ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 15, March 2016

ELECTRONIC MEDICAL RECORDS FOR DIABETES MELLITUS USING ASSOCIATION RULE MINING

[1]Ms. G.RAHINI
Second Year [M.E]
Department of Computer Science and Engineering
Renganayagi Varatharaj College of Engineering
Sivakasi, Tamil Nadu, India
rahinirahi37@gmail.com, Indian

[2]Mr.P.DINESH KUMAR
Assistant Professor
Department of Computer Science and Engineering
Renganayagi Varatharaj College of Engineering
Sivakasi, Tamil Nadu, India
dineshoutlook@gmail.com, Indian

Abstract- Diabetes is part of the growing plague of non-catching diseases, with a high burden for the society on developing countries in future. For suppressing the development of diabetes mellitus and the onset of complications to manage their healthcare or personal data the system aim to apply association rule mining to electronic medical records to discover sets of risk factors. The four methods summaries the high risk of diabetes. The extension to the bottom up summarization algorithm produced the most suitable summary. Adjusted for confounders, advancing age, rural-urban migration, physical inactivity, smoking, abstinence of alcohol, low intake of fruits-vegetables, family history of DM, refined sugar intake, high social class, high intake of animal fat and protein, and stress, were the independents determinants of all cases of DM.Extended four popular associa-tion rule set summarization techniques by incorporating the risk of diabetesinto the process of finding an optimal summary.The performance of the test is also lessthan ideal. As a specific example, consider theaccuracy of the FPG test at a threshold of 126mg/dL when using the OGT test as the determinant

Index terms – Data mining, association rules, survival analysis, association rule summarization

INTRODUCTION

Diabetes mellitus is a chronic illness that requires continuing medical care and on going patient self-management education and support to prevent acute complications and to reduce the risk of long-term complications. Diabetes care is complex and requires multi factorial risk reduction strategies beyond glycemic control. A large body of evidence exists that supports a range of interventions to improve diabetes outcomes. These standards of care are intended to provide clinicians, patients, researchers, payers, general treatment goals, and tools to evaluate the quality of care.

The success of developing such a methodology fundamentally depends on our ability to accurately quantify the effect of a treatment in aphenotype.Suppose we wiquantify the effect of statins on diabetes in a phenotype defined by the association pattern {hypertension,renal failure}.In association rule mining, a naïve and commonly used technique is to directly compare the prevalence of diabetes among those who take statins and those who do not,among the patients presenting with hypertension and renal failure.The current medication regimens of all CT participants were inventoried at baseline and at years 1, 3, 6, and 9. In the OS,medication data were inventoried at baseline and year 3. At eachinventory, the brand or generic name on the medication labelwas matched to the corresponding item in the Master Drug DataBase (Medi-Span, Indian a polis, Indiana). We sorted for statinuse as users or nonusers at baseline and year 3.

Given that Sat-tar et al9found a null effect of lipophilicity among statins, andin the absence of dose information, we determined statin cat-egories by relative potency of action to decrease low-density lipoprotein cholesterol. Accordingly, statins were designatedas low (fluvastatin, lovastatin, pravastatin) or high (simvasta-tin, atorvastatin) potency.The sensitivity analyses also attempt to discover andresolve detection and/or selection bias, but it is possiblethat such biases remain. Second, we did not have data on blood lipid,C-reactive protein, or hemoglobin A1clev-els to distinguish if those using statins were at higher.

This results in a complete deficiency of the insulin hormone. Some people develop a type of diabetes, named secondary diabetes, which is similar to the more common Type 1 diabetes, in which the beta cells are destroyed by some other condition, such as cystic fibrosis or pancreatic surgery. The treatment for type1 diabetes is focused upon providing insulin through injections and control of diet. Association rule mining to identify sets of risk factors and the corresponding patient subpopulations that are at significantly increased risk of progressing to diabetes.The advantage of our approach over the pattern profile approach is clearly demonstrated in the figure.Occurrence statistics of itemsets at the lowest level are used to construct an initial MRF. to

support counselling with respect to their applicability merits and demerits

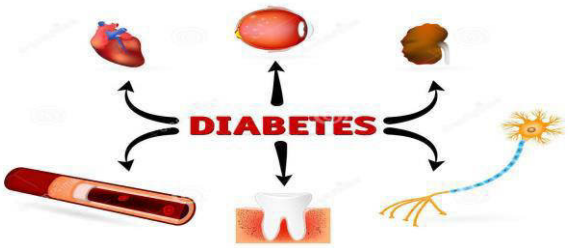

Figure 1 . Diabetes Mellitus

Diabetes mellitus is a growing epidemic, affecting more than 25 million people in the United States alone. In addition, an estimated 79 million people suffer from prediabetes14, defined by blood glucose levelsabove normal but below the threshold for the diagnosis of diabetes. Prediabetes is often accompanied by other comorbidities, such as obesity,hyperlipidemia and hypertension, which require appropriate treatment including the use of multiple drugs.In the case of hyperlipidemia,statintherapy is usually prescribed.While use of statins lowers cholesterol levels, and the overall risk of cardio vascular mortality there has been recent research indicating an increased risk of incident diabetes associated with their use.

