Medical Chatbot

PES1201700118 SAIONI CHATTERJEE
PES1201700847 DAKSHA SINGHAL

November 2019

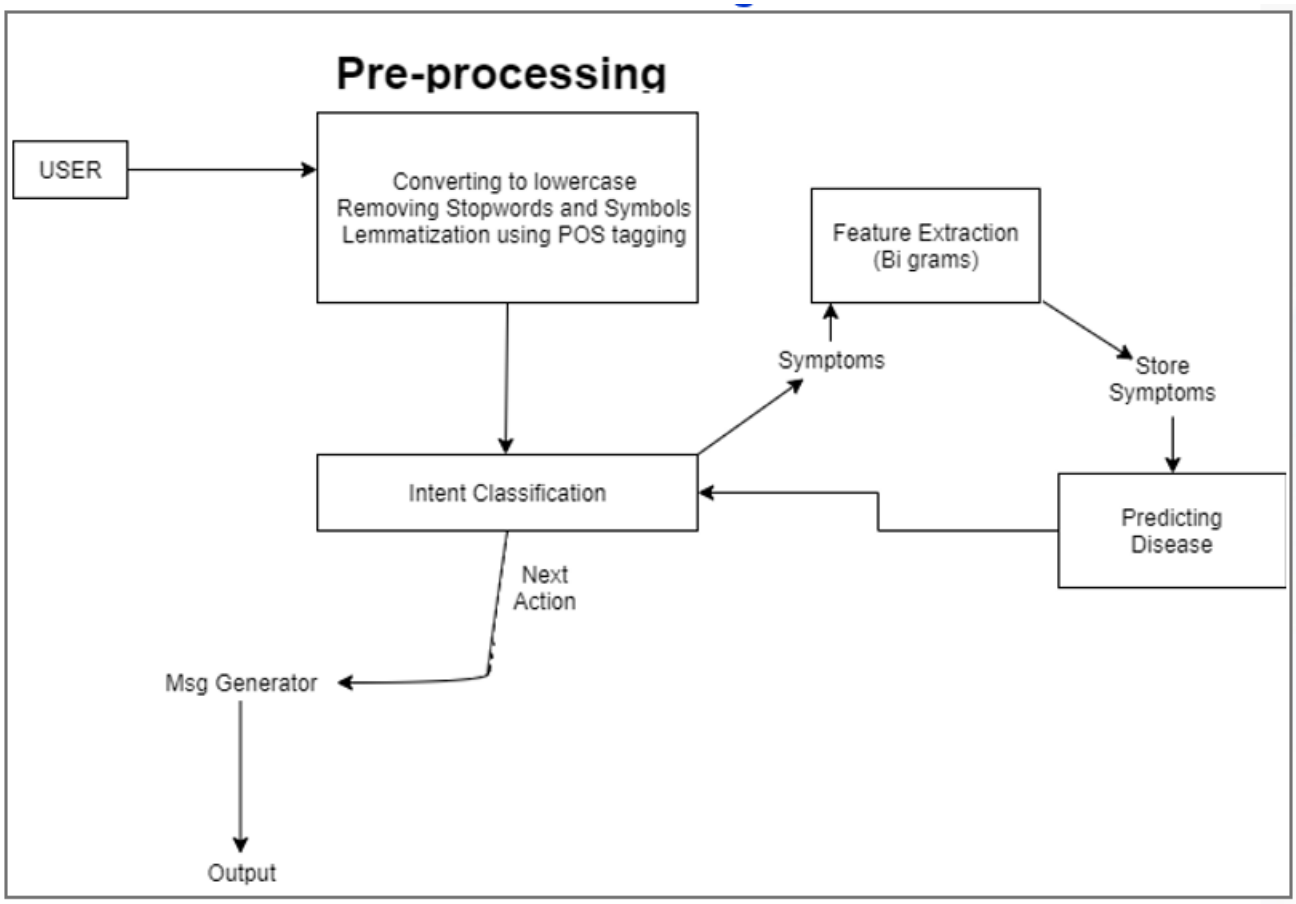
1. Introduction

In today's fast paced life people often tend to ignore their own health. In today's era we have some highly personalised virtual assistant such as Apple's siri, Microsoft Cortana, so similarly we can build a chatbot who can give health assistance to people. Chatbots, mainly designed using machine learning algorithms, have the ability to mimic human conversation. There can be two types of chatbots – general and retrieval based. General based where the chatbot have complete freedom to give answers on their own. There are no already existing repositories of datasets for training. The main disadvantage of this method can be sometimes it may produce irrelevant answers. The other type is retrieval based where the model is first trained on the pre-defined dataset, it checks for the response with the highest score, based on which answers are produced. So there is no chance of producing any irrelevant answers even if the chatbot is not trained for that particular question. The challenging task here is the choice of Machine Learning algorithm with best accuracy.

2. Literature Survey

A survey is made on the different design techniques used by the researchers and compared various approaches followed to build a chatbot. For example, Eliza, a artificial therapist which tries to find keywords from the conversation with the user and acts on that, if not found it responses with fixed phrases to continue the conversation. Pharmabot, a medical consultant bot was designed by researchers to provide safe medication for children. Another paper talks about finding the best algorithm from the four machine learning algorithms used to diagnose the disease