Credit Card Fraud Detection Using Frequent Pattern Mining Using FP-Modified Tree And Apriori Growth

Vipin kumar Choudhary[1], Er. Divya[2]
[1, 2] RPIIT Bastara, Karnal

Abstract- *The involvement of huge amounts of monetary transactions in the finance industry makes it most vulnerable to attacks from most hackers. All hackers follow some pattern of steps to break through the security of credit cards which is mostly a digital pin. So, one way to detect the fraudulent activity can be to study the pattern of activities of the fraudulent users and match them against any new user. We studied many different frequent pattern matching schemes and the advantages and disadvantages offered by each. The most commonly used schemes are Apriori method and FP tree methodology. The Apriori method has disadvantage of doing repeated database scans thereby increasing time complexity. The latter suffers from the complexity of its implementation. It performs recursive operations to generate a new tree every time making the process complicated. Another modification of FP Tree i.e FP Split tree reduces this complexity of FP Tree by avoiding the recursive scans. It also performs single scan of database unlike FP Tree which performs scan twice and Apriori which performs multiple scans. It also suffers in performance for long patterns formed. So, we finally aim to evolve a new methodology which is a hybrid of above mentioned methodologies and covers the negative aspects of both.*

Keywords- Apriori algorithm, FP Tree, Frequent patterns mining.

I. INTRODUCTION

Every year billions of Euros are lost world wide due to credit card fraud, thus, forcing financial institutions to continuously improve their fraud detection systems. In recent years, several studies have proposed the use of machine learning and data mining techniques to address this problem. However, most studies used some sort of mis-classification measure to evaluate the different solutions, and do not take into account the actual financial costs associated with the fraud detection process. Moreover, when constructing a credit card fraud detection model, it is very important how to extract the right features from the transactional data.

The use of credit and debit cards has increased significantly in the last years, unfortunately so has fraud. Because of that, billions of Euros are lost every year. According to the European Central Bank (European Central Bank, 2014), during 2012 the total level of fraud reached 1.33 billion Euros in the Single Euro Payments Area, which represents an increase of 14.8% compared with 2011. Moreover, payments across non-traditional channels (mobile, internet, etc.) accounted for 60% of the fraud, where as it was 46% in 2008. This opens new challenges as new fraud patterns emerge, and current fraud detection systems are less successful in preventing these frauds.

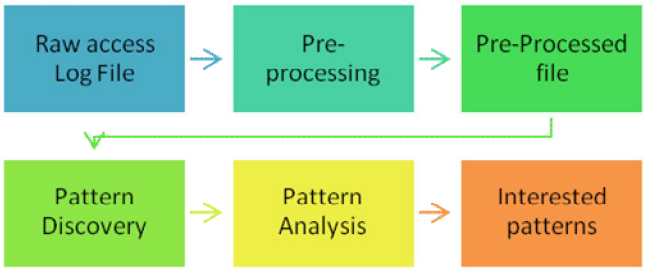

Fig. 1: Complete Process of fraudulent card usage detection

II. BASE RESEARCH

There have been numerous researches in the recent past for detection of fraudulent transactions. The following research by K.R. Seeja [5] uses frequent pattern mining by Apriori technique. This is a very obsolete technique and suffers from a major drawback of repeated database scans during itemset generation.

Seejaet. al [5] proposed an intelligent credit card fraud detection model for detecting fraud from highly imbalanced and anonymous credit card transaction datasets. The class imbalance problem was handled by finding legal as well as fraud transaction patterns for each customer by using *frequent itemset mining*. A matching algorithm was also proposed to find to which pattern (legal or fraud) the incoming transaction of a particular customer was closer and a decision was made accordingly. In order to handle the anonymous nature of the data, no preference was given to any of the attributes and each attribute was considered equally for finding the patterns. The performance evaluation of the proposed model was done on UCSD Data Mining Contest 2009 Dataset (anonymous and imbalanced) and it was found

that the proposed model had very high fraud detection rate, balanced classification rate,Matthews correlation coefficient, and very less false alarm rate than other state-of-the-art classifiers.

III. LITERATURE REVIEW

Bahnesonet. al in [1] expanded the transaction aggregation strategy, and proposed to create a new set of features based on analyzing the periodic behavior of the time of a transaction using the vonMises distribution. Then, using a real credit card fraud data set provided by a large European card processing company, the author compared state-of-the-art credit card fraud detection models, and evaluated how the different sets of features had an impact on the results. The author first proposed a new savings measure based on comparing the financial cost of an algorithm versus using no model at all. Then, an expanded version of the transaction aggregation strategy was proposed, by incorporating a combination criteria when grouping transactions, i.e., instead of aggregating only by cardholder and transaction type, the author combine it with country or merchant group. This allowed to have a much richer feature space. The research proposed a new method for extracting periodic features in order to estimate if the time of a new transaction is within the confidence interval of the previous transaction times. The motivation was that a customer was expected to make transactions at similar hours. The proposed methodology was based on analyzing the periodic behavior of a transaction time, using the vonMises distribution. By including the proposed periodic features into the methods, the results showed an average increase in savings of 13%. The research showed that by preprocessing the data in order to include the recent consumer behavior, the performance increased by more than 200% compared to using only the raw transaction information.