I. RELATED WORK

Association rules are implications that associate a set of potentially interacting conditions (e.g. high BMI and the presence of hypertension diagnosis) with elevated risk. The use of association rules is particularly beneficial because in addition to quantifying the diabetes risk, they also readily provide the physician with a "justification", namely the associated set of conditions. This set of conditions can be used to guide treatment towards a more personalized and targeted preventive care or diabetes management. A clinical application of association rule mining to identify sets of co-morbid conditions that imply significantly increased risk of diabetes. Association rule mining on this extensive set of variables resulted in an exponentially large set of association rules. The main contribution is a comparative evaluation of these extended summarization techniques that provides guidance to practitioners in selecting an appropriate algorithm for a similar problem.

Association rule mining to identify sets of risk factors and the corresponding patient subpopulations who are at significantly increased risk of progressing to diabetes.An excessive number of associationrules were discovered impeding the clinical interpretation of the results. For this method to be useful, the number of rules is used for clinical interpretation is make feasible. Many of these rules are slight variants of each other leading to the obfuscation of the clinical patterns underlying the ruleset. One remedy to this problem, which constitutes the main focus of this work, is to summarize the ruleset into a smaller set that is easier to overview.and finally, we extend these methods so that they can take a continuous outcome variable (the martingale residual in our case) into account.

II. SYSTEM DESIGN

A distributional association rule is defined by an itemset I and is an implication that for a continuous outcome y, its distribution between the affected and the unaffected subpopulations is statistically significantly different.To apply rule set summarization techniques namely APRX-COLLECTION, RPGlobal, TopK, BUS to compress the original rule set commonly available in electronic medical record (EMR) systems to predict the Relative Risk of Diabetics Milletus of patients in the subpopulation. In our third module to apply rule set summarization techniques namely APRX- COLLECTION, RPGlobal, TopK, BUS to predict the Risk of Diabetics Millets.

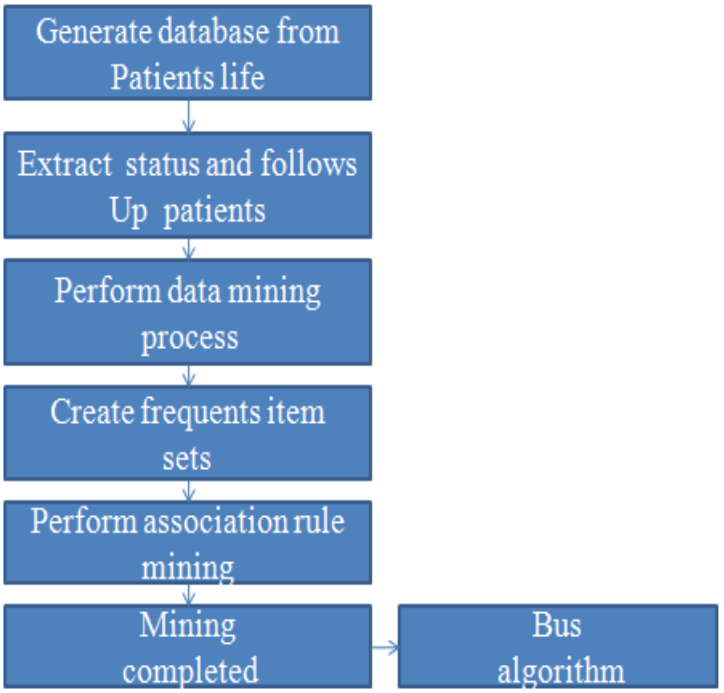

Figure 1: System Configuration

A. DATA LOADING

In our Process we have to load diabetes dataset to process.And then we have to insert the dataset on database dynamically.After that We also insert the new diabetes report on database.Dataset should be loaded after preprocessing automatically and also inserted into database newely whenever we run the process.

B. STATUS AND FOLLOW OF PATIENT

We have to extract the data based on status and follow up patient.Status patient are the people who caused by

diabetes in long year which based on dataset attributes. Follow Up Patient are the people who caused either diabetes at starting stage or not. Status Patient report are stored automatically because we have to find the high risk patient report for future purpose.

