Analysis of Credit Card Fraud Detection Methods

Analysis of Credit Card Fraud Detection Methods

V.Dheepa[1] , Dr. R.Dhanapal [2]
[1] Department of M.Sc -Information Technology,Velammal Engineering College, Chennai, Tamilnadu, India
Email: dsvdeepasaro@gmail.com
[2] Department of Information Technology and Research & Development, Vel Tech Multi Tech Dr. Rangarajan
Dr.Sakunthala Engineering College, Chennai, Tamilnadu, India.Email: drdhanapal@gmail.com
Research and Development Centre, Bharathiar University, Coimbatore, Tamilnadu, India.

Abstract—**Along with the great increase in credit card transactions, credit card fraud has become increasingly rampant in recent years. In Modern day the fraud is one of the major causes of great financial losses, not only for merchants, individual clients are also affected. Three methods to detect fraud are presented. Firstly, clustering model is used to classify the legal and fraudulent transaction using data clusterization of regions of parameter value. Secondly, Gaussian mixture model is used to model the probability density of credit card user's past behavior so that the probability of current behavior can be calculated to detect any abnormalities from the past behavior. Lastly, Bayesian networks are used to describe the statistics of a particular user and the statistics of different fraud scenarios. The main task is to explore different views of the same problem and see what can be learned from the application of each different technique.**

Index Terms—**Clustering, Credit card Fraud Detection, Bayesian Networks**

I. INTRODUCTION

The use of credit cards is prevalent in modern day society. Detecting credit card fraud is a difficult task when using normal procedures, so the development of the credit card fraud detection model has become of significance, whether in the academic or business community recently.

In this paper, three approaches to fraud detection are presented. The clustering model, the probability density estimation method and the model based on Bayesian networks. This paper investigates the usefulness of applying different approaches to a problem of Credit card fraud detection.

II. RELATED WORK ON CREDITCARD FRAUD DETECTION

From the work of view for preventing credit card fraud, more research works were carried out with special emphasis on data mining and neural networks. Sam and Karl [1] suggest a credit card fraud detection system using Bayesian and neural network techniques to learn models of fraudulent credit card transactions. Kim and Kim have identified skewed distribution of data and mix of Legitimate and fraudulent transactions as the two main reasons for the complexity of credit card fraud detection [2].This paper investigates the usefulness of applying different learning approaches.

III. THE CLUSTERING MODEL

Clustering helps in grouping the data into similar clusters that helps in uncomplicated retrieval of data [3]. Cluster analysis is a technique for breaking data down into related components in such a way that patterns and order becomes visible [4]. This model is based on the use of the parameters' data clusterization regions.

In this system 24 parameters of transactions are used for classification. (Few attributes are described in Table I).To finds regions of data clusterization a corresponding analysis should be provided at first. For example, if in the training data we observed values from Table II, then the regions of clusterization from Table III will be found.

TABLE I
INPUT ATTRIBUTES

Input attribute	Example of attribute value
Message type	1031
Type of transaction	700
Network identification	25
Day of registration in system	2
	720

In order to determine these regions of clusterization first its need to find the maximum difference ($DIFF_{max}$) between values of an attribute in the training data. This difference ($DIFF_{max}$) is split into $N_{interval}$ segments. $N_{interval}$ is the binary logarithm of the attribute values account N_{points}. In general, $N_{interval}$ can be found using another way of looking. Such calculation of $N_{interval}$ is based on the assumption that a twofold increase of N_{points} will be equal to $N_{interval}$ plus one.

TABLE II
VALUES OF ONE ATTRIBUTE

Number of value	Value
1	1200
2	1220
3	1260
4	1270
5	720
6	730
7	780
8	800

TABLE III
FOUND REGIONS OF DATA CLUSTERIZATION

Number of region	Center of region	Maximum deviation
1	757.5	42.5
2	1237.5	37.5

.

For each found segment the calculation of the average value and the corresponding deviation for hit attribute values is made. Thus $N_{interval}$ centers and corresponding deviations that describe all values of the certain attribute from the training data appears (Figure. 1).

