Global Journals Inc.

GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY
Volume 11 Issue 21 Version 1.0 December 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Association of Data Mining and Healthcare Domain: Issues and Current State of the Art

By Fawzi Elias Bekri, Dr. A. Govardhan

Jigjiga University, Jigjiga, Ethiopia

Abstract - Data mining has been used prosperously in the favorably perceived areas such as e-business, marketing and retail because of which it is now applicable in knowledge discovery in databases (KDD) in many industrial areas and economy. Data mining is mainly gaining its importance and usage in the areas of medicine and public health. In this paper the investigation of present methods of KDD, applying data mining methods for healthcare and public health has been discussed. The problems and difficulties related to data mining and healthcare in practice are also mentioned. In survey, the use of data mining has increased, along with examination of healthcare institutions so that the health policy prepared is the best, perceive disease causes and protect deaths in hospital and discover the dishonest insurance declaration.

Keywords : health problems; recognition; data mining; elderly; motion capture; mining methods and algorithms; medicine and science.

GJCST Classification : H.2.8

Association of Data Mining and Healthcare Domain: Issues and Current State of the Art

Fawzi Elias Bekri[α] , Dr. A. Govardhan[Ω]

Abstract - Data mining has been used prosperously in the favorably perceived areas such as e-business, marketing and retail because of which it is now applicable in knowledge discovery in databases (KDD) in many industrial areas and economy. Data mining is mainly gaining its importance and usage in the areas of medicine and public health. In this paper the investigation of present methods of KDD, applying data mining methods for healthcare and public health has been discussed. The problems and difficulties related to data mining and healthcare in practice are also mentioned. In survey, the use of data mining has increased, along with examination of healthcare institutions so that the health policy prepared is the best, perceive disease causes and protect deaths in hospital and discover the dishonest insurance declaration.

Keywords : health problems; recognition; data mining; elderly; motion capture; mining methods and algorithms; medicine and science.

I. Introduction

Data mining has been used prosperously in the favorably perceived areas such as e-business, marketing and retail because of which it is now applicable in knowledge discovery in databases (KDD) in many industrial areas and economy. Data mining is mainly used in the field of healthcare.

The rest of the paper organized as fallow. Section II explores the frequently quoted research work in recent literature that reveals the association of data mining and healthcare sector. The section III illustrates an example that concludes the impact of data mining in healthcare sector. Section IV explores the taxonomy of Healthcare issues that demands the data mining as critical requirement. Section V reveals the reckoned obstacles in Knowledge discovery from healthcare databases. In Section VI we discuss the current state of the art in "Data mining in healthcare sector". Section VII explores the conclusion of this paper that fallowed by references.

II. Data Mining In Health Sector

From past decades there occurs the use of definite information and proofs that encourage medical judgment (evidence based medicine or EBM). The father of modern epidemiology, John Snow utilized 1854 maps which consisted bar graphs to find out the flow of cholera to show that it passes through the water that flows beneath [15]. Snow numbered the deaths occurred and then as black bars he marked the sufferers address on the map. He noticed that most of the deaths occurred were surrounded around particular water well in London.

In 1855 polar-area diagrams was introduced by Florence Nightingale to explain that the deaths in army could be reduced to some extent by using hygiene clinical methods. To decrease the death rate, she utilized the diagrams so that the policy-makers can utilize it in their application reforms [1] [9].

During the occurrence of this problem Snow and Nightingale on their own they gathered the information, organized it and then examined the data since it was controllable. Though in present scenario the population is large, gadgets are used to analyze the victims of any disease but still they cannot succeed the result obtained by the investigator in the past. Ere data mining is considered useful as it is helpful in solving different issues that occur while gaining information related to the field of healthcare.

In recent studies the most considered topic is data mining and its uses in the field of medicine and public health. In 2003, Wilson et al studied all the past researches where KDD and data mining were used in the area of healthcare, which he felt to be confusing. Data mining was utilized by few authors for gaining information and others use them in statistical methods within the directory knowledge process [17].

Since the data mining definition is misunderstood in the medical field so the most preferred definition of data mining is a cluster of processes and methods for determining and illustrating models and developments in information [18].

