Exploration on Association Principles of Mining Based Scheme for Healthcare Repository

D.Ravi Kiran (Research Scholar)
Dr. S V N Srinivasu (Research supervisor)
Acharya Nagarjuna University,
Nagarjuna Nagar, Guntur (D.t)
Email: ravikirand2005@gmail.com

Abstract

Early discovery of patients with eminent risk developing conditions in case of numerous diseases is a challenging issue; in this connection Health caring organizations enhance the depository of relevant information of the diseased patients' in a growing method which would be more helpful by carrying out relational analysis. Data mining algorithms are confirmed to be quite useful in exploring useful correlations from larger data repositories. In this abstract we have explored the organization Rules mining based scheme for finding co-occurrences of infections passed by a patients using the healthcare repository. Prediction of clinical outcome (PCO) scheme which extracts data from patients' healthcare database, transforms the OLTP data into a Data Warehouse by generating association rules have also been proposed in this piece of writing. The PCO system helps to reveal the relations among the diseases. The system of PCO forecasts the other relevant connections of the diseases at first stage and the second stage.

Key words Disease Correlation, Association Mining, e-Health, Healthcare, Medicare

1. Introduction

Now a day's medical science has exposed that the frequency of one disease may cause/ result to several associated diseases. For instance, Heart-Block which may be caused for the possibility of being attacked by other ailments, such as Cardiac Arrest, Hypertension. This is, on the other hand, an appealing .problem just to understand how this medical attitude grasps from numerical viewpoint. Techniques connected to Data based such as association rule mining, have acquired fame among modern scientists to get an understandable way in relation to the body and scientific phenomenon. In this abstract, we implemented association rule mining to extort information from medical information for forecasting the relationship of diseases affected by the patients. From the perspective of scientific study, data mining is comparatively a new subject which was progressed by chiefly from the research done in a range of subjects such as calculating, marketing, statistics etc. In classical examination of data, problems of data mining and paralleled resolutions have origins. Most of the methods employing in data mining have brought from two areas of study: i) Improvement in machine open society (i, e. Artificial Intelligence) a ii) Progression of statistical society particularly in multivariate and computational statistics. Both of them have greatly contributed for getting a better understanding and appliance of data mining techniques.

2. Back ground

Computerized healthcare systems are accumulating huge amount of information about patients and their medical conditions every day. Unluckily, a small number of methods have been improved and

implemented to find out this concealed information. The cluster-analysis fundamental design is suggested and discussed. For transferring prostate cancer patients into same groups with the purpose of supporting future clinical treatment conclusions as description. An enhanced algorithm was initiated with a number of restrictions in order to search association rules with a loud and elevated data of medicinal depository.

The Ovarian Cancer, employing B.P.S i, e. Bio-marker Pattern Software was used as an arithmetical study on the basis of understanding of categorization for diagnosing. However such assignment has been completed on comparing the methods of data mining that supports diagnosis of tumor. However, to make a diagnosis of breast cancer by making use of digital mammograms the classifications of Association rule have also to be exploited. Neural Network also utilized for the similar purpose on the basis of understanding of classification. Association Mining applied on questionnaire responses related to human sleeping where questionnaire data and clinical summaries comprised a total of 63 variables including gender, age, and index of physic and Epworth and depression scores. Many Clinical Decision Support Systems (CDSS) have been developed. CHICA is a CDSS, developed to improve preventive pediatric primary care.

Acceding to the requirements of the patients energetic forms are produced and modified patients' on the basis of MLM i, e, Medical Logic Modules .A skilful administration framework for spreading healthcare arrangement has been projected to combine varied systems utilized by lot of sections from clinical care for organizing process. However, to apply those abovementioned improvements specifically are very difficult. Instead of enhancing an application limited to a specific purpose such as prostate cancer, skin cancer, etc. Whereas a PCO system with a new generic version which has the ability of working for all types of diseases in a related way and can produce the relations based on the input data., The structural design of the appliance and its operational procedures are mentioned in the next section

3. Framework and operational process

The PCO system will extract data from an OLTP system. Thus to implement the PCO system, we require the regular operational OLTP system to generate input data for the PCO system. The structural design of the system can be illustrated as follows.

