

Telecommunication Fraud Detection A Fuzzy Rule Based Approach

A. Kanimozhi¹, S.Karthick²
^{1,2}Department of Computer Science
^{1,2}VLBJCAS, Kovaipudur

I. INTRODUCTION

1.1 DATA MINING

Data Mining is the non trivial extraction of implicit, previously unknown and potentially useful information from the data. Data mining techniques are the result of a long process of research and product development. One of the major areas where data mining can be applied is the telecommunication industry. The telecommunications industry was an early adopter of data mining technology and therefore many data mining applications exist where one of the applications is fraud detection.

1.2 FRAUD DETECTION

Fraud can be described as any activity by which service is obtained without intention of paying bills and also the use of one's occupation for personal enrichment through the deliberate misuse or application of the employing organization's resources or assets.

Fraud detection in Telecommunication:

The telecommunications industry suffers major losses due to fraud. The various types of fraud may be classified into two categories: Subscription fraud - fraudsters obtain an account without intention to pay the bill. Superimposed fraud - fraudsters "take over" a legitimate account. In such cases, the abnormal usage is superimposed upon the normal usage of the legitimate customers.



The most prominent among existing fraud detection methods is based on the analysis of the usage patterns of mobile users. The Call Detail Records (CDR) are gathered and

analyzed all the relevant usage data such as calling time, geographic position of mobile, call duration and call frequency to recognize individual patterns of normal or fraudulent behavior.

Fuzzy Logic (FL) is a multivolume logic that allows intermediate values to be defined between conventional evaluations like true/false, yes/no, high/low. Notions like rather tall or very fast can be formulated mathematically and processed by computers as rules, in order to apply a more human-like way of thinking in the programming of computers.

Fuzzy inference system:

Fuzzy inference is the process of formulating the mapping from a given input to an output using fuzzy logic. After constructing the rules and also having the idea of what the output should look like, then begin working with the GUI tools to construct a fuzzy inference system for this decision process. GUI tools are: FIS editor, Membership function editor, rule editor and rule viewer.

The rule viewer displays a roadmap of the whole fuzzy inference process. It is based on the fuzzy inference diagram. The Rule Viewer shows one calculation at a time and in great detail. In this sense, it presents a sort of micro view of the fuzzy inference system.

1.3 OBJECTIVES

The objective of this work is to formulate set of fuzzy rules using the variables which is estimated based on the sudden deviation occurring in the user behavior. Then the fuzzy rules are implemented in a rule based system called Fraud Detection Pro 1.1 in the process of detecting fraud in mobile telecommunication.

1.4. THESIS ORGANIZATION

The thesis is organized as follows: In chapter 2, the problem of fraud detection in mobile telecommunication and the literature review of previous work are given. Chapter 3 presents an overview of the data mining as well as some

application domains. Chapter 4 focuses on the fraud detection and various areas in fraud. In chapter 5, the methodology and the implementation work is given.

II. DATA MINING

Data mining is the process of mining hidden patterns or knowledge from data. Knowledge mining is a shorter term may not reflect the emphasis on mining from large amounts of data. Nevertheless, mining is a vivid term characterizing the process that finds a small set of precious nuggets from a great deal of raw material (fig.3.1 [17]). Many other terms carry a similar or slightly different meaning to data mining, such as knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology and data dredging.



Data Mining — Searching For Knowledge.

III. MAJOR ISSUES IN DATA MINING

1. Mining different kinds of knowledge in databases

As different users are interested in different kinds of knowledge, data mining should cover a wide spectrum of data analysis and knowledge discovery tasks, including data characterization, discrimination, association, correlation analysis, classification, prediction, clustering, outlier analysis and evolution analysis (which includes trend and similarity analysis). These tasks may use the same database in different ways and require the development of numerous data mining techniques.