III. Impact of Data Mining in Health Sector

The event that happened at Rizal Medical Center in Pasig City of Philippines in October 2006 can be described to realize the contact of data mining in the area of healthcare. Due to lack of proper hygiene and cleanliness the hospital faced problem of deaths of new-born children because of neonatal sepsis (bacterial infection). Till the increase in deaths no one noticed about the cause and then hospital data was analyzed

Author [α] : Department of Computer Science & Engineering Jigjiga University, Jigjiga, Ethiopia. E-mail : fozbekri@gmail.com

Author [Ω] : Professor of Computer Science & Engineering Principal JNTUH of Engineering College, Jagityal, Karimnagar (Dt), A.P. India . E-mail : govardhan_cse@yahoo.co.in

where the Department of Health (DOH) originated that because of sepsis for instance, 12 out of 28 children born on October 4, died [12]. The DOH could find out the reason behind the cause and restrict them before things went out of control with the proper utilization of data mining to the past records.

IV. The Importance and Uses of Data Mining in Medicine and Public Health

The demand and want for data mining is more in field of healthcare, regardless of variations and conflicts in processes. Various discussions led to the demand of data mining in the field of healthcare which includes both public health as well private health. Many facts can be achieved from the past data stored in computers. Since the data is huge in quantity so it's a disadvantage for persons to examine the entire data and gain awareness [5]. Specialists consider that the improvement in medicals has reduced leading to complication of recent medical data. To overcome this drawback computers and data mining can be utilized.

Evidence-based medicine and prevention of hospital errors: On utilizing data mining on the available data much new informative and possibly life-rescuing information is achieved or else which would have left unutilized. For example, in recent research on hospitals and wellbeing it was originated that almost 87% of the deaths in the United States could have been reduced if the errors would have been lowered by the hospital staff [6].The preventive measurements could have been adopted by the hospitals and government supervisors using data mining to hospital data.

Policy-making in public health : To examine the relatedness among society health centers in Slovenia, Lavrac et al. [19] merged GIS and data mining via Weka with J48. The utilization of data mining in healthcare data helped health centers to determine methods that would lead to policy suggestions to the Public Health Institute. Decision making can be improved by proper utilization of data mining and decision support techniques.

More value for money and cost savings : Extra information can be obtained at less additional price by the firms and institutions using data mining. To know the information about scam in credit cards and insurance claims KDD and data mining is utilized [17].

Early detection and/or prevention of diseases : To obtain the before time recognition of heart problem which is the main public health issue through the world, Cheng et al mentioned the application of arrangement strategy.

Early detection and management of pandemic diseases and public health policy formulation: For before time identification and supervision of pandemics health specialists have preferred to utilize data mining. To obtain the reasons behind occurrence of diseases Kellog et al [3] discussed various methods which is a mixture of spatial modeling, simulation and spatial data mining. The output of examined data mining in the simulated situation can be utilized further to discover and organize disease causes.

V. The Reckoned Obstacles to Apply Data Mining in Medicine and Public Health Sector

Due to the unconventional behavior of the medical work utilization of data mining in the area of medicine is a difficult task. Different natural arguments among the traditional styles of data mining strategies and medicine were seen in the work of Shillabeer et al[11]. Data mining in the field of medical research begins with an assumption and then the outputs are altered so that the assumptions are achieved, which is different in case of standard data mining where data set begins without the assumptions in the beginning. Since traditional data mining is all worried about the development and models in data groups, data mining is more attracted in the alternative that do not match to the development and models. The main difference in methods is that most standard data mining neglects clarifying development and models and mainly focuses on relating. In the field of medicine explanation is a must for the reason that a distinction can lead to variation in the life or death of a human.

Let us consider an example of anthrax and influenza contributes to similar signs of respiratory troubles. When flu spreads the data mining test results in anthrax problem. The critical situation arises when the supposed flu problems is really anthrax epidemic [16]. In all the researches of data mining on disease and cure it is observed that the results were unclear and alerts, which included promoted outcomes but resulting in additional revise. This drawback points towards the present lack of credibility of data mining in the field of healthcare.

The definition of data mining is like a puzzle which results in further problems. The use of data mining in few researches has just explained the use of graphs which is not the correct meaning. According to Shillabeer [10] the confusion of data mining with inappropriate definition is widespread in the healthcare area.

