Outlier analysis for plastic card fraud detection a hybridized and multi-objective approach

Arturo Elías_Ramírez, , Alejandro Padilla_Díaz
[1] Universidad Autónoma de Aguascalientes
{aeliasr,apadilla}@correo.uaa.mx

(Paper received on September 09, 2010, accepted on October 20, 2010)

Abstract. Nowadays, plastic card fraud detection is of great importance to financial institutions. This article presents a proposal for an automated credit card fraud detection system based on the outlier analysis technology. Previous research has established that the use of outlier analysis is one of the best techniques for the detection of fraud in general. However, to establish patterns to identify anomalies, these patterns are learned by the fraudsters and then they change the way to make de fraud. The approach applies a multi-objective model hybridized with particle swarm optimization of typical cardholder's behavior and to analyze the deviation of transactions, thus finding suspicious transactions in a non supervised scheme.

Keywords: Credit Card Fraud, Outlier Detection, MOO, PSO, Unsupervised scheme.

1 Introduction

Fraud is an activity almost as old as mankind, which tries to take advantage of some kind, usually economic, by the fraudster with respect to shame. Specifically in the case of plastic card fraud there are several variants [1]. The total cost of plastic card fraud is high relative to other forms of payment. The first line of defense against fraud is based on preventive measures such as the Chip and PIN cards. Subsequent methods are used to identify potential fraud trying to minimize potential losses. These methods are called fraud detection systems (FDS) and usually employ a variety of proposals with the idea of detecting the majority of potential fraudulent behavior.

There are two major frameworks to detect fraud through statistical methods. If fraud is conducted in a known way, the pattern recognition techniques are typically used, especially supervised classification schemes[2]. On the other hand if the way in which fraud is done is not known, for example, when there are new fraudulent behaviors, outlier analysis methods are recommended[3]. Some studies show simple techniques for anomaly detection analysis to discover plastic card fraud.[4].

The idea of the proposal is to use time series in which transactions that have a similar behavior are grouped, so that subsequent transactions that deviate strongly from the clusters formed are candidates to be considered to have an anomalous behavior. The problem with this approach is not always abnormal behaviors are fraudulent, so a successful system must locate transactions that are detected as fraud, but they really are fraud and not only appear to be fraudulent, where time is a factor against it, because to reduce losses, fraud detection should be done as quickly as possible. With this in mind,

©G. Arroyo-Figueroa (Ed)
Special Issue in Advances in Artificial Intelligence and Applications
Research in Computing Science 51, 2010, pp. 99-108

the proposal is strengthened by a multi-objective approach in order to improve the performance of FDS.

The remainder of this paper is organized as follows: Section 2 describes the theoretical framework of the methodologies employed in the paper. In Section 3 we provide some basic concepts to establish the formulate hypothesis for this approach. The proposed method is formulated and discussed in Section 4; In Section 5 implications of the study are presented; Section 6 draws the conclusions with a few lines of open research.

2 Theoretical Framework

2.1 Clusters and Outliers

The clustering is primarily a technique of unsupervised approach, although the semi-supervised clustering has also been studied frequently in recent days[5]. Although often clustering and anomaly detection appear to be fundamentally different from one another, have developed many techniques to detect anomalies based on clustering, which can be grouped into three categories which depend on three different assumptions regarding[6]: a) Normal data instances belong to a pooled data set, while the anomalies do not belong to any group clustered. b) Normal instances of data are close to the cluster centroids, while anomalies are further away from these centroids. c) The normal data belongs to large, dense clusters, whereas the anomalies belong to small and sparse clusters.

Each of the above assumptions has their own forms of detect outliers which have advantages and disadvantages between them.

2.2Particle Swarm Optimization (PSO)

Swarm intelligence (SI) simulates the social and collective behavior of living creatures such as ants, bees and termites[7], but also develops global models of local interaction behavior of artificial agents. Diverse swarm intelligence algorithms have been designed and implemented; some algorithms like ant colony optimization (ACO) and particle swarm optimization (PSO) have been studied and applied in many studies and investigations[8].