2. Interactive mining of knowledge at multiple levels of abstraction

The data mining process must be interactive to know exactly what kinds of data's are discovered within a database. For databases containing a huge amount of data, appropriate

sampling techniques can first be applied to facilitate interactive data exploration. Interactive mining allows users to focus the search for patterns, providing and refining data mining requests based on returned results.

3. Incorporation of background knowledge

Background knowledge or information regarding the domain under study, may be used to guide the discovery process and allow discovered patterns to be expressed in concise terms and at different levels of abstraction. Domain knowledge related to databases, such as integrity constraints, deduction rules can help in focus and speed up the data mining process.

4. Presentation and visualization of data mining results

Discovered knowledge should be expressed in high-level languages, visual representations or other expressive forms so that the knowledge can be easily understood and directly usable by humans. This is especially crucial if the data mining system is to be interactive. This requires the system to adopt expressive knowledge representation techniques, such as trees, tables, rules, graphs and charts, cross tabs, matrices or curves.

5. Mining information from heterogeneous databases and global information systems

Local and wide area computer networks (such as the Internet) connect many sources of data, forming huge, distributed and heterogeneous databases. The discovery of knowledge from different sources of structured, semi-structured or unstructured data with diverse data semantics poses great challenges to data mining.

Major Issues handled in this thesis:

IV. TECHNIQUES / TOOLS OF DATA MINING

Fuzzy logic:

Fuzzy Logic has emerged as a profitable tool for the controlling and steering of systems and complex industrial processes, as well as for household and entertainment electronics, as well as for other expert systems and applications like the classification of data.

Researchers identify two fundamental goals of data mining: Prediction and Description. Prediction makes use of existing variable in the database in order to predict unknown or future values of interest and description focuses on finding patterns describing the data and the subsequent presentation

for user interpretation. The relative emphasis of both prediction and description differ with respect to the underlying application and the technique.

Verification Model:

In this model, the user formulates a hypothesis and tests the hypothesis on the data's to verify its validity.

Discovery Model:

The discovery model is a system that automatically discovering important information hidden in the data. The data is sifted in search of frequently occurring patterns, trends and generalizations about the data without intervention or guidance from the user. The manner in which the rules are discovered depends on the class of the data mining application.

The typical discovery driven tasks are:

- Discovery of association rules
- Discovery of classification rules
- Clustering
- Frequent episode discovery
- Deviation detection.

Discovery of association rules:

An association rule is an expression of the form $X \Rightarrow Y$, where X and Y are the sets of items. The intuitive meaning of such a rule is that the transaction of the database which contains X tend to contain Y . Given a data base, the goal is to discover all the rules that have the support and confidence greater than or equal to the minimum support and confidence, respectively.

Discovery of classification rules:

Clustering:

Clustering is a method of grouping data into different groups. Clustering constitutes a major class of data mining algorithms. The algorithm attempts to automatically partition the data space into a set of regions or clusters.

The objectives of clustering are

- To uncover natural grouping.
- To initiate hypothesis about the data.
- To find consistent and valid organization of the data.

Frequent Episode discovery:

Frequent episode discovery is a popular framework for mining data available as a long sequence of events. An episode is essentially a short ordered sequence of event types and the frequency of an episode is some suitable measure of the episode occurs in the data sequence.

Deviation Detection:

Deviation detection is to identify outlying points in a particular data set and explain whether they are due to noise or other impurities being present in the data or due to trivial reasons. The techniques/ tools used in this work:

- Customer Segmentation
- MATLAB
- Fuzzy inference system
- JSP
- SERLET
- Oracle

Customer Segmentation:

Customer segmentation [18] is the practice of dividing a customer base into groups of individuals that are similar in specific ways relevant to telecommunication industry, such as service plan, call service, credit rating and so on. Customer segmentation procedures include: collecting and integrating data from various sources; developing methods of data analysis for segmentation, establishing effective communication, implementing applications to effectively deal with the data and respond to the information it provides.

