Building Medical Chatbot system using Decision Trees

Abhishek Kuriyal

Abstract— **This report presents use of Decision Trees in modelling an intelligent chat based medical assistant that would determine kind of body infection or disease using symptoms entered by the user. The decision tree algorithm uses top to down searching approach to identify and diagnose the problem and suggests a possible infection along with some measures to eliminate it. The approach of model is question based where the user is queried about various symptoms and the model arrives at a final decision. We will explore the impact of decision trees along with limitations and possible improvements.**

I. INTRODUCTION

One of the man's greatest discovery within the field of Artificial Intelligence(AI) is constructing a model that has an ability to learn [1]. Such model will keep advancing and improve their performance provided necessary amount of data is fed into it. The data acts as heuristic for predicting results that are computationally hard for a normal mathematical algorithm. The idea of using an AI system to diagnose and provide remedy for daily health issues can save lots of time and money of visiting and waiting for doctors at clinics. However, building an efficient AI system is tricky and can become fatal if the decision made by the model is different than actual disease of patient.

In this project, the scope is to create such a model that can provide high accuracy and does not provide a radically different decision than actual. It should be taken into account that aim of the model is not to replace the purpose of doctor or medical expert but rather provide a friendly interface that could list all valid possible infections a patient might be suffering from.

II. METHODOLOGY

The architecture of model is simple to interpret. The model essentially uses Decision Trees[2] as the driving algorithm. A decision tree is a classifier expressed as a recursive partition of the instance space. The decision tree consists of nodes that form a rooted tree, meaning it is a directed tree with a node called "root" that has no incoming edges. All other nodes have exactly one incoming edge. A node with outgoing edges is called an internal or test node. All other nodes are called leaves (also known as terminal or decision nodes). In a decision tree, each internal node splits the instance space into two or more sub-spaces according to a certain discrete function of the input attributes values. Each leaf is assigned to one class representing the most appropriate target value. Alternatively, the leaf may hold a probability vector indicating the probability of the target attribute having a certain value. Instances are classified by navigating them from the root of the tree down to a leaf, according to the outcome of the tests along the path. Given this classifier, the analyst can predict the response of a potential customer (by sorting it down the tree), and understand the behavioral characteristics of the entire potential customers population regarding direct mailing. Each node is labeled with the attribute it tests, and its branches are labeled with its corresponding values. User will be queried with symptoms he is suffering with, which the model will use to navigate through the decision tree. The output will be determined when complete tree will be traversed from top to bottom.

Entire architecture can be visualized using following flow chart diagram:-

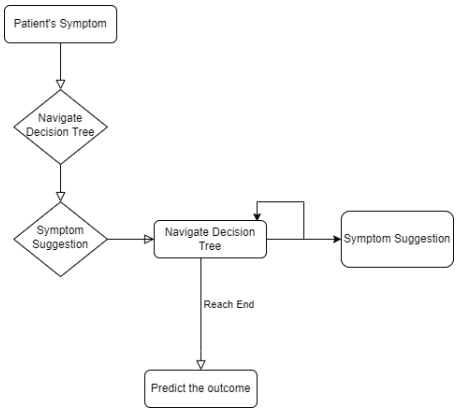

FIG 1: Model architecture

The decision tree search algorithm used in our project drives the perception based AI softbot questionnaire, as illustrated in figure 2. We get new perception at every input from the user, which determines the next level of question, therefore the use of tree search algorithms facilitates the system through match strategies (input), if there is no match, and then the system would continue the loop until end of tree is reached.

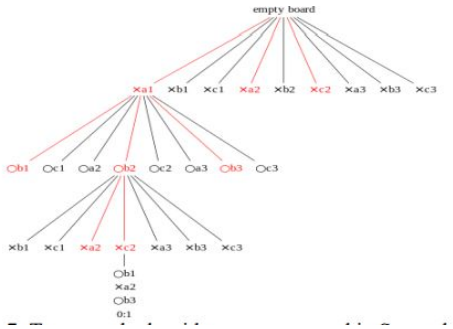

FIG 2: Tree search algorithm as used in model

Based on tree search, the smart doctor uses a question and answer approach. In this system, each question is mapped with an id and then cross-referenced with symptoms provided by the user input and then again based on this a new question is formed and asked for the user.

The decision algorithm used for simple disease uses a basic tree structure with minimum nodes as shown in figure 3.

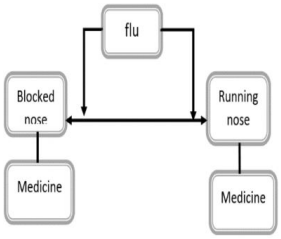

FIG 3: Basic Decision Tree

The decision tree logic for a complex disease with more symptoms is created with more complex node structure based on multiple user quires and responses as illustrated in figure 4.

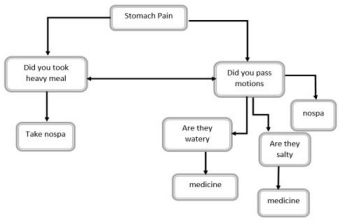

FIG 4: Complex Decision Tree

III. RESULT AND DISCUSSION

The final model is a working prototype using more than 40 diseases and their diagnostics at several layers of tree search. Table 1 describes the symptoms asked by patient from model and Table 2 describes the symptoms suggested by model to patient.

Serial No	Symptoms asked by Patient
1	I have pain in my neck
2	I have pain in my stomach
3	I have got sore throat

Table 1: Symptoms asked by patient from model

Serial No	Suggestions given by System
1	Do you have vomiting too
2	Do you have diarchic
3	Do you ear-ache

Table 2: Symptoms asked by system to the patients

The output of the model is shown in figure 5

```
You should take the consultation from doctor.
You may have  Arthritis
Arthritis is the swelling and tenderness of one or more
ness, which typically worsen with age. The most common
Please take following measures :
1 ) exercise
2 ) use hot and cold therapy
3 ) try acupuncture
4 ) massage
```

FIG 5: Output of the model

IV. CONCLUSION AND FUTURE WORK

Artificial Intelligence is becoming a norm of every technological development. Through this project the gap between Ai and health science has been narrowed down and a new possibilities have been introduced by which AI can be used to facilitate human development. The project uses a AI system to develop an efficient chatbot system based on rational agent with tree searching algorithm and knowledge base of medical data. The chatbot asks sever questions from the user or patient and then based on the replies suggests a possible remedy. In future, the system need to increase its database and machine learning capabilities for much improved diagnostics. The use voice as input would also be an effective feature. The system needs to go to deeper level of tree from 3 to 7 and be more accurate.

REFERENCES

[1] Michalski, R. S., Carbonell, J. G., Mitchell, T. M. (Eds.), "Machine learning: An artificial intelligence approach" Springer Science Business Media, pp.92, 2013.

[2] Lior Rokach, Oded Maimon, "The Data Mining and Knowledge Discovery Handbook" pp.165-192, 2005.