

Enhancing health care fraud detection system using fuzzy classification and hidden markov model.

¹ Survase Komal, ² Sangshetty Manisha, ³ Pawar Smita, ⁴ Waghire Rajani, ⁵ Prof.P.B.Jawalkar

¹UG student, ²UG student, ³UG student, ⁴UG student, ⁵UG student,

¹Computer Engineering,

¹JSPM'S BSIOTR, Pune, India

Abstract— In health care claims, relationships between doctors and patients could lead to potential fraud discoveries. Many of the health care insurance companies are existed to provide facility to the patients, but all of them are suffer from the headache of fraud insurance claims from the doctors. So as a tiny step towards this the proposed system develops a web application panel for both the doctors and insurance companies to identify the fraud claims of the doctors at insurance company's end using Hidden markov model which is powered with fuzzy classification. Majority of the methodology are having low performance, parameters which eventually effect health care fraud detection. This paper mainly focusing on different methodologies of health care fraud detection by considering numerous entities.

Index Terms— Data mining, Fraud detection, Health care, Clustering, Classification.

Introduction

Health care has become a major disbursement in the US but fraud and prostitution lead to additional expenses in health care system for united state and some other countries. Basically health care fraud is an misrepresentation of services made by the Doctor or other entity with the result in some kind of unauthorized benefit of doctor or entity. Hence health insurance fraud detection became a challenging task. There it is necessary to cease the fake claims and make health insurance company free from frauds. This paper discuss claims data from physician and processed at insurance company end to detect the fraud. There are many types of frauds but some of most common frauds are mentioned below:

- Billing services that is costlier than actual provided services.
- Billing of non-covered services as covered services.
- Billing for medical equipment more expensive than actual equipment.
- Misrepresentation of dates of services.
- Misrepresentation of location of services.
- Incorrect reporting of procedure.
- Two doctors treats the same patient.
- One doctor refer a patient to another doctor.

To detect all these frauds some of the data mining techniques are used. This paper proposing the C-means clustering algorithm and hidden Markov model.

Related Work

[1] Surveys different data mining techniques which is applied to the application. health insurance fraud is an intentional act which make financial benefit to individual or group. to detect all the fraud claim it uses data mining techniques which is divided ,such as supervised and unsupervised learning techniques. basically this paper concentrate only on advantages and disadvantages of data mining techniques. many fraud detection system uses either supervised or unsupervised learning algorithms. the supervised algorithms uses vector machines, and neural networks. but this algorithms failed to detect unexceptional conditions, so it does not gives satisfied result.

[2] Explains supervised and unsupervised algorithms to detect the fraud at insurance company end. the major drawback of both these algorithm is, they cannot classify or divide the fraud claims as per diseases. so in this section, uses Evolving Clustering Method (ECM) for clustering and support vector machine(SVM) for classification of data according to diseases type and then they are classified to detect duplicate claims.

Evolving Clustering Method (ECM):

ECM is used to cluster or divide the dynamic data. dynamic data means ,the data which keep changed with respect to time. when new data set comes in system, ECM clusters that data by modifying the size and position of cluster.

support vector machine(SVM):

The support vector machine is a supervised learning technique used for classification. It has an initial training phase where data that has already been classified according to the algorithm. After the training phase is finished, SVM can predict into which class the new incoming data will fall into both of these ECM and SVM methods are involved in system called as hybrid model.

[3] Focuses on frauds and uses techniques such as Data mining empowers a variety of insurance providers with the ability to predict which claims are fraudulent so they can effectively target their resources and recoup significant amounts of money. Data mining helps medical insurance company to focus, for example, on claims with high percentage of recoverable fraud, isolate factors which indicates a payment request has a high probability of fraudulence, develop rules to use them to flag only claims likely to be fraudulent, and ensure adjusters could review claims that are not only likely to be fraudulent but also have the greatest adjustment potential.

[4] Narrates insurance companies use human inspections and heuristic rules to detect fraud. First, it is impossible to detect all emerge constantly. health care fraud by manual inspection over large databases. Second, new types of health care fraud. The goal of this paper is to detect suspicious health care frauds from large databases. In order to achieve this goal, this paper applies two clustering methods, SAS EM and CLUTO , to a large real-life health care database and compares the performances of these two methods. In this study, SAS Enterprise Miner and CLUTO, are used for clustering analysis.

DATA UNDERSTANDING AND PREPROCESSING:

The storage and processing of this data was conducted within the insurance company's infrastructure in compliance with privacy regulations. A typical example is City and State addresses. They can be removed because zip code variable provides city and state information. Inclusion of redundant variables will increase computational complexity and slow down the clustering process.

EMPIRICAL CLUSTERING RESULTS:

Most clustering algorithms require the users to input the number of clusters they desires. Therefore, we conduct a series of clustering experiments and discuss the results with business experts to decide which numbers of clusters are desirable. Assessing cluster quality is the most challenging step in clustering analysis. Different clustering methods may generate very different clusters for the same dataset. Researchers have proposed various quality measures for clustering and most of them can be grouped into two categories: external quality measure and internal quality measures .