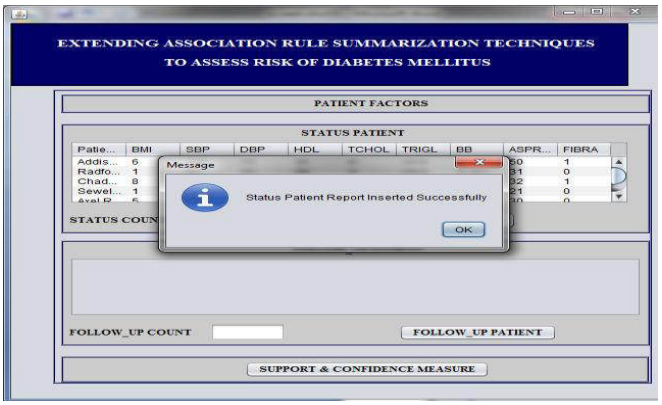

Figure 2.Status And Follow Of Patient

Classification

Assigning a type of diabetes to an individual often depends on the circum stances present at the time of diagnosis, and many diabetic individuals do not easily into a single class. For example, a person with gestational diabetes mellitus (GDM) may continue to be hyperglycemic after delivery and may be determined to have, in fact, type. Alternatively, a person who acquires diabetes because of large doses of exogenous steroids may become once discontinued, but then may develop diabetes many years later after recurrent episodes of pancreatitis.

Another example would be a person treated with t who develops diabetes years later. Because they seldom cause severe hyperglycemia, such individuals probably have type 2 diabetes that is exacerbated by the drug. Thus, for the clinician and patient, it is less important to label the particular type of diabetes than it is to understand the pathogenesis of the hyperglycemia and to treat it effectively.

C. ASSOCIATION RULE MINING

After finding the result of support and confidence to mining the report based on support count. And extract the resulting itemset from overall itemset. And then we extract the diabetes report based on itemset who are satisfy the condition and affected by symptoms.

a) Rules

A transaction t contains X, a set of items (item set) in if $X \subseteq t$. An association rule is an implication of the form is data mining technique based on machine learning. Basically classification is used to classify each item in a set of data into one of predefined set of classes or groups.

$X \rightarrow Y$, where $X, Y \subset I$, *and* $X \cap Y = \varnothing$

b) Clustering

Clustering is a data mining technique that makes meaningful or useful cluster of objects that have similar characteristic using automatic technique. Different from classification, clustering technique also defines the classes and put objects in them, while in classification objects are assigned into predefined classes. For example In prediction of heart disease by using clustering we get cluster or we can say that list of patients which have same risk factor.

c) Support and Confidence Measure

The support count of an item set X, denoted by $X.count$, in a data set T is the number of transactions in T that contain X. Assume T has n transactions. Given a transaction data set T, and a minimum support and a minimum confident, the set of association rules existing in T is uniquely determined.

$$support = \frac{(X \cup Y).count}{n}$$

$$confidence = \frac{(X \cup Y).count}{X.count}$$

To recap, in order to obtain A $\rightarrow$ B, we need to have support (A $\cup$ B) and support (A) All the required information for confidence computation has already been recorded in item set generation. No need to see the data T anymore.

D. RPC AND DATA COVERAGE METHOD

Relative Patient Coverage(RPC) can be extract from the status & follow up patient report who are caused by relative symptoms and affected by diabetes.This can be calculated through association rule mining and support and confidence measureData Coverage Method is based on RPC how many dataset are related and all these process are summazation to extract the data.The prediction as it name implied is one of a data mining techniques that discovers relationship between independent variables and relationship between dependent and independent variables. For instance, prediction analysis technique can be used in sale to predict profit for the future if we consider sale is an independent variable, profit could be a dependent variable. Then based on the historical sale and profit data, we can draw a fitted regression curve that is used for profit prediction.

E. BUS PROCESS

The Bottom Up Approch is the BUS Process that is we have to remove the related report in the dataset of patent in bottomwise.After calculate BUS,APRX,RPC all the three report result are merge to match the Status Patient report. And then extract the matched report based on result we have to finally get the High Risk Patient Report who are affected by diabetes in serious condition.

