Expand the Association Rule Mining Technique for Finding Related Medication to Risk of Diabetes

Yegireddy Anil Kumar
Final M.Tech Student,
Dept of CSE,
Sarada Institute of Science, Technology and Management (SISTAM), Srikakulam,
Andhra Pradesh.

Konni Srinivasa Rao
Assistant Professor,
Dept of CSE,
Sarada Institute of Science, Technology and Management (SISTAM), Srikakulam,
Andhra Pradesh.

Abstract:

Diabetes is a growing epidemic of non-communicable disease which affects most of the people in the world. We aim to find the risk factors of diabetes using association rule summarization techniques. A number of successful association rule set summarization techniques have been proposed. In this paper we are proposed an extended association rule summarization technique for generating rules. Usually association rule mining generates large volume of data set which we need to summarize for any medical records or any clinical use. The use of association rules is particularly beneficial because in addition to quantifying the diabetes risk, the also readily provide the physician with a justification namely the associated set of conditions.

These conditions can be used to guide treatment towards a more personalized and targeted preventive care or diabetes management. The main aim of rule set summarization technique is to represent a set of rules with smaller set of rules such that can be recovered from with minimal loss of information. By getting minimal loss of information we are implementing risk factor summarization technique for generating rule set. By implementing that technique we can also find out risk factors and also get related medication for reduce the risk over the diabetes.

Keywords:

Data mining, association rule mining, optimal feature extraction, frequent patterns.

I. INTRODUCTION:

Diabetes mellitus is growing epidemic disease which affects more than 25.8 million people and approximately 7 million of them do not know they have this disease. Usually diabetes is a group of diseases characterized by high blood sugar (blood glucose). When a person has diabetes the body either produces enough insulin or unable to use its own insulin effectively. When glucose gets build up in our blood, that glucose should be controlled or must be effectively used else it may to lead death. The risk of death of a person who has diabetes is twice as the person who does not have diabetes of same age. Early identification of patients at risk of developing polygenic disease may be a major health care want. Appropriate management of patients in danger with manner changes and or medications will decrease the chance of developing diabetes by half-hour to hour. Diabetes may be an sort of diseases characterized by High blood glucose level (blood sugar). If a person has diabetes mellitus, the body either doesn't manufacture enough hypoglycemic agent or the body is unable to use its own insulin. Aldohexose builds up within the blood and causes a condition that, if not controlled,

Volume No: 3 (2017), Issue No: 3 (August)
www. IJRACSE.com

August 2017

Page 11

will result in serious health complications such as stroke and even death. The chance of death for someone with diabetes mellitus is double the chance of someone of comparable age who doesn't have diabetes mellitus. Diabetes mellitus have three types. Type 1diabetes mellitus - the body doesn't manufacture hypoglycaemic agent. Some 100% of all diabetes mellitus cases area unit sort one. Type 2diabetes mellitus in this the body doesn't manufacture enough hypoglycaemic agent for correct operate. Some ninetieth of all cases of diabetes mellitus worldwide area unit of that sort. Gestational diabetes mellitus - that sort affects females throughout maternity. The most common diabetes mellitus symptoms embrace frequent voiding, intense thirst and hunger, weight gain, unusual weight loss, fatigue, cuts and bruises that don't heal male sexual pathology, symptom and tingling in hands and feet. Association rules are implications that relate a set of potentially interacting conditions (e.g. max Body mass index and the occurrence of hypertension diagnosis) with eminent risk.

The use of association rules is mainly favourable, because in addition to signifying the diabetes mellitus risk, they also readily provide the physician with a ―justification‖, namely the related set of conditions. Namely co-morbid sickness, laboratory results, tablets and demographic information those are commonly available in electronic medical record (EMR) systems. With such an extensive set of factors of risk, the set of invented rule for data grows combinatorially large, to a size that severely hinders interpretation. To overcome that challenge, we applied rule for data set summarization techniques to compress the original rule for data set into a most compact set that can be interpreted with ease. Association rule mining, one of the most important and well researched techniques of data mining. It aspires to extract interesting correlations, frequent patterns and associations among sets of items in the transaction databases.

Let an item be a binary indicator signifying whether a patient possesses the corresponding risk factor. Association rule mining is a method for discovering interesting relations among sets of items in large databases. Let an item be a binary indicator signifying whether a patient possesses the corresponding risk factor. E.g. The item htn indicates whether the patient has been diagnosed with hypertension. Let X denote the item matrix, which is a binary covariate matrix with rows representing patients and the columns representing items. An item set is a set of items: it indicates whether the corresponding risk factors are all present in the patient. If they are, the patient is said to be covered by the item set (or the item set applies to a patient).An association rule is of form I→ J, where I and J are both item sets. The rule represents an implication that if J is likely to apply to a patient given that I apply. The item set I is the antecedent and J is the consequent of the rule. The strength and "significance" of the association is traditionally quantified through the support and confidence measures.

