Comput Syst Sci & Eng (2020) 3: 141–149

International Journal of
Computer Systems
Science & Engineering

Classification-Based Fraud Detection for Payment Marketing and Promotion

Shuo He[1]*, Jianbin Zheng[1†], Jiale Lin[2‡], Tao Tang[1§], Jintao Zhao[1¶] and Hongbao Liu[1‖]

[1] *China UnionPay Research Institute of Electronic Payment. Shanghai 201201, China, P.R China*
[2] *Shanghai Jiao Tong University, Shanghai 200240, China, P.R China*

Nowadays, many payment service providers use the discounts and other marketing strategies to promote their products. This also raises the issue of people who deliberately take advantage of such promotions to reap financial benefits. These people are known as 'scalper parties' or 'econnoisseurs' which can constitute an underground industry. In this paper, we show how to use machine learning to assist in identifying abnormal scalper transactions. Moreover, we introduce the basic methods of Decision Tree and Boosting Tree, and show how these classification methods can be applied in the detection of abnormal transactions. In addition, we introduce a graph computing method, which implicitly describes the characteristics of people and merchants through node correlation, in order to mine deep features. Because of the volume of large data, we carried out reasonable block calculation, and succeeded in reducing a large amount of data to a series of segments, thereby decreasing the computational resources and memory requirements. Compared with other work on abnormal transaction detection, we pay more attention to creating and using the portraits of merchants or individuals to assist in decision-making. After data analysis and model building, we find that focusing on only one transaction or one day does not yield a comprehensive number of characteristics, and many characteristics can be obtained by examining the transactions of a person or a merchant over a period of time. Furthermore, a large number of characteristics can be obtained from transactions in a period of time. After GBDT (Gradient Boosting Decision Tree) based classification prediction and analysis, we can conclude that there is a clear distinction between abnormal trading shops and conventional shops, facilitating the clustering of abnormal merchants. By filtering transaction data from multiple dimensions, multiple sub-graphs can be obtained. After hierarchical clustering, the abnormal trading group is mined and classified according to its features. Finally, we build a scoring model and apply it to the big data platform of one of China's largest payment service providers to help enterprises identify abnormal trading groups and specific marketing strategies.

Keywords: Payment fraud detection, Outlier detection, Gradient boosting decision tree, Machine learning

1. INTRODUCTION

Fraud detection has been a central topic in data mining and machine learning, with wide application in different areas including auto insurance [1], telecommunication [2] credit card risk management [3], health insurance [4], etc. Among these applications, different machine learning techniques have been adopted and tailored to develop effective fraud detection systems and tools to pinpoint the risk of fraud by different parties which can cause significant losses for the stake holders, such as payment service providers especially when they are launching promotional activities.

In this paper, we develop supervised machine learning approaches and tools for fraud detection, in the context of mobile payment services, which is a relatively new scenario compared with traditional payment channels. In particular, via the new payment channels, especially in developing countries like China, many free coupons are given to users of payment Apps without protection against malicious attack, which can cause a significant economic loss to the payment service providers who launch these promotional events.

In fact, there are many payment channels such as Alipay, Cloud Quick Pass, and TenPay that are the mobile payment products offered by major players in China such as Ant Financial,

*shuohe@unionpay.com
†zhengjianbin@unionpay.com
‡Email:linjiale@sjtu.edu.cn
§tangtao2@unionpay.com
¶zhaojintao@unionpay.com
‖liuhongbao@unionpay.com

UnionPay, and Tencent. All these players together with other companies, have invested heavily in promoting their payment Apps. Examples of promotion strategies include a random discount on the goods and return of cash (namely back-cash) to users, as well as more complex sales promotions. As a result, some users can often easily obtain additional discounts or cashback by certain means. Originally, this was intended to increase the popularity of this payment method and the number of payment Apps users. To avoid deliberate arbitrage and ensure returns on the marketing and promotion investment, an effective and efficient technical approach is required to identify the attackers in a precise and timely manner. A notable fact is that most of the promotion costs are covered by the payment service providers; hence, they are highly motivated to support fraud detection during the promotion procedure. In the past, many rule-based approaches have been developed. Although they have been effective in certain cases, they still have fundamental limitations as discussed below.

First, the rules are often extracted and summarized by domain experts. However, their knowledge could be out-of-date which subsequently can produce poor performance in a dynamic environment. This can be exacerbated by the fact that the attackers can identify the fraud detection rules, and manipulate these rules to avoid detection. We have seen many examples in industry which can be regarded as a game between the two sides. Moreover, it is becoming more challenging for rule-makers to effectively discover and extract rules for fraud detection. One reason for this is that the attackers are becoming more and more resourceful in disguising themselves and avoiding scrutiny. On the other hand, the marketing forms and media are becoming increasingly complex, rendering the rule-based detection face the combinatorial explosion problem when all the dimensions are considered, especially in a dynamic environment whereby the features need be fused over time. Hence, the learning-based computational fraud detection module is a welcome replacement or supplement for traditional rule-based systems.

Typical fraud activities can occur in several ways. For example, an individual client may conspire with the merchant. In some cases, the payment Apps may launch a promotion that offers refunds of up to 50 percent on transactions conducted by the individual client and the merchant that provides the goods. Given this opportunity, the merchant and individual client could conduct ten fake transactions, each one worth 100 US dollars. Then the client receives a refund of $10 \times 50 = 500$ US dollars. Then the merchant and client share these ill-gotten gains. The fraudulent activity may be detected by an abnormal signal indicating that many concurrent transactions have been conducted between the client and the merchant with a notable amount of back-cash. In reality, the clients and merchants can disguise their activity in a more undetectable way. For example, clients may use multiple credit cards to make the payments, or a group of clients are involved in the fraud and share the back-cash. Hence, more advanced rules and technologies are required to detect these fraudulent transactions promptly and effectively.

