ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 5, Issue 6, June 2016

IDENTIFICATION OF MEDICINE BEHAVIOR WITH PREVENTION OF DISEASE

***Tiju.T**

***Student, Saveetha School of Engineering, Saveetha University, Chennai.**

+Anusuya S

+Associate Professor, Saveetha School of Engineering, Saveetha University, Chennai.

ABSTRACT

In earlier days it's a risk to evaluate diabetes to make it easier we apply association rule mining to electronic medical records (EMR) to discover sets of risk and their corresponding subpopulations that represent patients at particularly high risk of developing diabetes. Based on the affinity propagation the newly arrived objects were clustered. In each data sets were categorized into three major Varieties namely Symptoms (Name), Drug (Variety – Eg. Manufacturers), Dosage and period (Number of day) &. Based on these three clustering was formed an dataset. Based on data set we analysis whether the it create the following diseases like BP, Cholesterol, Diabetes. And we also recommend medicine for the particular diseases. If these are not fit into these three categories it will be considered as outlier and the data will not exceed to the user.

Key words- Aryl Hydrocarbon Receptor (AHR), Association rule, Cluster, Electronic medical records (EMR), Outlier.

I. INTRODUCTION

Association rule is a well researched method for discovering interesting relations between variables in large database. It is intended to identify strong rules discovered in databases using different measures of interestingness. The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. A cluster is a group of data objects that similar to one another within the same cluster and are unrelated to the objects in other cluster. The exist data objects that do not comply with the general manners or model of the data. Such data objects, which are grossly different from or inconsistent with the remaining set of data are called outliers.

II. RELATED WORK

This paper is based on determining which proteins affect the activity of Aryl Hydrocarbon Receptor (AHR) system. Learning a model that can accurately predict its activity when single genes are knocked out [1]. Experiments with results are presented when models are trained on a

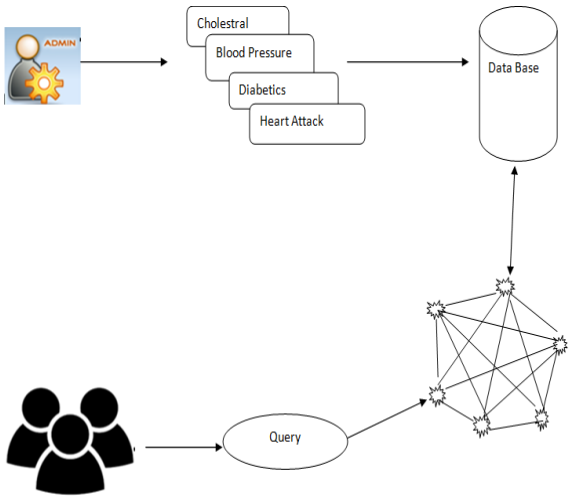
single source of information: abstracts from Medline (http://medline.cos.com/) that talk about the genes involved in the experiments. The results recommend that AdaBoost classifier with a binary bag-of-words representation obtains significantly better result.

Chien-Lung Chan et al.,used SPSS Clementine 8.1 package software to clean the data and Apriori algorithm is adopted to find the association rules[2]. It has been found that the prevalence of diabetes mellitus and metabolic syndrome will increase with age, especially in female patients. Chronic liver and the relation between Metabolic Syndrome and diabetes are the new findings from this research. From this research, it is proved that using data mining the relation of diseases that are similar with the results which are experimented from clinical.

The study is intended at finding out the characteristics that establish the existence of diabetes and to follow the maximum number of women suffer from diabetes[3]. Data mining concepts like clustering and attribute oriented induction techniques have been engaged to track the characteristics of the women suffering from diabetes. Information connected to the study was obtained from National Institute of Diabetes, Digestive and Kidney Diseases. The results were evaluate in five different clusters and they show that 23% of the women suffering from diabetes fall in cluster-0, 5% fall in cluster-1, 23% fall in cluster-2, 8% in cluster-3 and 25% in cluster-3. It was also bring into being that the characteristics seem to be varying for each cluster. This study helps to predict the state of diabetes and also helps in estimating the maximum number of women suffering from diabetes. [3]