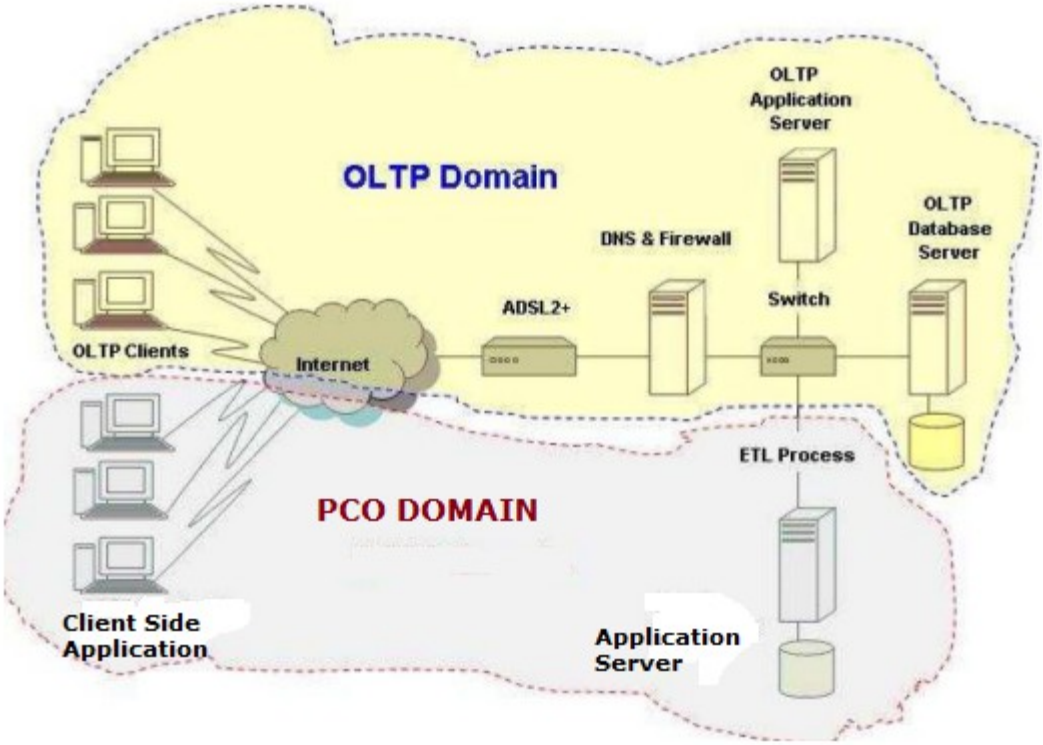

Figure 1 – Indicates a total structural design of the planned System for forecasting relationship among diseases by using healthcare data of association mining. The upper-port of the figure describes the design of the OLTP system the PCO system is illustrated by the lower part.

3.1 System Architecture

The disease diagnosis application proposed here based on two different software systems. One is OLTP i,e Online Transaction Processing System and another is prediction of clinical outcome (PCO) application. **OLTP Application:** Generally patients have to fill-up a particular form before they meet a doctor, therefore, such data from the prescribed from is planned to be gathered through web enabled OLTP system.

PCO Application

This application will import transactional records from OLTP application for further processing to generate correlations along with diseases employing Co-related Rules data mining. The OLTP system is a usual database application to capture patient's information and to preserve the records into a database repository. It is easy and no logical procedures have been included in this part and therefore we mainly concentrated on the PCO system in this paper and we believe that the OLTP system has been applied or

implemented effectively and information has been gathered consequently from healthcare organizations.

3.2. Association Rules Mining and Apriori Algorithm

In relation to the data mining, correlation rule is used to find out appealing relationships among different types of diseases. Support and Confidence are the two key factors which are to be employed to figure out the relations among variations can be described as follows:

Support: In a permanent number of transactions the happening of exacting event is the support of that event. For instance, T transactions among which Txy transaction hold the item set {X,Y} is Tx.