MATLAB:

MATLAB, which stands for mathematical laboratory, is a software package developed by Math Works, to facilitate numerical computations as well as some symbolic manipulation. Matlab optimized for working with vectors and matrices. It is a powerful tool for data analysis, plotting and much more.

Fuzzy inference system:

Fuzzy inference system is the overall name for a system that uses fuzzy reasoning to map an input space to an output space. The fuzzy inference involves: membership functions, fuzzy logic operators and if-then rules.

Java Servlet Pages:

JSP is widely used for developing dynamic web sites. JSP is used for creating database driven web applications because it provides superior server side scripting support.

- **It simplifies the process of development** - It allows programmers to insert the Java code directly into the JSP file, making the development process easier.
- **Portability** - the Java feature of 'write once, run anywhere' is applicable to JSP. JSP is platform independent, making it portable across any platform and therefore, multi-platform.
- **Efficiency** - As soon as the request is received, the JSP pages get loaded into the web servers' memory. JSP is used for generating dynamic web pages. It is essential that data in these systems be maintained efficiently and securely.

SERVLET:

SERVLET are the Java Technologies. They are programs which run on the server side and generate dynamic content. Using servlet, the Java Virtual Machine is always running, therefore starting a servlet creates a Java thread as opposed to a system process. Servlet easily share data and maintain information, making session tracking and other chores a breeze. Servlet can be run by the servlet engine in a restrictive sandbox, similar to web browser's sandbox for applets.

ORACLE:

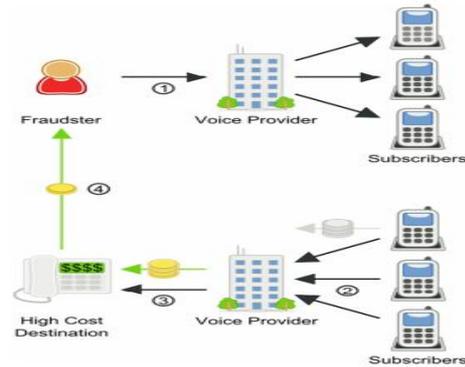
ORACLE is a fourth generation relational database management system. In general, a database management system (DBMS) must be able to reliably manage a large amount of data in a multi-user environment so that many users can concurrently access the same data. All this must be accomplished while delivering high performance to the users of the database.

V. FRAUD DETECTION

In the technological systems, fraudulent activities have occurred in many areas of daily life such as telecommunication networks, mobile communications, on-line banking and e-commerce.

Fraud is increasing radically with the expansion of modern technology and global communication, resulting in substantial losses to the businesses. So fraud detection has become an important issue to be explored.

5.1. FRAUD MANAGEMENT



Fraud prevention: is an essential part of the design of services for deployment on emerging networks.

Fraud prevention usually depends on the implementation of standard security facilities like authentication of identity and policy-based authorization and involves some kind of a priori processing.

5.2. DIFFERENT TYPES OF FRAUD

Credit card fraud:

In today's increasingly electronic society and with the rapid advances of electronic commerce on the Internet, the use of credit cards for purchases has become convenient and necessary. Credit card transactions have become the fact standard for internet and web based e-commerce.



Certainly, all transactions which deal with accounts of known misuse are not authorized.

Computer Intrusion:

Many intrusion detection systems base their operations on analysis of audit data generated by the operating system. An audit trail is a record of activities on a system that are logged to a file in chronologically sorted order. An intrusion detection system is needed to automate and perform system monitoring by keeping aggregate audit trail statistics.

Insurance fraud:

Fraud occurs when someone knowingly lies to obtain some benefit or advantage to which they are not otherwise entitled or someone knowingly denies some benefit that is due and to which someone is entitled. Depending on the specific issues involved, an alleged wrongful act may be handled as an administrative action by the Department or the Fraud Division may handle it as a criminal matter.

Telecommunication Fraud:

Fraud is a serious problem for telecommunication companies, leading to billions of dollars in lost revenue each year.

