International Journal of Advanced Trends in Computer Applications
www.ijatca.com

A Model System to Identify Health Care Frauds

Rupali Gorad[1], Archana Augustine[2]
[1,2]Department of Information Technology,
Pillai HOC College of Engineering & Technology, Rasayani,
Raigad, 410207, India
[1]rgorad2191@gmail.com, [2]aan.archana@gmail.com

Abstract: *As the human life march towards the modern amenities all the sectors of life become more and more advanced, Health care is not spared from this. The revolutionary health care policy concept eventually facilitates all the patients irrespective of any cast and creed to avail the best services of the doctors for their diseases. Many of the health care insurance companies are existed to provide this facility for the peoples, but all of them are suffer from the headache of fraud insurance claims from the doctors. Many systems are existed to deal with these kinds of fraud health insurance claims from the doctors, but most of them are not up to the mark to identify the proper fraud detection operandi. 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.*

Keywords: *Clustering algorithms, Matrix converters, Detection algorithms, Insurance, Algorithm design and analysis*

I. INTRODUCTION

Fraud detection is a area applicable to many industries such as banking and financial sectors, insurance, government agencies and law enforcement, and more. ... Millions of transactions can be searched through the use of sophisticated data mining tools, to spot patterns and detect fraudulent transactions.

There are different types of frauds in health care systems, such as drug abuses, counterfeit drugs, off-label marketing issues. In this paper, we will focus on the health insurance claims. When health services are provided, a set of claims is submitted to one or more insurers for reimbursements. Health insurance is like any other types of insurance that there is a claim processing system to adjudicate these claims to determine if a claim should be paid or by how much a claim should be paid.

Top00.00.....00.0.0.0.0.0.0.0.0.0000000000revent the possible frauds, there are multiple levels of edits within the claim processing systems. Some edits are implemented to prevent the incorrect payments and are part of pre-payment system adjudication. Some edits are implemented after the payments have been made to the health providers and they are the post-payment edits.

As the human life march towards the modern amenities all the sectors of life become more and more advanced, Health care is not spared from this. The revolutionary health care policy concept eventually facilitates all the patients irrespective of any cast and creed to avail the best services of the doctors for their diseases. Many of the health care insurance companies are existed to provide this facility for the peoples, but all of them are suffer from the headache of fraud insurance claims from the doctors.

Many systems are existed to deal with these kinds of fraud health insurance claims from the doctors, but most of them are not up to the mark to identify the proper fraud detection operandi. 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

II. RELATED WORK

Community Detection is a well studied area. There are so many researches that have been done in Community Detection algorithm development. [1] In this paper, they developed algorithms that target at one type of frauds. That is the suspicious provider communities that either share patients between or refer patients to each other. These communities are usually small and have exclusive relationships within the communities and no outside connections. The relationships between these communities are suspicious; however we couldn't be

100% confident that these communities are conducting fraudulent activities. There are other factors we need to consider, for example, incomplete data etc. These communities can be put on a watch list for further investigations and review. The additional review will help prevent incorrect payments to go out to these groups of providers or patients.[10] In "Credit Card Fraud Detection Using HMM" paper, They have proposed an application of HMM in credit card fraud detection. The different steps in credit card transaction processing are represented as the underlying stochastic process of an HMM. They have used the ranges of transaction amount as the observation symbols, whereas the types of item have been considered to be states of the HMM. They have suggested a method for finding the spending profile of cardholders, as well as application of this knowledge in deciding the value of observation symbols and initial estimate of the model parameters. It has also been explained how the HMM can detect whether an incoming transaction is fraudulent or not. Experimental results show the performance and effectiveness of our system and demonstrate the usefulness of learning the spending profile of the cardholders. Comparative studies reveal that the Accuracy of the system is close to 80 percent over a wide variation in the input data. The system is also scalable for handling large volumes of transactions.
One of the most popular algorithms is to use Modularity as the objective function to optimize the cluster assignments until it reaches to an optimal structure. These algorithms are effective to parse the whole physician network into smaller communities based on their similarities. The selection of similarities is another research topic. Some algorithms use the distance between nodes as the similarity measurements. Some similarities are based on the existence of connections. In this paper, we select to use the connection-based similarities rather than the distance-based similarities. In another research we conducted, we developed an algorithm based on spectral analysis and it can parse a network into similar communities using Fielder Vector.

