MEDICAL FRAUD DETECTION SYSTEM IN HEALTH INSURANCE SCHEMES USING LINK AND BASKET ANALYSIS ALGORITHM

Ogwueleka, Francisca Nonyelum
Department of Computer Science,
Federal University of Technology, Minna.
E-Mail: nonnyraymond@yahoo.co.uk
Phone No: +234-816-847-1775

Abstract

The loss by medical institutions due to fraudulent medical-healthcare transactions was noted to have amounted to a good percentage of the nation's yearly medical expenditure. Timely detection and prevention of fraud could aid in the recovery of lots of money in medical organizations. The designed system used link analysis and basket analysis algorithms, data mining techniques that provide the solutions required to uncover new fraud schemes and transform the new knowledge to medical business rules for real-time transaction screening through identification of unknown fraud patterns from the analysis of transactional data. These algorithms provided adequate means in discovering groups of providers sharing a large number of patients with valid patient identity as in the ghost patient billing fraud scheme and simultaneously uncovered groups of patients/patient IDs appearing in transactions performed by several providers with identified groups of providers rendering services to the same patients. The designed detection system is a main-memory optimized high performance Java cross-platform disk-based relational database management system (DBMS) featuring real-time replication and advanced disaster recovery. The results from the analysis prove that they are effective algorithms for accurate detection of medical fraud. especially in areas of provider-patient and ghost patient billing fraud.

Keywords: Data mining, medical fraud, ghost patient billing, provider-patient, link analysis, basket analysis

Introduction

Frauds have plagued medical institutions for a long time (Ogwueleka, 2011a). There are different types of frauds in the medical industry and these frauds cost them millions of naira per year. Fraud detection has become an imperative and vital task for the medical industry (Ogwueleka, 2009b). Currently, a number of methods have been implemented to detect frauds, from both statistical approaches, such as data mining and hardware approaches, such as firewalls, and smart cards (Ogwueleka, 2009a). Fraud detection is an incessantly growing subject and requires a tool that is intelligent enough to adapt to criminals strategies and ever changing tactics to commit fraud. Data mining is a popular way to fight frauds because of its effectiveness (Ogwueleka and Inyiama, 2009a)

Data mining is the search for new, valuable and nontrivial information in large volumes of data. It is a cooperative effort of humans and computers. Best results are achieved by balancing the knowledge of human experts in describing problems and goals with the search capabilities of computers (Han & Kamber, 2000). Timely information on fraudulent activities is strategic to the health care institutions. Medical fraud detection is very essential. Medical frauds are not so evident, and are also not easy to detect (Ogwueleka, 2008). Medical care fraud is produced when either the provider practices are inconsistent with business or medical practices, and result in an unnecessary cost or reimbursement of services that are not medically necessary or that fail to meet professionally recognized standards for health care (Ogwueleka, 2011a).

Medical fraud can take place at different stages. Fraud in health care industry can be grouped into illegal actions related to associate, medical professionals, staff and manager, and suppliers. According to the National Health Care Anti-Fraud Association (NHCAA, 2005), health care fraud is an intentional deception or misrepresentation made by a person, or an entity that could result in some unauthorized benefit to him or his accomplices. Typical medical fraud schemes include billing for services not actually performed; falsifying a patient's diagnosis to justify procedures that are not medically necessary; misrepresenting procedures performed to obtain payment for non-covered services (such as cosmetic surgery); "upcoding" (billing for a more costly service than the one actually performed); "unbundling"(billing each stage of a procedure as if it were a separate procedure); accepting kickbacks for patient referrals; waiving patient co-pays or deductibles and over-billing the insurance carrier or benefit plan (involves both the provider and the patient)(Megaputer, 2007).

