What is the Best Approach?: Evaluating the Quality of Aggregated and Personalized approaches in Credit card Anomaly Detection

John Richard D. Kho[1], Dr. Madhavi Devaraj PhD[2], [3]Joel C. De Goma
[1],[2,[3]Department of Computer Science, Mapua University, Manila, Philippines

[1]jrdkho@gmail.com,[2] mdevaraj@mapua.edu.ph, [3]jecdegoma@mapua.edu.ph

Article Info
Volume 83
Page Number: 5840 - 5845
Publication Issue:
May - June 2020

Article History
Article Received: *19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *17 May 2020*

Abstract

Utilization of cashless transactions such as credit card transactions has been increased over the past recent years. This reason forces the researchers to create personalized model for each credit card holder, to identify fraudulent transactions, using various approaches. This paper presents the effectiveness of aggregated and personalized approaches in fraud detection. Data set collected from bank transactions are effectively used to compare the predictions driven by the above said approaches. Naïve Bayes & Random Forest classifiers are effectively used in this research.

Keywords: *credit card; fraud detection; personalized model; aggregated model; Naïve Bayes; Random Forest; Mean Absolute Error; WEKA*

I. Introduction

For the past years, it has been observed that people are becoming inclined to performing cashless transactions [1]. Moreover, this behavior is expected to advance in the next 5 to 10 years [2] and the most dominant payment scheme is using the credit card. The Credit Card business model is very common to banks because of the profit that it can generate for the company. Different business strategies can be augmented from just maximizing the potential of the business model. However, fraudsters are able to formulate different modus-operandi to somewhat acquire these profits from the banks, merchants, and card holders.

Over the nine decades of existence, different credit card related crimes were already identified and published to the private and public sectors for precautionary measures. The five recognized credit card fraud types are (i) counterfeit credit cards, (ii) lost or stolen, (iii) no-card fraud (i.e., giving card information to non-legit telemarketer), (iv) stolen cards during mailing fraud, and (v) identity-theft fraud [3]. Detection of such cases will be achieved by monitoring the spending behavior of the stolen credit card; like checking the location of the transaction, the amount of the transaction, the frequency and the volume of the transactions that the card was used, and many more. However, not all credit card fraud cases can be detected using the behavioral approach. Like the "application fraud" where a person uses stolen or fake documents to open an account in another person’s name [4]. In this case, a thorough background check must be performed during the application period by the issuing bank.

Different security features have been applied in the credit card plastics just to minimize the fraudulent transactions caused by card skimming and or fabrication which pose high occurrences. Holograms, signature panel, and magnetic stripes are some of these features that will uniquely identify that the card is legitimate. However, this does not stop the criminal to explore other alternatives to gain control or obtain personal and or credit card information and alike. Thus, this study suggests handling of fraud detection through data mining approach.

A typical procedure for to data mining is to collect the data, do some data cleansing and normalization, perform training, and lastly to validate or test the effectiveness of the created model [5]. This type of streamline can be described as aggregated design wherein the data are collectively evaluated to form a model. On the contrary, the personalized design will only evaluate dataset pertaining to the specific credit card owner for the reason that spending behavior of each customer differ from one another. This paper will present thorough evaluation of these two approaches to determine which one is better in handling credit card fraud detection.

II. RELATED WORK

Understanding the essence of the dataset at hand can be achieved by selecting the proper model for clustering and or classification. Will it be a personalized or aggregated approach? Personalized approach will evaluate each cardholder's recent transactions and validate if the incoming transaction will deviate from its normal behavior [6]. While, the aggregated model will gather all same instances of a particular or different fraud types and be used to understand their patterns. Collected patterns will then be used to validate incoming transactions.

The paper [7] contradicts the study regarding the effectiveness of the personalized model in detecting fraudulent transactions using credit cards. They suggested that the quantity of the dataset available affects the effectiveness of a classifier to determine the difference between the legitimate and fraudulent transactions. In the paper, they also compared the accuracies of Naïve Bayes and Random Forest classifiers in building models for the aggregated approach. They concluded that the Random Forest fits the aggregated model while Naïve Bayes is suitable for the personalized model.