Beheraet. al in [2] proposed a novel approach towards credit card fraud detection in which the fraud detection was done in three phases. The first phase did the initial user authentication and verification of card details. If the check was successfully cleared, then the transaction was passed to the next phase where fuzzy cmeans clustering algorithm was applied to find out the normal usage patterns of credit card users based on their past activity. A suspicion score was calculated according to the extent of deviation from the normal patterns and thereby the transaction was classified as legitimate or suspicious or fraudulent. Once a transaction was found to be suspicious, neural network based learning mechanism was applied to determine whether it was actually a fraudulent activity or an occasional deviation by a genuine user. Extensive experimentation with stochastic models showed that the combined use of clustering technique along with learning helped in detecting fraudulent activities effectively while minimizing the generation of false alarms. The author used stochastic models for analysing the performance of the proposed system. The simulation yielded up to 93.90% TP and less than 6.10% FP. Based on the results, the author concluded that combinatorial use of fuzzy clustering and learning were the appropriate approaches for addressing this type of real world problems.

Halvaieeet. al in [3] addressed credit card fraud detection using Artificial Immune Systems (AIS), and introduced a new model called AIS-based Fraud Detection Model (AFDM). The author used an immune system inspired algorithm (AIRS) and improved it for fraud detection. AFDM improved detection rate up to 23%, decreased cost up to 85%, and training time up to 40%. It was possible to get higher than 50% detection rate while having a very low false positive rate (less than 2%), which was considerable. Also, implementing the parallel model only in a test environment showed fair decrease in training time, which was expected to be better in a real cloud computing system.

Pozolloet. al in [4] provided some answers from the practitioner's perspective by focusing on three crucial issues: unbalancedness, non-stationarity and assessment. The analysis was made possible by a real credit card dataset provided by an industrial partner. This paper aimed at making an experimental comparison of several state of the art algorithms and modeling techniques on one real dataset, focusing in particular on some open questions like:

- Which machine learning algorithm had to be used?
- Was it enough to learn a model once a month or it was necessary to update the model everyday?
- How many transactions were sufficient to train the model?
- Had the data to be analyzed in their original unbalanced form?

The author presented a way to create new features in the datasets that could trace the card holder spending habits. By doing this it was possible to present the transactions to the learning algorithm without providing the card holder identifier.

Seejaet.al in [5] proposed an intelligent credit card fraud detection model for detecting fraud from highly imbalanced and anonymous credit card transaction datasets. The class imbalance problem was handled by finding legal as well as fraud transaction patterns for each customer by using frequent item set mining. A matching algorithm was also proposed to find to which pattern (legal or fraud) the incoming

transaction of a particular customer was closer and a decision was made accordingly. In order to handle the anonymous nature of the data, no preference was given to any of the attributes and each attribute was considered equally for finding the patterns. The performance evaluation of the proposed model was done on UCSD Data Mining Contest 2009 Dataset (anonymous and imbalanced) and it was found that the proposed model had very high fraud detection rate, balanced classification rate, Matthews correlation coefficient, and very less false alarm rate than other state-of-the-art classifiers.

IV. PROPOSED METHODOLOGY

Credit card logs mining is one of the most significant fields in the area of data mining. There have been a large number of data mining algorithms rooted in these fields to perform different data analysis tasks. The top algorithms identified by the IEEE International Conference on Data Mining (ICDM) presented here are among the most influential algorithms for classification, clustering, statistical learning, association analysis, and link mining.

The proposed algorithm which is a hybrid of a modified FP Tree creation algorithm and Apriori Growth mining algorithm, is given below in two phases.

The first phase constructs the FP Split Tree which is more efficient way to create candidate sets than FP Tree since the latter involves two complete scans of the database while the former does it once. This impacts the efficiency almost 2 times better. More over the FP Split Tree created by our algorithm involves lesser use of pointers as we don't link each node in the Tree to its predecessor and successor. Rather we maintain a header list separately which maintains a list separately for each of the pages which points to the occurrence of these items in the final tree created.

The second phase involves mining the FP Split tree created using the Apriori growth algorithm. This algorithm is more efficient than FP Growth as it does not involve recreating the FP Split trees repeatedly every time in recursion as in FP Growth algorithm thereby reducing the time involved. The effectiveness of proposed technique shall be confirmed by performing comparison between the two algorithms – FP Tree algorithm and proposed hybrid technique.