Though the outcomes of data mining are useful the main disadvantage is to influence the health performers to modify their practices. Ayres [2] shows few incidents where the hospital staff when provided with proofs rejected to modify their hospital procedures. In few areas it was noticed that doctors ignored cleansing their hands after autopsy and cured patients with the same hands which lead to an increase in deaths of the patients.

As per Shillabeer [10] research he concluded that the hospital staff desires to accept the view of the head of the medical institution instead of the data mining outcome. His conclusion is suitable because in the medical institutions management's opinion is considered to be the best. One of the major disadvantages of data mining in healthcare is that the data of patients cannot be used for ethical and personal purpose. The amount of data should be less so that exact results can be obtained in case of data mining. The past data consists of private information which helps in knowing the disease and deaths can be prevented.

VI. Current State of The Art

WSARE was proposed by Wong et al [16], it is an approach to find out the before time cure for diseases. WSARE stands for "What's Strange About Recent Events" depends on the organizations policies and Bayesian methods. On simulation models WSARE is utilized which resulted in the exact guess for cure of simulated diseases. Before utilization of WSARE in actual life situations, safety measures must be considered.

Non-invasive diagnosis and decision support[13]: Every patient cannot afford for expensive, persistent and aching diagnostic and laboratory methods. For instance to discover cervical cancer using biopsy in women is a difficult task. K-means gathering algorithm was utilized by Thangavel et al [13] to detect the cervical cancer in women and he originated that the results of gathering data is more accurate then those of medical results. They even originated few factors which can help doctor in taking decision whether the patients suffering from cervical cancer must be suggested with biopsy or not.

Gorunescu et al[20] explained the use of data mining in the improvement of computer-aided diagnosis (CAD) and endoscopic ultrasonographic elastography (EUSE) to generate fresh non-invasive cancer identification. Ultrasound video was utilized by doctors to choose whether a patient must be suggested with biopsy or not in the traditional method.

Depending on the analysis of ultrasound video, the doctor's decision is biased. By utilizing data mining Gorunescu tried to solve this issue in other manner, he and his panel members paid attention on ultrasound videos rather than demographics. In the cases of malignant and benign tumors an organization algorithm was trained using a multi-layer perception (MLP). To differentiate between malignant and benign tumors, the pixels and the content of RGB was examined by the model. The outcome of the model was then utilized in other cases and it was noticed that the models output was appropriate and exact with very few variations in diagnosis.

Adverse drug events (ADEs): Few medicines and vaccines which were declared to be safe on humans are now realized to be injurious to humans when used for long periods. According to Wilson et al [17] he disclosed that data mining is used to investigate the drugs side effects in their data by US Food and Drug Administration. Around 67% of ADEs was discovered five years ago by Multi-item Gamma Poisson Shrinker or MGPS.

Medically Driven Data Mining Application : Recognition of Health Problems from Gait Patterns of Elderly: Bogdan Pogorelc et al [21] discovered a medically driven data mining application system for analyzing of walk models associated to the health problems of old people so that their independent living can be sustained.

The data gathered in this model is done using body antenna and RFID labels. Using motion capture gadgets the walking style of old people was captured, which included labels affixed to their body and antennas fixed in the building. To identify a particular health issue, a labels location is achieved by the antenna and then the responding time series of location directs are examined. To categorize the walking style of old people certain characteristics for training decision tree classifier and KNN classifier were introduced by the authors into:

a. normal
b. with hemiplegia,
c. with parkinson's disease,
d. with pain in the back and
e. with pain in the leg.

Bogdan Pogorelc et al [21] designed an automatic health-state identification, he introduced and examined 13 characteristics that were supported on the 12 labels which were affixed on the shoulders, elbows, wrists, hips, knees, and ankles of the old people. The introduced characteristics that are applied for modeling utilizing the machine learning processes are as shown:

i. Total variation among a) average space among right elbow and right hip and b) average space among right wrist and left hip.
ii. Right elbows normal angle.
iii. Proportion among maximum angle of the left knee and the right knee.
iv. Variation among the maximum and minimum angle of right knee.
v. Variation among a) maximum and minimum height of the left shoulder and b) maximum and minimum height of the right shoulder.
vi. Proportion among a) variation among maximum and minimum height of left ankle and b) maximum and minimum height of right ankle.
vii. Total variation among a) maximum and minimum speed of the left ankle and b) maximum and minimum speed of the right ankle.
viii. Total variation among a) average space among right shoulder and right elbow and b) average space among left shoulder and right wrist.
ix. Average speed of the right wrist.

x. Average angle among a) vector among right shoulder and right hip and b) vector among right shoulder and right wrist.
xi. Regularity of angle of the right elbow passing average angle of the right elbow.
xii. Variation among average height of the right shoulder and average height of the left shoulder.

