Fraud Detection in Health Insurance Using Expert Re-referencing

Fabida A.¹, Jasila E.K.²

¹PG scholar, ²Asst. Professor
¹,²Department of Computer Science and Engineering
MES College of Engineering, Kuttippuram
Kerala, India

Abstract—Fraud is widespread and very costly to the healthcare insurance system. Fraud involves intentional deception or misrepresentation intended to result in an unauthorized benefit. It is shocking because the incidence of health insurance fraud keeps increasing every year. In order to detect and avoid the fraud, data mining techniques are applied. Frauds blow a hole in the insurance industry. Health insurance is a bleeding sector with very high claims ratio. So, to make health insurance industry free from fraud, it is necessary to focus on elimination or minimization of fake claims arriving through health insurance. Here, supervised and unsupervised techniques is employed to detect fraudulent claims along with expert re-referencing is also employed to detect claims efficiently.

Keywords—data mining, fraud, health insurance fraud, supervised, unsupervised

I. INTRODUCTION

Today, human life is getting some unexpected incidents in their life such as accident, health problem. The accident will make a huge loss for human being so every persons want to save their own life and they need a back up to serve their life.

The insurance company helping peoples who met accident and suffering from health problem. There is a difference enclosed by fraud prevention and fraud detection. The fraud prevention describes measures to recoil fraud to occur. The fraud detection involves identifying fraud as fast as possible, once it has been committed.

Health insurance company that results in healthcare benefits being paid illegitimately to an individual or group is known as health insurance fraud. The main purpose of fraud is financial benefit. Insurance is a contract (policy) in which an individual or entity earn financial protection or reimbursement against losses from an insurance company. Insurance generally restricted into four types:

- home insurance
- life insurance
- motor insurance
- medical insurance

Amongst the motor insurance and medical insurance sector has more fraud problems. This paper is describing the details of medical insurance fraud detection. There are three major parties involved in the medical insurance system,
Service Providers: Includes the party who provide health care services to the insurance subscribers and in turn get the payments from insurance subscribers for the services provided. Service providers include doctors, hospitals and laboratories.

Insurance Subscribers: Includes the ones who receive regular premiums from their subscribers and pay them back the settlement amount. Insurance Carriers include governmental healthcare departments and private insurance companies.

Insurance Carriers: They render the services from the service providers and make the payments for those services. The subscribers in turn get the reimbursement (settlement amount) from the the insurance carriers. Insurance Subscribers include patients and patients’ employers.

The service providers including doctors, hospitals, ambulance companies and laboratories. The insurance subscribers including patients and patients employers. The insurance carriers who receive regular premiums from subscribers and pay health care cost on behalf of their subscribers. According to which party commits the fraud, fraud behaviours can be classified as follows:

Service Provider’s Fraud include:

– Billing of services which are not actually rendered.
– Unbundling Billing each stage of services as if it were a separate treatment.
– Upcoding Billing more costly services than actually performed.
– Performing unnecessary medical services for generating more payments.
– Falsifying patient’s diagnosis or the treatment history to justify tests, surgeries that are not medically necessary.

Insurance subscriber’s fraud include:

– Falsifying the records of employment for obtaining a lower premium rate.
– Filing claims for healthcare services which are not actually received.
– Using the coverage of some other person to illegally obtain the healthcare benefits.

Insurance Carrier’s Fraud include:

– Falsifying the reimbursements.
– Falsifying the benefits/service statements. The main focus in this paper is on detection of the above mentioned Service Provider’s fraud using data mining technology.

The rest of the paper is organized from section II to VII. Section II describes related works. Section III describes Existing system, section IV describes problem definition, section V describes proposed work are described in section. The performance of the paper is analysed and concluded in section VI and VII respectively.

II. RELATED WORKS

Researches are always been conducted to detect and avoid fraudulent pattern in health insurance. This paper briefly presents some of such effective approaches to detect fraudulent behaviour.
NMF algorithm is a clustering algorithm[1], for clustering medical treatment items, such as medicines or medical measurements, into several groups according to usage of different patients. Then each group is considered as a kind of medical treatment items for curing identical symptoms. If medical treatment item shifts from one cluster in this pace to another cluster in next pace, then this algorithm could classify the patient using this medical treatment item as a fraud suspicious. In the end, all these fraud-suspicious cases are submitted to medical experts for detailed careful detection.

The K-Means method is a very famous algorithm for clustering high-dimensional data[2]. Cluster centres are initiated, which assigns every data point to its nearest center, and then readjusts the centres, reassigns the data points, until it stabilizes. All medical claim records are entered into the system, the Euclidean distance between each pair of the claim forms is calculated. The user of the system sets the desired maximum distance for the records to belong to the dupe cluster. The sum claimed for all the claim forms is computed and the average amount determined. If the amount for a given claim overtake a set amount for that cluster, then that particular claim form is considers as fraud suspicious.

Bayesian co-clustering is a co-clustering methods for detection of conspiracy fraud[3]. Which describes the dyadic dynamic that connects providers and beneficiaries. Co-clustering enables us to group providers and beneficiaries synchronously, that is, the clustering is interdependent. The objective of the approach is to identify potentially fraudulent associations among the two parties.

Most common data mining techniques used for finding fraudulent records is anomaly detection[4]. This technique aims to detect outliers or anomalies which vary from the usual patterns.