Customer data acquired by the private sectors especially by the financial and banking industries are kept in utmost secrecy. Most countries even impose laws that protect bank customers (or clients) from disclosing their information without proper consent or authorization [8] [9] [10]. Consequently, it is very hard for data analysts to perform thorough research and study in applying data mining as fraud detection solution in the financial sector. The aforementioned research papers are not exempted from this dilemma although they have overcome this limitation by building a survey or questionnaires to mimic legitimate and fraudulent credit card transactions. With the formulated online data gathered from the participants, they validated the effectiveness of their created classification models. The alarming part here is that the dataset they have obtained from the participants is not the actual credit card transaction formulated by the bank's infrastructure – the results of the evaluation is highly doubtful. Hoping that this paper will somehow fortify their findings.

III. METHODOLOGY

There was an NDA (Non-Disclosure Agreement) between the proponents of this paper and the participating bank, therefore, dataset attributes will be fully concealed.

A. Data Preparation:

The data were derived from two different sources (i) recorded credit card fraud cases, and (ii) transaction logs found in the credit card host system. The latter data are composed of different types of transactions such as installment details, straight transactions, reversals, auto-debit arrangement, and many more. However, in this study, only straight transactions were utilized. All transactions coming from the payment terminals such as POS, ATM, online banking, and alike were forwarded to this table – transaction log table. Though communications between these terminals and card host systems conform to the financial transaction message format [11], records are already processed, segregated, and partly sanitized inside the transaction logs table.

Data from these sources have undergone some data cleansing such as removal of excessive spaces and special characters. Afterwards, data annotation was performed; mapping of the two (2) tables is shown in Figure 1. Data from the transaction logs table that exist in fraud cases table will be marked as "TRUE" for the target class value, otherwise "FALSE".

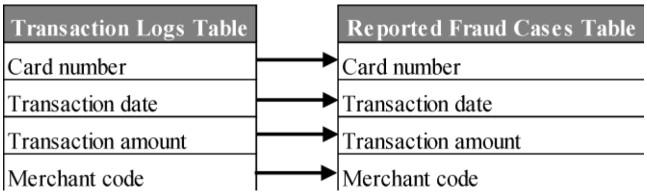

Figure 1. Data annotation mapping [12].

To improve the space and time complexity in the model creation, several pre-processing procedures may be utilized. However, this will be optional; certain measurements are still needed to be performed and confirmed. A pre-processing operation is binning-- where values will be thoroughly analyzed and divided into groups. These groups will be given a unique identifier or a value (e.g. mean or median based from the values included in the group). Another technique that can be implemented is the feature selection. The objective is to evaluate the importance of the attributes based on their values and to see if some attributes have correlations to other attributes. Gain Information and Correlation-based Feature Selections are some of the feature ranking and feature selection techniques were utilized in this paper. Table I contains the ranking of the transaction logs attributes; field with ranking of zero were automatically removed from the list.

Information Gain		**Schema**		
Grade	***Rank***	***Field***	***Length***	***Decimal***
0.4359	1	attr02	3	0
0.1023	2	attr25	3	
0.1019	3	attr18	3	
0.0846	4	attr36	11	0
0.0797	5	attr24	12	2
0.0771	6	attr11	4	0
0.0771	7	attr23	3	
0.0663	8	attr19	4	0
0.0451	9	attr07	4	0
0.0432	10	attr26	13	2
0.0369	11	attr14	5	0
0.0273	12	attr16	3	0
0.0205	13	attr33	3	0
0.0205	14	attr32	1	
0.0162	15	attr08	11	2
0.0124	16	attr17	5	0
0.0111	17	attr16	41	
0.0097	18	attr37	10	
0.0097	19	attr38	10	
0.0064	20	attr21	3	
0.0012	21	attr28	1	
0.0011	22	attr27	1	0
0.0011	23	attr39	16	
0.0007	24	attr35	4	0

TABLE I. FINAL DATA ATTRIBUTES [12]

B. Data Analysis

The data used in this study came from a bank that receives millions of transactions per month. However, we only requested for data from January to May 2016 due to limitation of the resources – handling such amount of data requires a high-performance server. The common issue encountered by the previous studies is that dataset of credit card is highly skewed or uneven with respect to ratio between normal and anomalous transactions [13]. To somehow resolve this dilemma, accounts that existed in the reported fraud cases were used as reference to extract those transactions that are present in the transaction logs dataset. However, this paper utilized only three (3) different customers from the dataset. The data count pertaining to the selected customers are seen in Table II while Table III contains their spending behavior.