Telecommunication fraud is the burglary of telecommunication service (telephones, cell phones, computers etc.) or the use of telecommunication service to commit other forms of fraud. Categories of fraud in mobile telecommunication are:

- Subscription fraud.
- Superimposed fraud.
- Superimposed internal fraud.

1. Subscription fraud

Subscription fraud occurs when a customer opens an account (with false identification) with the intention of never paying for the account charges. After the service account period ends, the customer will try to churn (move from the service to another service).

2. Superimposed fraud

Superimposition fraud involves a legitimate account with some legitimate activity, but also includes some “superimposed” illegitimate activity by a person other than the account holder.

3. Superimposed internal fraud

This type of fraud occurs internally in the service:

The legal customer gets the bill amount for the call which has not made on that particular date. The legal customers pin is used by the staff in the service.

The legal customer’s identification is made used to another customer by the service.

VI. SYSTEM ARCHITECTURE

The proposed system designed for detecting the fraud in mobile telecommunication includes two modules: segmentation process and fuzzy inference model. The segmentation process involves the user segmentation, estimation of variables and formulating the rules. Fuzzy inference model involves designing the inference model using the GUI tool in MATLAB for decision process. Then the rules are implemented in a rule based system.

VII. CONCLUSION

The system first formulates the segments according to the user behavior (service plan) and their profiles (credit rating). Then segment the users according to the formulated segments. Based on the sudden behavioral changes of the user variables are estimated and rules are formulated. Set of fuzzy inference model is designee using the variables for decision process.

The process in formulating the fuzzy rules involves: segmentation of users according to their profiles and if any sudden indication of changes in the user behavior of a particular segment, identify the indicators which tries to change.

Finally the rules are implemented in the proposed system to detect fraud. Future direction of the system as follows: changing the variables while constructing new rules and apply the rule based system in other areas in data mining.

REFERENCES

- [1] Fawcett and F. Provost, “Adaptive fraud detection”, *Data Mining and Knowledge Discovery*, Vol. 1, No. 3, pp. 291-316, 1997.
- [2] Wetxu, Yepang, Jian Ma, Shou-Yang Wang, Gang Hao, Shoo Zeng, Yu-Hua Qian, “Fraud detection in telecommunication- a rough fuzzy set based approach”, *Proc. SIMLCN*, Vol.3, pp.1249-1253, 2008.
- [3] P Burge, J Shawe-Taylor, C Cooke, Y Moreau, B Preneel, C Stoermann, “Fraud detection and management in mobile telecommunication networks “, *Proc. ECOS’ 97*, pp. 91-96, 1997.
- [4] Oluwagbemi O.O, “Predicting Fraud in Mobile Phone Usage Using Artificial Neural Networks”, *Journal of Applied Sciences Research*, 4(6), pp. 707-715, 2008.
- [5] Yves Moreau¹, Ellen Lerouge¹, Herman Verrelst, Joos Vandewalle¹, Christof St ormann, Peter Burge “A hybrid system for fraud detection in mobile communications” *Proc. ESANN’*, pp. 447-454, 1999.

- [6] Pablo A. Estévez *, Claudio M. Held and Claudio A. Perez “ subscription fraud prevention in telecommunication using fuzzy rules and neural network”, Expert Systems with Applications, Vol. 31, No. 2 , pp. 337-344, Aug 2006.
- [7] Yufeng Kou Chang-Tien Lu Sirwongwattana.S, Yo-Ping Huang, “ Survey of fraud detection technique”. Proc. NSC IEEE int’l Conf, pp.749-754, 2004.
- [8] PJ Almeida, MB Jorge, L Cortesão, M Vieira, P. Carvalho, “fraud management system in telecommunication: a practical approach”, Proc. 12th ICT, 2005.
- [9] Delena D. Spann ,Fraud Analytics: Strategies and Methods for Detection and Prevention, October 2013.