early to improve treatment efficiency. Preprocessing was carried out using Recurrent Neural Network, this was mainly to discuss about the challenges faced using Recurrent Neural Network. A problem solving chatbot was built using Deep Neural Network which was designed to remember various data structures and respond from a knowledge base.

3. Proposed System



We have designed a linear based retrieval based chatbot which acts as a tool of communication to predict disease based on the input, the symptoms given by the user.It has 2 phases of conversation i.e extraction of symptoms and diagnosis of the symptoms. The input given to the chatbot is in the form of unstructured text and thus we have used Natural language Processing to produce a structured output using stemming, lemmatization (using parts of speech), removing stopwards. Contractions are used to remove apostrophes from the word. Bag of words is used in order to extract meaning of the words. Bag of words ignores the syntax or the order in which the words are used to form the sentence, It just takes the count of number of occurrences of each words and thus forms a vector model which removes unnecessary words. These words are now passed to 3 machine learning model, the model shortlists a number of disease based on the input. Here one of the model used is decision tree whose root node returns the most likely disease. Another 2 model used is Gaussian Naive Bayes and K Nearest Neighbour Classifier in order to compare which model finds the best response based on the highest accurcay. To get a better understanding of the disease the chatbot responses with a number of questions and thus finally the most likely disease is predicted based on the number of symptoms matched. The front end of the system is designed using Flask, HTML and CSS.