In an attempt to find an efficient technique for search and optimization, the PSO algorithm was proposed [9] and later was explained in detail for them[10]. From the perspective of PSO, a swarm can be defined as '*a population of elements that interact and are able to optimize some objective function through collaborative search of a space*'[11]. PSO does not require information from a gradient of the objective function and method features are developed with simple mathematical operators. Ease of use have caused many designed variants of PSO which was applied in several studies.

In the PSO algorithm, particles fly in the course of a search space and each particle has a corresponding position and velocity at any instant of time. The position of a particle to the origin corresponds to one of the solutions to the problem of search.

The movement of particles in the search space to new positions means the generation of a set of solutions. The convenience of these solutions is quantitatively measured by a fitness function[12].

When starting the PSO algorithm, the positions of the particles and their velocities are randomly initialized. The position and current speed of the particles is represented by $X_i(t)$ and $V_i(t)$, respectively. As the algorithm evolves, the positions of the particles are influenced by positional information, $P_i(t)$. If the position information is evaluated using the complete neighborhood cluster of particles then the knowledge of the position is called the global best or simply gbest, P_g. Furthermore, as small neighborhoods used in the evaluation of the position of particles then the position information is the local best or lbest P_l .[13]

In one dimensional space problem *D*, in each iteration of PSO algorithm, particles using the positional information, P_l and P_g, to adjust their speeds and their subsequent positions as indicated by the following equations

$$V_{id}(t+1) = V_{id}(t) + c1.(P_{id} - X_{id}(t)) + c2.(P_{id} - X_{id}(t)) \qquad (1)$$

$$X_{\downarrow}id\ (t+1) = X_{\downarrow}id\ (t) + V_{\downarrow}id\ (t+1)) \qquad (2)$$

The update process of the particles is compensated by the random numbers, c_1 and c_2 whose values typically have an upper limit of 2.0[14].

2.3 Multi-Objective Optimization

In real life most of the problems facing not only an objective optimization, actually there are several objectives to be achieved to have the arguments needed to make a decision. These problems must be addressed as a multi-objective optimization (MOO), bearing in mind that generally, the improvement in the achievement of a goal causes deterioration of other or others objectives. So with regard to the problems of clustering approach there are also multi-objective clustering (MOC) proposals. These methods consist in decomposing a dataset into similar groups to optimize multiple objectives in parallel.[15]. Some researchers have suggested that multi-objective search and optimization could be a problem area where evolutionary algorithms (Multi-objective Evolutionary Algorithms (MOEA)) can achieve better performance compared to other search strategies[16], and later other bio-inspired mechanisms was proposed like social behavior [17]. The latest trends in bio-inspired strategies have been collective intelligence or swarm intelligence. For many years scientists have studied ants, wasps and bees because they reach high efficiency from their collective efforts, although a technique that has become popular is the particle swarm optimization. Thus, swarm intelligence has become a valuable tool to optimize operations in different businesses.

3 Research model and Hypotheses

This approach is conducted under premise of improving the efficiency for detect fraudulent activity on plastic card transactions. In order to do this; the system is developed with a foundation of multi-objective clustering, which places the problem of detecting fraud in an appropriate context to reality. In the same way, the system is strengthened through hybridization using PSO for the creation of clusters, then find the anomalies using Mahalanobis distance.

The research model consists of two main dimensions: an unsupervised approach and a hybridized clustering model to identify the standard behavior in card transactions. Fig. 1 describes these claims.