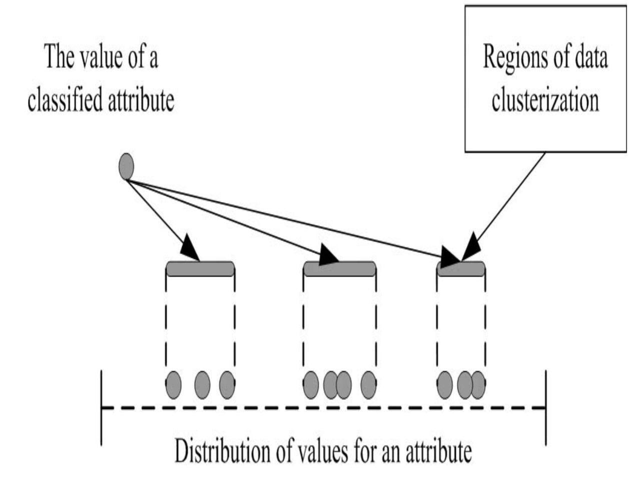
The value of a classified attribute
Regions of data clusterization
Distribution of values for an attribute

Figure 1. Regions of data clusterization

This information is collected for each attribute of a transaction during the learning process separately for legal and fraudulent transactions [5]. During classification, with regions of data clusterization for legal transactions should be compared. If the transaction is not typical for legal transactions then it is compared with regions of data clusterization for fraudulent transactions. The transaction is recognized as fraudulent if it was found as a typical fraudulent transaction.

Comparisons are made for each parameter of a transaction with found regions of data clusterization for this parameter. If a value of a transaction parameter hits into any region then this parameter is recognized as typical and as the result of classification for this parameter ($Class_i$) value 1 is returned. If not, then it is recognized as not typical and as the result of classification for this parameter ($Class_i$) value -1 is returned .The final result of classification of the whole transaction is the linear combination of classification results for each parameter:

$$Result = w_1 \text{ x } Class_1 + w_2 \text{ x } Class_2 + ... + w_n \text{ x } Class_n \qquad (1)$$

Here $Class_i$ is the result of comparing parameter i with corresponding regions of data clusterization, n is an account of parameters, w_i is a factor (weight factor) of the importance of classification result of parameter i for the whole transaction classification process [6].If Result is greater than 0 then the transaction is recognized as typical. If Result is less than 0 then the transaction is recognized as not typical. The absolute value of Result is the accuracy of the transaction classification.

IV. PROBABILITY DENSITY ESTIMATION METHODS

To model the probability density function, Gaussian mixture model is used, which is a sum of weighted component densities of Gaussian form. This is shown in Equation 2.

$$p(x) = \sum_{j=1}^{M} p(x \mid j)\, P(j) \qquad (2)$$

The $p(x \mid j)$ is the j^{th} component density of Gaussian form and the P(j) is its mixing proportion. The parameters of the Gaussian mixture model can be estimated using the EM algorithm (Computes maximum-likelihood estimates of parameters). This method specialize the general model by re-estimating the mixing proportions for each user dynamically after each sampling period as new data becomes available. Whereas the means and the variances of the user specific models are common, only the mixing proportions are different between the users' models.

In order to estimate the density of past behavior, it is necessary to retrieve the data from the last k days and adapt the mixing proportions to maximize the likelihood of past behavior. But this approach requires too much interaction with the billing system to be used in practice. To avoid this burdensome processing of data, this method formulates the partial estimation procedure using on-line estimation. The on-line version of the EM algorithm was first introduced by Nowlan [7].

$$P(j)^{new} = \alpha\, P(j)^{old} + P(j \mid x) \qquad (3)$$

Remembering that the new maximum likelihood estimate for P(j) is computed as the expected value of $P(j \mid x)$ over the whole data set with the current parameter fit, this model can easily formulate a recursive estimator for this expected value as can be seen in Equation 3. The decay term α determines the efficient length of the exponentially decaying window in the past. The approach performs statistical modeling of past behavior and produces a novelty measure of current usage as a negative log likelihood of current usage [8]. The detection decision is then based on the output of this novelty filter.