II. RELATED WORK:

In general, there is extensive literature on measuring the risk of diabetes mellitus. This section reviews about the some related work in order to explore the strengths and weakness of existing methods. C.Wang and S. Parathasarthy [1] In this paper, redundancy aware Top-k patterns extracts a large collection of frequent patterns. The Greedy algorithms are involved to provide the optimal solution with better performance bound O (log k). Top-K patterns apply association rule mining, clustering and indexing to evaluating the significance of various kinds of patterns and eliminate the redundancy among the closed patterns. The drawback of this paper is information loss. C.Wang and S. Parathasarthy, [2] this paper proposed the bus with cluster based approach and it achieves that reduces the original data while summarization.

Volume No: 3 (2017), Issue No: 3 (August)   **August 2017**
www. IJRACSE.com

Page 12

The two approaches are used to prevent the information loss are: first approach is an adaptation of clustering and the second approach makes use of frequent itemsets from the association analysis domain. It fails to relate the incorporate data. grouping of data when the boundaries variant group the groups are vague and ambiguous. Propose a novel fuzzy rule for data-depends grouping of data algorithm (FRBC). Aysel Ozgur, Pang-Ning Tan, and Viper Kumar [3] a framework for making regression models by using the rule for data for association of data. Propose a pruning scheme for redundant and insignificant rule for datas in the rule for data extraction step, and also a number of heuristics for making regression models. Jian-Ping Mei and Lihui Chen [4] the original FCM uses Euclidean normally used to measure the object-to centroid distance. To propose a new fuzzy grouping of data approach called LinkFCM where an additional term is added into a fuzzy c-mean grouping of data type approach. Existing systems intend to apply association rule mining to electronic medical records to determine sets of risk factors and their consequent subpopulations that symbolize patients at mainly elevated risk of increasing diabetes. Given the elevated dimensionality of EMRs, association rule mining generates a extremely huge set of rules which we have to summarize for simple medical use. We reviewed four association rule set summarization techniques and conducted a relative assessment to provide assistance concerning their applicability, strengths and weaknesses.

III.   PROPOSED SYSTEM:

We try to use association rule mining to the electronic medical record (EMR). All the risk factor about a patient namely co-morbid disease and laboratory results and medications are being available in the EMR, there are less chances to miss details about a patient with the extensive set of risk factors the set of discovered risk becomes extremely large to overcome this we use rule set summarization technique which is used to compress the original rule set into a compact set. The implementation procedure of rule set summarization technique is as follows.

Read the Electronic Medical Record and Generate Rule Set Factors:

In this module we can read information related each patient suffering with diseases related for diabetes. The information should contain patient testing result for the diabetes. By using that result we can identify risk factor diabetes over the each patient. Before identifying risk factors of each patient, we will take testing result and find out rule set of factors. By identifying rule set of factors we are using result comparative based approach. The implementation process of result comparative based approach is as follows.

1. Read each record from the database contains values related testing result of each patient.

2. Before reading testing result values we can also read training dataset related risk with scores and also medication of that risk.

3. Take the each risk value of each patient and compare that value into training dataset risk values.
4.If those risk values are greater that equal to training dataset risk values and get those risk until end the of record.

5. After completion of comparative process take those risks will form the one rule.

6. Repeat step1 to step 5 for the completion of all testing record of patients.

Generate Medication Based Rule Sets:

In this module we are taking risk rule set database and generate medication based rule set over the risks. By identifying medication based rule set we are also take the training dataset contains risk values and medication

related to particular risk. The generation of medication based rule set is as follows.

1. Take the risk based rule set data base and training database as on input.
2. Read each record from the risk result base rule set and get the each risk from that record.

3. Take that risk name and check that risk related medication in training dataset.

4. If found the risk related medication in a training data set, get those medication name and put into one variable.

5. This process repeats until all record of testing risk based result oriented records are completed.

6. After completion of step 5 we are get medication based rule set.

Finding Optimal Summary Based Rule Set:
In this module we are taking medication rule set as input and find out maximum risk factors over the diabetes. By finding maximum risk of diabetes we are using association rule mining approach. In this paper we are using probability based association rule mining technique for finding maximum risk factors of diabetes. The implementation of procedure of probability based association rule mining technique is as follows.

1. Take the medication risk rule set as input of given technique.

2. Read the all risks and medication from the rule set.

3. Take the each risk name and scan all medication rules for finding count.
4. Take another risk value and scan the entire rule sets for finding count, this process will repeat for completion of signal iteration.

