ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Identifying Diabetic Risk Factor Using Association rule Mining Based on Bottom up Summarization

Jyoti Bhabad, Kishorshedge

M.E. Student, Department of Computer Engineering, SVIT, Nasik, India

Assistant Professor, Department of Computer Engineering, SVIT, Nasik, India

ABSTRACT:Developing diabetes mellitus is critical to the improvedprevention and overall clinical management of this patient is early detection of patients with elevated risk. To discover sets of risk factors and their corresponding subpopulations that represent patients at particularly high risk of developing diabetes, to apply association rule mining to electronic medical records (EMR). In this four association rule set summarization techniques and conducted a comparative evaluation and proposed extensions to incorporate risk of diabetes into the process of finding an optimal summary. In this extension to the Bottom-Up Summarization (BUS) algorithm produced the most suitable summary. The subpopulations identified by this summary covered most high-risk patients, had low overlap and were at very high risk of diabetes.

KEYWORDS: Data mining, association rules, survival analysis, association rule summarization

I. INTRODUCTION

DIABETES mellitus is a growing epidemic that affects 25.8 million people in the U.S, and approximately 7 million of them do not know they have the disease including ischemic heart disease, stroke, nephropathy, retinopathy, neuropathy and peripheral vascular disease, Diabetes leads to significant medical complications a major healthcare need, early identification of patients at risk of developing diabetes. The risk of developing diabetes by 30% to 60%, appropriate management of patients at risk with lifestyle changes and/or medications can decrease. A large proportion of the population, multiple risk factors have been identified affecting. For example, prediabetes is present in approximately 35% of the adult population and increases the absolute risk of diabetes 3 to 10 fold depending on the presence of additional associated risk factors, such as obesity, hypertension, hyperlipidemia, etc.

In theoretical aspect, Afrati, A. Gionis, and H. Mannila, [1]describe One of the most well-studied problems in data mining is computing the collection of frequent item sets in large transactional databases. R. Agrawal and R. Srikant [2] problem of discovering association rules has received considerable research attention and several fast algorithms for mining association rules have been developed.

II. RELATED WORK

Association rules are a fundamental class of patterns that exist in data. The key strength of association rule mining is its completeness. It finds all associations in the data that satisfy the user specified minimum support and minimum confidence constraints. This strength, however, comes with a major drawback. It often produces a huge number of associations. This is particularly true for data sets whose attributes are highly correlated. The huge number of
The technique first prunes the discovered associations to remove those insignificant associations, and then finds a special subset of the unpruned associations to form a summary of the discovered associations. We call this subset of associations the *direction setting* (DS) *rules* as they set the directions that arefollowed by the rest of the associations. Using this summary, the user can focus on the essential aspects (or relationships) of the domain and selectively view

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

the relevant details. The approach is effective because experiment results show that the set of DS rules is typically very small. They can be analyzedmanually by a human user .The proposed technique has also been applied successfully to a number of real-life applications.

III. PROPOSED ALGORITHM

A. *Architecture:*
Following gives the description of blocks usedin architecture.

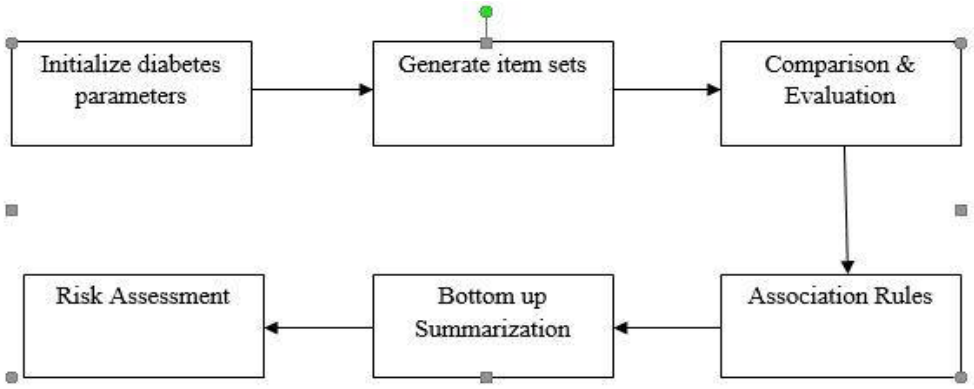

Fig. 1 proposed system architecture

1) Initialize Diabetes parameters: At first researchcenter sends the request to hospitals it reaches the hospitals for that request hospital admin replies whether it can access or not. If the hospital gives an access we can able to access the hospital database. Then the Observed patients in dataset database are fetched into the research center database with privacy preservation. Fetched database consists only needed information in clear manner now we can able to see the Observed patients in dataset medical information and id.

2) Discover Item sets and association rule:Groupingdata elements can belong to more than one cluster. In fuzzy grouping of data every point has a degree belonging to as in fuzzy logic rather than belonging completely to just one same cluster. Next step is finding of association rules by using the apriori hybrid algorithm. Used the apriori hybrid algorithm, a variant of the well-known Apriori algorithm that discovers candidate set of items that contain specific items the item corresponding to the (binary) diabetes

3) Comparison & Evaluation: Next step is to find theunsummarized rule for data. It consists of the comparative risk and complete risk of Observed patients in datasets. These values are calculated depends on the sugar level, BP, BMI,Tablets etc. Every value consists of some of particular defined value depends on the gender and age by summing that values we can calculate the comparative risk and absolute risk .

