

EXTENDING ASSOCIATION RULE CHARACTERIZATION TECHNIQUES TO ASSESS THE RISK OF DIABETES MELLITUS

M.Lavanya¹, Mrs.P.M.Gomathi²

Research Scholar, Head of the Department,

P.K.R. College for Women, Gobi.

Abstract

Diabetes leads to significant medical complications including ischemic heart disease, stroke, nephropathy, retinopathy, neuropathy and peripheral vascular disease. Early detection of patients with elevated risk of developing diabetes mellitus is critical to the improved prevention and overall clinical management of these patients. Apply association rule mining to Electronic Medical Records (EMR) to discover sets of risk factors and their corresponding subpopulations that represent patients at particularly high risk of developing diabetes[1]. The high dimensionality of EMRs, association rule mining generates a very large set of rules which need to summarize for easy clinical use. The concept reviewed four association rule set summarization techniques and conducted a comparative evaluation to provide guidance regarding their applicability, strengths and weaknesses[5]. The proposed system extensions to incorporate risk of diabetes into the process of finding an optimal summary. Evaluated these modified techniques on a real-world pre diabetes patient cohort and found that all four methods produced summaries that described

subpopulations at high risk of diabetes with each method having its clear strength. The Bottom-Up Summarization (BUS) algorithm produced the most suitable summary [5]. The subpopulations identified by this summary covered most high-risk patients, had low overlap and were at very high risk of diabetes [8]. In this paper is used to find the high risk patients itemsets and discover it to the efficient manner.

Keywords:

***Datamining, Association rule mining
and associative classification, methods.***

I.INTRODUCTION

Diabetes mellitus is a group of metabolic diseases characterized by high blood sugar (glucose) levels, which result from defects in insulin secretion, or action, or both. Diabetes mellitus, commonly referred to as diabetes (and in this article, will be referred to as "diabetes"), was first identified as a disease associated with "sweet urine," and excessive muscle loss in the ancient world. Elevated levels of blood

glucose (hyperglycaemia) lead to spillage of glucose into the urine, hence the term sweet urine. Normally, blood glucose levels are tightly controlled by insulin, a hormone produced by the pancreas. Insulin lowers the blood glucose level. When the blood glucose elevates (for example, after eating food), insulin is released from the pancreas to normalize the glucose level. In patients with diabetes, the absence or insufficient production of insulin causes hyperglycaemia. Diabetes is a chronic medical condition, meaning although it can be controlled, it lasts a lifetime.

Over time, diabetes can lead to blindness, kidney failure, and nerve damage. Diabetes is also an important factor in accelerating the hardening and narrowing of the arteries (atherosclerosis), leading to strokes, coronary heart disease, and other blood vessel diseases. Diabetes affects 15 million people (about 8% of the population) in the United States. In addition, an estimated 12 million people in the United States have diabetes and don't even know it. From an economic perspective, the total annual economic cost of diabetes in 1997 was estimated to be 98 billion dollars in the United States. The per capita cost resulting from diabetes in 1997 amounted to \$10,071, while to health care costs for people without

diabetes incurred a per capita cost of \$2,699. During this same year, 13.9 million days of hospital stay were attributed to diabetes, while 30.3 million physician office visits were diabetes related. Remember, these numbers reflect only the population in the United States. Globally, the statistics are staggering.

II.DATAMINING

Data mining is a challenging area in the field of medical research. Extraction of useful knowledge from the database and providing scientific decision making for the diagnosis and treatment of diseases increasingly becomes necessary. Data mining in medicine can deal with this problem. Medical data mining has great potential for exploring the hidden patterns in the data sets of the medical domain Data mining is a logical process that is used to search through large amount of data in order to find useful data. A wide availability of huge amount of data and the need to convert such data to useful information necessitates the use of data mining techniques. It has become an established method for improving statistical tools to predict future trends. The necessity for using data mining techniques to develop and improve risk models arises from the need for clinicians to

improve their prediction models for individual patients.

Various data mining techniques are available with their suitability dependent on the domain application. Data mining applications in health can have a tremendous potential and usefulness.

III.ASSOCIATION RULES MINING AND ASSOCIATIVE CLASSIFICATION

There are different data mining techniques including supervised classification, association rules mining or market basket analysis, unsupervised clustering, web data mining, and regression. One technique of data mining is classification. The goal of classification is to build a model of the training data that can correctly predict the class of unseen or test objects. The input of this model learning process is a set of objects along with their classes (supervised training data). Once a predictive model is built, it can be used to predict the class of the objects of test cases for which class is not known. To measure the accuracy of the model, the available dataset is divided into training and test sets. The training set is used to build the model and test set is used to measure its accuracy.