By applying decision tree classifier and k-nearest neighbor classifier the tests were conducted. The main aim of test was to examine the categorization exactness of models, built applying the machine learning processes. By utilizing stratified 10-fold cross justification the tests appropriateness were achieved. The information for decision tree classifier was obtained from 7 labels and 5mm standard is the variation of sound. The information for KNN classifier was obtained from 8 labels and 0-20mm standard range is the variation of sound. Bogdan Pogorelc et al [21] reported that KNN results are exact when compared to decision tree results. The decision tree obtained 95% of exactness where as KNN obtained more than 99% of exactness.

Observation : In this paper the examining of the walking style of old people is done in connection to the health related issues so that they are maintained with their independent living. At the beginning stage (no sound, all labels) it was found that the decision tree obtained 90.1% exactness and k-nearest neighbor obtained 100% exactness. The elder people were provided by protection and assurance which resulted in less needless ambulance prices. Based on the results we conclude that k-nearest neighbor has achieved exactness of 99% with only 8labels and sound of 0-20mm standard when compared to decision tree which has achieved only 95% of exactness.

The implication of the characteristics applied to model the machine learning is not calculated and the elements picked to instruct the classifier is also not acceptable. So from the results we can say that the researches were just conducted to evaluate the presentation of decision tree and KNN tree considering the training constraints and opted elements. The study must validate the implication of the training characteristics and elements linked to healthcare field. Data mining is utilized in the area of before time identification of diseases, rescuing patients from deaths, enhancement of diagnosis and identification of dishonest health declarations. Data mining can be utilized in healthcare with few warnings.

Signaling Potential Adverse Drug Reactions from Administrative Health Databases: Huidong Jin et al [22] introduced an ADR indicating method that indicates sudden and irregular models feature of ADRs. To carry on the method Huidong Jin et al[22] discussed that all the present post market ADR indicating methods depend on unplanned ADR case results, which undergo grave underreporting and latency data. Because of ADRs there is an increase in hospitalization and deaths universally. On the other hand the administrative health data is gathered universally and regularly. This method consists of a domain-driven facts illustration Unexpected Temporal Association Rule (UTAR), its attractiveness measure, unexlev, and a mining strategy MUTARA (Mining UTARs given the Antecedent). It also proposed HUNT to emphasize the irregular and sudden models by evaluating their grades based on unexlev with those based on established influence.

Observation : Huidong Jin et al[22] proposed two interestingness measures, unexlev and rankratio, in the circumstance of indicating irregular and sudden models features of ADRs from organizational health information. He also introduced two easy but successful mining strategies MUTARA and HUNT to detect pair wise UTARs from the connected organizational health information QLDS. On evaluation to MUTARA the HUNT indicated a small number of strange ADR models.

Predictive Data Mining to Learn Health Vitals of a Resident in a smart Home: Vikramaditya Jakkula[23] mentioned about the practice of analytical data mining to discover health importance of a person staying in a smart home. He introduced a process where tools of smart home are discovering and acquiring their guess capabilities in the course of adjusting to smart home tenants.

During the method of analyzing with introduced model the gathering of data was done using a set of array of motion antennas that gathers information with the help of Argus antenna network. The gathered information is then improved with significant health information gathered by applying digital gadgets. The information gathering process continued for one hundred and fifty days period, which was done on a single tenant of the building.

