ASIAN JOURNAL OF PHARMACEUTICAL AND CLINICAL RESEARCH

INNOVARE ACADEMIC SCIENCES
Knowledge to Innovation

Online - 2455-3891
Print - 0974-2441

Special Issue (April)

Full Proceeding Paper

# ASSOCIATIVE RULE LEARNING FOR ANOMALISTIC BEHAVIORAL MODELING IN BANKING FRAUD APPLICATIONS

**ANKUSH RAI*, JAGADEESH KANNAN R**

**School of Computing Science & Engineering, VIT University, Chennai, Tamil Nadu, India.**
**Email: ankushressci@gmail.com**

***Received: 13 December 2017, Revised and Accepted: 07 April 2017***

**ABSTRACT**

Over the years, banking sector has suffered severe loss due to several fraudulent schemes and techniques. Development of a rapid behavioral modeling method in banking sectors is need of the hour. In this study, we present the solution for such fraudulent by availing real-time anomalistic behavioral modeling in banking scenario using the associative rule learning. The presented technique is tested for its validity on the publicly available data sets for performance review.

**Keywords:** Online transactions, Associative rule learning, Fraud detection.

 DOI: http://dx.doi.org/10.22159/ajpcr.2017.v10s1.19655

## INTRODUCTION

The increasing familiarity of e-commerce with the vendors and consumers is seeing rise in popularity. For quite some time, there is a noticeable rise in the banking ethics and cyber fraudulent [1-3]. The illegal or unauthorized access to the credit card or its details is a criminal offense, and there is nonexistence of an effective way to inhibit such thieveries until the past decade [4-7]. The security and surety for the making of secured transaction has been an extensive research topic for the growth of business, financial institution, and electronic commerce [7-9]. In the few other studies, the credit card fraud is divided into certain types such as [10]:

1. Bankruptcy fraud
2. Theft fraud/counterfeit fraud
3. Application fraud
4. Behavioral fraud.

Out of this four types of fraud, the second one which implies toward the illegal accusation of the credit card and making personal transaction is what has been in rise and there were few studies made on it to define an approach to inhibit such types of electronic frauds [11-13]. As per the study conducted by Euromonitor International in 2006; let alone Germany in 2004 had faced over 345 billion pounds of credit card fraud [14]. In this study, we discussed such techniques which have been proven successful against the cyber frauds over European markets. This techniques aid in detecting such frauds by learning the pattern of the previous transaction of the card holder and each time verifying it with the previously trained pattern to allow the next transactions to take place.

Several techniques had been proposed to overcome the issue. Among those attempted, Ghoosh and Reily used neural networks for labeled credit card account transactions for its training [15]. Other significant works include that of Syeda *et al.*, who used parallel granular networks to facilitate the speedup of execution in knowledge discovery with comparatively larger attributes from the databases [16]. Stolfo *et al.* introduced a metal earning technique for learning-based fraud detection system [17]. In the latter studies, they used Java agents for the data mining applications [18]. In the study conducted by Aleskerov *et al.*, they introduced CARDWATCH which used neural networks with several interfaces to the credit card database [19]. The following study proposed an innovative fraud detection database with real-time applications using the cascaded neural network (CNN) in a modified morphology of water filling algorithm to adjust with the several attributes made in credit card transaction.

## METHODOLOGY: FRAUD DETECTION USING ASSOCIATIVE LEARNING

### Experimental set-up

The proposed model is prototyped over MATLAB R2012a under Windows platform. The experiments are conducted over the machine with hardware configurations of Intel's seventh generation 8-core microprocessor, 8GB RAM giving a fine clocking speed of 2.7 GHz. As the emphasis of our work is over the recognition of fraudulent transaction states and its correlation for over a larger group of attributes. The consolidated databases available online are used as test data sets for the algorithm [20,21]. The two databases used are, namely German Credit Fraud Data set and Australian Credit Approval Data Set comprising 567 and 690 instances, respectively (Table 1).

### The model

However, we are still far from being remodel the sampling time for a database transaction in real time without any sever delay which gives anomalistic situation where dependency of the several parameters cannot be modeled for the expert system to form an adaptive network to perform intelligent operations. Thus, the two essential properties of an expert system or artificial intelligence lie within two components, i.e., feedback and swarm behavior for ranging associative parametric evaluation at every instant in the due process [22-24]. This is where our proposed algorithm comes into play which remodel the CNN for increasing its effectiveness against multivariable dependencies of time-dependent sampling transactions with several attributes to data mine in a logical sequencing the architecture of the proposed data mining algorithm is presented in the Fig. 1. Furthermore, the modeled algorithm using a hybrid of CNN is formalized below:

Algorithm: Unsupewrvised Cascaded Profile Classifier (UCPC)

Input: List of t transactions made & $CC(i,j)$, i.e., template classes of the trained vectors for the t-1 transactions.