There are several problems from a wide range of domains which can be cast into classification problems. Therefore there is always a need of algorithms for building comprehensible and accurate classifiers. Association Rules Mining (ARM) is another important data mining technique. It is used to find strong and interesting relationships among data items present in a set. A typical example of ARM is market basket analysis. In market basket analysis each record contains a list of items purchased by a customer. This research work focuses to find out the set of items that are frequently purchased together.

The objective is to search for interesting habits of customers. The sets of items occurring together can be written as association rules. These association rules can be written as “IF THEN” statements. IF part is called the antecedent of rule and THEN contains the consequent of the rule. In ARM the antecedent and consequent are sets of data items called item-set. An item set that contains k items is called k item set. An association rule is written as $A \Rightarrow B$, where A and B are set of items.

IV.EXISTING SYSTEM

In an existing system, a statistical modeling technique that constructs predictive models on time-to-event data

under censoring the patient records manually. Censoring takes place when we fail to obtain full information about a patient. For example, if a patient drops out of the study, we may know that he did not develop diabetes during the time period we could observe him, but do not know whether he ultimately developed diabetes by the end of the study. The ability to use such partial information and the ability to take time into account are the key characteristics of survival analysis making it a mainstay technique in clinical research.

V. PROPOSED SYSTEM

ALGORITHMS

The discovery of distributional association rules consists of two steps. In the first step, a suitable set of item sets is discovered and in the second step, the set of item sets is filtered so that only the statistically significant ones are returned as distributional association rules.

- **Item set Discovery:** Most if not all item set enumeration algorithms can be used to discover item sets. We used the Reorder algorithm, a variant of the well-known Apriori algorithm that only discovers candidate item sets that contain specific items the

item corresponding to the (binary) diabetes outcome in our case

- **Testing Statistical Significance:** For each discovered item set, we have to test whether the outcome distribution in the affected and the unaffected subpopulations are indeed different. A generic statistical test, the Wilcoxon test, is always applicable; however, when the outcome variable follows a known distribution more powerful tests can be applied. For example, when y represents the number of diabetes events (count outcome following Poisson distribution), the Poisson test is a better choice.

Advantages

- Reduce the Human Error.
- Easy to Process and reduce the time.
- Reduce the cost.

VI. METHODS

- Diabetic Retinopathy.
- Fetching Database Collection in EMR and APRX and RP Global Summarization.
- Top and BUS Summarization.

6.1.Diabetic retinopathy:

- Proposed a new constraint for optic disk detection where we first detect the major blood vessels and use the intersection of these to find the approximate location of the optic disk.
- Exudates and micro aneurysms and hemorrhages can be detected quite accurately using different morphological operations applied appropriately.
- The optic disk is detected using blood vessel convergence and high intensity property. In order to determine the features and classification method to be used for a given lesion, a Bayesian probabilistic system is used.
- This method is used to find diabetic whether present or not.

6.2.Fetching Database Collection in EMR and APRX and RP Global Summarization:

- Method two collect the emr data from suitable hospital or research field gather the data to process to write the suitable format here use xlsx format to align the all the data.

- Load process system button load all the data at very effectively process system in order to all the detail which correctly suitable software format.
- EMR data process to field used next bottom summarization process.

6.3.Bus summarization:

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 association rules 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.

This detail will gather to previous method retinopathy and emr detail loading process used to arrange the bottom up approach diabetic patients.

VII.CONCLUSION

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. For example, used distributional 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 association rules were discovered impeding the clinical interpretation of the results. Mainly this methods to be useful, the number of rules needed to be reduced to a level where clinical interpretation is feasible.

REFERENCES

- 1]. R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in Proc. 20th VLDB, Santiago, Chile, 1994.
- 2]. 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 impaired fasting glucose," in Proc. AMIA Annu. Symp., 2011.
- 3]. D. Guo and P. Richardson, "Automatic vessel extraction from angiogram images", *IEEE Computers in Cardiology*, vol. 25, pp. 441-444, 1998.
- 4]. F. Zana and J.C. Klein, "Robust segmentation of vessels from retinal angiography", in *IEEE International Conference on Digital Signal Processing*, vol. 2, pp. 1087-1090, 1997.
- 5]. V. Chandola and V. Kumar, "Summarization – Compressing data into an informative representation," *Knowl. Inform. Syst.*, vol. 12, no. 3, pp. 355–378, 2006.
- 6]. M. A. Hasan, "Summarization in pattern mining," in *Encyclopedia of Data Warehousing and Mining*, 2nd ed. Hershey, PA, USA: Information Science Reference, 2008.
- Blond L The road ahead in diabetes management, *Am J Manag Care* 11(5), S167-S169, 2005.
- 7]. Chand J.L., Abrahamson M.J., "Pharmacological management of type 2 diabetes mellitus Rashnale for rational use of insulin", *Mayo Clin. Proc*, 78 459467, 2003.
- 8]. Varun Kumar V, "Image processing based wound assessment system for patients with diabetes using six classification algorithms C", Malathy2016 *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*.