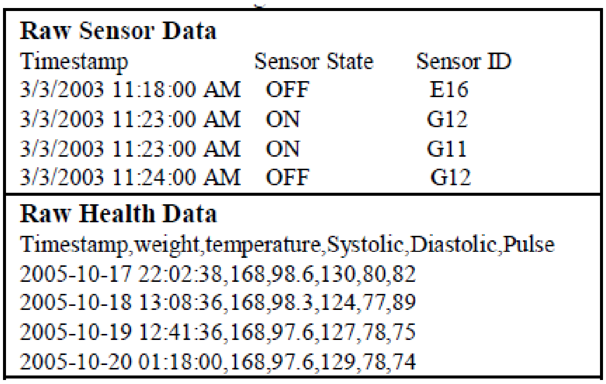
The information obtained from the motion antennas were applied directly as the information can be applied directly. Data samples can find in the fig 1.

WEKA [24] conducted the tests. The information of 150days was divided and grouped into training and testing data. Among the two tests carried out, the former test forecast health importance sign values.

The first test aimed on forecast examinations. The forecast enhancement is reliant of the classifier to educate. The presentations of different classifiers were analyzed beside the time series health information gathered from the occupants in smart home.

The test conducted concluded that KNN performed very well when compared to other practices like SMO regression, LazyL WL and Multi Layer observation.

The second test aimed on forecasting whether the next given time structure is odd or not?.. in the first test it was concluded that KNN is the best with 51% exactness and in the second test it is concluded that KNN is doing well in forecasting with 85% exactness.



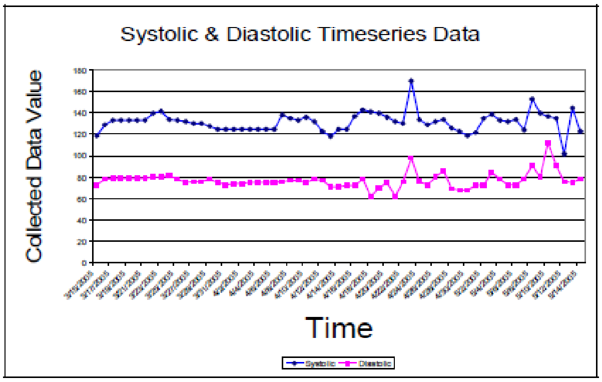
Raw Sensor Data

Timestamp	Sensor State	Sensor ID
3/3/2003 11:18:00 AM	OFF	E16
3/3/2003 11:23:00 AM	ON	G12
3/3/2003 11:23:00 AM	ON	G11
3/3/2003 11:24:00 AM	OFF	G12

Raw Health Data

Timestamp,weight,temperature,Systolic,Diastolic,Pulse
2005-10-17 22:02:38,168,98.6,130,80,82
2005-10-18 13:08:36,168,98.3,124,77,89
2005-10-19 12:41:36,168,97.6,127,78,75
2005-10-20 01:18:00,168,97.6,129,78,74

(a) Sensor readings as collected in a smart home.



(b) Systolic and Diastolic time series data plot.

Figure 1 : Data samples collected from sensors.

Observation : The tests indicate the prospective of data mining utilization in smart home study. In this process various learning strategies were evaluated and the conclusion was KNN performed well with 51% exactness for forecasting important health sign values and also has 85% exactness of forecasting of odd periods. Main emphasis was laid on forecasting the exactness of categories and the drawback is the validation of constraints opted as information features.

Patient Histories derived from Electronic Health Records: Jeremy Rogers et al[26] proposed data mining model called CLEF Chronicle to derive Patient Histories from Electronic Health Records, which is a representation of how a patient's illness and treatments unfold through time. Its primary goal is efficient querying of aggregated patient data for clinical research, but it also supports summarization of individual patients and resolution of co-references amongst clinical documents.

Properties of CLEF Chronicle: The CLEF Chronicle for an individual patient seeks to represent their clinical story entirely as a network of typed instances and their interrelations. Figure 1 illustrates the general flavor of what we are trying to represent: a patient detects a painful mass in their breast, as a result of which a clinic appointment occurs, where drug treatment for the pain is arranged and also a biopsy of the (same) mass in the (same) breast. The first clinic arranges a follow-up appointment, to review the (same) biopsy, which finds cancer in the (same) mass. This (same) finding is in turn the indication for radiotherapy to the (same) breast.