Epidemics of seasonal influenza are a major public health concern, causing tens of millions of respiratory illnesses and 250,000 to 500,000 deaths worldwide each year1. In addition to seasonal influenza, a new strain of influenza virus against which no prior immunity exists and that demonstrates human-to-human transmission could result in a pandemic with millions of fatalities2. Early detection of disease activity, when followed by a rapid response, can reduce the impact of both seasonal and pandemic influenza3,4. One way to improve early detection is to monitor health-seeking behavior in the form of online web search queries, which are submitted by millions of users around the world each day. Here we present a method of analyzing large numbers of Google search queries to track influenza-like illness in a population. Because the relative frequency of certain queries is highly correlated with the percentage of physician visits in which a patient presents with influenza-like symptoms, we can accurately estimate the current level of weekly influenza activity in each region of the United States, with a reporting lag of about one day. This approach may make it possible to utilize search queries to detect influenza epidemics in areas with a large population of web search users.[4]

III. PROPOSED WORK



The architecture diagram explains that the admin of the system provide the enough information in the server, first the machine learns all the information provided then based on the frequent item sets the data are grouped into clusters, similar objects are grouped together i.e. Symptom, Drug and Dosage. The application must be used by the patient by registering their particulars they must be provide by the login id. When the patient query their symptoms as the input the searching process must be done in the clustering. The association rule mining take place in the searching process and the appropriate disease will be displayed using the hadoop tool. If the query are not fit into these three categories it will be considered as outlier and the data will not exceed to the user.

Methodology of association Rules

Association rule minig is a form of I→J. Here we using two types of rule mining techniques. They are predictive rule mining and regressive rule mining. In predictive rule mining, initialy the items were stored and from that the outcome is predicted.

Regressive and quantitative rule minig is a expanded form of predictive rule mining. In regressive rule mining the continious outcome to serve as a consequent rule.

$$RR = .y(I)/.y(-I)$$

Y(I) - affected subpopulation

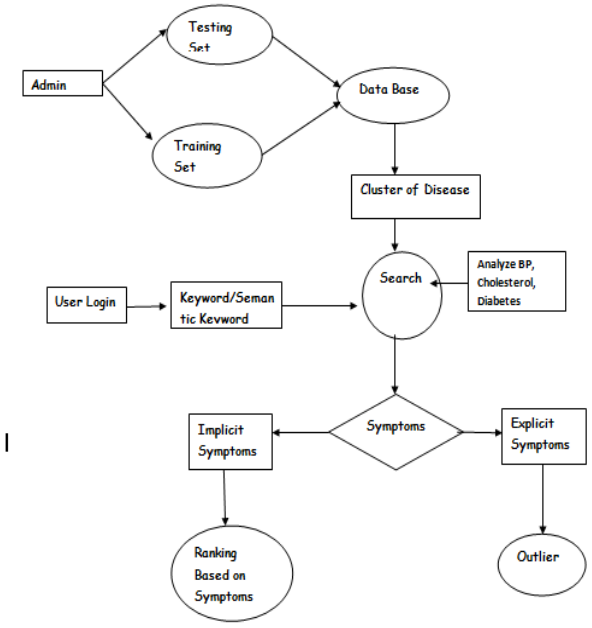
Y(-I) - unaffected subpopulation

R.R- Relative Risk. In some cases, the difference in the outcome between these two subpopulations cannot be suitably captured with the mean outcome, sometimes the distributional form of the outcome also plays a role. Distributional association rule can capture such differences. Distributional rule mining is defined by an itemset I and the continous outcome y. It's a distribution between affcted and unaffected sub population. The discovery of distributional rule consists of two steps. They are suitable itemset is descovered and itemset is filtered.

IV. OVERALL DESIGN PROCESS

In data mining, association rules are practical for analyzing and predicting customer behavior. Programmers use association rules to construct programs capable of machine learning. Machine learning is a type of artificial intelligence (AI) that seeks to construct programs with the capability to become more capable without being explicitly programmed. we apply association rule mining to mine knowledge from clinical data for predicting correlation of diseases carried by a patient [14]. Healthcare systems are accumulating huge quantities of information about patients and

their medical conditions everyday. The Cluster Analysis based model is suggested and discussed [16] for assigning prostate cancer patients into homogenous groups with the aim to support future clinical treatment decisions as an illustration.