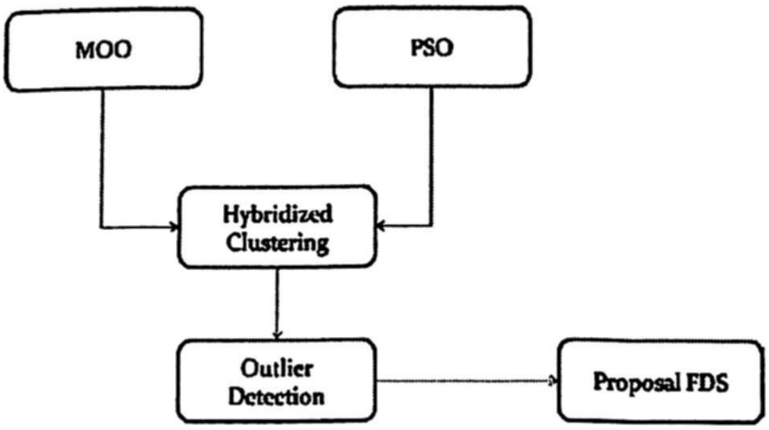

Fig. 1 Research Model

3.1Precision

Overall accuracy is simply the percentage of correct predictions of a classifier on a test set of "ground truth". TP means the rate of predicting "true positives" (the ratio of correctly predicted frauds over all of the true frauds), FP means the rate of predicting "false positives" (the ratio of incorrectly predicted frauds over those test examples that were not frauds, otherwise known as the "false alarm rate".)[18].

Other two types of rates are considered for the results delivered by FDS, FN means the rate of predicting "false negatives" (the ratio of no predicted frauds over all the true frauds) and TN means the rate of predicting "true negatives" (the ratio of normal transactions detected). Table I shows the classification rate of results obtained by the FDS after analyzing a transaction.

The accuracy of the FDS is represented as the fraction of total transactions (genuine and fraudulent) that are detected as correct, which can be expressed as follows [19]

$$\boldsymbol{Presicion} = \frac{\#\ of\ TN + \#\ of\ TP}{Total\ of\ carry\ out\ transaction} \tag{3}$$

Accuracy is the key part of the proper functioning of a FDS. We thus formally hypothesized:

Hypothesis 1 **(H1).** The level of accuracy determines the success of a FDS.

Table I. Classification rate of results

Outcome	Classification
Miss	False Negative (FN)
False Alarm	False Positive (FP)
Hit	True Positive (TP)
Normal	True Negative (TN)

3.2 Hybridization

As in many aspects of artificial intelligence to detect abnormalities very current trend is the hybridization. The reason is because many developed algorithms do not follow entirely the concepts of a simple classical metaheuristic[20], to solve this problem is looking for the best from a combination of metaheuristics (and any other kind of optimization methods) that perform together to complement each other and produce a profitable synergy, to which is called hybridization[21],[22].

Some possible reasons for the hybridization are: 1.- improve the performance of evolutionary algorithms, 2.- improve the quality of solutions obtained by evolutionary algorithms and 3.- incorporate evolutionary algorithms as part of a larger system.[23].

The different instances of hybridization of metaheuristics with *bio-inspired and evolutionary algorithms* (BEAs) can be grouped into different categories. The first two groups are derived from well developed taxonomy for hybridization of metaheuristics[21], which developed on the basis of their control strategies: collaborative hybrid metaheuristics and integrative hybrid metaheuristics. A third method to construct hybridization of metaheuristics with evolutionary algorithms is through the incorporation of the intensification and diversification (I&D) which are the two biggest issues when designing a global search method [20] where a component I&D is defined as any algorithmic or functional component having an effect of identification or diversification in the search process. Hybridization is a technique that aims to improve the performance of metaheuristics and BEAs. We thus formally hypothesized:

Hypothesis 2 **(H2).** The hybridization affect positive the process of cluster construction.

3.3 Multi-objective Pareto Front Clustering

Bio-inspired and evolutionary algorithms have been the most frequently used for clustering. However previous research in this respect has been limited to the single objective case: criteria based on cluster compactness have been the objectives most commonly employed, as the measures provide smooth incremental guidance in all parts of search space.

In recent years there has been a growing interest in developing and applying BEAs in multi-objective optimization[24].

The recent studies on BEAs have shown that the population-based algorithms are potential candidate to solve multi-objective optimization problems and can be efficiently used to eliminate most of the difficulties of classical single objective methods such as the sensitivity to the shape of the Pareto-optimal front and the necessity of multiple runs to find multiple Pareto-optimal solutions.

