Enhanced Anomaly Detection in Imbalanced Credit Card Transactions using Hybrid PSO

N. Sivakumar
Research Scholar
PG and Research Department of Computer Science,
J.J. College of Arts and Science,
Bharathidasan University, Tamil Nadu, India

R. Balasubramanian, PhD
Research Supervisor
PG and Research department of Computer Science
J.J. College of Arts and Science, Bharathidasan University, Tamil Nadu, India

ABSTRACT
Anomaly detection is one of the major requirements of the current age that witnesses a huge increase in online transactions. Data imbalance also poses a huge challenge in the detection process. This paper presents a hybrid metaheuristic algorithm that performs effective anomaly detection on highly imbalanced data. Particle Swarm Optimization is used as the operating algorithm. This algorithm is hybridized by modifying the probabilistic selection using Simulated Annealing. A comparison study was carried out and it was observed that the simulated annealing based PSO showed much prominence when operated on both dominant and submissive data.

Keywords
PSO; Simulated Annealing; Credit Card Fraud Detection; Data Imbalance; Anomaly Detection.

1. INTRODUCTION
Increase in e-commerce and the ease of online transactions and payments has led to an exponential increase in the number of people opting for online purchases. This has automatically led to an increase in the number of fraudsters trying to exploit the transparence involved in online transactions. The security of online transactions clearly lag behind when compared to the increase in growth of the users. It was identified by the European Central Bank (ECB) that the value of credit card frauds increased by 14.8% in 2012 [1]. Due to the flexibility provided in 'card-not-present' transactions, this area has grabbed the attention of the fraudsters. Anomaly detection is one of the major requirements in the financial sector due to the increase in the number of online financial transactions worldwide. This work focuses on automatically detecting online fraudulent transactions with an increased awareness towards the imbalance nature of the data. A major disadvantage when dealing with online transactions is the imbalanced nature of the transactions. When one class in a data set dominates the other classes with a huge ratio difference of 1:10, 1:100 or in most online transactions, it is of the form 1:1 million, the dataset is said to be imbalanced. The imbalance nature acts as a huge downside by providing better training to the majority classes and very low training to the minority classes. This makes the classifier biased towards the majority classes. Since credit card transactions are of this form, it becomes mandatory for the anomaly detection algorithm to handle imbalance effectively to improve the accuracy and reliability of the algorithm.

2. RELATED WORKS
Credit card fraud detection technique, being a mature technique has several contributions to its credit. This section discusses some of the most prominent and recent techniques in this area.

A credit card fraud detection mechanism that performs the process of fraud detection using several intrinsic features from the characteristics of incoming transactions is presented in [2]. This technique uses customer's behavior patterns such as recency, frequency and monetary properties to predict frauds. Further, several network based features are used to derive a time-dependent suspiciousness score for a transaction to identify its legitimacy. Other unsupervised techniques that work on the basis of customer's spending history are [3-6]. A clustering based spending behavior analysis is presented in [3] that raises an alarm when a transaction does not fit into the normal cluster group. Another peer group method is described in [4] that identifies anomalies based on the grouping behavior of transactions. Self-organizing maps is another grouping strategy used in [5, 6]. Artificial Neural Networks (ANN) being one of the mostly used machine learning techniques, is one of the most suitable mechanisms for identifying anomalies. This area has witnesses a huge contribution towards anomaly detection [7-13, 21]. Several ensemble methods that work well in such applications include random forests [14], SVM [15] genetic algorithms [16] and hidden Markov models [17].

A Modified Fischer Discriminant Analysis based anomaly detection method is presented in [18]. This method claims to be the first under its category, this method modifies the Fischer Discriminant function to provide more emphasis to the minor classes, hence reduce the number of false positives. An Artificial Immune System (AIS) based fraud detection model is presented in [19] that utilizes the concept of AIS to identify fraudulent transactions. An ensemble based classifier that utilizes the results of several methods to identify fraudulent transactions is presented in [20]. Though several other methods exist for identifying fraudulent transactions in credit card data, most of them do not consider imbalance as a property in the detection process, and methods considering imbalance tend to exhibit high algorithm complexities. This paper is presented in order to overcome all these problems to provide a fast and reliable detection scheme.