[5] Introduce an effective medical claim fraud/abuse detection system based on data mining used by a Chilean private health insurance company. There is a difference between fraud prevention and detection Data mining which is part of an iterative process called knowledge discovery in databases (KDD) can assist to extract this knowledge automatically.

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. and fuzzy logic and genetic algorithms are used. The following few methods are used [A] Entities and Medical Claim Data,[B].

Business and Data Understanding,[C]Data Preparation, [D]Modeling [E]Incorporation into Fraud Detection Workflow.

[6] In this paper, we introduce different methods and techniques to detect it. There' use multilayer perceptron neural network, Neural Nets (NN), Bayesian Nets (BN), Naive Bayes (NB), Artificial Immune Systems (AIS), Decision Trees (DT), this are the Different Techniques for Fraud Detection. The aims of this paper are to assess the use of technique of decision trees. In combination with the management model CRISP-DM, this technique helps in the prevention of bank fraud . presents a system which is able to prevent subscription fraud in fixed telecommunications with the high impact on long distance carriers. So classification module, in divided into four different categories: subscription fraudulent, otherwise fraudulent, insolvent and normal. and On the classification module, this was implemented using fuzzy rules.

System Overview

To enhance the process of fraud claims detection of the doctors at the insurance company's end proposed method put forwards an idea of identifying fraud claims by clustering the claims based on the protocols by using the C-means clustering technique which is then poweredwith Hidden markov model to extract the fraud list and this process is catalyzed by fuzzy logic classification.

C-means Clustering method works by assigning membership to each data point. More the data near to cluster center more is its membership towards the particular cluster center. Clearly, summation of membership of each data point equal to one. After each iteration membership and clusters center are updated according to some formulas. Another one is a Hidden Markov Model(HMM) is a powerful statistical tool for modeling generative sequences that can be characterised by an underlying process generating an observable sequence. HMMs have found application in many areas interested in signal processing, and in particular speech processing, but have also been applied with success to low level NLP tasks such as part-of-speech tagging, phrase chunking, and extracting target information from documents. Andrei Markov gave his name to the mathematical theory of Markov processes in the early twentieth century but it was Baum and his colleagues that developed the theory of HMMs in the 1960s. Markov Processes figure 1 depicts an example of a Markov process. The model presented describes a simple model for a stock market index. The model has three states, Bull, Bear and Even, and three index observations up, down, unchanged. The model is a finite state automaton, with probabilistic transitions between states.

Given a sequence of observations, example: up-down-down we can easily verify that the state sequence that produced those observations was: Bull-Bear-Bear, and the probability of the sequence is simply the product of the transitions, in this case $0.2 \times 0.3 \times 0.3$.

Hidden Markov Models figure 2 shows an example of how the previous model can be extended into a HMM. The new model now allows all observation symbols to be emitted from each state with a finite probability. This change makes the model much more expressive.