In general, the goal of a multi-objective optimization algorithm is not only to guide the search towards the Pareto-optimal front but also to maintain population diversity in the set of the Pareto optimal solutions. In this way the following three main goals need to be achieved: (i) maximize the number of elements of the Pareto optimal set found; (ii) minimize the distance of the Pareto front produced by the algorithm with respect to the true (global) Pareto front (assuming we know its location); (iii) maximize the spreads of solutions found, so that we can have a distributions of vectors as smooth and uniform as possible[25].

Currently PSO has been presented as an efficient population-based heuristic technique with a flexible and well balanced mechanism to enhance and adapt the global and local exploration capabilities. So, the relative simplicity of PSO and its population based approach have made it a natural candidate to be extended for MOO [17].

Hypothesis 3 **(H3).** The multi-objective Pareto front solution using PSO affect positive the process of cluster construction.

4 Research Methodology

The FDS is running on the plastic card issuing institution. When a transaction arrived is sent to the FDS to be verified. The FDS receives the card details and purchase value to verify if the transaction is genuine, by calculating the anomalies, based on the expenditure profile of each cardholder, purchasing and billing locations, time of purchase, etc. When FDS confirms that the transaction is malicious, it activates an alarm and the financial institution decline the transaction. The cardholder concerned is contacted and alerted about the possibility that your card is at risk.

To find information dynamically observation for individual transactions of the cardholder, stored transactions are subject to a clustering algorithm. In general, transactions are stored in a database of the financial institution, which contain too many attributes. In this paper we analyze three factors, the amount spent, time and location where the transaction takes place. So, if the purchase amount exceeds a certain value, the time between the uses of the card is low or the locations where different transactions are distant are facts to consider activating the alarm.

All this required the calculation of anomalies through the clustering of transaction information through a multi-objective Pareto front with the support of PSO.

Based on the population nature of PSO, it is desirable to produce several (different) non-dominated solutions with a single run. So, as with any other evolutionary algorithm, the three main issues to be considered when using PSO to multi- objective optimization are: (i) how to select g_{best} particles in order to give preference to non-dominated solutions over those that are dominated? (ii) how to retain the non- dominated solutions found during the search process in order to report solutions that are non-dominated with respect to all the past populations and not only with respect to the current one? Also it is desirable that these solutions are well spread along the Pareto front; (iii) how to maintain diversity in the swarm in order to avoid convergence to a single solution?

When solving single-objective optimization problems, the g_{bcst} that each particle uses to update its position is completely determined once a neighborhood topology is established. However in the case of multi-objective optimizations problems, each particle might have a set of different g_{bcst}s from which just one can be selected in order to update its position. Such set of g_{bcst}s is usually stored in a different place from the swarm that we will call external archive denoted as EX_ARCHIVE. This is a repository in which the non-dominated solutions found so far are stored. The solutions contained in the external archive are used as global bests when the positions of the particles of the swarm have to be updated. Furthermore, the contents of the external archive are also usually reported as the final output of the algorithm.

Once clusters are established, new transaction is entered and evaluated in the FDS, to see if it belongs to a cluster set or is outside of him, seeing the transaction as an anomaly and becoming a candidate to be fraudulent.

The idea of the proposal is to work at the level of cardholder's account, keeping in main that the transaction flow of transaction logs is complex (more than 60 fields), including a unique account number. For the i_{th} account transaction has the following sequence:

$$X_i = \{x_t | x_t \in \mathbf{R}^N, t = 1, 2, \dots\} \quad (4)$$

Where, X_i represents the i_{th} account, while x_t is the sequence of transactions for that account at time t.

Outliers detection conform the aim of this approach. Computing the outliers is a process based on Mahalanobis distance.

Transactions outside of clusters are candidates to be considered fraudulent, however as mentioned above the accuracy of the system is a factor to be considered, which is expected to maximize in order to increase the functionality of the FDS. Fig. 2 shows the idea of the full flow of the process proposed for the FDS.

