Significant Pattern Mining of Clinical Data for Diagnosis of Diseases

Mrs. Pushpavalli K
Department of Information Technology, Agni College of Technology, Chennai
pushpavalli.it@act.edu.in

Mr.Sachin S.
Department of Information Technology, Agni College of Technology, Chennai
sachinsmpth@gmail.com

Mr. Lokeshwaran R.
Department of Information Technology, Agni College of Technology, Chennai
lokeshramu16@gmail.com

***Abstract*— The chances of a perfectly good lifestyle mainly depends up on healthy we are. The sooner we diagnose a disease, the more the healthier we can be. A proper diagnosis will help to treat the disease accordingly towards a healthy lifestyle. In this project, the diseases can be predicted based on the analysis from their symptoms and the report is generated from the systematic analysis of particular disease. It is a multilayered method which uses significant pattern mining process using iterative search technique to build a disease prediction system which predicts various diseases. It is efficient, user friendly, time and cost saving.**

Keywords—significant pattern mining; Iterative search; disease prediction

I. INTRODUCTION

In the pattern mining algorithmn,we use iterative search to find the exact data from the unorganised data pool. Disease Prediction plays an important role in data mining. Data Mining is used intensively in the field of medicine to predict diseases such as heart disease, lung cancer, diabetics etc. Medicinal data mining has high potential for exploring the unknown patterns in the data sets of medical domain .These patterns can be used for medical analysis in raw medical data. The death rate of many countries increased as many diseases were not cured during its earlier stage. Half of the deaths occurring in the developing countries like India are due to various common diseases and people were not aware of these diseases. The main purpose of the project is to detect or predict the diseases based on the symptoms obtained from the user and the report is generated from that predicted analysis. This helps in detection of a person’s previous history for any common diseases before going for clinical and lab tests which is cost and time consuming. A feedback option is provided to the user and the feedback is incorporated into a classification model that predicts relevant to the user data objects. An initial sample set is selected and the iterative data exploration is initiated: the user provides his feedback on the sample set by labeling data objects as either relevant or irrelevant to him (User Relevance Feedback). The labeled samples are the training set of a classification algorithm that generates a model to characterize the user interests (Data Classification). In the next iteration, new labeled samples are incorporated to the training set and a new classification model is built. Given the current classification model, we identify promising data to be sampled further and can improve the effectiveness of our system (Space Exploration). Finally, we retrieve the next sample set from the new exploration data areas and we present them to the user for feedback (Sample Extraction).

II. DISEASE PREDICTION FRAMEWORK

In this section we tell about existing system, proposed system and introduce our technological background, system model and the algorithms used.

A. Existing System

In the existing system, data exploration was limited only to data mechanisms which is difficult to learn. A user can read a stale data for a period of time. Obtaining interesting results with traditional pattern mining methods can be tough and time- consuming job. Background knowledge of the domain expert is not taken into account. As a solution to the redundancy problem in pattern mining, a recent trend is to mine pattern sets instead of individual patterns. It can deal with interesting measures that are completely objective. Careful tuning of the algorithm parameters and manual filtering of the results is necessary. Algorithms still need to be tuned by running the algorithm, waiting for the final results, changing the parameters, re-running, waiting for the new results, etc.

B. Proposed System

Our approach offers an iterative exploration model in each iteration the user is prompted to provide her feedback on a set of sample objects as relevant or irrelevant to her exploration task. This technique leverage the properties of classification models to identify single objects of interest, expand them to more accurate areas of interests and progressively refine the prediction of these areas. It employs a unique combination of machine learning, data exploration, and sample acquisition techniques to deliver highly accurate predictions of linear patterns of user interests with interactive performance. This iterative process aims to generate a user model that identifies all relevant objects while eliminating the misclassification of irrelevant ones.

C. Technological Background

The common features of all the disease prediction systems that are designed to provide the patients specific information with an application that provides an interface to enter patient's information to access the knowledge base according to entered data and finally provides the required information. It can be implemented using a variety of platforms (i.e. Internet-based and LAN based).

