A Review on Health Insurance Claim Fraud Detection

Faseela V. S[1], Dr.P.Thangam[2]
[1]PG Scholar, CSE Department, Coimbatore Institute of Engineering and Technology, India
[2]Associate Professor, CSE Department, Coimbatore Institute of Engineering and Technology, India

Abstract— The anomaly or outlier detection is one of the applications of data mining. The major use of anomaly or outlier detection is fraud detection. Health care fraud leads to substantial losses of money each year in many countries. Effective fraud detection is important for reducing the cost of Health care system. This paper reviews the various approaches used for detecting the fraudulent activities in Health insurance claim data. The approaches reviewed in this paper are Hierarchical Hidden Markov Models and Non Negative Matrix Factorization. The data mining goals achieved and functions performed in these approaches have given in this paper.

Keywords: *Hidden Markov Models, Non Negative Matrix Factorization*

I. INTRODUCTION

In several countries fraudulent behavior in health insurance is a major problem. Data mining tools and techniques can be used to detect fraud in large sets of insurance claim data. One of the most common data mining techniques used for finding fraudulent records is anomaly detection.

This paper aims to review various approaches used for Health insurance fraud detection. There are three major parties involved in the entire system,

(1) Service Providers
(2) Insurance Subscribers
(3) Insurance Carriers

The Service Providers including doctors, hospitals, ambulance companies and laboratories. The Insurance Subscribers including patients and patient's employers. The Insurance Carriers who receive regular premiums from subscribers and pay health care cost on behalf of their subscribers.

There is a difference between fraud prevention and fraud detection. The fraud prevention describes measures to avoid fraud to occur. The fraud detection involves identifying fraud as quickly as possible, once it has been committed.

According to the National Health care anti-fraud association, health care fraud is the misrepresentation of Claims for gaining some shabby benefits. The health industry in India is losing approximately Rs.600 crores on "false claims" every year. So to make health insurance feasible, there is a need to focus on eliminating or reducing fraudulent claims.

Generally there are two types of frauds.

First one is Hard fraud: This is a deliberate attempt either to point an event or an accident, which requires hospitalization or other type of loss that would be covered under a medical insurance policy.

Second one is Soft fraud: Which occur when people purposely provide false information such as claim fraud, application fraud and eligibility fraud sources and then put to use by data miners to achieve the desired results.

The rest of the paper is organized as follows. Two approaches are explained in section II. Comparative study presented in section III. Concluding remarks are given in section IV.

II. HEALTH CARE FRAUD DETECTION TECHNIQUES

Data Mining for Healthcare Management is an emerging potential area with respect to its impact on improving healthcare as a result of discovering new patterns and trends in voluminous data generated by healthcare transactions. Some of the existing approaches of data mining for health insurance fraud management are been listed below.

II.A. Using Hierarchical Hidden Markov Models

In this approach [1], first decomposes the dataset into groups of claimants of similar age since the age contributes to a

patient's medical conditions. The approach applies recursively the Gaussian mixture clustering and HMM procedures on randomly selected samples from the training set until the classification errors converge to a prescribed minimum threshold is observed. The entire process as follows,

(i) FEATURE EXTRACTION

To extract the temporal behaviors of the claimant, for example unique personal identifier, date of claim, age of the claimant and total claims per day.

(ii) CREATION OF COHORTS

The claimants of similar age range forms an age cohort. The above two steps are preprocessing steps which prepare the data for the application of pattern discovery techniques.

(iii) CLUSTERING

Gaussian mixture clustering is applied to identify cluster in the data for each age cohort.

(iv) PATTERN DISCOVERY

The pattern discovery is accomplished by using Hidden Markov models (HMM).This model commonly used in temporal behavioural pattern discovery. The steps (iii) & (iv) executed recursively until convergence occurred. This recursive process yields a set of HMMs which are hierarchically organized.

II.B. Using Non Negative Matrix Factorization

This paper [2] proposes a Non-Negative Matrix Factorization (NMF) method for fraud detection, which introduces a technique for clustering medical treatment items such as medicines or medical measurements in to several groups according to usage of different patients.

Then each group is considered as a kind of medical treatment items for curing similar symptoms. If a medical treatment item shifts from one cluster in this month to another cluster in next month, then this algorithm could classify the patient using this medical treatment item as a fraud suspicious patient.

In the end, all these fraud suspicious patients are submitted to medical experts for detailed careful detection. The factorization can be used to compute a low rank approximation of a large sparse matrix along with preservation of natural data Non-Negativity.

Each vector component is given a positive value (or weight) if the corresponding medical treatment item is used by the patient and a zero value otherwise, the resulting matrix is always non-negative.

III. COMPARITIVE STUDY

Table 1: Comparative study of the Healthcare Fraud Detection Systems

TITLE	TECHNIQUE USED	MERITS	DEMERITS
Hierarchical Hidden Markov Models An Application to Health Insurance Data	Hierarchical Hidden Markov Models(HHM)	Recursive training of HMMs provides a mechanism to detect redundancy in the dataset	Computational loads are high when the training datasets are large
Health Care Fraud Detection Using Nonnegative Matrix Factorization	Nonnegative Matrix Factorization(NMF)	Identification of unknown types of fraud	A distributed NMF is needed for a large dataset

IV. CONCLUSION

In conclusion, this paper reviews two approaches for detecting fraudulent behavior in health insurance claim. By analyzing the aforementioned techniques, we will get a clear idea for the future work in health insurance claim fraud detection. In India, we have three levels of health care network, namely primary, secondary, and tertiary. It provides an opportunity for data miners to use the huge amount of data. The main task is to integrate data from different sources and then put to use by data miners to achieve the desired results.

REFERENCES

[1] Ah Chung Tsoi, Shu Zhang, Markus Hagenbuchner, "Hierarchical Hidden Markov Models An Application to Health Insurance Data", 2006.

[2] Shunzhi Zhu, Yan Wang, Yun Wu, "Health care Fraud Detection Using Nonnegative Matrix Factorization", *The 6th International Conference on Computer Science & Education*, 2011.

[3] Pedro A.Ortega ,Cristian J. Figueroa , Gonzalo A. Ruz, "A Medical Claim Fraud/Abuse Detection System based on Data mining: A Case Study in Chile", *International Conference on Data mining*, USA, 2006.

[4] Qi Liu ,MiklosVasarhelyi, "Health fraud detection: A survey and a clustering model incorporating Geo-location information", 2013.

[5] Melih Kirlidog, Cuneyt Asuk,"A fraud detection approach with data mining in health insurance", 2012.

[6] Dallas Thornton, Roland M. Mueller, Paulus Schoutsen, Jos van Hillegersberg, "Predicting Healthcare Fraud in Medicaid: A Multidimensional Data Model and Analysis Techniques for Fraud Detection", 2013.

[7] H. Lookmansithic,T.Balasubramanian, "Survey of Insurance Fraud Detection Using Data Mining Techniques", *International Journal of innovative Technology and Exploring Engineering*, 2011.

[8] Anand Sharma, Vibhakar Mansotra, "Emerging applications of Data mining for Health care Management-A Critical Review", *IEEE* Explore, 2011.

[9] Jiawei Han, Michelinekamber, Jian pie, *Data Mining Concepts and Techniques*, Third Edition 2012