III. SYSTEM OVERVIEW

3.1 System Architecture

The Proposed System Steps depicted in figure 1can be elaborated as below mentioned steps**.**

Step 1: Data collection.
Here in this step all the data in between the given dates by the insurance admin is gathered in a two dimensional array.

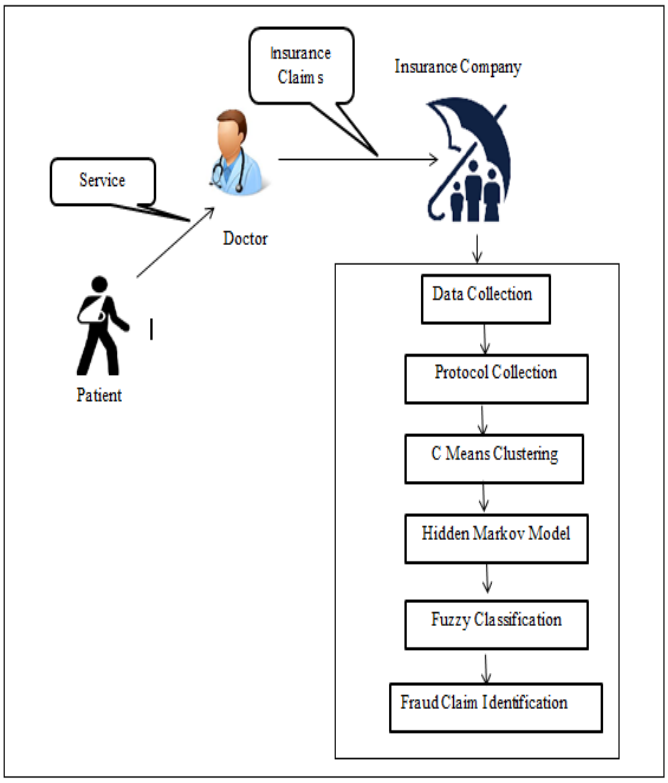

Figure 1: System Architecture

The Proposed System Steps depicted in figure 1can be elaborated as below mentioned steps**.**

Step 2: Protocol Collection
Here all the protocols that are set for the fraud detection which were stored in the database is retrieved and gathered in the form of a object of multi-dimension list.

Step 3: C-means Clustering.
Here all the data that is been collected for the calming of insurance is clustered logically using c means clustering with the following technique. This algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center. Clearly, summation of membership of each data point should be equal to one. After each iteration membership and cluster centers are updated.

Algorithmic steps for Fuzzy c-means clustering

Let $X = \{x_1, x_2, x_3 ..., x_n\}$ be the set of data points and $V = \{v_1, v_2, v_3 ..., v_c\}$ be the set of centers.

1) Randomly select *'c'* cluster centers.

2) Calculate the fuzzy membership $'\mu_{ij}'$.
3) Compute the fuzzy centers $'v_j'$
4) Repeat step 2) and 3) until the minimum 'J' value is achieved or $\|U^{(k+1)} - U^{(k)}\| < \beta$.

where, *'k'* is the iteration step.
'β' is the termination criterion between [0, 1].
'U = $(\mu_{ij})_{n*c}$' is the fuzzy membership matrix.
J' is the objective function.

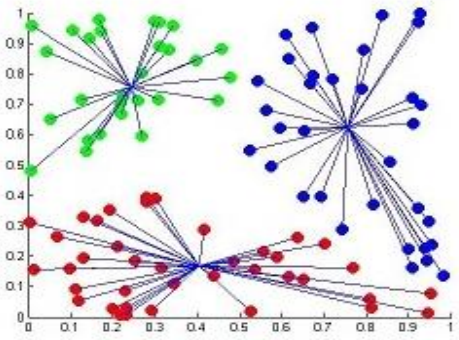

Figure 2: Result of Fuzzy c-means clustering

Step 4: Hidden Marko Model

The Hidden Markov Model (HMM) is a powerful statistical tool for modeling generative sequences that can be characterized 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 3 depicts an example of a Markov process

