Privacy Preserving Mining Based Framework for Analyzing the Patient Behavior

Dhanashree M. Patil[1], Dnyaneshwar A. Rokade[2]

Department of Computer Engineering, JSPM's Imperial College of Engineering and Research, Pune, India

Abstract: *The data driven mining technology was applied in the most of the existing data mining algorithms. The limitations of this method are that expert analysis is required before the derived information can be used. To analyze data for predicting the patient's behavior here we assume the strategy of domain driven data mining and utilize association's rules, clustering and decision trees. The proposed system is named as the Combined Mining based patient Behavior Prediction System (CM-PBPS). In this paper, we are implementing domain –driven data mining strategy and exploit clustering, decision trees, and association rules to explore data of patient's behavior. This system specifies the patients who are infected by which diseases. For the implementation, initially for analyze the patient behavior the associations rules are used. After that for patient segmentation the clustering algorithm is used. For reducing the data imbalanced we delete the clusters of the patient who are not infected by diseases. At last decision tree was utilized to predict and analyze the rest of the data by using the derivative attributes and the attributes provided.*

Keywords: Association rules, clustering, decision trees, Domain-driven data mining.

1. Introduction

Medical industries have developing quickly and spread not much more efforts to maximize their profit by wining more peoples by providing best medical services and use of improved technologies efficiently. But still they have to face from various negligence form Patients due to various reasons. CARE, a Collaborative Assessment and Recommendation Engine, which depends only on a patient's available medical history using ICD- 9-CM codes for predicting risk related to future diseases. CARE combines filtering methods like clustering and collaborative with clustering to predict each individual patient's greatest disease risks which mainly depend on their own medical treatment and that of similar patients. Here we also mention an Iterative version, ICARE, which incorporates group concepts for better performance. These systems will not require specialized information and provide it will predictions for medical behavior or conditions of all types in a one single run. We present experimental outcome on a big Medicare dataset, representing that CARE and ICARE perform well at capturing future disease risks

Thus [14], here we discussed about the behavior of the patients who are infected by various different diseases. Health care, thus, needs to be more proactive in recognition of disease and related risk. How- ever, the problem caused by the various disease factors and the available previous medical history of a patient is very complicated that not even single health care specialized can fully understand it all. Currently, doctors first looking for health and family history and previously done physical examination to estimated the risk of patient [8]. However, these qualitative and irregular 'risk assessments' generally concentrates on only a few diseases and are limited by a particular doctor's time, memory and experience. Therefore, current medical care is reactive, stepping in once the symptoms of a disease have emerged, rather than proactive, treating or eliminating a disease at the earliest signs

Data driven mining methodology is very essential for people related with information Technology but it is not that much useful for persons relate with business. Taking example of association rule mining, some rules only denotes very basic knowledge and they are not interested in business logics. E.g. If cold drink is bought then snacks are brought is proved with very high confidence and support. This is very famous rule according to Aprior algorithm.
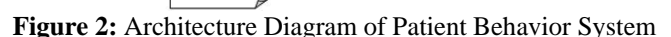
This technique is not suitable for business because the given rule is like common scene knowledge as well there may be possibility of misleading of persons who make decision because the rule "If the cold drink is bought , then snacks are not bought "may also present, if we consider not bought simultaneously. And if mentioned techniques only focus on common scene information, interestingness may also be able to be derived.

In recent times, Longbing Cao invented the domain driven Data mining concept. (D3M) Cao, 2010; Cao et al., 2010, and attached it with industry to collect various important information from data mining. They highlight this topic for data mining in real world and recommend the trends from data-centered hidden pattern discovery to domain driven actionable knowledge discovery (AKD) [10]. Here ''actionable'' denotes that the mentioned knowledge format can provide important place to decision makers in Business for allowing proper actions as well as for Delivering predictable outputs to business. From that innovation, Many related practices were performed (Du & Ling, 2010;Jin et al., 2010; Mansingh, Osei-Bryson, & Reichge lt, 2011; Marinica & Guillet, 2010; Xu, Lin, &Xu, 2009).

2. Literature Review

Privacy Preserving Data Mining (PPDM) is a big field related with the privacy driven from personally identifiable information when considered for data mining. So PPDM is becoming important field of research. A number of method and techniques have been developed for privacy preserving data mining. So PPDM is now become significant field of

investigation. There are various techniques and methods available for Privacy preserving data mining. This paper presents full review on PPDM and other related techniques like data restriction techniques, data modification and data partition. This can preventaccesses the data from unauthorized users.

2.1 Domain-Driven Knowledge Discovery

At the conference of SIGKKD 2003 and 2007[2][3], actionable knowledge discovery is focused by technicians as challenge for future data mining. Algorithm for data mining extracts patterns from converted data based on expert's theory. Data mining is automated process which produce algorithm automatically also creates tools without introduction of human.

There are different types of data mining like financial data mining which is very highly constraint based.[4], [5]constraints can be of economical, social or technical aspects. Real word business requirements and problems are related with process with expertise and business rules. Pattern which are actionable to business are generally hidden with source distribution, complex structure and dynamics. Frequently sometime these patterns are not that much actionable to business even they are interested to research [7]. So that it is not possible to develop such patterns in real world if they are incorporated with regulations processes and business rules. Constraints such as rule constrain, knowledge constraint and level constraints [11].

