to act as a friendly healthcare assistant that helps in all the work that people need to take care of their health.

Keywords— Friendly Healthcare Assistant, Machine Learning, Medical term identifier, Natural language processing, Normalization, Optical Character Recognition, Question-Answer System, Smart Bot, Voice based.

I. INTRODUCTION

Nowadays, people are more curious about their health. In the recent surveys it is found that people search on Internet for small problems related to health; specially those above 20. Due to the heavy workloads and busy schedules people try to avoid taking appointment from medical specialists and do searching for home remedies over Internet. Mobile phones make it easier for users to have global information in their hands. Rather than wasting time in browsing, this bot provides a handy-solution to user where they can ask query related to their health issues. Pre-processing system consist of three modules: 1. Noun-Phrase Extraction [1], 2. Medical term identifier [1][2], 3. Normalization [1][3]. When the user problem terms are extracted, they are stored in a dataset and system will check whether user have included all information required by it. System will check what type of health issue user is suffering through. If user haven't inserted the required information like duration of disease or its severity, etc. Then system will generate question for the missing entities.

This job is done by the Question-Answer System (QAS). QAS is responsible for generating question as well as search for the respective remedy of disease. When all the system requirements are filled then the precise answer will be generated from the answer-map. Output to the user is given in the voice format on their respective devices.

II. RELATED WORK

There are many online websites where people can get a solution for their health-related queries [9] [10]. Also, there are some websites where they can directly ask question on website and get an expert advice. However, this is a time-consuming process since user has to spend more time online, seeking for their health information. Moreover, these websites do not provide accurate remedies to user, they list up all the disease possibilities, even if the probability of that disease is very low. Hence, this application proposes a question answering system where user can ask questions and get the remedy for that problem. In the approach adopted by Varsha Bhoir in [7], three major steps are introduced: pre-processing, token identification and answer extraction. The pre-processing of text can also be done by using the concept of local mining [1]. Three basic steps are involved in local mining: 1) Noun phrase extraction:In this stage all the nouns are extracted from given input. 2) Medical term Identifier: This phase includes extraction of all medical terms. 3) Normalization: In this phase, the terms are normalized to medical concepts. For a system to generate questions, the sentence is simplified by splitting the sentence and its compressed. Then, certain rules are mentioned for every interrogative pronoun and the answer phrase is replaced with this. In this way, many questions are generated and to evaluate the correct question, question ranking system is used [3].

Some work [4], have introduced concept for generating questions called nine factor estimation for English language. Here, Yuya Yokoyama eliminates the presence of mismatches between the respondent and questioners' intentions. Therefore, giving good result for all factors. By defining similar factor concept the proposed application gives accuracy in QA set. The work of [2] proposes approach to generate question to define a specific motivation for that question. Initially the input statement is categorized, then specific motivations are defined for each category and accordingly a question is generated. In restricted domain Question Answering system, question type, answer type, question keyword and question headword is extracted from the given input.

Later, Liqiang Nie[1] stated that answer is extracted from search engine and corresponding sentence is extracted as an answer by using certain answer extraction algorithm. In this way, answer is generated for the given question input.

III.   PROPOSED ARCHITECTURE

The proposed architecture contains two major sections: 1) pre-processing system and 2) question answer system.In this system user will enter the medical history and ask questions related to health. The medical history will be helpful to generate precise result to the query.
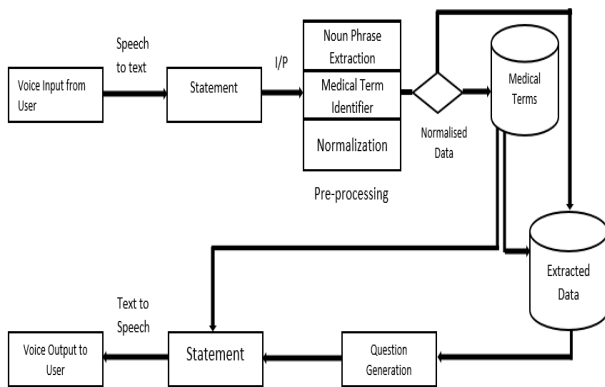

Fig.1 Architecture of the proposed system

A.   Pre-Processing System

Here, the user will provide a voice input to the device. Consider for example:" I am having headache for two days". Now, with the help of "Julius" tool, the voice input is converted into text. This text is given to the system as input. Now, the given sentence is tokenized with the help of Stanford Part-of-Speech tagger (POS Tagger). Tokenization process will give result in the form of word-tag pair. Further, noun phrase extractor is used to extract nouns from the tokens.The system is medical domain specific [8][6] so the terminologies which are related to medical concepts are retrieved with the help of medical term identifier. Medical term identifier is the second stage of pre-processing. In this stage, all the general noun-phrases are discarded. From the [8][5] Impurity finder is used, where it is assumed that those concepts which are more related to medical domain will occur more frequently as compare to the non-medical terms. So, for any concept to be determined its Impurity finder is given as below:

$$IF(m) = -\sum_{i=1}^{2} P(Ci|m) \log P(Ci|m)$$

where m = concept to determine.
C1, C2 = Medical corpus and General Domain corpus.
P(Ci—m) represents the probability of concept m related to specified domain Ci. This is calculated as:

$$P(Ci|m) = \frac{count(m, Ci)}{count(m)}$$

This method helps to detect more precise medical concept. Although medical terms are extracted here but it may happen that they are not standardized. So, to improve this

normalization is done.In this process, more precise terms related to the medical are extracted. These terms are much more accurate as compared to previous terms. All terms related to the given keyword are extracted. These terms are much more authenticated. All the extracted information is then stored in the new database. For example, only the term" Headache" and "2 days" as mentioned in previous example are required by the system [8]. This information is stored in an array, which will help the system to understand which required fields are absent.