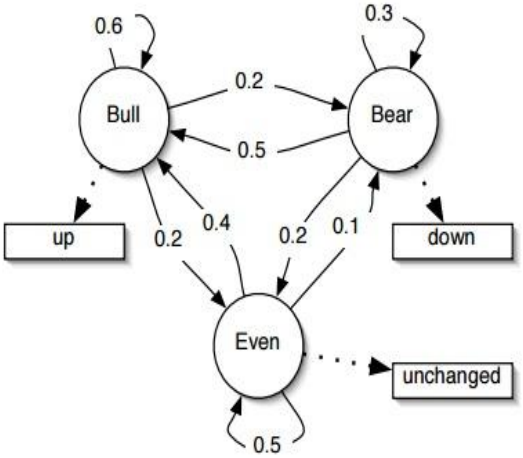

Figure 3: Markov process example

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 × 0.3 × 0.3.

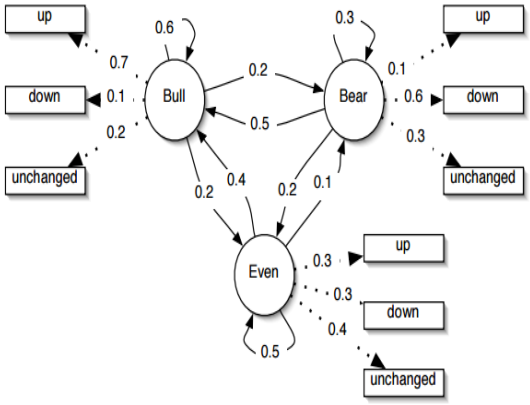

Figure 4: Hidden Markov model example

Hidden Markov Models figure 4 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 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.

Step 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 fall in between very high and high are having the tendency of doing fraud claims.

IV. CONCLUSION

Fraud detection is a topic applicable to many industries including banking and financial sectors, insurance, government agencies and law enforcement, and more. ... Through the use of sophisticated data mining tools, millions of transactions can be searched to spot patterns and detect fraudulent transactions. 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 powered with Hidden markov model to extract the fraud list and this process is catalyzed by fuzzy logic classification theory.

REFERENCES

[1] S. Chen and A. Gangopadhyay, "health care fraud detection with community detection algorithms" 2016 IEEE.
[2] S. Chen and A. Gangopadhyay, "A novel approach to uncover health carefrauds through spectral analysis," in Healthcare Informatics (ICHI), 2013IEEE International Conference on. IEEE, 2013, pp. 499–504.
[3] A. Gangopadhyay, S. Chen, and Y. Yesha, "Detecting healthcare fraud through patient sharing schemes," in Information Systems, Technology and Management. Springer, 2012, pp. 421–426.
[4] M. E. Newman, "Modularity and community structure in networks "Proceedings of the national academy of sciences, vol. 103, no. 23, pp.8577–8582, 2006.
[5] Internet Banking Fraud detection using Hidden Marcov Model IEEE-20180 Mr. Sunil S Mhamane M.E(C.S.E) Walchand Institute of Technology Solapur, India. Mr. L.M.RJ Lobo Associate Professor, Dept.Of Computer Science and Engg. Walchand of Institute of Technology Solapur, India ,july,2012.
[6] "Fraudulent Internet Banking Payments Prevention using Dynamic Key" By Osama DandashYiling Wang and Phu Dung Leand Bala Srinivasan. ""JOURNAL OF NETWORKS, VOL. 3, NO. 1,JANUARY 2008".
[7] Jing Li. Kuel-Ying Huang, Jionghua Jin, Jianjun Shi (2007) A Survey on statistical methods for health care fraud detection, Health Care Management Science.
[8] Yang WS, Hwang SY (2006) A process-mining framework for the detection of health care fraud and abuse. Expert SystAppl 31:56–68.
[9] S. Fortunato, "Community detection in graphs," Physics reports, vol. 486,no. 3, pp. 75–174, 2010.
[10] L.R. Rabiner " A Tutorial on Hidden Markov Models and Selected applications in speech recognition" Proceeding of the IEEE, Vol. 77, No 2, Feb 1989 .
[11] NHCAA (2005) The Problem of Health Care Fraud: A serious and costly reality for all Americans, report of National Health Care Antifraud Association (NHCAA).