Output: $CC'(i,j)$, final state of the transactions (1 or 0 for validation).

*for* $b_{jq}$*//for each transaction instances*

*//Run the conventional water filling algorithm for set CC(i,j), get water level W*

$W=P(CC(i,j)=P_{t+1}|P_{t+n}=S_j),1\leq i<N, N\leq j\leq 1$

*Where P is the probability of states and S the number of states* $S=\{S_1,S_2,...S_n\}$ *& N is the number of water channels.*

*for t = 1 to X:*

$$H_n=tanh(w_{HX}X_N+w_{HH}H_{N-1}+B_N)$$

$$k_i = w_{HH}.X_N + B_N + \sum_{i+1}^{N}\tanh(\delta_H.(1-y_{N-1})$$

Move subchannels to temporal set

$$CC''(i,j)=\{CC\left|\sum_{1}^{N}H_N^{-1}\leq w_{HH}-\frac{\Delta W}{N}\right\}$$ *//for each hidden layer*

*while j<X*

$$k_j = w_{HO}.X_N + \sum_{i+1}^{N}\tanh(\delta_O\cdot y_{N-1})$$

$O_N=w_{OH}H_N+B_O$

*Update CC'(i,j)(attribute classes) with the channel N &* $\Delta w_{XX}$ *as weighted subchannels to buffer the state sequences:*

$$\Delta w_{XX}=\eta.\delta_O.k_i.O_N+\frac{\Delta W}{N}$$

$$CC''(i,j)=\sum_{i=1}^{m}\sum_{j=1}^{n}sgn\left(\Delta w_{XX}\cdot O(k_i)+\Delta W.\ Y\left(t',k_j\right)\right)+\sum_{i,j}^{m,n}P(i,j)$$

//attribute classes

*Where, m, n belongs to runtime transactions made in real time.*

*end loop*

*end while loop*

$$if\ P(i,j)=\frac{CC''(i,j)-CC(i,j)}{|W|}<N$$

$$CC'(i,j)=\prod_{i=1}^{N}\Delta w_{XX}t_{f_t}(R_i,L_i)$$

*else*

*Print "Transaction Invalid"*

*end//if*

*end loop*

Given a sequence of input vectors ($X_1$, $X_2$, $X_3$,..., $X_n$) from *CC(i,j)*, a sequence of hidden states ($H_1$, $H_2$, $H_3$,..., $H_n$) and a sequence of outputs ($O_1$, $O_2$, $O_3$,..., $O_n$) are generated in due process. Notions in these equations are, namely, $w_{HX}$ is the input-to-hidden weight matrix, $w_{HH}$ is the hidden-to-hidden (or recurrent) weight matrix, $w_{OH}$ is the hidden-to-output weight matrix, and the vectors $B_N$ and $B_O$ are the biases. The expression replaces the inputs received from feedback loops with a special initial bias vector checked for nonlinearity while ensuring that the training is done coordinate-wise. η is the learning rate, *t'* is the time of the next frame, $k_i$ is the local induced field of activation potential for the $i^{th}$ neuron, $k_j$ is the coactivation neuron field for the next sequence of activation units, and $\delta_H$ and $\delta_O$ are the pointer variable for the field and subfield trace of an emotion, respectively. Where the integrated co-emotional involvement with the principal and subsequent emotional states derived from RCCC.

**Table 1: Enlisted database used in the experiment**

| Database | Number of instances | Area | Number of attributes |
|---|---|---|---|
| German credit fraud data set | 567 | Financial | 16 |
| Australian credit approval data set | 690 | Financial | 14 |