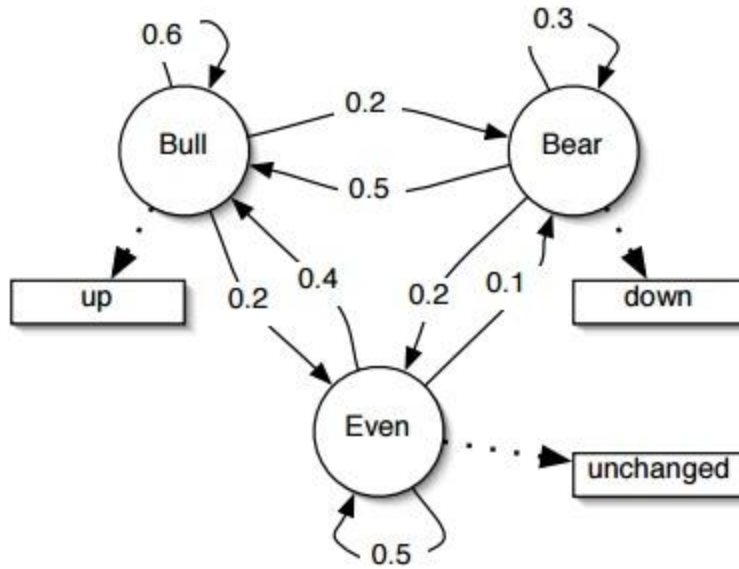


Figure 1: Markov process example

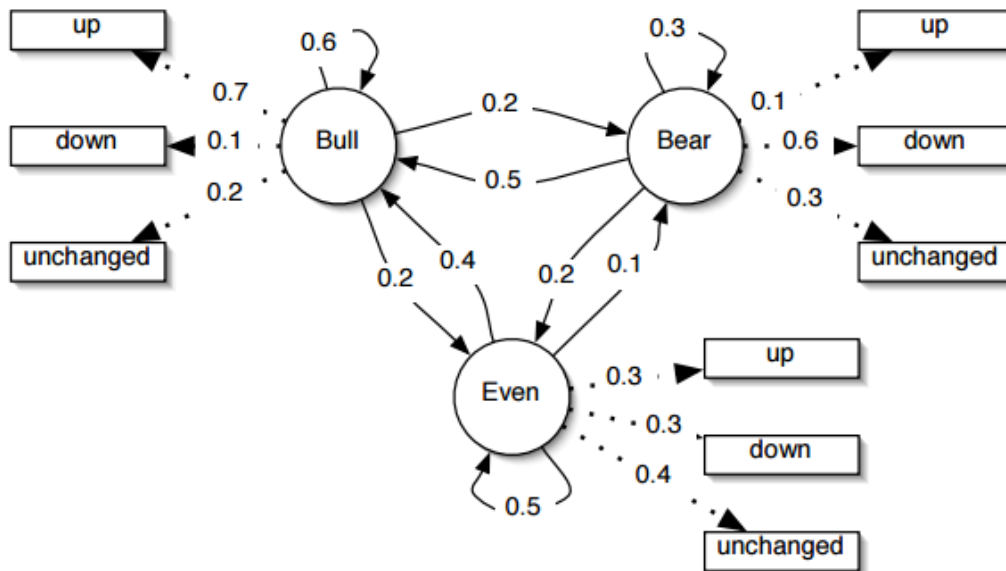


Figure 2: Hidden Markov model example

and able to better represent our intuition, in this case, that a bull market would have both good days and bad days, but there would be more good ones. The key difference is that now if we have the observation sequence up-down-down then we cannot say exactly what state sequence produced these observations and thus the state sequence is ‘hidden’. We can however calculate the probability that the model produced the sequence, as well as which state sequence was most likely to have produced the observations. The next three sections describe the common calculations that we would like to be able to perform on a HMM. The formal definition of a HMM is as follows $\lambda=(AB,\pi)$(1)

S is our state alphabet set, and V is the observation alphabet set:

$$S=(s_1, s_2, \dots, s_N) \dots \dots \dots (2)$$

$$V=(v_1, v_2, \dots, v_M) \dots \dots \dots (3)$$

We define Q to be a fixed state sequence of length T, and corresponding observations O:

$$Q=q_1, q_2, \dots, q_T \dots \dots \dots (4)$$

$$O=o_1, o_2, \dots, o_T \dots \dots \dots (5)$$

Fuzzy Classification

Fuzzy Logic - The aim of Fraud Detection of insurance claim is based on extraction Hidden models of insurance claims. One of the methods to get the appropriate claims is to consign some numerical measure of a claims for the fraud known as Fraud weighting and then select the best ones.

The system involves of the following core Steps:

Step A: In the fuzzifier, crisp inputs are taken, which are result of the HMM hidden model features.

Step B: After fuzzification, the inference engine refers to the rule base containing fuzzy IFTHEN rules.

Step C: In the last step, we get the final fraud score. In inference engine, the most important part is the definition of fuzzy IF-THEN rules. The essential claims are extracted from these rules according to our fraud criteria. Sample of IF-THEN rules are described below.

Here in our rules we mentioned the claims score in between 0 to 1 as follow.

VERY LOW ----- 0 TO 0.2

LOW -----0.21 TO 0.4

MEDIUM -----0.41 TO 0.6

HIGH -----0.61 TO 0.8

VERY HIGH -----0.81 TO 1.0

So any claims falls in between very high and high are having the tendency of doing fraud claims.

Conclusion

Due to increase in awareness of the health in everyone's life health insurance plays a vital role. Now days this health care sector industry is the one of the leading sector in the business. So this obvious gives a reason for scammers and fraudsters to get their means. So this paper mainly concentrates on the study of different aspects of health care fraud aspects and their detection system narrated by the other authors. And this study leads us to think about the other innovative options about the health care fraud detection techniques like by using machine learning methodologies and fuzzy classification theory which will be described in our next edition.

References

- [1] "Survey on Hybrid Approach for Fraud Detection in Health Insurance", Punam Devidas Bagul, Sachin Bojewar and Ankit Sanghavi, [2016].
- [2] "Fraud Detection in Health Insurance using Data Mining Techniques", Vipula Rawte and G Anuradha, [2015].
- [3] "Data Mining for Fraud Detection", Manjunath K.V and Patharaju S.D, [2015].
- [4] "Application of Clustering Methods to Health Insurance Fraud Detection", Yi Peng, Gang Kou, Alan Sabatka, Zhengxin Chen, Deepak Khazanchil and Yong Shi
- [5] "A Medical Claim Fraud/Abuse Detection System based on Data Mining: A Case Study in Chile", Pedro A. Ortega, Cristi'an J. Figueroa and Gonzalo A. Ruz
- [6] "Survey on Fraud Detection Techniques Using Data Mining", Muhammad Arif and Amil Roohani Dar.
- [7] "Using Data Mining to Detect Health Care Fraud and Abuse", Hossein Joudaki, Arash Rashidian, Behrouz Minaei-Bidgoli, Mahmood Mahmoodi, Bijan Geraili, Mahdi Nasiri, [2015].
- [8] "DEA restricts narcotic pain drug prescriptions", Radnofsky, L., and Walker, J., 2014.