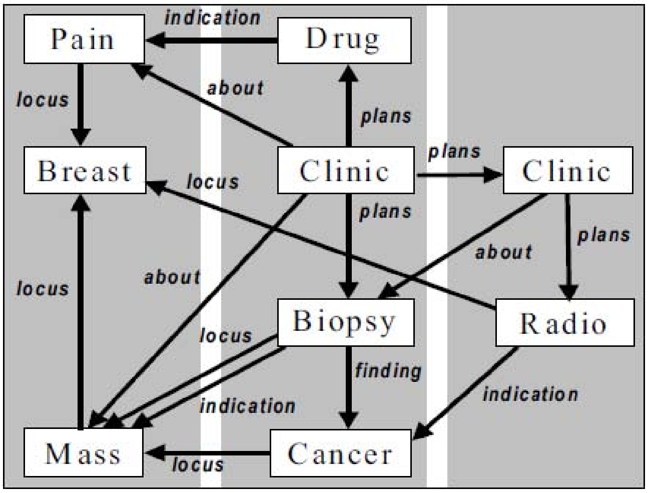


Figure 1[26]: Informal View of Patient-History Fragment (NOTE: Time-flow is roughly left ⍰ right)

In addition to the obvious structural difference between this representation and that of traditional electronic records, any clinical content represented as a CLEF Chronicle should have two central properties: Parsimony – traditional patient records contain multiple discrete mentions of relevant instances in the real world (the tumor, the breast etc). A CLEF Chronicle should have only one occurrence of each.

Explicitness – traditional patient records may only imply clinically important information (e.g. the fact of relapse). This must be explicit within a CLEF Chronicle Representation.

Functions of CLEF Chronicle :

- The CLEF Chronicle is intended to support more detailed, and more expressive, querying of aggregations of patient stories than is currently possible whilst at the same time improving the efficiency of complex queries. More detailed, because the Chronicle is more explicit: we can now ask e.g. how many patients 'relapsed' within a set period of treatment. More expressive, because the typing information associated with each Chronicle instance is drawn from a rich clinical ontology, such that queries may be framed in terms of arbitrarily abstract concepts: we can ask how many cancers of the lower limb were recorded, and expect to retrieve those of all parts of the lower limb. This is more efficient, because the traditional organization of patient records tends to require much serial or nested processing of records.
- An individual Chronicle can serve as an important knowledge resource during its own reconstruction from available electronic sources of traditional clinical records. In particular, a Chronicle can help resolve the frequent co-references and repeated

references to real-world instances such as characterize traditional records. For example, heuristic and other knowledge linked to an ontology of Chronicle data types can be used to reject any request to instantiate more than one identifier for a [Brain], or any attempt to merge two mentions of [Pregnancy] separated by more than 10 months.

- The Chronicle is intended to serve as a knowledge resource from which summarizing information or abstractions may be inferred. For example, deducing 'anaemia' from a run of discrete low blood counts obtained from the traditional record, or 'remission' from several years of clinical inactivity and 'relapse' when this is followed by a flurry of tests and a new course of chemotherapy. Similarly, where the record does not explicitly say why drug X was given, reasoners browsing a Chronicle may identify condition Y because the drug has no other plausible context of use.
- The CLEF Chronicle is intended to support automatic summarization of patient records. Given the sometimes chaotic nature of real patient records, manual case summarization is recognized as good clinical practice. Manual derivation of such summaries from the content of the record is, however, notoriously time consuming whilst the result of such labours is notoriously out of date whenever it would be most clinically valuable.

Observation: The idea of representing clinical information as some form of semantic net, particularly focusing on why things were done, is not new: echoes of it can be found in Weed's work on the problem oriented record [27]. Ceusters and Smith more recently advocated the resolution of co-references in clinical records to instance unique identifiers (IUIs) [28]. The semantic web initiative offers new possibilities for implementing such an approach, but the lack of any suitable clinical data severely constrains any practical experimentation. The model CLEF discussed here effective to provide a useful means to explore some of the computational and representational issues that arise.

Detecting Non-compliant Consumers in Spatio-Temporal Health Data [25]: K.S. Ng et al[25] attempt to describe their experience with applying data mining techniques to the problem of fraud detection in spatio-temporal health data in Medicare Australia. A modular framework that brings together disparate data mining techniques was adopted by authors. Several generally applicable techniques for extracting features from spatial and temporal data are also discussed. The system was evaluated with input from domain experts and observed high hit rates. Finally they concluded some conventions drawn from the experience.