		COUNT	
Instance	**Class value**	**per class**	**per instance**
customer-1	false	77	85
	true	8	
customer-2	false	174	185
	true	11	
customer-3	false	103	111

	true	8	
TOTAL		381	

TABLE II. DATA COUNT PER CUSTOMER

	TRANSACTION AMOUNT		
Instance	**Min**	**Max**	**Avg**
customer-1	100.00	81,000.00	2,900.00
customer-2	130.00	71,000.00	3,700.00
customer-3	62.00	68,000.00	4,200.00

TABLE III. SPENDING BEHAVIOR OF EACH SPECIMEN

With the given information, customer-1 had the lowest count of transactions yet have spent more than the other selected profiles. While customer-2 had the highest instance of spending activities but did not exceed nor fall behind to both customer-1 and customer-3's maximum transaction amounts.

C. *Modeling and Testing*

Prior to modeling, various preprocessing steps were performed in the final dataset such as data sanitation, noise reduction, feature selection and etc. Table I shows the number of transaction instances of each customer and from here, datasets were formulated. The personalized dataset contained their specific number of records reduced by four (4) were used as the training dataset; the aggregated instance contained all the records of these customers. To summarize, there were 4 datasets (which has 81, 181, and 107 for customer-1, customer-2, and custermer-3 respectively) for training and 3 datasets (which has 4 transactions each with 2 positives and 2 negatives class) for testing.

Models were built from Naïve Bayes (NB) or Random Forest (RF); these classifiers were evaluated based on their detection rate. The four (4) datasets were cross-evaluated by the selected classifiers imposing a 10-folds cross validation. Results were tabulated and presented in Table IV and Table V for personalized and aggregated approach.

	Customer-1		Cutstomer-2		Customer-3	
	NB	**RF**	**NB**	**RF**	**NB**	**RF**
Accuracy (%)	**95.06**	95.06	98.89	**99.45**	99.07	**100.00**
Kappa	**0.69**	0.57	0.89	**0.94**	0.92	**1.00**
Precision (W.A.)	**0.96**	0.95	0.99	**0.99**	0.99	**1.00**
Recall (W.A.)	**0.95**	0.95	0.98	**0.99**	0.99	**1.00**

TABLE IV. CLASSIFIER EVALUATION USING PERSONALIZED DATASET

	NB	**RF**
Accuracy (%)	94.31	**97.83**
Kappa	0.62	**0.79**
Precision (W.A.)	0.97	**0.98**
Recall (W.A.)	0.94	**0.98**

TABLE V.CLASSIFIER EVALUATION USING AGGREGATED DATASET

Both test results have shown that the Random Forest outmatched the Naïve Bayes in all aspects of the given measurements.

During the testing phase, a model was created using the Random Forest classifier utilizing the four (4) datasets. For the personalized model, they were tested with their own respective test set while the aggregated model processed all the three test sets (cross-evaluation). Table VI shows results to both personalized and aggregated models.

	Customer-1	**customer-2**	**customer-3**
Accuracy (%)	100.00	100.00	100.00
Kappa	1.00	1.00	1.00
Precision (W.A.)	1.00	1.00	1.00
Recall (W.A.)	1.00	1.00	1.00

TABLE VI.TEST RESULTS FOR BOTH PERSONALIZED AND AGGREGATED MODELS

The results show that both approaches produced a perfect detection rate. However, looking deeply on the confidence value of the prediction it was revealed that the personalized approach is better compared to aggregated approach – which can be supported by evaluating the mean absolute error values of each test as presented in Table VII.

The mean absolute error (MAE) measures the model's prediction with respect to actual value. It tells how confident the selected classifier in providing its prediction – lower the value, the better. Although the personalized model failed to provide low MAE value in the customer-1 test result, it surpassed most of the remaining test instances (seen in customer-2 and customer-3 results).