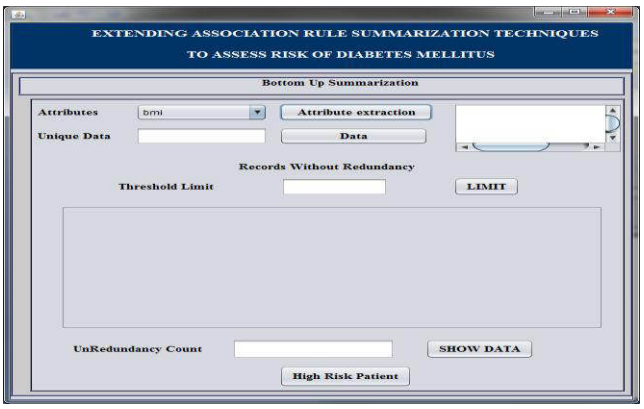

Figure 3.Bottom Up Summarization Process

III. CONCLUSION

Association rule mining to identify sets of risk factors and the corresponding patient subpopulations that are at significantly increased risk of progressing to diabetes. The system found that the most important differentiator between the algorithms is whether they use a selection criterion to include a rule in the summary based on the expression of the rule or based on the patient subpopulation that the rule covers. Between TopK and BUS, we found that BUS retained slightly more redundancy than TopK, which allowed it to have better patient coverage and better ability to reconstruct the original data base. There is direct evidence that the incidence of diabetes can be reduced in people at high risk of the future development of type2 diabetes who may be identified as a result of activities directed towards diabetes detection.

FUTURE WORK

In proposed system, the system proposes extensions to incorporate risk of diabetes into the process of finding an optimal summary. This system presents a clinical application of association rule mining to identify sets of co-morbid conditions that imply significantly increased risk of diabetes. The system evaluates some modified techniques on a real-world patient cohort. This proposed method extends those techniques to in corporate information about continuous outcome variables .we predict the type2 disease for high risk patient and then we give prescription for each patient and avoid the future risk. Then we will increase the accuracy by implementing SVM process. Then also identify the insulin based on stress, gene etc.

REFERENCES

[1] G. Fang et al., "High-order SNP combinations associated withcomplex diseases: Efficient discovery, statistical power and func-tional interactions," PLoS ONE , vol. 7, no. 4, Article e33531,2014.

[2] H. S. Kim, A. M. Shin, M. K. Kim, and N. Kim, "Comorbiditystudy on type 2 diabetes mellitus using data mining," KoreanJ. Intern. Med., vol. 27, no. 2, pp. 197–202, Jun. 2013.

[3] 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 patientswith impared fasting glucose," inProc. AMIA Annu. Symp. , 2012.

[4] Centers for Disease Control and Prevention. "Nationaldiabetes fact sheet: National estimates and general infor-mation on diabetes and prediabetes in the United States,"U.S. Department of Health and Human Services, Centers forDisease Control and Prevention,2011[Online].Available:http://www.cdc.gov/diabetes/pubs/factsheet11.html.

[5] G. S Collins, S. Mallett, O. Omar, and L.-M. Yu, "Developing riskprediction models for type 2 diabetes: A systematic review ofmethodology and reporting,"BMC Med. , 9:103, Sept. 2011.videos," IEEE Trans. Biomed. Eng., vol. 58, no. 10, pp. 2777–2786, Oct. 2011.

[6] T. M. Therneau and P. M. Grambsch, "Modeling survival data:Extending the cox model," inStatistics for Biology and Health.Springer, 2010.

[7] R.Jin,M.Abu-Ata,Y.Xiang,andN.Ruan,"Effective and efficient pattern summarization:Regression-based approach," inProc. ACM Int. Conf. KDD , Las Vegas, NV, USA, 2008.

[8] M. A. Hasan, "Summarization in pattern mining," in Encyclopediaof Data Warehousing and Mining, 2nd ed. Hershey, PA, USA:Information Science Reference, 2008.

[9] C. Wang and S. Parthasarathy, "Summarizing itemset patternsusing probabilistic models," inProc. ACM Int. Conf. KDD ,NewYork, NY, USA, 2006.

[10] D. Xin, H. Cheng, X. Yan, and J. Han, "Extracting redundancy- awaretop-k patterns," inProc. ACM Int. Conf. KDD, Philadelphia,PA, USA, 2006.

[11] Yixuan Yuan, Baopu Li, and Max Q.-H. Meng, Fellow, association rule mining" Automation science and engineering, IEEE transactions on. pp.1545-5955, March 2005

G.Rahini received the B.E Degree in Computer Science and Engineering from Theni Kammavar Sangam College Of Technology,Theni,Tamilnadu, India. which is affiliated to Anna University, Chennai in 2012 and she is currently pursuing the M.E degree in Computer Science and Engineering in Renganayagi Varatharaj college of Engineering Affiliated to Anna University, Chennai. Her area of Interest is Data Mining and Java.

Mr.P.Dinesh Kumar is currently working as assistant professor in computer science and engineering department in Renganayagi Varatharaj College of Engineering. he received his B.E degree from P.S.R Engineering College, Sivakasi, Tamilnadu, India which is affiliated to Anna University, Chennai in 2008..He has a experience of above three years in teaching filed. His area of interest is soft computing, data mining