The prediction system model is presented in Fig.1.

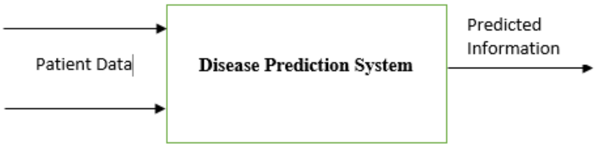


Fig.1. Prediction System Model

D. System Model

The workflow of our exploration framework is depicted in Fig. 2. The model presents to the user sample database objects and requests her feedback on their relevance to her exploration task, i.e., characterize them as relevant or not. For example, in the domain of evidence-based medicine, users are shown sample clinical symptoms and they are asked to review their abstract and their attributes (e.g., year, outcome, patience age, medication dosage, etc.) and label each sample symptom as present or not. The model allows also the user to explain samples that are similar (in some attribute) but not match exactly her interest, by marking them as "similar" symptoms. Finally, the user can modify her feedback on previously seen symptoms, however this could prolong the exploration process. The iterative steering process starts when the user provides the feedback by labelling samples are relevant or not.

The relevant and irrelevant samples are used to train a binary classification model that characterizes the user's interest, e.g., it predicts which clinical trials are relevant to the user based on the feedback collected so far (Data Classification)

In each iteration, more samples (e.g., records of clinical trials) are extracted and presented to the user for feedback.

The model leverages the current user model as well as the user's feedback so far to identify promising sampling areas (Space Exploration) and retrieve the next sample set from the database (Sample Extraction). Optionally, the model "translates" the classification model into a query expression (Query Formulation). This query will retrieve objects characterized as relevant by the user model (Data Extraction Query).

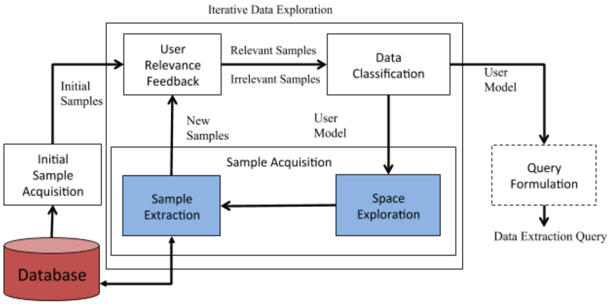


Fig.2. Architecture Diagram

E. Classification Algorithm

Pattern mining using Iterative search

It is a classification algorithm concerned with finding statistically relevant patterns between data examples where the values are delivered in a sequence. It is usually presumed that the values are discrete, and thus time series mining is closely related, but usually considered a different activity. Sequential pattern mining is a special case of structured data mining.

Let us assume a decision tree classifier that predicts relevant and irrelevant diseases based on the specific attributes that are mentioned.

- Symptom A – Symptom B – Symptom D -> Disease 1
- Symptom A – Symptom C – Symptom F -> Disease 2
- Symptom A – Symptom C – Symptom E – Symptom G -> Disease 3