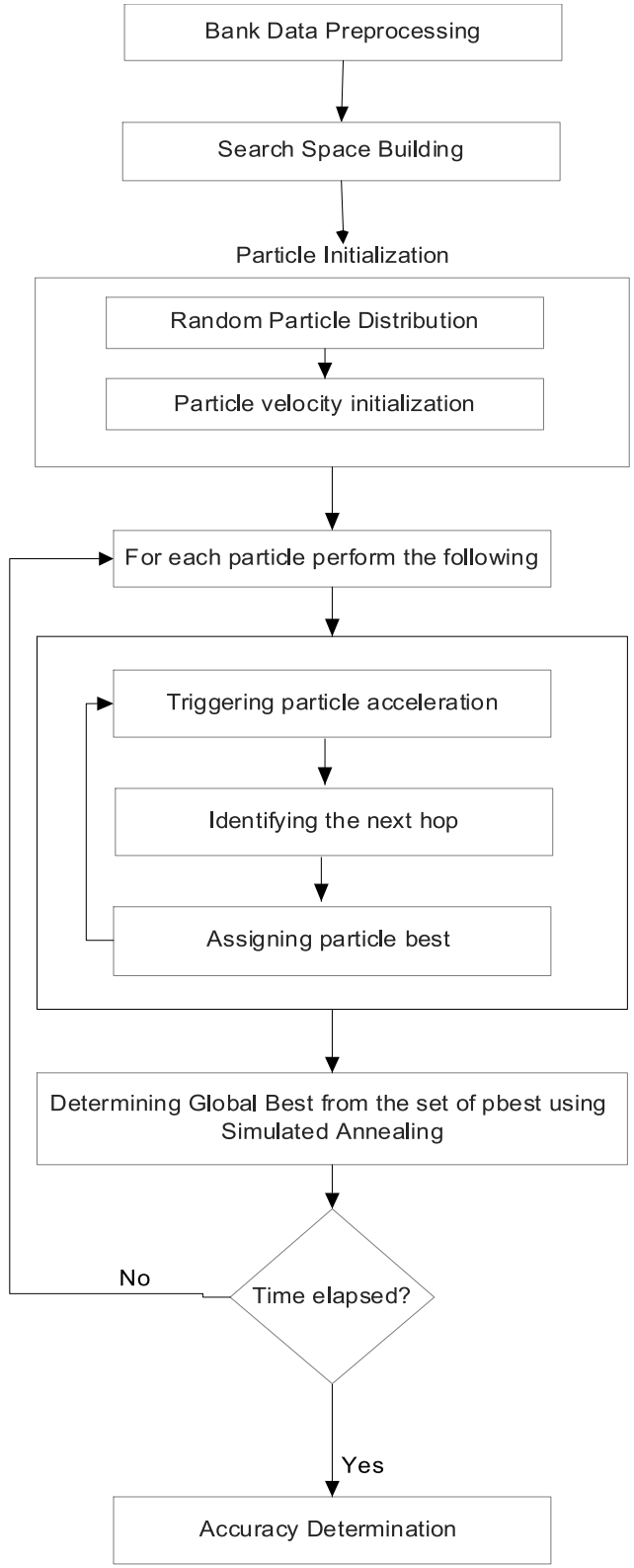

Fig 1: Anomaly Detection – Architecture

3. SYSTEM ARCHITECTURE

The architecture in Figure shows the process of anomaly detection in credit cards transactions. The process begins by preprocessing the bank data to identify for missing values. These values are eliminated and the search space is built using the transaction data. Particles are distributed in this search space and movements are triggered according to the conventional rules of PSO. After completing an iteration, the global best values are crosschecked with the current particle best values and the values exhibiting the minimum distance are considered as the current global best values. This process is repeated until time elapse or until the best solution is obtained. The final global best values are considered as the final solutions.

This process is continued for a specific time interval or until a defined stagnation behaviour. The current approach sets a maximum number of iterations as the stopping criteria. When the stopping criteria is reached, the final g*best* value is

4. ENHANCED ANOMALY DETECTION IN IMBALANCED CREDIT CARD TRANSACTIONS USING HYBRID PSO

Enhanced anomaly detection in imbalanced credit card data uses a modified form of PSO to provide better and effective predictions. Bank transactions are machine generated and hence tends to contain missing values. The initial preprocessing phase analyzes the data to eliminate all the inconsistencies. This phase generates data that is in a usable format for PSO. The search space of PSO comprises of this data. Every tuple in the data is described as a dimension in PSO.