5 Implications

Evaluate FDSs for plastic cards' using real data is too complex. Banks are generally not agree to share their data with researchers, as well as the absence of a data set for comparison (benchmark) available for experiments[1, 26]. Therefore large-scale simulation is developed to prove the effectiveness of the system. Simulator is used to generate a mixture of genuine and fraudulent transactions. The number of fraudulent transactions in a defined amount of mixed transactions is normally distributed with mean and standard deviation specified by the user, taking the cardholder's spending behavior in his account. The mean specifies the average number of fraudulent transactions in a given transaction mix. In a typical scenario, the FDS of the card issuing institution receives a large number of genuine transactions mixed moderately with fraudulent transactions, where legitimate transactions are generated from profiles of cardholders.

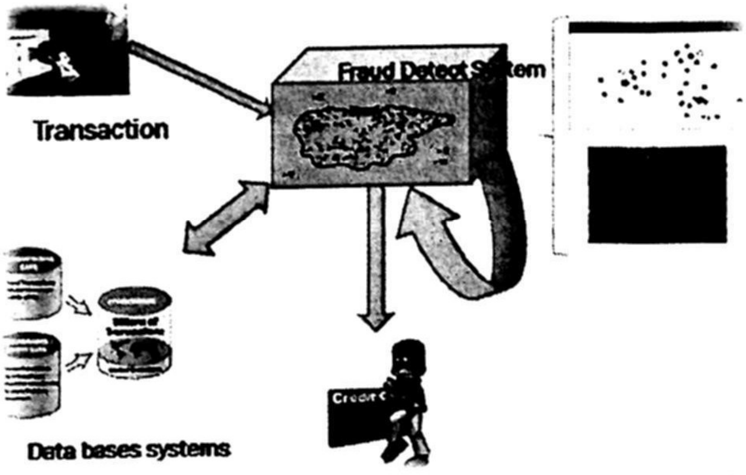

Fig. 2 Full flow of the process

6 Conclusions

It is noteworthy that achieve exact replication of the problem is not possible; although it is assumed that many of the more general features of the data used could be reproduced in subsequent data. Fraud, however, is a known phenomenon changing in response to market conditions, as well as to measures taken by financial institutions against it, so it is quite possible that not only this but any proposal made in this regard not be accurate and universally applicable, although a fairly good approximation, which is reliable in many cases.

The methodologies for the detection of fraud have their own strengths and weaknesses characteristics. The overall strength of FDS using anomaly detection is the adaptability to new patterns fraudsters, in the particular case of this study is strengthened with the application of hybridization clustering processes giving a greater dynamism to the system and making it look like a promising component within the fraud detection systems with potential advantages in regard to: upgrade and management of the heterogeneity of customers and their transactions, achieving a better accuracy in the results, and greater dynamism in the system.

Additionally, the multi-objective approach place it in a better position compared to other systems, due to the characteristics of fraud detection problem where there are several factors to consider for best results.

Future work establishing the need for FDS to be increasingly proactive in order to adapt to the greatest extent possible so changing the behavior presented by fraudsters.

References

1. Sánchez, D., et al., *Association rules applied to credit card fraud detection.* Expert Systems with Applications: An International Journal, 2009: p. 3630-3640.
2. Whitrow, C., et al., *Transaction aggregation as a strategy for credit card fraud detection.* Data Mining and Knowledge Discovery, 2009: p. 30-55.