5. After completion of first iteration take two risk values and scan the entire rule set for finding count.

6. This process also repeat remaining two risk values and also find the count.
7. After counting we can find out support count of each risk value by using following formula.

Sup (count) = total no. of risk count/total number no rule set.

8. Repeat Step6 to step7 until maximum length of rule set and also find frequent rule set.

9. Before finding frequent rule sets we can compare support count value to minimum support value.

10. If the support value of each rule set is greater than are equal to minimum support value then the rule set is frequent.

By using probability based association rule mining technique we can find out maximum risk that patients are suffered and also choose the related medication for those risks. So that by implementing those concepts we can easily find out most the patients will suffered what type of risk and also find out related medications are used for the recovery of that risk. So that by providing those concepts we can improve the efficiency for finding frequent risk.

IV.   CONCLUSIONS:
In this paper we are proposed an efficient association rule mining technique for identifying frequent risk factors over the diabetes and also find out related medication for those risks. Before performing those operations we can take testing result of each patient and also take training dataset related to risk values and medication. By using training data set we can identify risks based on testing result values. After identifying risks take those risk oriented testing result and performing generation of medication rule set process.

Volume No: 3 (2017), Issue No: 3 (August)   August 2017
www. IJRACSE.com

Page 14

By performing that process we can get medication based rule based on result oriented of risks. After generating medication based rule set, take that rule set and perform the optimal summary based rule set for finding frequent risk rule. By applying that process we can also get risk related medication and also get frequent risk over the patient are suffering from the diabetes. By implementing those concepts we can improve more efficiency for finding risks and also get risk related medication.

REFERENCES:

[1].C.Wang and S. Parathasarthy,2006, " Summarizing Itemset Patterns Using Probabilistic Models," in Proc. ACMInt. Conf.KDD, New York, NY, USA

[2].C.Wang and S. Parathasarthy,2006, " Summarizing Itemset Patterns Using Probabilistic Models," in Proc. ACMInt. Conf.KDD, New York, NY, USA

[3].D. Xin, H. Cheng, X. Yan and J. Han,2006, " Extracting Redundancy Aware Top-K Patterns," in proc. ACMInt. Conf. KDD, Philadelphia, PA, USA.

[4].Jian-Ping Mei, Lihui Chen. ―A Fuzzy approach for multitype relational data clustering,‖IEEE transactions on Fuzzy systems, vol 20, No.2, April 2012.

[5].Xiaoxin Yin and Jiawei Han. CPAR: Classification based on predictive association rules. In SIAM International Conference on Data Mining (SDM), 2003.

[6].Pedro J. Caraballo, M. Regina Castro, Stephen S. Cha, Peter W. Li, and Gyorgy J. Simon. Use of association rule mining to assess diabetes risk in patients with impared fasting glucose. In AMIA Annual Symposium, 2011.

[7].Bing Liu, Wynne Hsu, and Yiming Ma. Integrating classification and association rule mining. In ACM International Conference on Knowledge Discovery and Data Mining (KDD), 1998.

[8].Bing Liu, Wynne Hsu, and Yiming Ma.‖Integrating classification and association rule mining‖. In ACM International Conference on Knowledge.

[9].Gyorgy J. Simon, Member, IEEE, Pedro J. Caraballo, Terry M. Therneau, Steven S. Cha, M. Regina Castro, and Peter W. Li (2015),"Extending Association Rule Summarization Techniques to Assess Risk of Diabetes Mellitus," vol. 27, no.1.

[10].Diabetes Prevention Program Research Group, "Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin," N. Engl. J. Med., vol. 346, no. 6, pp. 393–403, Feb. 2002.

[11].Centers for Disease Control and Prevention. "Nationaldiabetes fact sheet: National estimates and general informationon diabetes and prediabetes in the United States,"U.S. Department of Health and Human Services, CentersforDisease Control and Prevention, 2011 [Online]. vailable:
http://www.cdc.gov/diabetes/pubs/factsheet11.htm

BIOGRAPHIES:

Yegireddy Anil Kumar is student in M.tech (CSE) in Sarada Institute of Science Technology and Management, Srikakulam. He has received his B.tech (I.T) from Sarada Institute of Science Technology and Management, Srikakulam. His interesting areas are data mining, network security and cloud computing

Konni Srinivasa Rao is working as an Assistant Professor in Sarada Institute of Science, Technology and Management, Srikakulam, Andhra Pradesh. He Receive his B.tech from gmrit engineering college, rajam. He received his M.Tech (cse) from Pragati engineering college, Kakinada, East Godavari, Andhra Pradesh. His research areas include Network Security and Computer Networks.

Volume No: 3 (2017), Issue No: 3 (August)   **August 2017**
www. IJRACSE.com

Page 15