4) Association Rules: The general strategy for handlingsuch outcomes in distributional association rule mining is to construct a survival model and to use the residuals of the model as a continuous outcome for distributional

association rule mining.

5)Apply summarization techniques: The goal of *rule set(Bottom Up) summarization* is to represent a set *I* ofrules with a smaller set *A* of rules such that *I* can be recovered from *A* with minimal loss of information. Since a rule is defined by a single itemset, we will use ‘itemset’ in place of ‘rule’ meaning the ‘itemset’ that defines the rule’. Data set summarization is a slightly different problem

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

B. *Description of the Proposed Algorithm:*

The main idea behind the BUS algorithm is to incrementally select best candidates from the candidate set such that at each step, for a certain gain in compaction, minimum information loss is incurred. The definition of a "best" candidate is based on a heuristic decision and can be defined in several different ways as we will describe later.

Input: T : a transaction data set.
W : a set of weights for each feature.
l: size of final summary
Variables :Sc: current summary
Output : S: the final summary

Method:

1. Generate Cc = {All closed frequent itemsets of T } + T
2. While (|Sc| != l)
3. Cbest = select_ best(Cc,Sc,T)
4. Sc = Sc – {Summaries in Sc covered by Cbest}+Cbest
5. End While
6. S = Sc
7. End

IV. SIMULATION RESULTS

Following fig 2. Shows the result grap of patients coverage

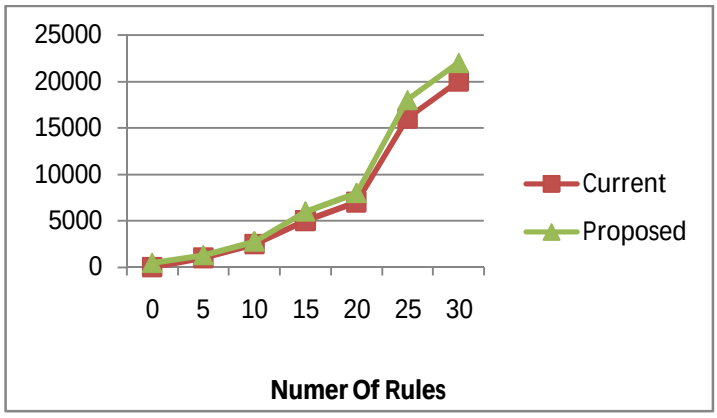

Fig. 2. Result Graph

The Graph in figure 2.shows Number of rules on X axisand Patients Coverage on y axis. The comparision between the current system and proposed system. The result increasescompared to current system

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

	Current	Proposed
0	0	500
5	1000	1300
10	2500	2800
15	5000	6000
20	7000	8000
25	16000	18000
30	20000	22000

Fig. 3.Result Table

V. CONCLUSION AND FUTURE WORK

The electronic data generated by the use of EMRs in routine clinical practice has the potential to facilitate the discovery of new knowledge. Association rule mining coupled to a summarization technique provides a critical tool for clinical research. It can uncover hidden clinical relationships and can propose new patterns of conditions to redirect prevention, management, and treatment approaches.

REFERENCES

1. F. Afrati, A. Gionis, and H. Mannila, "Approximating a collection of frequent sets," in *Proc. ACM Int. Conf. KDD*, Washington, DC, USA, 2004
2.]R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proc. 20th VLDB*, Santiago, Chile, 1994.
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 patients with impared fasting glucose," inProc. AMIA Annu. Symp., 2011.
4. G. S Collins, S. Mallett, O. Omar, and L.-M. Yu, "Developing risk prediction models for type 2 diabetes: A systematic review of methodology and reporting,"BMC Med., 9:103, Sept. 2011
5. B. Liu, W. Hsu, and Y. Ma, "Integrating classification and association rule mining," inProc. ACM Int. Conf. KDD,NewYork,NY
6. B. Liu, W. Hsu, and Y. Ma, "Pruning and summarizing the discovered associations," inProc. ACM Int. Conf. KDD,NewYork,NY, USA, 1999
7. D. Xin, J. Han, X. Yan, and H. Cheng, "Mining compressed frequent-pattern sets," in Proc. 31st Int. Conf. VLDB, Trondheim,Norway, 2005. [16] R. Srikant, Q. Vu, and R. Agrawal, "Mining association rules with item constraints," in Proc. AAAI, 1997.
8. T. M. Therneau and P. M. Grambsch, "Modeling survival data:Extending the cox model," inStatistics for Biology and Health.Springer, 2010
9. J. Tuomilehto et al., "Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impared glucose tolerance," N. Engl. J. Med., vol.344, no. 18, pp. 1343–1350, May 2001
10. C. Wang and S. Parthasarathy, "Summarizing itemset patterns using probabilistic models," inProc. ACM Int. Conf. KDD,New York, NY, USA,2006.

BIOGRAPHY

Miss. JyotiKisan Bhabad has received B.E degree in Information Technology from SND COE, Nashik. She is currently pursuingMaster Degree in Computer Engineering at SVIT, Chincholi, Nasik, India. Her research interest includes Data Mining ,DBMS

Prof. Kishor N. Shedge presently working as assistantprofessor in the Department of Computer Engineering at SVIT, Chincholi,Nasik.He received B.E.degree in ComputerScience And Engineering from Government College.