3. Kou, Y., et al. *Survey of fraud detection techniques*. in *Proccedings IEEE International Conference on Networking, Sensing and Control*. 2004. Taipei: IEEE press.
4. Juszczak, P., et al., *Off-the-peg and bespoke classifiers for fraud detection.* Computational Statistics & Data Analysis, 2008: p. 521-532.
5. Basu, S., M. Bilenko, and R.J. Mooney. *A probabilistic framework for semi-supervised clustering*. in *Proceedings of the Tenth ACM SIGKDD international Conference on Knowledge Discovery and Data Mining*. 2004. Seattle, WA: ACM, press.
6. Chandola, V., A. Banerjee, and V. Kumar, *Anomaly detection: A survey*. ACM Computing Surveys, 2009: p. 1-58.
7. Beni, G. and U. Wang. *Swarm intelligence in cellular robotic systems*. in *NATO advanced workshop on robots and biological systems*. 1989. Tuscany, Italy: No editorial.
8. Özçift, A., et al., *Swarm optimized organizing map (SWOM): A swarm intelligence based optimization of self-organizing map*. Expert Systems with Applications: An International Journal, 2009: p. 640-648.
9. Kennedy, J. and R. Eberhart. *Particle Swarm Optimization*. in *Proceedings of IEEE International Conference on Neural Networks. I*. 1995. Piscataway, NJ: IEEE press.
10. Eberhart, R., Y. Shi, and J. Kennedy, *Swarm Intelligence*. 2001, San Francisco, CA: Morgan Kaufmann.
11. Eberhart, R.C. and Y. Shi. *Comparison between genetic algorithms and particle swarm optimization*. in *Proceedings of the 7th International Conference on Evolutionary Programming VII*. 1998. San Diego, CA: Berlin: Springer-Verlag.
12. Cui, X., T.E. Potok, and P. Palathingal. *Document Clustering using Particle Swarm Optimization*. in *Proceedings Swarm Intelligence Symposium, 2005*. 2005. New York: IEEE, press.
13. Omran, M.G., A.P. Engelbrecht, and A. Salman, *A color image quantization algorithm based on particle swarm optimization*. Informatica 2005: p. 261-269.
14. Kennedy, J. *The particle swarm: social adaptation of knowledge*. in *Proceedings of IEEE International Conference on Evolutionary Computation*. 1997. Indianapolis, IN: IEEE, press.
15. Jiamthapthaksin, R., C.F. Eick, and R. Vilalta, *A Framework for Multi-Objective Clustering and Its Application to Co-Location Mining*, in *Advanced Data Mining and Applications*. 2009, Springer Berlin / Heidelberg: Beijing, China. p. 188-199.
16. Zitzler, E. and L. Thiele, *Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach*. IEEE Transactions on Evolutionary Computation, 1999: p. 257-271.
17. Coello, C.A.C. and M.S. Lechunga. *MOPSO: A proposal for multiple objective particle swarm optimization*. in *Proceedings of the 2002 Congress on Evolutionary Computation*. 2002. Hawaii: IEEE Press.
18. Stolfo, S., et al. *Credit Card Fraud Detection Using Metalearning: Issues and Initial Results*. in *Proceeding AAAI Workshop AI Methods in Fraund and Risk Management*. 1997. Columbia: AAAI Press.
19. Stolfo, S.J., et al. *Cost-Based Modeling for Fraud and Intrusion Detection: Results from the JAM Project*. in *Proccedings DARPA Information Survivability Conference & Exposition*. 2000. Hilton Head: IEEE, Press.

20. Lozano, M. and C. García-Martínez, *Hybrid metaheuristics with evolutionary algorithms specializing in intensification and diversification: Overview and progress report.* Computers and Operations Research, 2010: p. 481-497.
21. Raidl, G.R. *A unified view on hybrid metaheuristics.* in *Proceedings of Hybrid Metaheuristics, Third International Workshop.* 2006. Berlin: Springer Verlag.
22. Talbi, E., *A Taxonomy of Hybrid Metaheuristics.* Journal of Heuristics, 2002: p. 541-564.
23. Grosan, C. and A. Abraham, *Hybrid evolutionary algorithms: methodologies, architectures, and reviews*, in *Hybrid evolutionaryalgorithms.* 2007, Springer Verlag-Heidelberg: Berlin. p. 1-17.
24. Deb, K., *Multi-objective optimization using evolutionary algorithms.* 2001, Chichester, Uk: John Wiley and Sons.
25. Dehuri, S. and S.B. Cho, *Multi-criterion Pareto based particle swarm optimized polynomial neural network for classification: A review and state-of-the-art.* Computer Science Review, 2009: p. 19-40.
26. Srivastava, A., et al., *Credit Card Fraud Detection Using Hidden Markov Model.* IEEE Transactions on Dependable and Secure Computing, 2008: p. IEEE Press.