The admin is responsible for the database to update the data inthe data base. The data were of two types they are testing set and training set. In database the disease are grouped into cluster based on their symptoms. To predict the disease the user must have their own login id and password. This is done by creating their own account by giving their basic particulars like name, gender, age etc into the application and then give their symptoms as the input, the search process is done in the cluster it will analyse the bp, cholestrol, sugar etc level. The serching process is done by association rule mining and the disease is mapped using hadooptool. If it's a implicit symptom the disease will be shown or if the symptoms are explicit it will lie under outlier.

V. IMPLEMENTATION

In software, a module is a part of a program. Programs are composed of one or more independently developed modules that are not combined until the program is linked. A single module can contain one or several routines.

1. User Registration
2. Research server
3. Pre Stored Data Comparison
4. Predictive disease Analysis
5. Suggestive Alternative Drug

a. USER REGISTRATION

In this module we are going to create an User application by which the User is allowed to access the data from the Server. Here first the User want to create an account by providing their basic details and then they are allowed to access the Network. Once the User create an account, they are allowed to login into their account to access the application. Based on the User's request, the Server will respond to the User. All the User details will be stored in the Database of the Server. In this Project, we will design the User Interface Frame to Communicate with the Server through Network Coding using the programming Languages like Java/ .Net.

b. RESEARCH SERVER

The research server provides secure storage space for certain project data stemming from externally funded research and other sponsored activites.Server space allocations and utilization are managed and monitored by the research officer, in collaboration with hospital to ensure the efficient use of

resources and limit excess unused capacity. Server maintenance and support is managed. In this module we designed a data server which store all the medical information regarding the facts given by the doctor and by the reasearch done by people who may have medical knowledges. The informations were about the disease and their drugs with respect to the symptoms and basic particulars about the patient.

c. PRE STORED DATA COMPARISON

In this module doctor will import all the details about the medicine i.e. what are the symptoms, dosages and drug. And how will store more about of the medicine so that we can make some use of it for example we can give awareness to the society. we store the all the data in the clustering format so that data can spitted and stored in the different clusters. Similar data were grouped into one cluster and different were stored in another. And the drugs were grouped based on their dosage and symptoms. So that it will easily to classify the data for the research.

d. PREDICTIVE DISEASE ANALYSIS

In this module we implement predictive disease analysis system in which the data will be analysis so that we can predictive the disease based on the symptoms. This module interact with server to analysis, the analysiation is done by the researchers. So they get the data from the server to make analysis to find the disease based on the symptoms. For an example if a patient may have a symptoms like continious headache it may because of tensin or brain tumer or some other disease. But the correct disease may predict using the basic particulars which may provided by the patient. If the basic particulars were not provides the machine may predict the disease with respect to the symptoms it may lead to the wrong prediction.

e. SUGGESTIVE ALTERNATIVE DRUGS

In this we automate the suggestion of the alternative drugs. So that we can provide the medicine to diseases i.e. a preferred medicine. So the researcher will analysis the dosage of the drug and the symptoms. The drugs for the same disease for each patient may get differ based on their symptoms, body conditions and also by the reaction of the drugs to the patient, because all the drugs were wont cure all the patients some patients may affect by the side effects of particular drugs based on their body conditions. If in this it gives the best result in the analysis they prefer the best medicine for the list of disease mentioned in the proposed sytem.

VI. CONCLUSION:

The electronic data generated by the use of EMRs in routine clinical practice has the potential to make possible the discovery of new knowledge. Association rule mining coupled to a summarization technique provide a critical tool for clinical research. It can uncover hidden clinical relationships and can propose new patterns of conditions to forward prevention, management, and treatment approaches. For this method to be useful, the number of rules required to be reduced to a level where clinical interpretation is feasible. To this end, we studied four methods to review these rules into sets of 10-20 rules that clinical investigators can evaluate. While all four methods created reasonable summaries, each method had its clear strength. However, not all of these strengths are necessarily

beneficial to our application. We establish that the most important differentiator between the algorithms is whether they use a selection criterion to include a rule in the outline based on the expression of the rule or based on the patient subpopulation that the rule covers.