Experimental Objectives that the authors considered was

i. Is the characterization of prescription shoppers given is accurate?
ii. Can Consumer RAS be used to identify prescription shoppers that do not rigidly fit the strong criteria. The false identification of genuinely ill patients that exhibit certain characteristics of prescription shopping as prescription shoppers have in the past been an issue for Medicare Australia. Can Consumer RAS avoid making such errors?

To conduct the experimental study, the data was picked randomly from a populous postcode in a major capital city of Australia in which they try to identify possible fraudulent activities.

The software that was used is LOF implementation in the dprep package in R for our analysis. The modified Huff model and the temporal feature extraction scheme was implemented in-house using C++.

Experiment I The first experiment seeks to verify the accuracy of our quantitative characterization of prescription shoppers. Only 12 people in the chosen postcode satisfy our criteria. These are passed on to domain experts for evaluation. We are interested in the percentage of these people that are true prescription shoppers.

Experiment II The second experiment seeks to verify whether it is feasible to use outlier detection technique to identify, with low false positive rate, prescription shoppers that do not fit the criteria identified. To do that, we remove all consumers identified in Experiment I from the data, perform a LOF analysis on the rest, and then pick out consumers that have high volumes of drugs of concern for evaluation by the experts. Fourteen consumers were picked this way. Some of these consumers have genuine needs for their drugs. In the aim to know, whether prescription shoppers tend to exhibit higher LOF scores compared to patients with genuine needs? In other words, do prescription shoppers show up as statistical outliers in the data? We excluded older consumers in this experiment because they are more likely to have genuine medical conditions and we do not want to spend time analyzing such patients. For the LOF analysis, the data are normalized and we compute the minimum LOF value with k ∈ [20, 50]. People identified in Experiment I were removed from the LOF analysis because we do not want to miss people who look normal with respect to that suspicious group.

Observation: Further work is required to assess the potential application of this work within the Medicare Australia compliance framework. Though the authors concluding a high degree of confidence in the validity of methodology for detecting prescription shoppers, it is presently unclear the extent to which the system could be used as a standalone and only method for identifying all prescription shoppers within a population. It is likely that the value of approach lies in targeting higher risk prescription shoppers. The true extent of our false negative rate with respect to the entire population, not just the targeted subset, needs to be quantified.

The main limitation for the system is not being able to see the MBS side of the story to augment what it can infer from a consumer's PBS record. Such restrictions on linking MBS and PBS claims data are due to legislative requirements and will continue to pose difficulties for us.

A second limitation is that the system was designed from the beginning to look at individual consumers. From a cost-benefit perspective, detection of a colluding group of consumers is clearly more useful.

VII. Conclusion

The research of data mining utilization in medicine and public health offered only synopsis of present observations and disputes. Healthcare institutions and firms would utilize this process of data mining to gain more facts and knowledge from the information that is already present in their institution records. An institution must mention all rules and strategies on the safety and confidentiality of victim's data, before boarding on data mining. The same rule must be said to its partners and implied to other institutions and branches. Public health related issues like pandemic occurrence, the want to identify the commencement of disease in a non-invasive, easy manner and the want to be more reactive towards the patients. All these factors must be considered important and the desire for health institutions to combine information and data mining must be utilized to examine this information.