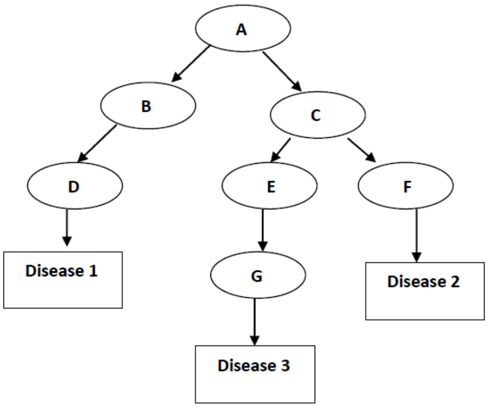


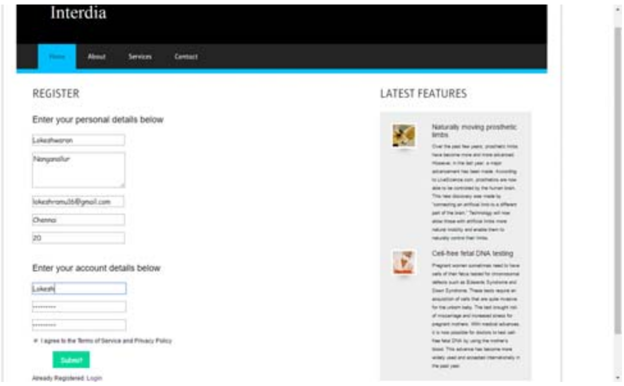
Fig.3. Tree showing Frequent Pattern of Symptoms with Diseases

III. IMPLEMENTATION

The system has been identified to have the following modules

A. Registration

The Registration is the process of making the user to register on our sites. If the user gets registered in our sites then the prominent user will allow to proceed the other Process. Once they choose the Symptoms from the list of symptoms then they will be followed with that plan.Only registered users can access the site. The user will get the notication when they logout the session.



REGISTER

Enter your personal details below

Lokeshwaran

Nanganallur

lokeshramu36@gmail.com

Chennai

20

Enter your account details below

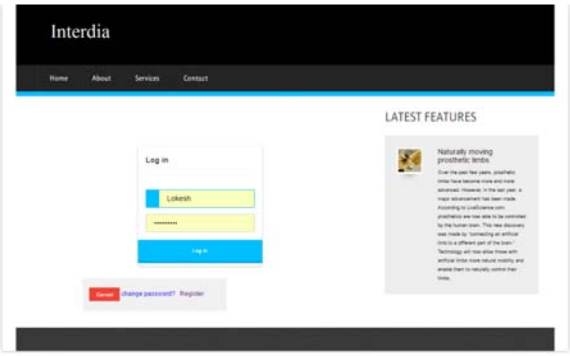
Lokesh

☑ I agree to the Terms of Service and Privacy Policy

Already Registered: Login

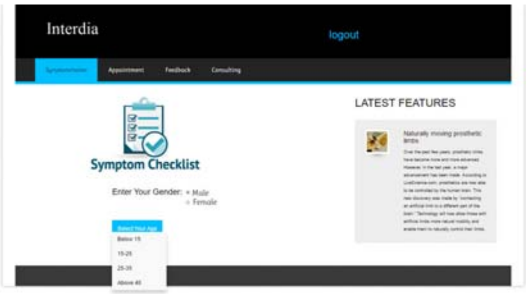
B. Login

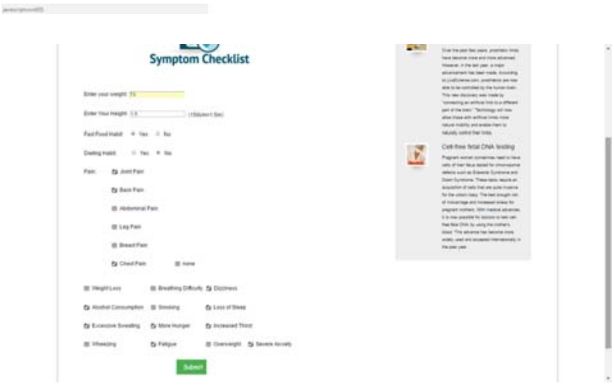
The users who register their details to the site and accessing it are called login. They have the following permissions. The user can book the appointment, can send report to the suggested doctors and can view comments from doctor.