Confidence: Confidence is a relative support. For example there are T transactions among which Txy transactions contain itemset {X, Y} and Tx transactions contain item X and thus confidence of occurring X and Y together is Txy/ Tx. The Apriori Algorithm is used here for Association Rule Mining to discover regular dataset that gratifies the early defined support at minimum level and confidence from a specified database. As computers are handy now-a-days, a viable PCO system can be arranged to gather large amount of information and to stored in the database simultaneously. This sort of data includes the transaction records of clinics, hospitals, supermarkets, banks, stock markets, telephone companies etc. In the rest of few parts, we tried to discover some hidden information from a sample transactional dataset.

3.2 Sample Dataset

Generally a variety of patients visit hospitals but many of them would go to a get a treatment for certain disease, which has been stated here as primary disease. After interviewing and checking up and noting down the other problems of patients, doctors or their associates make an interview with the patient and note down other problems (diseases), which is to be specified here as secondary or connected diseases they will also be put in into a database. The main objective of the paper is to discover the relations among the first stage of disease and other second stage of disease. The following table signifies an example for Medicare database which comprises information of disease patient-wise.

Table 1. - Sample for Data Transactional

Patient Id	Disease
P000000001	Bradycardia
P000000001	Cardiac Arrest
P000000001	Hypertension
P000000001	Myocarditis
P000000002	Bradycardia
P000000002	Cardiac Arrest
P000000002	Hypertension
…	…
P000001000	Cardiac Arrest

Table 2: Patient records with multiple diseases

Patient	Count	Diseases
P000000001	4	Heart-Block, Hypertension, Cardiac-Arrest, Bradycardia
P000000002	3	Heart-Block, Hypertension, Cardiac-Arrest
…	…	…
P000001000	1	Hypertension

3.3 Producing Itemsets

A set of diseases obtained by each patient presented here together with the number of diseases counted by the algorithm

Algorithm 1 - CountDisease

```
1   PROCEDURE CountDisease
2     FOR each p in P
3         D_s ⇐ ""
4         c ⇐ 0
5         find records for p
6         FOR each r in R
7             c ⇐ c+1
8             D_s ⇐ D_s + r + ","
9         NEXT r
10        D_s ⇐ D_s without comma at the end
11        INSERT (p, c, D_s) in to database table
12    NEXT p
13  END CountDisease
```

Figure 2 - Pseudo code to count patients' diseases

3.4 Counting Support of a disease(Item) in the sampling dataset

The frequency of every item in all the transactions has been calculated in the below table implementing algorithm and for a 1-item itemsets for the first pass.

```
    Algorithm 2- FindSupport
1   PROCEDURE FindSupport
2     d ⇐ disease
3     db ⇐ database in considarat ion
4     r ⇐ record
5     rs ⇐ recordset
6     s ⇐ support
7     BEGIN
8           FOR each d in db
9             s ⇐ 0
10        find records for d
11            FOR each r in rs
12              s ⇐ s + 1
13            NEXT r
14        UPDATE Database with s
15          NEXT d
16    END
17  END FindSupport
```

Figure 3 - Pseudo code to calculate support.

3.5 Generation of Association Rule:

The rule data has been produced by the PCO system as indicated in Table 3 below exploiting the model transactions of Table 1, after second pass (maxpass=2)

Table 3- Candidate Sets for stop-level 2

Itemset	Supp (%)	Conf (%)
{Bradycardia, Cardiac-Arrest}	2.60	8.52
{Bradycardia, Heart-Block}	0.90	2.69
……	….	….
{Heart Block, Hypertension}	1.00	1.82
{Heart Block, Myocarditis}	1.10	3.29

4. Conclusion

In this piece of writing, I have applied system-prototype, named PCO system, by implementing the association rules of data mining methods to a patients' (assumed) database for discovering patterns of diseases possessed by patients. As an innovative thought of mining the data capturing process can further be modified in the clinics as well as in the data-warehouses which should further be involved to enhance the PCO system we have proposed. The recognized pattern by this implementation definitely can improve the healthcare services along with medical researchers for further exploring trends of diseases that are correlated to be sure of certain strapping national economy and bio-security.