Billing for services never rendered and charging for more expensive procedures are just two ways that fraudulent health care providers affect patients (Ogwueleka, 2011a). The sad fact remains that health care providers have also been known to perform unnecessary medical services for the singular purpose of collecting insurance payments. Fraudulent providers also falsify medical treatment histories or diagnoses of medical conditions and use up patients' health care benefits, putting people's lives at risk. The possibility to drain a patient's private insurance benefits means that when they might really be needed, a patient may not have access to the appropriate insurance amounts required for adequate treatment. If a patient's medical insurance is depleted, that may affect future treatments and in serious cases lead to premature death. For health care payers, the long-term benefits of implementing a fraud detection solution offset high the initial implementation costs.

In monitoring medical billing fraud, one of the most common services provided by medical billing review companies is the ability to detect Current Procedural Terminology (CPT) code unbundling. CPT codes are an established list of five-digit numbers used to identify the medical procedures and services provided by physicians. The unbundling problem occurs when doctors submit their bills listing charges for routine procedures that should be classified under a specific CPT code, but instead are broken up and filed as a combination of several separate CPT codes. Physicians

do this because they can get more reimbursement money for the sum of the individual procedures than for the single composite service. This behaviour is illegal and constitutes insurance fraud, but it occurs on a regular basis. Companies performing review of medical billing for insurance purposes often claim to perform data mining on the submitted bills to look for these filing patterns. In actual sense, they usually run some low-level expert systems or neural networks that have been programmed to look for specific types of known patterns. Again, the data mining involving the discovery of the suspicious patterns initially occurred offline and the known patterns are incorporated into a set of rules to be matched automatically against online data (Westpal & Blaxton, 1998).

The result of extreme fraudulent claims is excessive billing amounts, excessive per-doctor patients, higher per-patient costs, higher per-patient tests, etc (Ogwueleka, 2011a). This excess can be identified using special analytical tools. Provider statistics include total amount billed, total number of patients, total number of patient visits, per-patient average billing amounts, per-patient average visit numbers, per-patient average medical tests, per-patient average medical test costs, per-patient average prescription ratios of specially monitored drugs, etc. When abusive claims are repeated frequently, the consequent is higher provider statistics. Various provider statistics can be used to identify fraudulent claims. Statistical analytic techniques can reveal excessive providers who might be completely unintelligent but it will be difficult to identify modest level fraud activities. Sophisticated techniques are thereby applied.

Medical fraud detection involves account auditing and detective investigation. Careful account auditing can reveal suspicious providers and policyholders (Rosella, 2008). It is better to audit all claims carefully one after another. Auditing all claims is not possible by any practical means and it is extremely difficult to audit providers without substantial clues. A practical approach is to develop short lists for scrutiny and perform auditing on providers and patients in the short lists. Various analytic techniques can be employed in developing audit short lists. Excessive fraudulent claims lead deviations in aggregate claims statistics. Fraudulent claims also develop into patterns that can be detected using predictive models Ogwueleka, 2009b).

Health care agencies are seeking automated tools to aid them in identifying and flagging suspicious activities and sorting of the valid transactions from the fraudulent ones. Massive volume and complexity of healthcare transactions complicate the task of their timely and accurate validation of transactions and prevention of fraud (Ogwueleka, 2011a).

The task of fighting fraud involves two major analytical steps, namely discovering unknown patterns and relations signifying new fraud schemes as solving this task will help to find past offenders and their fraudulent transactions; and incorporating discovered knowledge as business rules for screening new transactions, flagging and blocking fraudulent transactions in real time (Megaputer, 2007). Automatic fraud detection helps to reduce the manual parts of a fraud screening/checking process becoming one of the most established industry/government data mining applications (Phua et al, 2005).