The next process is the initialization of particles in the search space. The number of particles selected in the search space acts as the test data to identify the accuracy of the algorithm in place. Hence the number of particles is set to 25% of the total data contained in the search space. A uniform random function is used to initialize the particle to its corresponding node in the search space. Every particle is assigned to a different node. Hence a single pass of all the particles in the search space will provide a complete result set which is then fine-tuned to obtain the final results. The initial velocity of the particle is calculated at this phase, to aid in the acceleration.

$$V_i \sim U(-|b_{up} - b_{lo}|, |b_{up} - b_{lo}|) \quad (1)$$

where V_i is the velocity, b_{up} and b_{lo} are the upper and lower bounds of the search space respectively.

At this point, the particle best (*pbest*) and global best (*gbest*) values are calculated. Using the initial velocity obtained from equation (1) particle acceleration is triggered and the particles shift from the current position to the new position represented by $X_{i,d}$. The new *pbest* and *gbest* values can be calculated using

$$V_{i,d} \leftarrow \omega V_{i,d} + \varphi_p r_p (P_{i,d} - X_{i,d}) + \varphi_p r_p (g_d - X_{i,d}) \quad (2)$$

Where $P_{i,d}$ and g_d are the parameter best and the global best values, r_p and r_g are the random numbers, $X_{i,d}$ is the value of current particle position and the parameters ω, φ_p, and φ_g are selected by the user. PSO operates on continuous domain, but the current problem demands a discrete nature of working, hence the current points are discretized using the following function,

$$P' = \min\left(\sum_{j=1}^{n}\left(\sum_{k=1}^{d}\sqrt{(P_{ik} - N_{jk})^2}\right)\forall i = 1\ to\ p\right) \quad (3)$$

Where P_{ik} refers to the particle i's current location corresponding to dimension *k*, N_{jk} refers to the k^{th} dimension of node N_i.

obtained by passing all the *pbest* values and the current *gbest* value to the Simulated Annealing module. This module identifies the current *gbest* value from the set, which is considered as the final result. Final results are cross referenced with the ideal results to obtain the accuracy of the system.

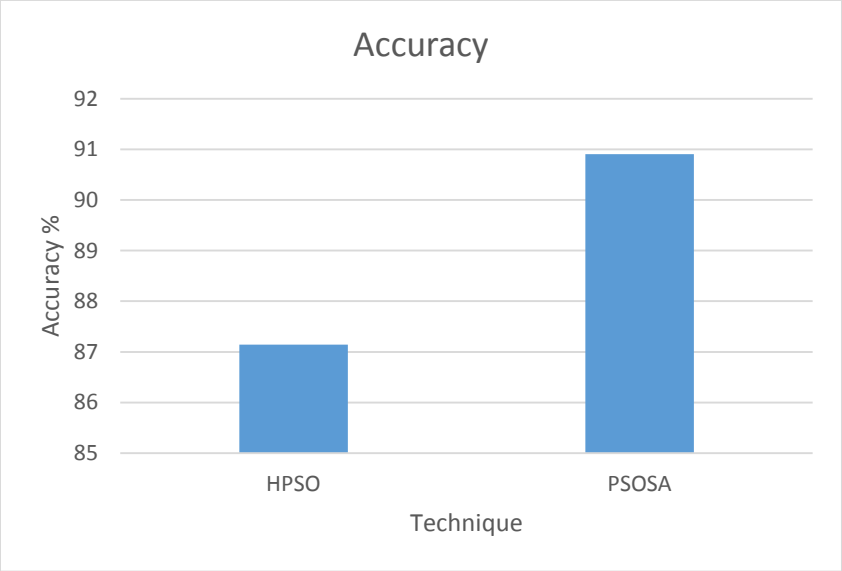

Fig 2: Accuracy

5. RESULTS AND DISCUSSION

Experiments were conducted on the credit card fraud detection dataset from a Brazilian Bank [19] with an imbalance level of 25. The data was processes using Simulated Annealing based PSO (PSOSA) and the results obtained were compared with the previous approach of the authors, HPSO. Fig 2 presents the accuracy levels obtained by HPSO and PSOSA and it was observed that PSO SA exhibited better accuracy when compared to HPSO.