### III. EXISTING SYSTEM

Hybrid method[5] is used for detecting health insurance frauds. In this approach, the insurance claims are clustered accordant to the illness type and then they are classified to detect any duplicate claims. For this, Evolving Clustering Method (ECM) and Support Vector Machines (SVM) are used for clustering and classification.

Hybrid approach is the combination of both clustering and classification were, clustering is an exploratory data analysis mechanism, it can sort different materials or properties into groups in such a way that the degree of association between two objects is maximal if they follow to the same group and minimal otherwise. And, cluster analysis can be integrated with high-throughput experimentation for fastly screening combinatorial data.

Classification is one of the core tasks of Data Mining. A classification technique is a systematic approach to building allocation models from training and testing data sets. Several classification models such as Decision Tree Classifier, Rule-Based Classifier, Neural Network Classifier, naive Bayesian Classifier, Neuro-Fuzzy classifier, Support Vector Machines. Each procedure employs a learning algorithm to analyse a model that best fits the relationships between the quality set and class label of the input data. The model achieved by a learning algorithm should both fit the input data well and correctly conclude the class labels of data set that has never seen before. Therefore, the key objective of the data mining algorithm is to build frames with good generalization capacity.

ECM is used to cluster dynamic data. Dynamic data are those which keep on changing with respect to time. As and when new data point comes in, ECM clusters them by modifying the position and
size of the cluster. There is a parameter known as radius associated with each cluster that determines the boundaries of that cluster. Initially, the cluster radius is set to zero. The radius of the cluster increases as more data points are added to that cluster.

A support vector machine is a supervised learning technique used in classification. It has an initial training phase where data that has already been classified. After the training phase is finished, SVM can predict into which class the new incoming data will fall into. In training and classification phase, two class labels are defined then choose support vectors that separates the claim into two classes and identify new incoming claims in to either legitimate or fraudulent class.

1) Training (Preprocessing Step):
   * Define two class labels viz. "legitimate” or ”fraudulent”.

2) Classification:
   * Identify the new incoming claims into either ”legitimate” or ”fraudulent” class.

Steps in hybrid model construction are as follows:

* Doctor bills patients for the services/equipment given to them during their treatment.
* Patients files claims to the insurance company and the claims are submitted to the hybrid framework.
* There is an expert who detect the claims for further investigation with the insurance company.
* Legitimate claims are further passed to the insurance company and those claims are paid to the patients.

IV. PROBLEM DEFINITION

Hybrid approach is the effective method to determine fraudulent claims. There is an expert who flags the fraudulent claims for further investigation with the insurance company. The limitation of hybrid approach is that expert efficiency calculation or re-referencing is not done here.

V. PROPOSED SYSTEM

In this section, We propose expert re-referencing method for detecting health insurance fraud. Hybrid approach, is the method were the insurance claims are clustered according to the disease type and then they are classified to detect any duplicate claims. Expert re-referencing is the modification that is done for identifying claims efficiently with the help of group of experts.

A. EXPERT RE-REFERENCING

In addition to expert decision regarding the claim is fraud or not, re-referencing of expert decision is proposed. Re-referencing is the process by which the expert decision is given to hub which consist of group of experts who compares the decision and finally the best decision is given to insurance company. Experts are the person who gives best decision regarding claims fraud or not with insurance company.

B. SYSTEM OVERVIEW
Fig. 1. System Model

The figure 1 represents the modified view of overall system in which patient apply for the claims and claims are submitted to hybrid framework and there is an expert who take the decision regarding claim is fraud or not. Then the decision of expert is given to hub which consist of group of experts who compares the decisions. Finally the best decision is given to the insurance company then, the legitimate (satisfied) claims are further passed to the patients.

Steps in system model are as follows:

- Patient files claim.
- Claims are submitted to hybrid framework were clustering follows classification.
- Decision given to expert who detect claim as fraud or not.
- Current expert decision is given to hub which consist of group of experts.
- Compares the decisions of current expert and group of experts.
- Best decision is given to insurance company and satisfied claims are paid to patients.

VI. RESULT

Fraud detection using hybrid approach is conducted by combining clustering and classification. Legitimate and fraudulent claims are identified by experts so that insurance company can make good decisions. Although this method have a better performance, further enhancements done for making the system more efficient.

Fraud detection using expert re-referencing is the enhancement that have been done for making system more efficient. Because in expert re-referencing not only one expert decision is considerd. By giving the expert decision to hub which consist of group of experts. Group of expert decision considerd and comparing the decisions, the best decision is given to the insurance company.
VII. CONCLUSION

Fraud detection in health insurance is not an easy task. Fraud detection can be done by different methods such as, Non-negative matrix factorization, K-means clustering technique, Bayesian co-clustering, Anomaly detection technique and Hybrid approach. Hybrid approach combines the advantages of supervised and unsupervised technique and it clusters and classifies the claims. Expert re-referencing is an enhancement to hybrid approach and which is a promising task to detect fraud efficiently. As fraud becomes more sophisticated and the volume of data grows, it becomes more difficult to recognize fraud from bulk of data. Fraud detection may not eliminate fraud but it can be surely reduced. The major task of Expert re-referencing is to detect fraudulent claims with concern of group of experts so that insurance company can make sensible decisions. The main issue of the expert re-referencing is that expert efficiency calculation is not done and future scope of work is related in this area.

REFERENCES