References:

[1] Lovell NH, Magrabi F, Celler BG, Huynh K, and Garsden H, Web-based acquisition, storage, and retrieval of biomedical signals, IEEE Eng. Medicine and Biology, vol. 20(3), pp. 38-44, 2001.
[2] Mazzi C, Ganguly P, and Kidd M, Healthcare application based on software agents, in Med info 2001 Proceedings, 2001, pp. 136-140.
[3] Takeuchi H, Kodama N, Hashiguchi T, and Hayashi D, Automated Healthcare Data Mining Based on a Personal Dynamic Healthcare System, the 28th IEEE EMBS Annual International Conference, New York City, USA, 2006.
[4] Giudici P, Applied Data Mining: Statistical Methods for Business and Industry: Wiley Publications, 2003.
[5] Mannila H, Data mining: machine learning, statistics, and databases, in Eighth International Conference on Scientific and Statistical Database Management, 1996, pp. 1-8.
[6] Kantardzic M, Data Mining: Concepts, Models, and Algorithms. NJ: Wiley Interscience, 2003.
[7] Prather J, Lobach D, Goodwin L, Hales J, Hage M, and Hammond W, Medical data mining: knowledge discovery in a clinical data warehouse, in Proc AMIA Annu Fall Symp., 1997, pp. 101-5.
[8] Churilov L, Bagirov A, Schwartz D, Smith K, and Dally M, Improving Risk Grouping Rules for Prostate Cancer Patients with Optimization, the 37th International Conference on System Sciences, Hawaii 2004.
[9] Ordonez C, Santana C, and Braal LD, Discovering Interesting Association Rules in Medical Data, in ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, 2000, pp. 78-85.
[10]Vlahou A, Schorge JO, Gregory BW, and. Coleman RL, Diagnosis of Ovarian Cancer Using Decision Tree Classification of Mass Spectral Data, Journal of Biomedicine and Biotech-nology, vol. 5, pp. 308–314, 2003.
[11]Grzymala-Busse JW and Hippe JS, Data Mining Methods Supporting Diagnosis of Melanoma, the 18th IEEE Symposium on Computer-Based Medical Systems (CBMS 2005), 2005, pp. 371–373.
[12]Za¨ıane OR, Antonie ML, and Coman A, Mammography classification by an association rulebased classifier, in MDM/KDD: Inter-national Workshop on Multimedia Data Mining (with ACM SIGKDD 2002), 2002, pp. 62–69.
[13]Land WH, Masters JT, Lo JY, McKee DW, and Anderson FR, New results in breast cancer classification obtained from an evolutionary computation/adaptive boosting hybrid using mammogram and history data in Industrial Applications, Proceedings of the 2001 IEEE Mountain Workshop on Soft Computing, Virginia Tech, Blacksburg. Virginia, 2001, pp. 47–52.
[14]Laxminarayan P, Ruiz C, Alvarez SA, and Moonis M, Mining associations over human sleep time series, the 18th IEEE Symposium on Computer- Based Medical Systems (CBMS), 2005, pp. 323–328.

[15]Anand A, Biondich PG, Liu G, Rosenman M, and Downs SM, Child health improvement through computer automation: The CHICA system, in MEDINFO, 2004.

[16]Kazemzadeh RS and Sartipi K, Interoperability of Data and Knowledge in Distributed Health care Systems, in The 13th IEEE International Workshop on Software Technology and Engineering Practice, 2005.

[17]Agrawal and Srikant, Fast algorithm for Mining Association Rules, in The 20th VLDB conference Santiago, Chile, 1994.

[18]A Healthier Future For All Australians. vol. ISBN: 1- 74186-940-4 Publications Number: P3 -5499: Commonwealth of Australia, 2009. [19]Medicare Australia, www.medicareaustralia.gov.au, last access: Oct, 2009.