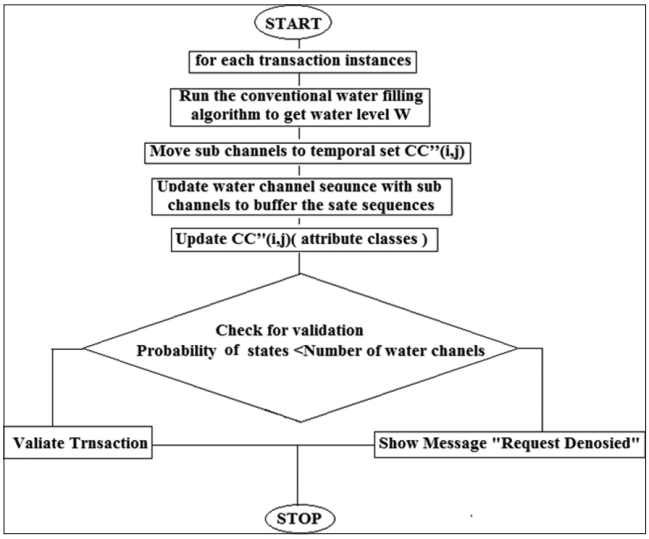


**Fig. 1: Flowchart of the proposed algorithm describing the workflow process**

As shown in the snippets of transaction (from German credit card database) at Table 2 of the sample attributes of the transaction, six classes have been defined as the principal attributes of the parameter which are encoded through the cascaded logical network into a sequence. Here, place of transaction made for represents the locality of the company/mall/hotel, etc., for whom the transactions are made to pay; this in turn may be the subset of the region of transaction. The plot representation of this data is shown in Fig. 2a which shows the total cascaded length of the sequence in bits required to be made for the German credit card database. This shows the length of rise in cascaded sequence is ideally increases in linear fashion. Fig. 2b shows the plot for testing time versus sequence generated for the CNN. Here, the time required to train the network is comparatively low than the previous methods cited in the literature, thereby giving high computational processing for large credit card transactions database with a promise for real-time application. Fig. 2c shows the transaction profiles of the four transactions ids portrayed in Table 2. The nodes in the transaction profiles show problematic transactions having high chances of being malicious. Furthermore, the curve shown in the plot of transactions Profile 3 and 4 are instantly made at the same time with the same machines and for the same reasons, thereby giving an overlap but since the distribution of the transactions are concorded thereby did not show much variance in water filling channels and hence are considered to legitimate whereas upon comparison with that of transaction Profile 2 and 1, the trough observed in the curve shown in Fig. 2c represents high chances of transactions being malicious but with the similar pattern and thus corresponding to the activities perused by the same criminal in the same regional locality.

**Table 2: Classification of transaction based on attributes divided into classes for water filling estimation in combination with the sequence validated by cascaded neural network**

| Transaction Id | Class 1 (transaction amount) | Class 3 (place of transaction made for) | Class 3 (purpose of transaction) | Class 4 (region of transaction, Class 3 ⊂ Class 5) | Class 5 (region for transaction) | Class 6 (cascaded sequence) |
|---|---|---|---|---|---|---|
| Transaction Profile 1 | 2000000 | V | P | V | X | 11010101 |
| Transaction Profile 2 | 9828282 | X | Q | Y | X | 10101001 |
| Transaction Profile 3 | 18265268 | Y | R | X | V | 10101010 |
| Transaction Profile 4 | 67188281 | Z | S | X | Y | 11001010 |

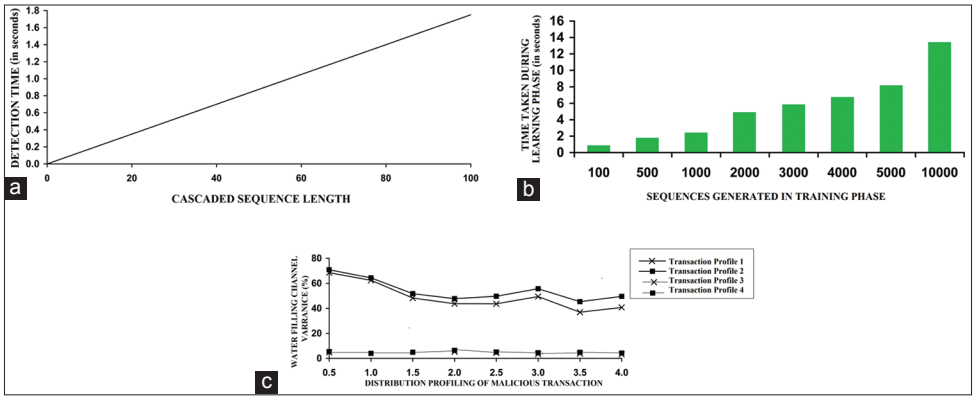


**Fig. 2: (a) Fraud detection time with the help of proposed algorithm with the rise in complexity of the cascaded sequence. (b) Plot for testing time versus sequence generated for the cascaded neural network. (c) Plot of variance in water filling channel for the consequent transaction made by same person and the distribution of the malicious transaction represented by nodes**

## CONCLUSION

In this study, we have presented a novel data mining algorithm for detection of credit card fraud. The learning model proposes an associative cascaded learning network with several attributes taken in association through a water filling algorithm. The different process in the processing of huge volumes of credit transaction has been listed. We have tested our detection framework with the German credit card database and also with the Australian credit card database. The performance analysis of the study gives an effective way of using such hefty process in relatively lower computational time bounds. The proposed UCPC algorithm greatly automates the data mining process with no human intervention at both the training and testing phases. There is still a huge room for improvement with this framework to increase the accuracy to the best it can possibly offer.