Data mining, as a process that uses a variety of data analysis tools to discover patterns and relationships in data can be used for making valid prediction. The work of data mining is to analyze a huge amount of data and to extract some practical information that can be interpreted for future uses. In doing this, the purpose of data mining has to be defined and the right structure of possible model or patterns that fit to the given data set found out. When the right model for the data have been gotten then the model can be used for predicting future events by classifying the data (Ogwueleka, 2008). Data mining can be used in different kinds of databases (e.g. relational database, transactional database, object-oriented database and data warehouse) or other kinds of information repositories (e.g. spatial database, time-series database, text or

multimedia database, legacy database and the World Wide Web) (Han & Kamber, 2000). Therefore, data to be mined can be numerical data, textual data, graphics or audio.

Data mining techniques are used to discover hidden knowledge, unknown patterns and new rules from large data sets, which maybe useful for a variety of decision making activity. With the increasing economic globalization and improvements in information technology, large amounts of medical data are being generated and stored. These can be subjected to data mining techniques to discover hidden patterns and obtain predictions for trends in the future and the behaviour of the medical institutions. With the immediacy offered by data mining, latest data can be mined to obtain crucial information at the earliest. This in turn would result in improved medical institution responsiveness and awareness of fraud leading to reduced costs and increased revenue. Data mining techniques start by sampling or selecting some of the training data like 20 percent or less of the total. An algorithm is then applied to explore the training data, seeking patterns in it. Patterns are then tested and refined on data, which have been kept aside for this purpose, called the "test" data. In addition to the training and test sets, it uses a "validation" set to estimate generalization error, in order to see how well the model performs under conditions of actual use.

Link and basket analysis algorithm installed on computers are used to spot unusual patterns in health care claims. It flags any indicative of possible fraud including - medical provider charging far more than peers for particular services, medical providers that provide more tests or procedures per patient than peers, medically improbable procedures, such as one patient having dozens of the same tests, given the choice of similar treatments, billing for the more expensive one more often than peers, high percentage of patients traveling long distances for routine services or tests, and high prices for medical equipment or supplies that can be purchased for far less. Association algorithms are types of data mining algorithm used to find correlations between different attributes in a dataset. The most common application of this kind of algorithm is for creating association rules, which can be used in a market basket analysis. The study was limited to two types of medical fraud, which include provider-patient fraud and ghost patient billing

The objectives of this study are as follows: automating the process of finding relationships and patterns in raw data and utilization of the results in an automated decision support system; the use of data mining technique on the large volumes of data available in the medical industry to identify fraudulent patterns and trends in their transactions; the use of the identified patterns and trends relating to fraudulent activities to provide training to the designed model, which will then learn to detect such fraudulent operations, and the use of data mining techniques to break barriers in business transactions by helping the medical industry become an agile competitor able to harness strategic business opportunities.

Literature Review

Data mining provides the technology to analyze mass volume of data and/or detect hidden patterns in data to convert raw data into valuable information (Chye et al, 2002). In healthcare, data mining is becoming increasingly popular, if not increasingly essential. Several factors have motivated the use of data mining applications in healthcare. The existence of medical insurance fraud and abuse, for example, has led many healthcare insurers to attempt to reduce their losses by using data mining tools to help them find and track offenders (Christy, 1997). There have been reports of successful data mining applications in healthcare fraud and abuse detection (Milley, 2000). Another factor is that the huge amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods. Data mining can improve decision-making by discovering patterns and trends in large amounts of complex data (Biafore, 1999). Data can be a great asset to healthcare organizations, but they have to be first transformed into information.

The goal of data mining is to learn from data, and there are two broad categories of data mining strategies: supervised and unsupervised learning (Matkovsky & Nauta, 1998). The method used in this study is unsupervised modeling. The attributes and models of fraud are not known, but the patterns and clusters of data uncovered by data mining can lead to new discoveries.

The concept of fraud detection has been founded on data mining techniques such as association rules and classification. Research on fraud detection has been focused on pattern matching in which abnormal patterns are identified from the normality. Some of theses are the Instruction Detection Framework and algorithms for pattern comparison proposed by Lee et al (1999). Major & Riedinger (2002) presented a tool for the detection of medical insurance fraud. They proposed a hybrid knowledge/statistical-based system, where expert knowledge is integrated with statistical power. Another framework presented for the detection of healthcare fraud, is a process-mining framework by Yang & Hwang (2006). The framework is based on the concept of clinical pathways where structure patterns are discovered and further analyzed.