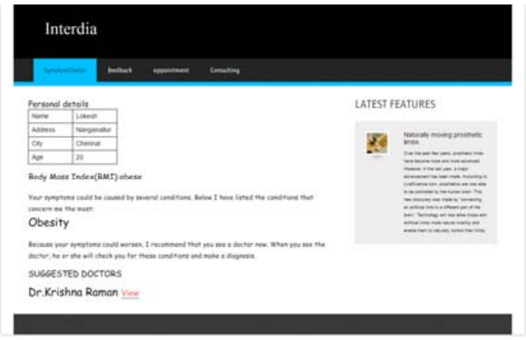


C. *Disease Prediction*

Here the user check the symptoms in the checklist according to his/her age. Additionally the weight and height also taken to know their obesity. After checking the symptoms in the list, the result is shown as a report with respective doctor for consulting. The disease is predicted based on Significant Pattern Mining using iterative search algorithm. The Pattern is formed according to the symptoms given. This pattern leads to a particular disease as result. This disease is given in the report to the user with the specialist name in that disease. The report also has the Body Mass Index (BMI) that gives the users BMI by calculating from their weight and height given. The personal details are given in the report that is generated. This helps to provide the doctor with all the necessary details pertaining to the disease and a suitable prognosis procedure to be identified by the doctor. Since there are various disadvantages over providing a possible cure for the predicted disease, we have decided to omit the same and leave the prognosis part to the concerned doctor. Therefore when the doctor logs in to his account, a report containing the details of the patients will be available along with the option of replying the possible prognosis along with other feedbacks.







Personal details

Name	Lokesh
Address	Nanganallur
City	Chennai
Age	20

Body Mass Index(BMI):obese

Your symptoms could be caused by several conditions. Below I have listed the conditions that concern me the most:

Obesity

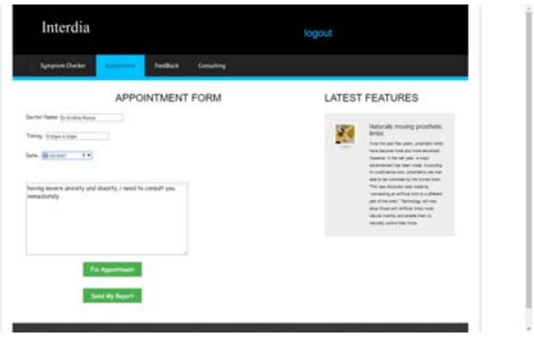
Because your symptoms could worsen, I recommend that you see a doctor now. When you see the doctor, he or she will check you for these conditions and make a diagnosis.

SUGGESTED DOCTORS

Dr.Krishna Raman

D. *Appointment Fixation*

User will fix the appointment with the individual doctor for Report consultation for the Particular day. Doctor needs to view the appointment for the particular day and check with their appointments.Then the token will be automatically generated to the user. The Doctor can view the entries of patients. The admin has the ability to provide excellent service with a commitment to users.

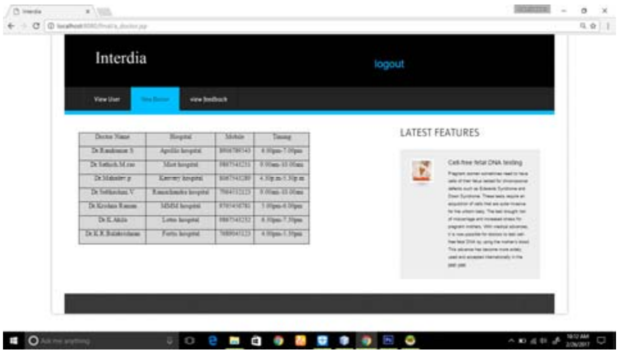


APPOINTMENT FORM

having severe anxiety and obesity, i need to consult you immediately

E. *Admin view Implementation*

Admin has the responsibility to maintain their user details. They also have privileges to create an announcement that can be viewed by all the users. Their roles and responsibilities are viewing the users and doctors details.