## REFERENCES

1. Aleskerov E, Freisleben B, Rao B. CARDWATCH: A neural network-based database mining system for credit card fraud detection. Proceeding of the IEEE/IAFE on Computational Intelligence for Financial Engineering; 1997. p. 220-226.
2. Anderson R. The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation. New York: Oxford University Press; 2007.
3. APACS, Association for Payment Cleaning Services, No date. Card Fraud Facts and Figures. Available from: http://www.apacs.org.uk/resources_publications/card_fraud_facts_and_figures.html. [Last accessed on 2007 Dec].
4. Bellis M. No date. Who Invented Credit Cards-the History of Credit Cards? Available from: http://inventors.about.com/od/cstartinventions/a/credit_cards.htm. [Last accessed on 2008 Oct].
5. Bentley P, Kim J, Jung G, Choi J. Fuzzy Darwinian Detection of Credit Card Fraud, Proceeding of 14th Annual Fall Symposium of the Korean Information Processing Society; 2000.
6. Bolton R, Hand D. Statistical fraud detection: A review. Stat Sci 2002;17:235-49.
7. Bolton R, Hand D. Unsupervised profiling methods for fraud detection. Credit Scoring and Credit Control VII; 2001.
8. Brause R, Langsdorf T, Hepp M. Credit card fraud detection by adaptive neural data mining. Internal Report 7/99. Frankfurt, Germany: J. W. Goethe-University, Computer Science Department; 1999a.
9. Pandey LB, Choubey S. Comparative study of different feature selection algorithm in small dataset among KNN, FUZZY and GENETIC algorithm. Int J Adv Res Comput Eng Technol IJARCET 2012;1(7):41-3.
10. Caminer B. Credit card fraud: The neglected crime. J Crim Law Criminol 1985;76;746-63.
11. Chan P, Fan W, Prodromidis A, Stolfo S. Distributed data mining in credit card fraud detection. IEEE Intell Syst 1999;14:67-74.
12. Chan P, Stolfo S, Fan D, Lee W, Prodromidis A. Credit card fraud detection using meta learning: Issues and initial results. Working Notes of AAAI Workshop on AI Approaches to Fraud Detection and Risk Management; 1997.
13. Chepaitis E. Information ethics across information cultures. Bus Ethics Eur Rev 1997;6(4):195-9.
14. Euro monitor International. Financial cards in Germany; 2006. Available from: http://www.euromonitor.com/Financial_Cards_in_Germany. [Last accessed on 2006 Nov].
15. Ghosh S, Reilly DL. Credit card fraud detection with a neural-network. Proceeding 27th Hawaii International Conference System Sciences: Information Systems: Decision Support and Knowledge-Based Systems. Vol. 3. 1994. p. 621-30.
16. Syeda M, Zhang YQ, Pan Y. Parallel granular networks for fast credit card fraud detection. Proceeding IEEE International Conference Fuzzy Systems; 2002. p. 572-7.
17. Stolfo SJ, Fan DW, Lee W, Prodromidis AL, Chan PK. Credit card fraud detection using meta-learning: Issues and initial results. Proceeding AAAI Workshop AI Methods in Fraud and Risk Management; 1997. p. 83-90.
18. Stolfo SJ, Fan DW, Lee W, Prodromidis AL, Chan PK. Cost-based modeling for fraud and intrusion detection: Results from the JAM project. Proceeding DARPA Information Survivability Conference and Exposition. Vol. 2. 2000. p. 130-44.
19. Aleskerov E, Freisleben B, Rao B. CARDWATCH: A neural network based database mining system for credit card fraud detection. Proceeding IEEE/IAFE: Computational Intelligence for Financial Engineering; 1997. p. 220-6.
20. German Credit Fraud Dataset. Available from: http://www.weka.8497.

n7.nabble.com/file/n23121/credit_fruad.arff.
21. Statlog (Australian Credit Approval) Data Set. Available from: http://www.archive.ics.uci.edu/ml/datasets/Statlog+%28Australian+Credit+Approval%29.
22. Rai A. Secure two party computation. J Adv Shell Program 2014;1(2):5-6.
23. Rai A. Computational modeling study of synfire chains from multiple plasticity mechanisms for model development at neural level: Introducing an evolving digital micro-brain. Res Rev J Comput Biol 2014;3(2):9-15.
24. Rai A. Automation in computation over linux integrated environment. J Adv Shell Program 2014;1(1):18-20.