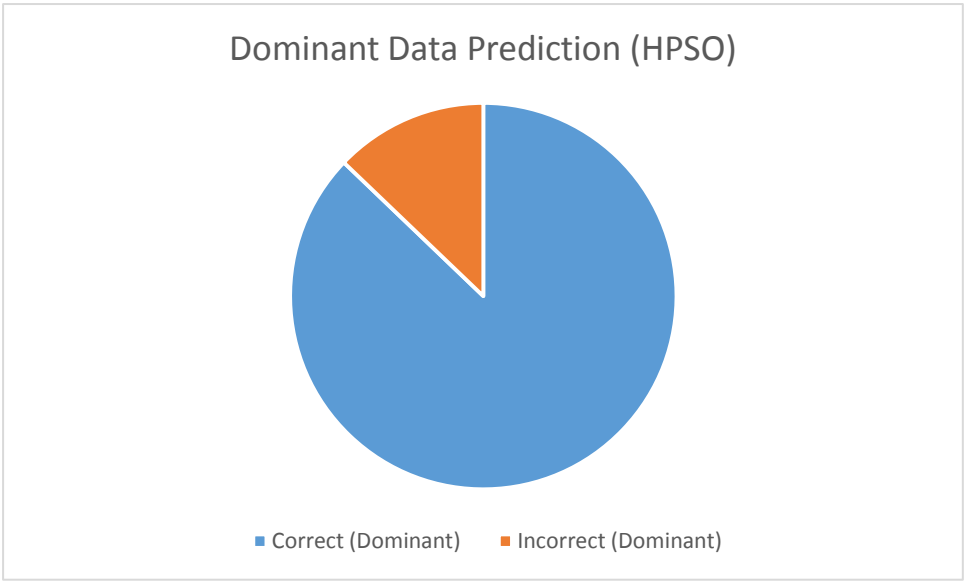

Fig 3: Dominant Data Prediction (HPSO)

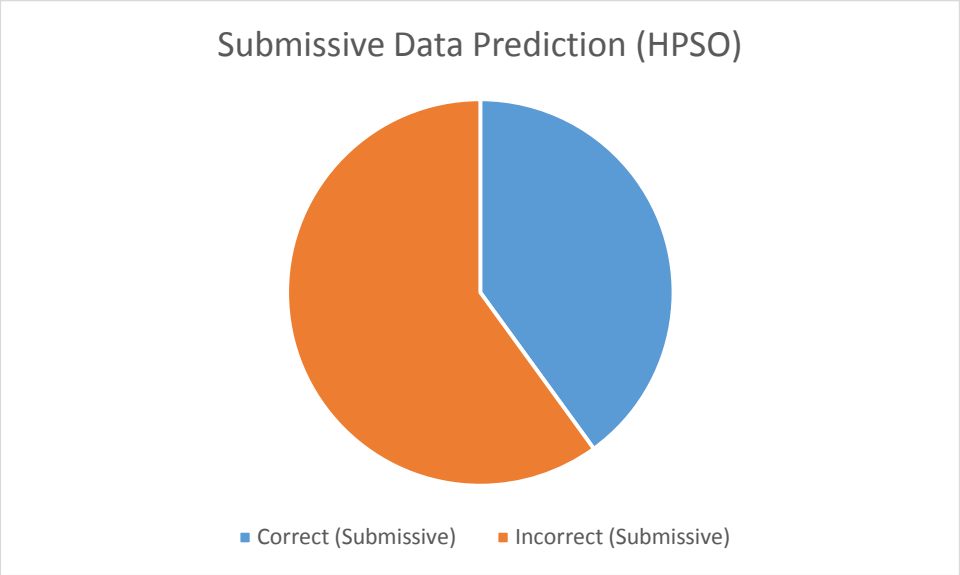

Fig 4: Submissive Data Prediction(HPSO)

Figures 3 and 4 represents the level of dominant and submissive data predicted by HPSO. It was observed that HPSO exhibits better prediction rates in dominant data rather than submissive data.