Phua et. al (2005) highlights fraud committed in insurance industry as one of the most studied in terms of the number of data mining-based fraud detection publications, existing four sub-groups of insurance fraud detection: home, crop, automobile and medical insurances. In (Yamanishi et al, 2004), an on-line discounting learning algorithm was used to indicate whether a case has a high possibility of being a statistical outlier in data mining applications such as fraud detection is used for identifying meaningful rare cases in health insurance pathology data from Australia's Health Insurance Commission (HIC). The performance of a k-Nearest Neighbor (kNN) algorithm with the distance metric being optimized using a genetic algorithm was applied in a real world fraud detection problems faced by the HIC (He et al, 1999). The hot spots methodology that entails the use of clustering and rule induction techniques has been used to identify possible frauds in the Australian Governments public health care system, Medicare (Williams, 1999). Becker et al (2002) identified the effects of fraud control expenditures and hospital and patient characteristics on upcoding, treatment intensity and health outcomes in the Medicare and Medicaid programs. A data mining framework that uses the concept of clinical pathways (or integrated care pathways) was utilized for detecting unknown fraud and abusive cases in a real-world data set gathered from the National Health Insurance (NHI) program in Taiwan (Yang & Hwang, 2005). Another model that uses attributes mainly derived from various expense fields of claims by experts' consultants was also designed to detect suspicious claims in the Taiwan NHI program (Chan & Lan, 2001).

In this study, Link Analysis algorithms were used as they prove to be efficient tools for identifying frauds such as patient-provider fraud while additional analytical algorithms are used for detecting other types of fraudulent transactions such as ghost patient billing. The Basket Analysis algorithm developed provided an adequate means to discovering groups of providers sharing a large number of patients. These algorithms were used because they gave the most appropriate and accurate results in terms of patient-provider medical fraud and providers sharing larger number of patients.

Methodology

Medical fraud detection requires compilation of potentially huge data, involving complex computation and sorting operations. The data mart platform for medical fraud detection was based on the designed architecture where the claim payment records are first transformed and loaded into healthcare fraud data mart as shown in Figure 1. Data is added into data mart, either monthly or quarterly basis. Summary information is created for providers, doctors and policyholders. Expert systems engines are used to analyze, score and detect potentially risky providers and claims. The auditors and investigators finally analyze the data.

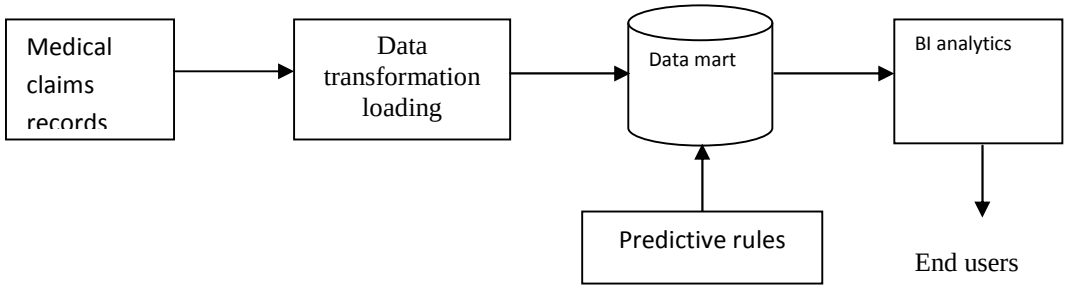

Figure 1: Architecture of the designed data mart platform for healthcare fraud detection

Medical-healthcare claims data marts proposed contain potentially huge amount of information. Asthe complexity in detecting fraudulent claims makes fraud detection extremely challenging, the features that contributes to the proposed system success are fast database management system, predictive modeling, expert systems, and chart and report writing.