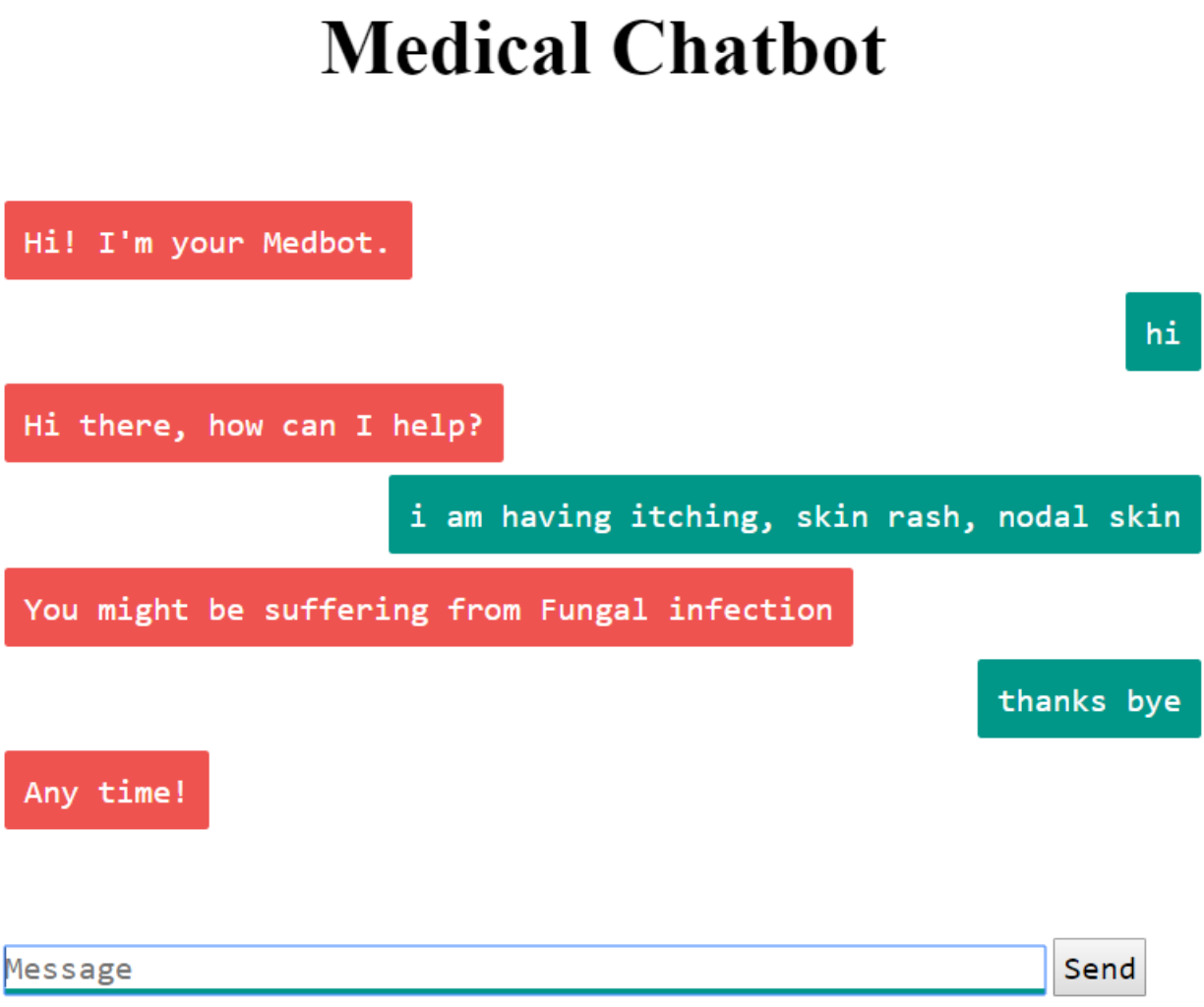
3.1. Symptom Extraction

Symptom extraction is performed to identify symptoms from the user's text input.A result of misspellings may lead to no or wrong symptoms mapping and may not always map to the database's symptoms. In order to handle this a variety of symptoms are extracted using bi-grams on the pre-processed text and compared with the symptoms in the database using fuzzywuzzy with a comparison threshold of 0.6.

3.2. Diagnosis of Symptoms

Various machine learning models were trained on the disease-symptom database. Once symptoms are extracted from the user, the Decision Tree model was applied to the set of symptoms and the predicted disease is specified to the end user by the chatbot.

3.3. Result and Discussions



The chatbot's conversation flow is as follows: The user will have a text to text converation with the chatbot and the chatbot will identify the intent of the user(greeting,goodbye,danger,symptoms etc). If the intent is 'symptom' then the chatbot will identify the disease. The above figure shows the conversation of the chatbot with the user.

3.4. Conclusions

We successfully built a working medical bot that is user friendly who knows how to type in the English language and provides personalized diagnosis of the user's health.It heavily relies on machine learning algorithms as well as the data to train the chatbot as well as predict diseases. In the future , the the conversational flow can be improved with better API's. Given, the disease is predicted accurately , the chat-bot can prescribe medicine to the user .A voice based input can also be added to the chatbot to make it more user-friendly.The database can be improved by adding all sorts of diseases and symptoms to make it more easy to use.

Referencias

[1] Gajendra Prasad K.C, satvik ranjan, Tathagat Ankit, Vivek Kumar, .A personalised medical chatbot: MediBot"

[2] Benilda Eleonor V. Comendador, Bien Michael B. Francisco, Jefferson S. Medenilla, Sharleen Mae T. Nacion, and Timothy Bryle E. Serac, "PharmaBot: A Pediatric Generic Medicine Consultant Bot"

[3] Krishnendu Rarhi, Abhishek Mishra, Krishnasish Mandal, .Automated Medical chatbot"

[4] Mrs Rashmi Dharwadkar, Dr. Mrs. Neeta A DEshpande, .A Medical Chatbot"