V. CONCLUSION AND FUTURE SCOPE

In modern digitized world, the one thing which has evolved the most is finance industry. It has seen some revolutionary changes in technology like ATMs, banking automation software and credit cards. The involvement of huge amounts of monetary transactions in this industry makes it most vulnerable to attacks from most hackers. So, there is always a requirement for getting better of these intruders and prevent any fraudulent activities.This research is focused on the prevention of fraudulent activities against credit card security. All hackers follow some pattern of steps to break through the security of credit cards which is mostly a digital pin. So, one way to detect the fraudulent activity can be to study the pattern of activities of the fraudulent users and match them against any new user. If the pattern matches, the banking system can impose some extra security measures, to prevent any theft. We studied many different frequent pattern matching schemes and the advantages and disadvantages offered by each. The most commonly used schemes are Apriori method and FP tree methodology. The Apriori method has disadvantage of doing repeated database scans thereby increasing time complexity. The latter suffers from the complexity of its implementation. It performs recursive operations to generate a new tree every time making the process complicated. Another modification of FP Tree i.e FP Split tree reduces this complexity of FP Tree by avoiding the recursive scans. It also performs single scan of database unlike FP Tree which performs scan twice and Apriori which performs multiple scans. It also suffers in performance for long patterns formed. So, we finally aim to evolve a new methodology which is a hybrid of above mentioned methodologies and covers the negative aspects of both.The effectiveness of proposed technique can be demonstrated by comparing the results with traditional techniques like FP Tree Algorithm.

REFERENCES

[1] Alejandro Correa Bahnsen∗, DjamilaAouada, AleksandarStojanovic, BjörnOttersten, "Feature engineering strategies for credit card fraud detection", Expert Systems With Applications, Elsevier 2016.

[2] Tanmay Kumar Behera and SuvasiniPanigrahi, "Credit Card Fraud Detection: A Hybrid Approach Using Fuzzy Clustering & Neural Network", Second International Conference on Advances in Computing and Communication Engineering, IEEE 2015.

[3] NedaSoltaniHalvaiee and Mohammad KazemAkbari, "A novel model for credit card fraud detection usingArtificial Immune System", Applied Soft Computing 24 Elsevier (2014) 40–49.

[4] Andrea Dal Pozzolo, Olivier Caelen, Yann-A El Le Borgne, Serge Waterschoot And GianlucaBontempi, "Learned lessons in credit card fraud detection from a practitioner perspective", Expert Systems with Applications 2015.

[5] K. R. Seeja and MasoumehZareapoor,"FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent ItemsetMining優e Scientific World Journal Volume 2014, Article ID 252797.
[6] Priyanka S. Panchal, Prof. Urmi D. Agravat, "Hybrid Technique for User's Web Page Access Prediction based on Markov Model", 4th ICCCNT 2013 || July 4-6, 2013 || Tiruchengode, India
[7] Harendra Singh, Ashish Kumar Srivastava, SitendraTamrakar, "A Modified FP-Tree Algorithm for Generating Frequent Access Patterns", JECET || June – August-2013; Vol.2.No.3 || 730-740 || E-ISSN: 2278–179X
[8] Omer Adel Nasser & Dr. Nedhal A.AL Saiyd, The Integerating Between Web Usage Mining and Data Mining Techniques,5th Conference on CSIT,IEEE(2013)
[9] ShipraKhare, Prof Vivek Jain, Prof ManojRamaiya, Implementation of Web Usage Mining with Customized Web Log Using FP Growth Algorithms, International Journal of Engineering & Managerial Innovations (IJEMI) || ISSN: 2321-693X || Volume I (II), September (2013)
[10] Jun Yang, Z. Li, Wei Xiang, An Improved Apriori Algorithm Based on Features, IEEE (2013)
[11] BinaKotiyal, Ankit Kumar, Bhaskar Pant, R.H. Goudar, ShivaliChauhan and SonamJunee, "User Behavior Analysis in Web Log through Comparative Study of Eclat and Apriori", Proceedings of 7th International Conference on Intelligent Systems and Control (ISCO 2013) || 978-1-4673-4603-0
[12] MajaDimitrijevic, TanjaKrunic, "Association rules for improving website effectiveness: case analysis", Online Journal of Applied Knowledge Management || Volume 1, Issue 2, 2013
[13] Kirti S. Patil, Sandip S. Patil, "Sequential Pattern Mining Using Apriori Algorithm & Frequent Pattern Tree Algorithm", IOSR Journal of Engineering (IOSRJEN) || e-ISSN: 2250-3021, p-ISSN: 2278-8719 || Vol. 3, Issue 1 (Jan. 2013), ||V4|| PP 26-30
[14] YunLong Song and RamWei, Research on Application of Data Mining based on FP Growth Algorithm for Digital Library, IEEE (2011)
[15] Sandeep Singh Rawat, Lakshmi Rajamani, "Discovering Potential User Browsing Behaviors Using Custom-Built Apriori Algorithm", International Journal of Computer Science & Information Technology (IJCSIT) || Vol.2, No.4, August 2010
[16] Goswami D.N., ChaturvediAnshu, Raghuvanshi C.S., "An Algorithm for Frequent Pattern Mining Based On Apriori", (IJCSE) International Journal on Computer Science and Engineering || Vol. 02, No. 04, 2010 || 942-947
[17] RakeshAgrawal, RamakrishnanSrikant, "Fast Algorithms for Mining Association Rules", Proceedings of the 20th VLDB Conference, Santiago, Chile,