The proposed system has a big main-memory with optimized high performance cross-platform database management system and also very fast sort engine which drives sorting and aggregation operation. It uses up to 512 gigabyte main memory. This is essential for medical-healthcare data marts since its BI platform supports a number of advanced predictive modeling methods such as neural network, decision tree, regression, rule-based expert systems engine, and others such as link and basket analysis. Fraud patterns are transformed into audit and screening rules, and applied in screening and detecting fraudulent claims. The system has a built-in charting and report-writing engines incorporated with predictive modeling and expert systems engines.

In identifying and reducing the number of fraudulent medical transactions covering a state in Nigeria, data obtained from the medical organization specializing in medical billing solutions, which developed business rules for capturing known fraud mechanisms based on their background knowledge was used. Indiscriminate manual screening of data demonstrated that the system was able to identify about three quarter of medical fraudulent transactions, with the one-quarter being concealed by different fraud mechanisms, which could not be caught by a set of predefined business rules.

A system capable of identifying unknown fraud schemes directly from the analysis of medical transactional data, which will help in the increase rate of detecting fraudulent transactions, was developed. The data for the analysis had a standard medical data format listing patient name, provider name, date of service, diagnosis, type of procedure, billed and paid amounts for each procedure. The system was able to identify and flag suspicious providers and collections of medical transactions, which is an indicative of fraud and therefore required further investigation and decision-making.

The approach involves the following steps, which is also illustrated in Figure 2 (Ogwueleka, 2008):

1. Select an appropriate algorithm
(i) Implement the algorithm in software
(ii) Test the algorithm with known data set
(iii) Evaluate and refine the algorithm as you test with other known data sets
(iv) Publish the results

There are four basic steps, which were used in order to complete the implementation of the proposed medical fraud detection system. They are data selection, data transformation, applying algorithms and results interpretations. The objective of data selection was to determine the type of information and the way it is organized. Some part of the medical data available from the source data file was required and it helped in quick collection of relevant data after identification. The required medical data was sampled and the sample was mined. After data selection, the data was transformed, then algorithms were applied using two data mining techniques in order to extract the required information, which will aid in achieving the required objective. The result of applying data mining algorithms was interpreted. The result was also analyzed using a visualization and decision support tool. The designed model was validated and tested. There was need in some aspect to refine the data and repeat the process sequence again.

