

The Application of Recently Developed Association Rule Mining Algorithms in Healthcare

O.Jamsheela, Assistant Professor,

Department of Computer Science, EMEA College of Arts and Science, Kondotty.

Abstract: Technology is increasingly becoming an indispensable part of people's daily life, facilitating to do a variety of useful responsibilities. Data Mining has been already using in almost all areas to make things smooth and more beneficial. Healthcare Data mining integrates Data mining tools and medical data to expand the capabilities and benefits of Data mining to mine valuable information. In healthcare, data mining technologies are widely using and is becoming increasingly popular. The first application of data mining generally approved as the mining of valuable information from huge data set. The huge amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods. Data mining approach provides the technology and tools to transform these huge data sets into useful information for decision making. Association rule mining(ARM) is a popular data mining technology. This paper explores the application of recent association rule mining technology in healthcare. In particular, it discusses association rule mining and its applications in patients' data.

I. Introduction.

Data mining in Healthcare is an important and very crucial task that needs to be executed accurately. It attempts to solve real world health problems in diagnosis and treatment of diseases. Data Mining techniques are used to find hidden knowledge from huge data set. Data Mining is related to database technology, statistics, pattern recognition, machine learning, data visualization, and expert systems etc. Data mining can be used to get answers for complicated queries. Data mining techniques can be implemented on massive data in an automated matter, whereas traditional statistical methods used in epidemiology require custom work by experts. The knowledge discovery in database (KDD) is dealing with development of methods and techniques for making use of data. Data Mining is one of the most important step of the KDD. Data mining is the process of knowledge discovery and extraction from huge amount of data is involved. Both the data mining and healthcare industry can be used together for finding reliable early detection systems and other various healthcare related systems from the clinical and diagnosis data. In regard to this emerge, we have included two recent Association rule mining methods and its applications in medical data. Section 2 contain some related work in this field, the two recent ARM methods and included in section 3 and its applications on medical data contains in section 4. Section 5 concludes the paper.

II. Related work

Many research papers have been published based on the application of data mining techniques in health related data. Jyothi et al., published a survey paper to this emerge and they have reviewed the various paper involved in this field in terms of method, algorithms and results [1]. Another work is proposed by Koh et al., to explore data mining applications in healthcare. In particular, the authors discussed data mining and its applications within healthcare in major areas such as the evaluation of treatment effectiveness, management of healthcare *etc.*[2]. The paper also gives an illustrative example of a healthcare data mining application involving the identification of risk factors associated with the onset of diabetes. Another survey paper is presented by Yoo et al., the survey explores the utility of various Data Mining techniques such as classification, clustering, association, regression in health domain. In this paper, the authors present a brief introduction of these techniques and their advantages and disadvantages. [3]. Many researchers have published survey papers on the same research area [4],[5],[6],[7],[8]. Another study explains utilization of medical data mining in determination of medical operation methods. The study conducted with the data collected from 80 pregnant women[9]. Srinivas et al., proposed a study to predict the likelihood of patients

getting a heart disease using medical profiles such as age, sex, blood pressure and blood sugar [10]. Many patients died due to insufficient amount of knowledge. Another research is proposed to solve real world health problems in diagnosis and treatment of diseases. The authors attempted to find out interesting patterns from data of heart patients. Three algorithm such as, Decision Tree, Neural Network and Naïve Bayes, are used in their study with two different scenarios[11]. By analyzing the related work we can see that many data mining tools are used in healthcare.

III. Application of Recent Association Rule Mining Algorithms(ARM) in Healthcare

As Healthcare industry produces a huge amount of data, the data mining tools can be used to find hidden patterns and interesting knowledge that may help in effective and efficient decision making.

A study is proposed by Azra, et al. to identify risk patterns for type 2 diabetes incidence using association rule mining (ARM) [12]. The study is conducted by using the data of 6647 individuals without diabetes and aged ≥ 20 years. The study is conducted for 10-12 years, to analyze risk patterns for diabetes occurrence. The authors presented the result of the study as in the case of women, the results showed that impaired fasting glucose (IFG) and impaired glucose tolerance (IGT), in combination with body mass index (BMI) ≥ 30 kg/m², family history of diabetes, wrist circumference > 16.5 cm and waist to height ≥ 0.5 can increase the risk for developing diabetes. For men, a combination of IGT, IFG, length of stay in the city (> 40 years), central obesity, total cholesterol to high density lipoprotein ratio ≥ 5.3 , low physical activity, chronic kidney disease and wrist circumference > 18.5 cm were identified as risk patterns for diabetes occurrence. Finally the authors concluded that the study showed that ARM is a useful approach in determining which combinations of variables or predictors occur together frequently, in people who will develop diabetes. The ARM focuses on joint exposure to different combinations of risk factors, and not the predictors alone.

Huang et al presented a paper by proposing to use association rule mining methods to investigate prescribing of smoking cessation medication in the UK primary care and to identify the characteristics of numerically important groups of patients who typically do, or do not, receive cessation therapy[13]. The authors used an association rule mining study using the Health Improvement Network Database. The study included 282433 patients aged greater than or equal to 16 years, who were registered with the practice throughout 2008 and recorded as a current smoker during that year. The authors have used the variable such as Age, gender, lifestyle indicators and co-morbidity etc. They have concluded the study by stating that theirs novel approach identified sizeable and easily definable groups of patients who are systematically failing to receive support for smoking cessation in primary care. Association rule mining can be used to identify key numerically important groups at high or low risk of receiving treatment and hence potentially to improve healthcare delivery.