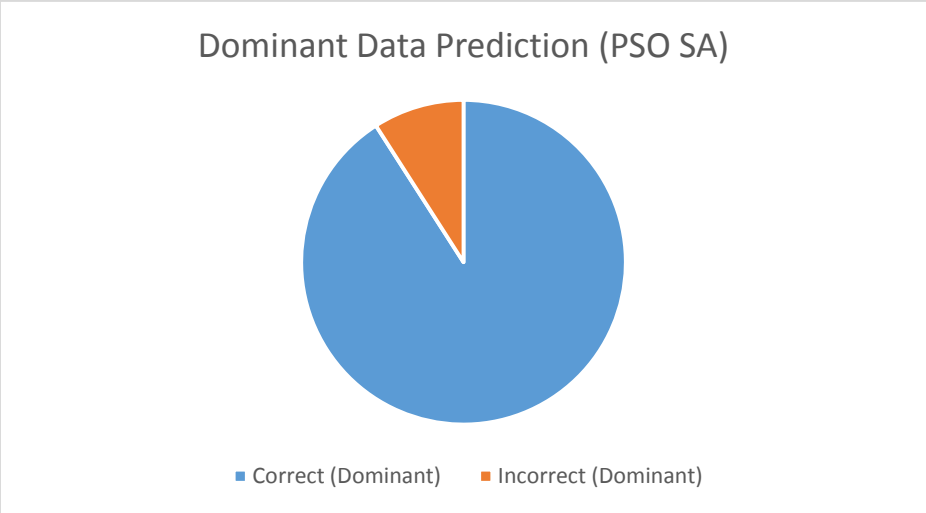

Fig 5: Dominant Data Prediction (PSO SA)

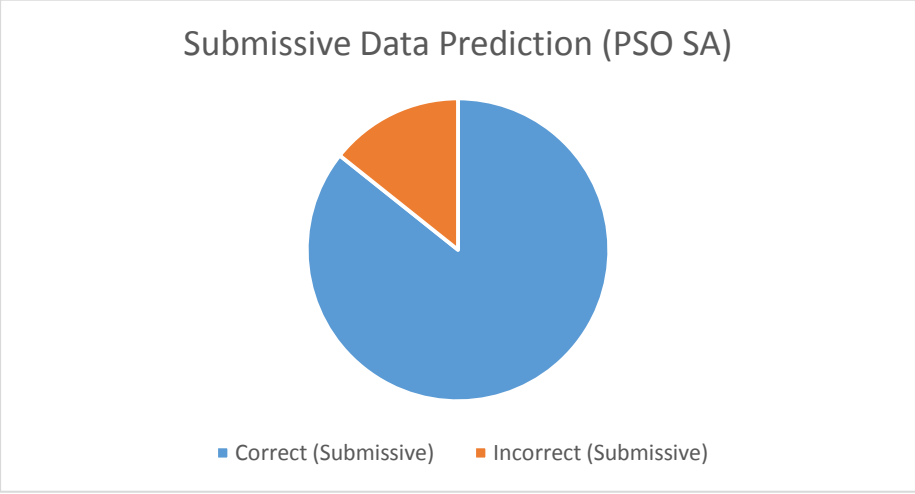

Fig 6: Submissive Data Prediction (PSO SA)

Figures 5 and 6 show the prediction rates of PSOSA on dominant and submissive data. It could be observed that PSOSA exhibits almost similar prediction rates when it comes to both dominant and submissive data.

6. CONCLUSION

The major concern in any online transaction remains to be the occurrence of fraud. Hence anomaly detection has become one of the major necessities for any system handling online financial transactions. The current approach presents a hybrid PSO based technique for identifying anomalies in credit card transactions. Regular PSO is modified by replacing its local search mechanism with simulated annealing. Limitations of this approach is that the mechanism requires missing valued entries be eliminated. Handling missing values is not supported, as every dimension requires its corresponding data to be available for the particles to navigate through them effectively. Future directions include enhancing this mechanism using parallel techniques to improve the processing speed.

7. REFERENCES

[1] ECB February 2014. Third Report on Fraud, European Central Bank.

[2] Véronique Van Vlasselaer, Cristián Bravo, Olivier Caelen, Tina Eliassi-Rad, Leman Akoglu, Monique Snoeck, Bart Baesens. 2015. APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions,Decision Support Systems, Volume 75 , Pages 38-48.