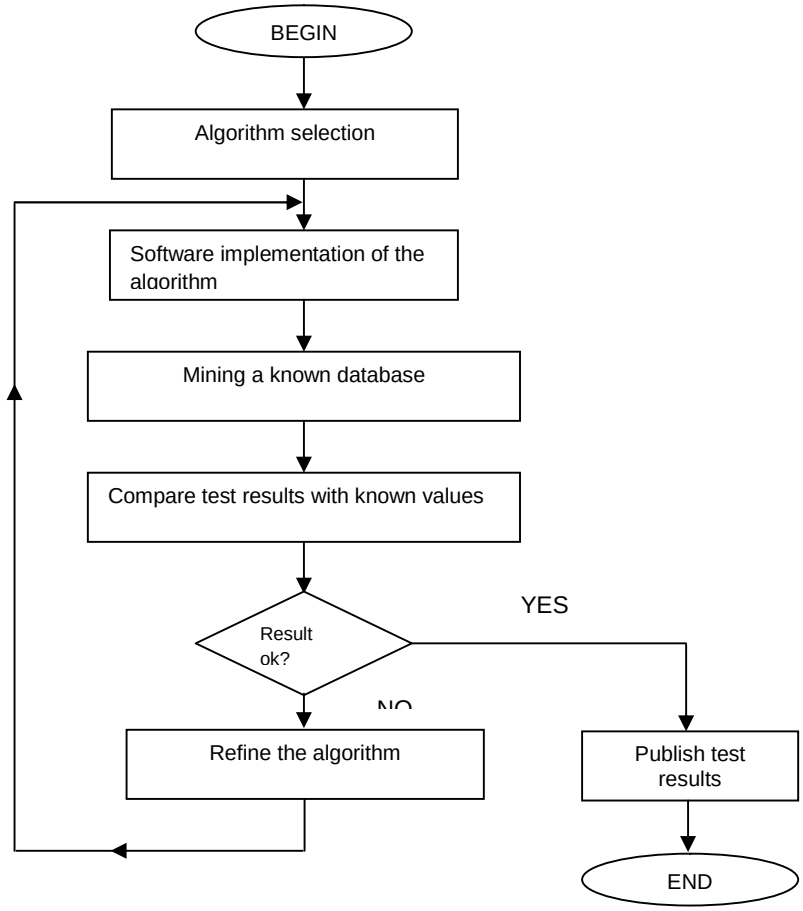

Figure 2: Steps used in the data mining approach

Results

To reduce the volume of medical data to be analyzed and concentrate on possible patients for fraud, the records of patients who received more than 70 medical transactions during one calendar year was isolated and the records studied. The resulting data contained about 220,000

transactions, which were explored with the help of advanced analytical algorithms of the data mining system.

The algorithm for the statistics summary demonstrated that the remaining records corresponded to 800 patients, 320 types of performed medical transactions, and 750 providers, 10 of which were larger hospitals performing more than 3,000 individual transactions yearly.

In detecting different medical fraud occurring for medical institutions, Link Analysis algorithms are very efficient tools for identifying frauds such as patient-provider fraud while additional analytical algorithms are best suited for detecting other types of fraudulent transactions such as ghost patient billing. This is noted from the model results obtained after testing.

The Basket Analysis algorithm provided adequate means in discovering groups of providers sharing a large number of patients and/or just valid patient IDs as in the ghost patient billing fraud scheme. This algorithm simultaneously uncovered groups of patients/patient IDs appearing in transactions performed by several providers, and also identified groups of providers rendering services to the same patients.

The designed detection system is a large main-memory optimized high performance Java cross-platform disk-based relational DBMS, featuring real-time replication, advanced disaster recovery, MM-DBMS level speed, embedded operation and supports SQL. It is optimized for large main-memory. Its single buffer architecture allows not only caching a large number of data pages, up to 512GB, but also handles large sorting data. The algorithm was written in Java. It can run on various platforms, such as, Windows, Mac OS X, and Linux. The databases created can be ported onto other platforms without costly conversions. As it was Java 100%, it can be integrated into the medical institution's J2EE web-servers or Java business applications. Embedding the designed DBMS is as simple as adding an archive file and calling start/stop APIs. The embedding greatly improved the performance of the system by virtually eliminating communication cost.

The detection system provides various data recovery mechanism and replication. It does not only prevent major data loss from disasters and computer virus attacks, but it also removes the need to make daily or periodic backups. The system employed advanced locking and concurrency control mechanisms.

Discussion

Medical transactions can be so complicated that a medical fraud investigator has to work very hard to separate fraudulent transactions from legitimate ones. The presence of a large number of overlapping patients should definitely raise doubt/suspicion for an analyst and require further investigation. The aim of a fraud detection system is not only to discover past fraudulent transactions and create a list of the corresponding offenders, but also to screen every new transaction with the discovered business rules and prevent processing suspicious transactions in real time. An analyst should schedule the most suspicious transactions for a more detailed manual scrutiny.

To automate the process of knowledge discovery and applying business rules for verification of transactions, a unique platform for rapid visual development of reusable push-button analytical solutions was utilized. This system allows an analyst to create and distribute advanced analytical scenarios throughout the organization. The system automatically executes these scenarios when certain predefined conditions become true. The medical institution gains the capability to quickly and consistently identify new fraud schemes and monitor the validity of submitted transactions in real time.