V. BAYESIAN NETWORKS

For the purpose of fraud detection, two Bayesian networks to describe the behavior of user are constructed. First, a Bayesian network is constructed to model behavior under the assumption that the user is fraudulent (F) and another model under the assumption the user is a legitimate (NF). The 'fraud net' is set up by using expert knowledge. The 'user net' is set up by using data from non fraudulent users. During operation user net is adapted to a specific user based on emerging data. By inserting evidence in these networks and propagating it through the network, the probability of the measurement x less than two above mentioned hypotheses is obtained. This means, it gives judgments to what degree observed user behavior meets typical fraudulent or non fraudulent behavior. These quantities we call $p(x \mid NF)$ and $p(x \mid F)$. By postulating the probability of fraud P (F) and P (NF) = 1-P(F) in general and by applying Bayes' rule, it gives the probability of fraud, given the measurement x,

$$P(F \mid x) = \frac{P(F)\, p(x \mid F)}{p(x)} \qquad (4)$$

Where the denominator p(x) can be calculated as

$$P(x) = P(F)p(x \mid F) + P(NF)p(x \mid NF) \qquad (5)$$

The fraud probability $P(F \mid x)$ given the observed user behavior x can be used as an alarm level. On the one hand, Bayesian networks allow the integration of expert knowledge, which we used to initially set up the models[9]. On the other hand, the user model is retrained in an unsupervised way using data. Thus our Bayesian approach incorporates both, expert knowledge and learning.

VI. CONCLUSION AND FUTURE WORK

Credit card fraud has become more and more rampant in recent years. To improve merchants' risk management level in an automatic and effective way, building an accurate and easy handling credit card risk monitoring system is one of the key tasks for the merchant banks. One aim of this study is to identify the user model that best identifies fraud cases. The models are compared in terms of their performances. To improve the fraud detection system, the combination of the three presented methods could be beneficial. It is possible to use Bayesian Networks based on the input representation method and the developed clustering model in the real fraud detection system. In the future, these models can extend to use in health insurance fraud detection.

REFERENCES

[1] Sam Maes, Karl Tuyls, Bram Vanschoenwinkel, Bernard Manderick. Credit Card Fraud Detection Using Bayesian and Neural Networks First International NAISO Congress on Neuro Fuzzy Technologies, Havana, Cuba. 2002.

[2] M.J. Kim and T.S. Kim, "A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection," Proc. International Conference on Intelligent Data Engineering and Automated Learning, Lecture Notes in Computer Science, Springer Verlag, no. 2412, pp. 378-383, 2002.

[3] Dr.R.Dhanapal, "An intelligent information retrieval agent", Elsevier International Journal on Knowledge Based Systems 2008

[4] Binu Thomas and Raju, "A Novel Fuzzy Clustering Method for Outlier Detection in Data Mining", International Journal of Recent Trends in Engineering, Vol.1, No.2, May 2009

[5] Dr.R.Dhanapal, P.Sarasu and D.Sharmili, "An intelligent system for visualization of intelligence of learners", International Journal of Recent Trends in Engineering,Vol.1,No.1,May 2009

[6] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," The Annals of Statistics, vol. 28, no. 2, pp. 337–407, 2000.

[7] S.J. Nowlan. Soft Competitive Adaptation: Neural Network Learning Algorithms based on Fitting Statistical Mixtures. PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, 1991.

[8] Jaakko Hollmén, "Novelty filter for fraud detection in mobile communications networks". Technical Report A48, Helsin ki University of Technology, Laboratory of Computer and Information Science, October 1997

[9] L. Mukhanov, "Using bayesian belief networks for credit card fraud detection," in Proc. of the IASTED International conference on Artificial Intelligence and Applications, Insbruck, Austria, Feb. 2008, pp. 221– 225.