Doctor Name	Hospital	Mobile	Timing
Dr.Ramkumar S	Apollo hospital	[illegible]	4.30pm-7.00pm
Dr.Sathish.M.rao	Miot hospital	[illegible]	9.00am-10.00am
Dr.Mahadev p	Kauvery hospital	[illegible]	4.30p.m-5.30p.m
Dr.Sethuraman.V	Ramachandra hospital	[illegible]	9.00am-10.00am
Dr.Krishna Raman	MMM hospital	[illegible]	5.00pm-6.00pm
Dr.K.Akila	Lotus hospital	[illegible]	6.30pm-7.30pm
Dr.K.R.Balakrishnan	Fortis hospital	[illegible]	4.00pm-5.30pm

IV. CONCLUSION

This paper presents a prediction system that is capable of predicting the various different diseases and generate reports based on that to have consultation appointments with the respective doctors. This architecture supports the management needs of the new services, simplifying the service operation.

V. FUTURE ENHANCEMENT

Server provides the ability to group servers together in a cluster so that the application can be protected from the failure of a single server that the application workload can spread out across a number of equivalent servers. The service integration bus is also configurable within the application server cluster in a variety of configurations depending upon whether you are clustering for high availability, workload management or both. For example you can choose how many messaging engines are configured in the cluster (from one up to the number of servers in the cluster) and you can choose the server to which a given messaging engine fails over if its primary server fails. The accuracy of the prediction system can also be increased by increasing the number of patterns that are most unlikely to be used.

VI. REFERENCES

[1] Veerash Patel, Trishansh Bhardwaj, Mohammad Adhil, Asoke K. Talukder, “Big Data Analytics of Genomic and clinical Data for diagnosis and prognosis of cancer” in Sustainable Global Development.

[2] Kyriaki Dimitriadou , Olga Papaemmanouil and Yanlei Diao, “AIDE: An Active Learning-Based Approach for Interactive Data Exploration” in Computer Society 10.1109/TKDE.2016.2599168.

[3] Smitha T, Sundharam V, “An Efficient Prediction of Diseases using Pattern Mining” in Internation Journal for Computer Applications.

[4] Manju.K.K, Srinitya.G,”Analysis and Prognosis of Cancer with Big Data Analytics” in IJRASET.

[5] Maja Hadzic, Elizabeth Chang “Ontology based support for Human Disease Study” in Hawaii International Conference on System Sciences.

[6] Kameswara Rao N.K, Saradhi Varma G.P, “A Hybrid Algorithm for Epidemic Disease Prediction with Multi Dimensional Data” in ijarcsse V4I3-0489.

[7] Felipe Llinares-Lopez, Mahito Sugiyama, Laetitia Papaxanthos “Fast and Memory Efficient Significant Pattern Mining via Permutation Testing” in Mining Applications, BSSE.

[8] Prasan Kumar Sahoo, Suvendu Kumar Mohapatra, Shih-Lin Wu, “ Analyzing Healthcare BigData with Prediction for Future Health Condition” 2169-3536.

[9] Feldman B, Martin EM, Skotnes T, ‘BigData in Healthcarre Hype and Hope” October 2012. Dr. Bonnie 360,2012.

[10] Goh K, Cusik M, Valle D, Childs B, Vidal M, “The human disease network” in National Academy of Sciences 104:8685.

[11] Natarajan N, Tewari A, woods J. O, Dhillon I. S, Marcott E. M., Singh-Blom U. M. “Prediction and Validation of Gene-Diseases Associations Using Methods Inspired by Socail Network Analyses” in PULOS one.

[12] U. S. N. L. o. M. (NLM), “MEDLINE Fact Sheet,” [online]. www.nlm.nih.gov/pubs/factsheets/medline.html.

[13] Terada A, Tsuda K, Sese J, “Premutation Procedure for combinational regulation Discovery, In IEEE International Conference on Bioinformatics and Biomedicine.

[14] Gopakumar S, Tran T, Nguyen T. D., Phung D, Venkatesh S, “Stabilizing high –dimentional prediction models using feature Graphs” IEEE journal of Biomedical and Health Informatics. 1044-1052.

[15] Ram S, Zhang W, Williams M, Pengetnze Y, “Predicting asthma related emergency department visits using Big Data”, IEEE journal of Biomedical and Health Informatics. 1216-1223