REFERENCE

[1]Oana Frunza and Diana Inkpen, "Textual information for predicting functional properties of gene" Current Trends in Biomedical Natural Language Processing, pages 110–111, Columbus, Ohio, USA, June 2008.

[2] Chien-Lung Chan, Chien-Wei Chen, Baw-Jhiune Liu. "Discovery of Association Rules in Metabolic Syndrome Related Diseases" 978-1-4244-1821-3/08/$25.00 c2008 IEEE.

[3]P.Padmaja, SrikanthVikkurty, NiloferInazSiddiqui, Praveen Dasari, BikkinaAmbica, V.B.V.E.VenkataRao, MastanValiShaik, V.J.P. RajuRudraraju, Department of Information Technology Gitam University, Visakhapatnam, Andhra Pradesh, INDIA "Characteristic evaluation of diabetes data using clustering techniques".

[4]. F. Afrati, A. Gionis, and H. Mannila, "Approximating a collection of frequent sets," in Proc. ACM Int. Conf. KDD, Washington, DC, USA, 2004

[5]. Xiaofeng Zhao School of Economics and Management Hebei University of Engineering Handan, China Qboy_best@163.com Liyan Jiao, Jinping An, Li Wang Affiliated hospital Hebei University of Engineering Handan, China Jiaoliyan88@163.com Lipin Jia Medical School Hebei University of Engineering Handan, China Jiaoliyan88@163.com."A Data Mining Method on the Study of Medical Information"

[6].Emirhan Gülçin Yldma, Adem Karahocaa, Tamer Uçara * a Bahcesehir University, Software Eng. Dept., "Dosage planning for diabetes patients using data mining methods" Istanbul 34353, Turkey

[7]. R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in Proc. 20th VLDB, Santiago, Chile, 1994.

[8] Y. Aumann and Y. Lindell, "A statistical theory for quantitative association rules," in Proc. 5th KDD, New York, NY, USA, 1999.

[9] P. J. Caraballo, M. R. Castro, S. S. Cha, P. W. Li, and G. J. Simon, "Use of

association rule mining to assess diabetes risk in patients with impared fasting glucose," in Proc. AMIA Annu. Symp., 2011.

[10]. K. C. Tan, Member, IEEE, Q. Yu, and T. H. Lee, Member, IEEE "A Distributed Evolutionary Classifier for Knowledge Discovery in Data Mining" 1094-6977/$20.00 © 2005 IEEE.

[11].Emirhan Gülçin Yldma, Adem Karahocaa, Tamer Uçara * a Bahcesehir University, Software Eng. Dept., "Dosage planning for diabetes patients using data mining methods" Istanbul 34353, Turkey

[12] D.Senthil Kumar1, G.Sathyadevi2 and S.Sivanesh3 "Decision Support System for Medical Diagnosis Using Data Mining " IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 1, 2011

[13] Hye Soon Kim1, A Mi Shin2, Mi Kyung Kim1, and Yoon Nyun Kim1,2 "Comorbidity Study on Type 2 Diabetes Mellitus Using Data Mining" september 14, 2011.

[14] Eleni Georga, Vasilios Protopappas, Member, IEEE, Alejandra Guillen, Giuseppe Fico, Diego Ardigo, Maria Teresa Arredondo, Themis P. Exarchos, Student Member, IEEE, Demosthenes Polyzos, and Dimitrios I. Fotiadis, Senior Member, "Data Mining for Blood Glucose Prediction and Knowledge Discovery in Diabetic Patients: The METABO Diabetes Modeling and Management System" IEEE 978-1-4244-3296-7/09/$25.00 ©2009 IEEE .

[15] Mahmood A. Rashid1 , Md Tamjidul Hoque2 , Abdul Sattar1 1 Institute for Integrated and Intelligent Systems (IIIS), 2Discovery Biology, Eskitis Institute for Cell & Molecular Therapies, Griffith University Nathan, QLD, Australia. "Association Rules Mining Based Clinical Observations".

[16] Churilov L, Bagirov A, Schwartz D, Smith K, and Dally M," Improving Risk Grouping Rules for Prostate Cancer Patients with Optimization", the 37th International Conference on System Sciences, Hawaii 2004.