B.   Question-Answer System

As shown in the example, only the disease and its duration are mentioned in the input statement. So, in the data structure these entities will be marked as one, but since the severity is not mentioned so the system will generate question for the such entities. [5] Questions are mapped from the question-sets for these entities and question related to the severity is retrieved from severity-question set. Here, Question-set is used under which all the possible questions related to each entity is stored. When theuser has inserted all the required fields like duration, severity, etc. then system will check whether the medical-terms from pre-processing system are same as that of symptoms. When similarity is found they are mapped on to the QA matrix and the respective solution is generated from the matrix set. QA matrix is the data set for each disease where multiple possibilities of each entity is saved. While generating the remedy, it is first compared with the user's history to avoid any side-effects of the remedy. This helps the system to understand the better requirements of user and generate precise solution. Hence, it can be said that from the proposed algorithm more accurate solutions are provided to the user.

IV.   RESULT AND DISCUSSION

- The bot can scan the symptoms from the user input data and inform the user concerning the attainable illness, the precautions to be taken and medications helpful to prevent the same.
- The bot also can give remedies in line with the user's interest in any particular medical field like home remedies, Ayurveda, homeopathy and allopathy.
- It can monitor the duration of the symptoms of the user and store the history to refer later and conjointly suggest a specific specialist doctor.
- The reports and prescription can be stored and also the functionality to share it with the doctor or any friends or relations can be provided.
- The bot also can provide information about the prescribed medicines. this may be later done by scanning the prescription directly, extracting the text out of it and look for that medicine's use and alternative connected information.
- The bot can develop a time chart to take dose based on the frequency and timing of the medicines.

V. CONCLUSION

The proposed system is a voice-based medical assistant question answer system. The aim of the proposed system is to increase health awareness among people through the answer provided by the system to the user's question. This system accepts health- related queries in voice format rather than text-format.

The pre-processing system and the question-answer system are the crucial elements of the proposed system. The question generation is performed using QA matrix. Further, the response of the system is reached to the user in voice format. This system provides accurate answer to the user. The high response capability and fast processing are the key features of the system. The system uses natural language processing techniques to provide precise remedy to the user's medical problems. The remedies suggested by the system include remedies for the mentioned symptom or diseases by the user.

VI. FUTURE SCOPE

In the future, a system that relies on deep learning, a neural network that enables the system to match a submitted photo of an injury with a vast database of iconic images from specialized websites like WebMD devoted to healthcare can be built. Crowd sourcing can be used to prepare a collection of images to train the neural network. By referring specialized doctor to the user, the system can be enhanced further. User can be recommended through the system for specific medical tests for the aforementioned symptom or disease. Also,a reminder of medicine timings can be given to the user. The system can also have daily newsletters feature for the user for healthy diets and precautions for diseases endangering lives.

REFRENCES

1. LiqiangNie, Yi-Liang Zhao, Mohammad Akbari, Kialie Shen and Tat-Seng Chua, "Bridging the Vocabulary Gap between Health Seekers and Healthcare Knowledge", IEEE Transactions on Knowledge and Data Engineering, Vol. 27, No. 2, February 2015.
2. LiqiangNie, Meng Wang, Luming Zhang and Shuicheng Yan, "Disease Inference from Health-Related Questions via Sparse Deep Learning", IEEE Transactions on knowledge and Data Engineering, Vol.27, No. 8, August 2015.
3. Mi-Young Kim and Randy Goebel," Detection and Normalization of Medical Terms Using Domain-specific Term Frequency and Adaptive Ranking", IEEE,978-1-4244-6561-3 2010.
4. Yuya Yokoyama, Teruhisa Hochin and Hiroki Nomiya," Estimation of Factor Scores from Feature Values of English Question and Answer Statements", ICIS Japan, June 2016.
5. Zhou Zhao, Lijun Zhang, Xiaofei He and Wilfred Ng," Expert Finding for Question Answering Via Graph Regularized MatrixCompletion", IEEETransactions on Knowledge and Data Engineering, Vol. 27, No.4, April 2015.
6. Paya J Biswas, Aditi Sharan and Nidhi Malik," A Framework for Restricted Domain Question Answering System", IEEE International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), September 2014.
7. Varsha Bhoir and M. A. Potey," Question Answering System: A Heuristic Approach", 978-1-4799-225 September 2014.
8. Ming Liu, Vasile Rus and Li Liu," Automatic Chinese Factual Question Generation", IEEE Transactions on Journal Name
9. WebMD (URL): http://www.webMD.com,2016.
10.]MedLinePlus (URL):http://www.nlm.nih.gov/medlineplus, 2016.

AUTHORS DETAILS

AASHITA TIWARI, SRM Institute of Science and Technology

AHITAGNEE PAUL, SRM Institute of Science and Technology

P.SUGANYA, SRM Institute of Science and Technology