[3] Bolton, R. J. Hand, D.J. 2001. Unsupervised profiling methods for fraud detection, Proceedings of the VII Conference on Credit Scoring and Credit Control, pp. 235–255 (Edinburgh, United Kingdom).

[4] Weston, D. J., Hand, D. J., Adams, N. M. Whitrow, C., Juszczak, P.2008. Plastic card fraud detection using peer group analysis, ADAC 2 (1) 45–62.

[5] Quah, J. T., Sriganesh, M. 2008. Real-time credit card fraud detection using computational intelligence, Expert Syst. Appl. 35 (4) 1721–1732.

[6] Zaslavsky, V., Strizhak, A. 2006. Credit card fraud detection using self-organizing maps, Inf. Secur. 18 48.

[7] Aleskerov, E., Freisleben, B. Rao,.B . 1997. Cardwatch: a neural network based database mining system for credit card fraud detection, Computational Intelligence for Financial Engineering (CIFEr), Proceedings of the , pp. 220–226.

[8] Brause, R., Langsdorf, T., M. Hepp,M . 1999. Neural data mining for credit card fraud detection, Proceedings. 11th IEEE International Conference on Tools with Artificial Intelligence, pp. 103–106.

[9] Dorronsoro, J. R., Ginel, F., Sgnchez, C., Cruz, C. 1997 .Neural fraud detection in credit card operations, IEEE Trans. Neural Netw. 8 (4) 827–834.

[10] Ghosh, S., Reilly,D.L. 1994. Credit card fraud detection with a neural-network, Proceedings of the Twenty-seventh International Conference on System Sciences, vol. 3, pp. 621–630.

[11] Maes, S., Tuyls, K. Vanschoenwinkel, B. Manderick, B.2002 Credit card fraud detection using bayesian and neural networks, Proceedings of the 1st International NAISO Congress on Neuro Fuzzy Technologies.

[12] Shen, A., Tong, R., Deng, Y. 2007. Application of classification models on credit card fraud detection, Service Systems and Service Management, International Conference on. pp. 1–4.

[13] Syeda, M., Zhang, Y.Q Pan, Y. 2002. Parallel granular neural networks for fast credit card fraud detection, Proceedings of the IEEE International Conference on Fuzzy Systems, vol. 1, pp. 572–577.

[14] Henderson, K. Gallagher, B., Li, L. Akoglu, L. Eliassi-Rad, T., Tong, H. Faloutsos, C. 2011. It's who you now: graph mining using recursive structural features, Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM , pp. 663–671.

[15] Bhattacharyya, S., Jha, S., Tharakunnel, K., Westland, J. C. 2011. Data mining for credit card fraud: a comparative study, Decis. Support. Syst. 50 (3) 602–613.

[16] Duman, E., Elikucuk, I. 2013. Solving credit card fraud detection problem by the new metaheuristics migrating birds optimization, in: I. Rojas, G. Joya, J. Cabestany (Eds.), Advances in Computational Intelligence. Vol. 7903 of Lecture Notes in Computer Science, Springer, Berlin Heidelberg, pp. 62–71.

[17] Srivastava, A. Kundu, A. Sural, S.Majumdar, A.K. 2008. Credit card fraud detection using hiddenmarkov model, IEEE Trans. Dependable Secure Comput. 5 (1) 37–48.

[18] Nader Mahmoudi, Ekrem Duman, 2015. Detecting credit card fraud by Modified Fisher Discriminant Analysis,Expert Systems with Applications, Volume 42, Issue 5, Pages 2510-2516.

[19] Neda Soltani Halvaiee, Mohammad Kazem Akbari, 2014. A novel model for credit card fraud detection using Artificial Immune Systems,Applied Soft Computing, Volume 24, Pages 40-49.

[20] Masoumeh Zareapoor, Pourya Shamsolmoali, 2015. Application of credit card fraud detection: Based on bagging ensemble classifier",Procedia Computer Science, Volume 48, Pages 679-685

[21] Jarrod West, Maumita Bhattacharya. 2016. Payment Card Fraud Detection Using Neural Network Committee and Clustering, Computers & Security, Volume 57, Pages 47-66.