References Références Referencias

1. Audain, C. 2007. Florence Nightingale. Online: http://www.scottlan.edu/lriddle/women/nitegale.htm. Accessed 30 July 2009.
2. Ayres, I 2008. Super Crunchers. New York: Bantam Books.
3. Bailey-Kellog, C. Ramakrishnan, N. and Marathe, M. Spatial Data Mining to Support Pandemic Preparedness. SIGKDD Explorations (8) 1, 80-82.
4. Cao, X., Maloney, K.B. and Brusic, V. 2008. Data mining of cancer vaccine trials: a bird's-eye view. Immunome Research, 4:7. DOI:10.1186/1745-7580-4-7
5. Cheng, T.H., Wei, C.P., Tseng, V.S. 2006 Feature Selection for Medical Data Mining: Comparisons of Expert Judgment and Automatic Approaches. Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems (CBMS'06).
6. Health Grades, Inc. 2007. The Fourth Annual HealthGrades Patient Safety in American Hospitals Study.
7. Kou, Y., Lu, C.-T., Sirwongwattana, S., and Huang, Y.-P. 2004. Survey of fraud detection techniques. In Networking, Sensing and [8]Control, 2004 IEEE International Conference on Networking, Sensing and Control. (2) 749-754.
8. Nightingale, F 1858. Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army.
9. Shillabeer, A 29 July 2009. Lecture on Data Mining in the Health Care Industry. Carnegie Mellon University Australia.
10. Shillabeer, A. and Roddick, J 2007. Establishing a Lineage for Medical Knowledge Discovery. ACM International Conference Proceeding Series. (311) 70, 29-37.
11. Tandoc, E.S 14 October2006.http://services.inquirer .net/print/print.php?article_id=26612.
12. Thangavel, K., Jaganathan, P.P. and Easmi, P.O. Data Mining Approach to Cervical Cancer Patients Analysis Using Clustering [14]Technique. Asian Journal of Information Technology (5) 4, 413-417.
13. Tufte, E. 1997. Visual Explanations. Images and Quantities, Evidence and Narrative. Connecticut: Graphics Press.
14. Wong, W.K., Moore, A., Cooper, G. and Wagner, M . What's Strange About Recent Events (WSARE): An Algorithm for the Early Detection of Disease Outbreaks, 2005. Journal of Machine Learning Research. 6, 1961- 1998.
15. Wilson A., Thabane L., Holbrook A 2003). "Application of data mining techniques in pharmacovigilance". British Journal of Clinical Pharmacology. (57) 2, 127-134.
16. Witten, I. H. and Frank, E. Data mining : practical machine learning tools and techniques. Morgan Kaufmann series in data management systems. Morgan Kaufman 2005.
17. Nada Lavrac, Marko Bohanec, Aleksander Pur, Bojan Cestnik, Marko Debeljak, Andrej Kobler: Data mining and visualization for decision support and modeling of public health-care resources. Journal of Biomedical Informatics 40(4): 438-447 2007.
18. Kenneth Revett, Florin Gorunescu, Abdel-Badeeh M. Salem: Feature selection in Parkinson's disease: A rough sets approach. IMCSIT 2009: 425-428.
19. Pogorelc, B.; Gams, M.; , "Medically Driven Data Mining Application: Recognition of Health Problems from Gait Patterns of Elderly," Data Mining Workshops (ICDMW), 2010 IEEE International Conference on , vol., no., pp.976-980, 13-13 Dec. 2010.
20. Huidong Jin; Jie Chen; Hongxing He; Kelman, C.; McAullay, D.; O'Keefe, C.M.; , "Signaling Potential Adverse Drug Reactions from Administrative Health Databases," Knowledge and Data Engineering, IEEE Transactions on , vol.22, no.6, pp.839-853, June 2010.

21. Jian Xu; Maynard-Zhang, P.; Jianhua Chen; , "Predictive Data Mining to Learn Health Vitals of a Resident in a Smart Home," Data Mining Workshops, 2007. ICDM Workshops 2007. Seventh IEEE International Conference on , vol., no., pp.163-168, 28-31 Oct. 2007.
22. I.H. Witten and Eibe Frank, Data Mining: Practical machine learning tools and techniques, 2nd Edition, Morgan Kaufmann, San Francisco, 2005.
23. Ng, K.S.; Shan, Y.; Murray, D.W.; Sutinen, A.; Schwarz, B.; Jeacocke, D.; Farrugia, J.; , "Detecting Non-compliant Consumers in Spatio-Temporal Health Data: A Case Study from Medicare Australia," Data Mining Workshops (ICDMW), 2010 IEEE International Conference on , vol., no., pp.613-622, 13-13 Dec. 2010.
24. J. Rogers; C. Puleston; A. Rector; , "The CLEF Chronicle: Patient Histories Derived from Electronic Health Records," Data Engineering Workshops, 2006. Proceedings. 22nd International Conference on , vol., no., pp.x109, 2006
25. Weed LI (1969) Medical records medical education, and patient care. The problem-oriented record as a basic tool. Cleveland, OH: Case Western Reserve University.
26. Ceusters W, Smith B. (2005) Strategies for referent tracking in Electronic Health Records. Journal of Biomedical Informatics (in press).