The data mining process used the following six basic steps: defining the problem, preparing data, exploring data, building models, exploring and validating models, and deploying and updating models. As creating a data-mining model is a dynamic and iterative process, the six steps were utilized. After exploring the available medical data, it was noted that the data was insufficient to create the appropriate mining models and that more data will be looked for. Several models were built before realizing that they do not answer the problem posed when the problem is defined and so the problem was redefined. The models were updated after they have been deployed because more medical data became available. Most steps in the data mining process were repeated as many times as needed to create a good model.

The most commonly used techniques in data mining are artificial neural networks (non-linear predictive models that learn through training and resemble biological neural networks in structure), decision trees (tree-shaped structures that represent sets of decisions which generate rules for the classification of a dataset), genetic algorithms (optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution), and rule induction (the extraction of useful if-then rules from data based on statistical significance). The scope of this study limited it to only link and basket analysis.

Analyzing links in data mining establishes relationships between the records in the database which would otherwise be impossible to find because they cannot be predicted and so cannot be found other than by accident. It is a relatively recent technique, which has become well known through shopping basket analysis, which indicates popular combinations..

Link analysis is performed by investigators in many areas, from epidemiology to fraud detection, from criminal investigations to the study of social networks. Linkage data is typically modeled as a graph, with nodes representing entities of interest to the domain, and links representing relationships or transactions. An example is collection of medical data subpoenaed for a criminal investigation. The links as well as nodes may have attributes specific to the domain or relevant to the method of collection, such as, link attributes indicating the certainty or strength of a relationship, or probability of a fraud. Link analysis is distinct from techniques that construct connectionist networks, Bayesian belief networks, and association rules. These techniques discover and represent associations based on the aggregate statistical characteristics of a sample of instances drawn from some population, while link analysis begins with data that can be represented as a network and attempts to infer useful knowledge from the nodes and links of that network.

Link analysis helped in this study to ask and answer questions of: which nodes are crucial or central to the network, which links can be strengthened to enhance the operation of the network effectively, can the existence of undetected links or nodes be inferred from the known data, are there similarities in the structure of sub-parts of the network which may indicate an underlying relationship, what are the relevant sub-networks within a much larger network, and what data model and level of aggregation best reveal certain types of links and subnetworks.
Analyzing medical information revealed valuable medical fraud information. Basket analysis helped to determine what aspect of the medical fraud that occurs at the same time or at different times.

Conclusion

Medical institutions have to be able to distinguish fraudulent activities from legitimate business transactions. Fraud schemes are rapidly changing and medical institutions need to be able to differentiate new fraud patterns without an explicit prior knowledge of these patterns. Data

mining tools capable of processing large volumes of data, determining rules for separating fraud from legitimate transactions, and detecting unusual events deviating from normal operation patterns should be utilized by the medical institutions.

The two regular medical fraud methods, provider-patient fraud and ghost patient billing can be discovered using link analysis and basket analysis algorithms. These two methods cannot be exposed through the use of simple predefined business rules. Link analysis as a subset of network analysis, explores associations between objects. Link analysis helped in this study by providing the crucial relationships and associations between many objects of different types that are not apparent from isolated pieces of information. The designed automatic computer-based link analysis employed will aid the medical industry greatly in reducing fraudulent medical activities.

The problem of medical fraud detection has several key characteristics, such as, fraud being rare, and reliability of the medical data. Typically less than 0.1% of the transactions are fraudulent in the case of medical fraud. This means that there are very few examples of fraud in the data. The data is highly skewed and this is the root cause of many of the problems associated with this class of problem. The marking/tagging of transactions as being fraudulent is usually a manual task and is often subject to various sources of error. These errors effectively introduce a level of 'noise' into the data. This was a major problem. Despite these limitations, the results obtained from the use of these two algorithms – link and basket analysis shows their effectiveness in this area.