Nahar et al., proposed a An Association Rule Discovery Approach for Significant Cancer Prevention Factor Extraction[14]. The authors introduced a research proposal to extract the significant prevention factors for particular types of cancer. The authors first constructed a prevention factor data set with an extensive literature review on bladder, breast, cervical, lung, prostate and skin cancer, to find out the prevention factors. Subsequently, they employed three popular and effective association rule mining algorithms, Apriori, Predictive apriori and Tertius algorithms in order to discover most of the significant prevention factors against these specific types of cancer. In their experimental results, they state that the Apriori is the most useful association rule-mining algorithm to be used in the discovery of prevention factors.

Nahar et al., proposed another application of Association rule mining to detect factors which contribute to heart disease in males and females. The proposed method investigates the sick and healthy factors which contribute to heart disease for males and females. To identify these factors, Association rule mining, the popular data mining intelligence approach, is used. A biological database is considered along with the three rule generation algorithms – Apriori, Predictive Apriori and Tertius. They concluded that analyzing the information available on sick and healthy individuals and taking confidence as an indicator, females are seen

to have less chance of coronary heart disease than males. The authors also state that factors such as chest pain being asymptomatic and the presence of exercise-induced angina indicate the likely existence of heart disease for both men and women. However, resting ECG being either normal or hyper and slope being flat are potential high risk factors for women only. For men, on the other hand, only a single rule expressing resting ECG being hyper was shown to be a significant factor. This means, for women, resting ECG status is a key distinct factor for heart disease prediction. Comparing the healthy status of men and women, slope being up, number of coloured vessels being zero, and oldpeak being less than or equal to 0.56 indicate a healthy status for both genders.

IV. Conclusion

The paper is presented to introduce the application of some recent Association Rule Mining technology in Healthcare. Overall, all the above mentioned researches have demonstrated and proved that the use of Association rule mining to discover interesting knowledge is very effective. This paper has focused on the application of computational intelligence, in particular, association rule mining-based data mining techniques, to identify the key factors hiding in huge data sets.

V. References.

1. Jothi, Neesha, and Wahidah Husain. "Data mining in healthcare—a review." *Procedia computer science* 72 (2015): 306-313.
2. Koh, Hian Chye, and Gerald Tan. "Data mining applications in healthcare." *Journal of healthcare information management* 19.2 (2011): 65.
3. Tomar, Divya, and Sonali Agarwal. "A survey on Data Mining approaches for Healthcare." *International Journal of Bio-Science and Bio-Technology* 5.5 (2013): 241-266.
4. Yoo, Illhoi, et al. "Data mining in healthcare and biomedicine: a survey of the literature." *Journal of medical systems* 36.4 (2012): 2431-2448.
5. Tekieh, Mohammad Hossein, and Bijan Raahemi. "Importance of data mining in healthcare: a survey." *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. 2015.
6. Sow, Daby, Deepak S. Turaga, and Michael Schmidt. "Mining of sensor data in healthcare: A survey." *Managing and mining sensor data* (2013): 459-504.
7. Sharma, Anand, and Vibhakar Mansotra. "Emerging applications of data mining for healthcare management—a critical review." *2014 International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE, 2014.
8. Mahindrakar, Prakash, and M. Hanumanthappa. "Data mining in healthcare: A survey of techniques and algorithms with its limitations and challenges." *Int. J. Eng. Res. Appl* 3.6 (2013): 937-941.
9. Gharehchopogh, Farhad Soleimani, Peyman Mohammadi, and Parvin Hakimi. "Application of decision tree algorithm for data mining in healthcare operations: A case study." *International Journal of Computer Applications* 52.6 (2012).

10. Srinivas, K., B. Kavihta Rani, and A. Govrdhan. "Applications of data mining techniques in healthcare and prediction of heart attacks." *International Journal on Computer Science and Engineering (IJCSE)* 2.02 (2010): 250-255.
11. Shafique, Umair, et al. "Data mining in healthcare for heart diseases." *International Journal of Innovation and Applied Studies* 10.4 (2015): 1312.
12. Ramezankhani, Azra, et al. "An application of association rule mining to extract risk pattern for type 2 diabetes using tehran lipid and glucose study database." *International journal of endocrinology and metabolism* 13.2 (2015).
13. Huang, Y., Britton, J., Hubbard, R., & Lewis, S. (2013). Who receives prescriptions for smoking cessation medications? An association rule mining analysis using a large primary care database. *Tobacco control*, 22(4), 274-279.
14. Nahar, J., Tickle, K. S., Ali, A. S., & Chen, Y. P. P. (2011). Significant cancer prevention factor extraction: an association rule discovery approach. *Journal of medical systems*, 35(3), 353-367.
15. Nahar, J., Imam, T., Tickle, K. S., & Chen, Y. P. P. (2013). Association rule mining to detect factors which contribute to heart disease in males and females. *Expert Systems with Applications*, 40(4), 1086-1093.