2.2 Related Data-Mining Approaches

Data mining mainly aims to fetch useful patterns and knowledge from available data set to resolve different issues. Until now it is used in different to date, it has been used in approaches like network intrusion (Tajbakhsh, Rahmati, & Mirzaei, 2009), stock market analysis (Au & Chan, 2003; Hadavandi, Shavandi, & Ghanbari, 2010) and shopping cart analysis (Agrawal, Imielinksi, & Swami, 1993). Commonly it is used in mining of different association rules from various transaction data.
Association rule can be represents as C? D, where C and D is general product and rule of association states that if C is purchased than Product D is purchased along with it. There are two gauges which are used to measure validity of confidence, support and association rule. The first rule mining was recommended by Agrawal et al. (1993), following are the main steps 1. Produce frequent item sets based on minimum support 2. Pro-duce candidate item sets 3. Produce frequent item sets based on minimum confidence. Data which is based on data similarity is known as clustering method.

3. Patient Behavior System

This system predict the behavior of patient i.e. by entering the symptoms the system predict the behavior of the patient. Now we will discuss the architecture diagram of the system and also see the implementation flow of this system.

Figure 1: Data Mining Implementation Flow

In the above figure 1, we will show the data implementation flow of the patient behavior system. In the system the data of the patients are selected, the process of clustering is done for removing the data of the normal patient. The decision tree is build for the predicting the patient data.

Figure 2: Architecture Diagram of Patient Behavior System

Paper ID: OCT141076   1022

The overall system execution flow is shown in Fig 2, and it includes two parts: (1) the domain-driven mining phase and (2) the model tuning phase. In the first part, ETL (extraction transformation loading) is utilized to derive patients data. Next, for analyze the patient behavior the associations rules are used. After that for patient segmentation the clustering algorithm is used. For reducing the data imbalanced we delete the clusters of the patients who are not infected by diseases. At last decision tree was utilized to predict and analyze the rest of the data by using the derivative attributes and the attributes provided. In the second part, after the model is constructed, its efficiency has to be evaluated to maintain its predictive power. The system automatically verifies and compares the accuracy rate of each rule from patient behavior. If the accuracy rate of the rule is lower than the set threshold value, the system will highlight it for to review the rule. Apart from this, use new data to create rules from the constructed model, and if a new rule is verified, the system will add the rule to the database, thus maintaining Predictive power.

4. Conclusion

In this paper, we introduce framework for patient behavior that is Combined Mining Based Patient Behavior Prediction System (CM-PBPS). This is related with data mining methodology with strategy like domain driven data mining methodology to guess which patients are infected with which disease. To achieve various goals CM-PBPS model is introduce. The correctness of stated model was about 78.53% which is much higher than current model which was 65.60%.While plan is being implemented, the patient behavior record for previous treatment was inserted to the patient behavior model, and the correctness and recall of the available model is enhanced.

References

[1] R.Natarajan1, Dr.R.Sugumar, M.Mahendr. Anbazhagan, "A survey on Privacy Preserving Data Mining" International Journal of Advanced Research in Computer and Communication Engineering Vol.1, Issue 1, MARCH2012

[2] Ankerst, M., Report on the SIGKDD-2002 panel the perfect data mining tool: Interactive or Automated. ACMSIGKDD Explorations Newsletter, 4(2):110-111, 2002.

[3] Fayyad, U., Shapiro G., Uthurusamy R., Summary from the KDD-03 panel –Data mining: the next 10 years. ACM SIGKDD Explorations Newsletter, 5(2):191-196, 2003.

[4] Han, J., Towards Human-Centered, Constraint-Based, Multi-Dimensional Data Mining. Invited talk at Univ. Minnesota, Minneapolis, Minnesota, Nov. 1999.

[5] J-F. Boucicault, B. Jeudy. Constraint based data mining. The Data Mining and Knowledge Discovery Handbook, O. Maimon and L. Rokach (Eds.), Springer, pp. 399-416, 2005.

[6] Jin, H., Chen, J., He, H., Kelman, C., McAullay, D., & O'Keefe, C. M. (2010). Signalling potential adverse drug Reactions from administrative health Databases. IEEE Transactions on Knowledge and Data Engineering, 22(6), 839–853.

[7] Cao, L. (2010). Domain-driven data Mining: challenges and prospects. IEEE Transactions on Knowledge and Data Engineering, 22(6), 755–769.

[8] W. T. C. Consortium. A national Ambulatory medical care survey: 2001Summary. Nature, 447:661–678, 2007.

[9] Zhang, C., Zhang, Z., Cao, L., Agents and Data Mining: Mutual Enhancement By Integration, LNCS 3505, 50-61, 2005.

[10] Longbing Cao Fac. of Eng. & Inf. Technol., Univ. of Technol., Sydney, NSW, Australia Yanchang Zhao ; Huaifeng Zhang ; Dan Luo ; Zhang, C. ; Park, E.K., "Flexible Frameworks for Actionable Knowledge Discovery" Published in: Knowledge And Data Engineering, IEEE Transactions on (Volume: 22, Issue: 9)

[11] He, J., Zhang, Y., Shi, Y., & Huang, G. (2010). Domain-driven classification Based On multiple criteria and multiple constraint-level programming For intelligent credit scoring. IEEE Transactions on Knowledge and Data Engineering, 22(6), 826–838

[12] Cahill, M. H., Lambert, D., Pinheiro, J.C., & Sun, D. X.(2002). Detecting Fraud in the real World. Handbook of Massive data sets (massive computing4). Kluwer Acadamic Publishers (pp. 911–929). Kluwer Academic Publishers

[13] Raja Kumar, R. ; Department of Computer Science and Engineering, College of Engineering, Ann University, Chennai Â¿ 600025Tamilnadu, India; Indumathi, J. ; Uma, G.V. , "Optimizing PPDM in Asynchronous sparse data using Random" published in Information Reuse and Integration, 2008. International Conference on July 2008.

Paper ID: OCT141076   1023