This study has shown that the implementation of data mining tools and techniques can increase the quality and timeliness of detecting medical fraud, and the real-time flagging of suspicious transactions. Medical institutions can be saved by timely discovery and elimination of fraudulent transactions in healthcare.

References

Becker, D., Kessler, D. & McClellan, M. (2002). Detecting medicare abuse. *Journal of Health Economics, 24, 189–210.*

Biafore, S. (1999). Predictive solutions bring more power to decision makers. *Health Management Technology, 20(10), 12-14.*

Chan, C. L. & Lan, C. H. (2001). *A data mining technique combining fuzzy sets theory and Bayesian classifier – An application of auditing the health insurance fee.* Proceedings of the International Conference on Artificial Intelligence, 402–408.

Chye, K. H., Leong G., & Chan K. (2002). Data mining and customer relationship marketing in the banking industry. *Singapore Management Review*.

Christy, T. (1997). Analytical tools help health firms fight fraud. *Insurance & Technology*, *22*(3), 22-26.

Han, J. & Kamber, M. (2000). *Data mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann.

He, H., Graco, W., & Yao, X. (1999). *Application of genetic algorithms and k-nearest neighbour method in medical fraud detection.* Proceeding of SEAL1998, pp. 74– 81.

Lee, W., Stolfo, K., Salvatore J. & Mok, K. W. (1999). *Algorithms for mining system audit data,* computer science department, Columbia University.

Matkovsky, I.P. & Nauta K.R. (1998). *Overview of data mining techniques.* Presented at the Federal Database Colloquium and Exposition: San Diego, CA.

Megaputer (2007). *Medical fraud detection - megaputer case study in data mining.* Moscow: Megaputer intelligence, Ltd.

Milley, A. (2000). Healthcare and data mining. *Health Management Technology*, *21*(*8), 44-47.*

Major, J. & Riedinger, D. (2002). EFD: A hybrid knowledge/statistical based system for the detection of fraud. *Journal of Risk and Insurance, 69*(3), 309-324.

National Health Care Anti-Fraud Association (NHCAA). (2005). *Health care fraud: A serious and costly reality for all Americans,* 2005 REPORT, http://www.nhcaa.org

Ogwueleka, F. N. & Inyiama, H. C. (2009a). Credit card fraud detection using artificial neural networks with a rule-based component*. IUP University Journal of Science and Technology*, *5(1), 40-47.*

Ogwueleka, F. N. (2011a). *Medical fraud detection system using link and basket analysis algorithm.* 3rd International Conference on Mobile e-Services. Nigeria.

Ogwueleka, F. N. (2011b). Data mining application in credit card fraud detection system. *Journal of Engineering, Science and Technology, 6(3), 311-322.*

Ogwueleka, F. N. (2009b). Applications of data mining techniques in healthcare data. *Botswana Journal of Technology*, 18(2).

Phua, C., Lee V., Smith, K., & Gayler, R. (2005). A comprehensive survey of data mining-based fraud detection research. *Artificial Intelligence Review*. Monash University, Clayton, Australia. Pp 24-21.

Rosella predictive knowledge and data mining (2008). *Rosella BI Platform for healthcare data mart solutions.* Pp 4-6.

Westpal, C. & Blaxton, T. (1998). *Data mining solutions methods and tools for solving real-world problems*. New York: Wiley Computer Publishing. Pp 8-15, 34-40.

Williams, G. (1999). *Evolutionary hot spots data mining: An architecture for exploring for interesting discoveries.* In Proceeding of PAKDD99. pp 15.

Yang, W. S. & Hwang, S. Y. (2006). A process-mining framework for the detection of healthcare fraud and abuse. *Expert Systems with Applications*, *31*, 56-68.

Yamanishi, K., Takeuchi, J., Williams, G. & Milne, P. (2004). On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms*. Data Mining and Knowledge Discovery, 8, 275–300.*