	Mean Absolute Error	
	Personalized	Aggregated
customer-1	0.1775	0.1250
customer-2	0.0025	0.1250
customer-3	0.0530	0.1250

Table VII.Mean absolute scores for Personalized and Aggregated models

IV. CONCLUSION AND RECOMMENDATION

In the practice of data mining, it is most of the time performing an aggregated approach in creating any model. Datasets are prepared containing the history of the subject under observation to produce a model that will provide good prediction of the incoming data. However, the application of this approach is still on a case to case basis – which was presented in this study. Fraud detection in Credit Card can be performed mostly in aggregated approach wherein the previous fraud case instances are evaluated to produce rules that can predict the outcome of the incoming transaction. This paper has shown that the personalized approach is better than aggregated because it does not consider the spending behavior of the other customers. It only evaluates the transaction history of the customer without the influence of the other customers.

The aggregated and personal models can also be used specifically to detect certain type of credit card fraud case. For example, aggregated approach can be applied in detecting BIN attack because uninterrupted transactions using specific BIN value would suggest that an attack is in progress. Combining and monitoring all transactions will immediately identify the attack. While personalize approach is effective in detecting deviation from the normal credit card usage such as time of transaction, place, amount, frequency and many more pertaining to a particular card holder.

For future work, the proponent suggests performing the same evaluation using the data from other card host systems. In the card industry, several vendors offer different systems that are capable of managing the card business, operations, and transactions. These card management solutions have their own formatting, dataset segmentation, and field data types implementations; such analysis will benefit the entire credit card business in understanding the credit card data. Collating signatures from other banks using their systems will provide a more effective credit card fraud detection model. Furthermore, when implementing such an approach, this paper recommends a multi-tier design of model since each classifier has its own advantages and disadvantages – such design might compensate the weakness of other classifiers. The procedures and approaches presented in this paper will still be effective in selecting other classifiers when integrated in the proposed design.

V. References

[1] American Consumer Credit Counseling, Inc. (04 August 2015). "INFOGRAPHIC: CASH VS. CARD". Retrieved from http://www.consumercredit.com/financial-education/infographics/infographic-cash-vs-card.aspx

[2] The Nilson Report, (27 February 2018) "Charts & Graphs Archive". Retrieved from https://www.nilsonreport.com/publication_chart_and_graphs_archive.php?1=1&year=2018

[3] Statistic Brain Research Institute (2014, July 12). Credit Card Fraud Statistics (2014). Retrieved from http://www.statisticbrain.com/credit-card-fraud-statistics/

[4] VISA (2015, May 30). Identity Theft (2015). Retrieved from http://usa.visa.com/personal/security/learn-the-facts/identity-theft.jsp

[5] Han J. & Kamber M. 2006. Data Mining: Concepts and Techniques. Amsterdam: Elsevier

[6] R. C. Chen, S.-T. Luo, X. Liang and V. C. Lee, "Personalized Approach Based on SVM and ANN for Detecting Credit Card Fraud," ICNN&B '05 Conference on Neural Networks and Brain, p. 810 – 815, 2005.

[7] M. I. Alowais and L.-K. Soon, "Credit Card Fraud Detection: Personalized or Aggregated Model,"

2012 Third FTRA International Conference on Mobile, Ubiquitous, and Intelligent Computing, pp. 114 - 119, 2012.

[8] Banking laws of the Philippines – The laws on secrecy of bank deposits (01 November 2015). Retrieved from http://www.baiphil.org/wpcontent/uploads/2010/03/SBD_Booklet_Official22012.06.13.pdf

[9] Republic Act No. 1405 – An act prohibiting disclosure of or inquiry into, deposits with any banking institution and providing penalty therefore (01 November 2015). Retrieved from http://www.pdic.gov.ph/index.php?nid1=10&nid2=3

[10] Republic Act No. 10173 – An act protecting individual personal information in information and communications systems in the government and the private sector creating for this purpose a national privacy commission, and for other purposes (15 August 2012). Retrieved from www.gov.ph/2012/08/15/republic-act-no-10173/

[11] IBM Knowledge Center (21 July 2017). ISO8583 messaging standard. Retrieved from https://www.ibm.com/support/knowledgecenter/en/SSMKHH_9.0.0/com.ibm.etools.mft.doc/bd34064_.htm

[12] J. R. D. Kho and L. A. Vea, “Credit Card Fraud Detection Based on Transaction Behavior” in Bridging the Gap: Proceedings of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017. Pp 1880 – 1884.

[13] P. K. Chan, W. Fan, A. Prodomidis, and S. Stolfo, "Distributed Data Mining in Credit Card Fraud Detection". Copyright 1999