In this paper, we develop a practical tool for this purpose and, in particular, we develop and deliver such which have been used by a major mobile payment player in China. More specifically, our study covers two typical settings: merchant and individual fraud detection, from the broader perspective of the ecosystem. Our method, in general, involves an ensemble approach which we term 'Gradient Boosting Decision Tree' (GBDT) [5], which incorporates both ensembles learning [6] and decision tree [7] in the procedure. Unlike the popular Adaboost method [8], GBDT does not use the result of the last iteration as a linear weight; instead, it adopts the forward stage-wise additive model. The essence of the GBDT method is a fusion of the decision tree model, ensemble learning and various loss functions. In general, GBDT can often achieve more accurate results than the Support Vector Machine (SVM) [9] in many practical problems, and it also has little dependency on the selection of the hyperparameters. All these features make it an appealing technique for machine learning for real-world applications. For this reason, we also adopt this technique in our pipeline, while the main innovation lies in the effective design of relevant features as the inputs to the model, and in the detailed description of the special treatment of two fraud detection targets: merchants and individual Apps users.

In a nutshell, the main contributions of the paper are as follows:

- We have developed individual user level fraud detection based on a gradient boosting decision tree (GBDT) model. The input features include individual profiles, physical groups which have been detected by the transaction association, as well as the virtual group information which explores the data of transaction frequency, discount rate, payment target category variance (i.e. food, clothes and others), individual average transaction count, etc.

- We also develop the GBDT classifier using the input at the merchant level, complementing to the individual-level fraud detection engine.

- We present case studies of payment fraud detection at both individual level and merchant levels, and give several examples of the performance of quantitative detection using real-world data. The experimental results show the efficacy of our detection engine.

The rest of the paper is organized as follows. In section 2, the work related to different techniques and application scenarios is discussed. In section 3 and section 4, we present the proposed analytics pipeline for the detection of payment fraud, for individuals and merchants respectively. The detailed design and the motivation empowered with domain knowledge is discussed. The case study is presented in section 5 with lessons learned from these real-world applications; section 6 concludes the paper.

2. RELATED WORK

In this section, we discuss related works on fraud detection in a broader setting to encompass mobile payment as the latter has emerged relatively recently, although the underlying methodology and technology are very similar and easily referenced.

2.1 Fraud Detection Scenarios

Fraud detection has been a long-standing problem which has attracted intensive studies in recent years. Here we provide some examples from the business and application perspectives. Fraud is common and widespread, having been detected in various areas including auto insurance [1], telecommunication [2], [10], credit card risk management [3], and health insurance [4].

2.2 Fraud Detection Technology

From the technical perspective, fraud detection has been a challenging task partly due to the fact that the supervised learning-based models require adequate numbers of positive samples of transactions, individual clients, or merchants, in comparison with the normal samples. Because many frauds occur without being detected, the positive samples are often a subset of the actual total set of fraud samples. This leads to two issues: 1) the distribution of the positive samples (fraud) and negative samples can be biased; 2) the number of positive samples becomes even smaller as some are not detected, especially compared with the unlimited number of normal samples.

In the literature, different methods have been adopted for fraud detection in both supervised and unsupervised learning-based ways. With supervision, two rather popular machine learning methods, support vector machine (SVM) and Logistic regression (LR) have been applied in fraud detection.

While in reality, supervised information is often difficult to acquire, in such cases, unsupervised methods have attracted much attention, with clustering methods being the dominant techniques [11], [12].

Last, there are several predictive models which aim to detect and prevent the fraud or failure. In fact, dynamic failure modeling and prediction [13] have been central to preventative maintenance. Another thread is based on time series detection of activity over time, whereby the multi-dimensional streaming data is collected from different sensors.

2.3 Summary and Discussion

As discussed above, although fraud detection has been a long-standing problem in different financial areas, no study has fully explored fraudulent activities in the context of mobile payment Apps. In particular, it is an emerging area and China has been one of the main countries to adopt and promote new payment tools.

Hence, we intend to conduct an in-depth study of our real-world business needs (as one of the biggest payment service providers in the world) based on rich transaction-related data. In the following sections, we describe our methods and case studies in detail.

3. BUILDING LEARNING MODELS FOR INDIVIDUAL FRAUD DETECTION

In this section, we discuss the machine learning-based pipeline used to detect fraudulent activities of individual users who take advantage of the payment tools for dishonest gain.

In summary, for individual fraud detection, the procedure can be described as follows, and the corresponding working flowchart in Fig. 1 shows that there are multiple layers. As the input, the transaction network consists of red individual client nodes and blue merchant nodes. The raw network systematically undergoes four different and sequential filtering mechanisms to generate four versions of filtered networks from which rich features are extracted for the monomer trading characters, internal group characters, and group contribution characters. These extracted features are used as input to the classification and regression model (GBDT) for fraud detection for each individual and the corresponding abnormal probability. Similarly, the clustering model is used to discover the latent patterns of individual clients in terms of different fraudulent behaviors as well as normal behaviors.

The clustering results are illustrated in Fig. 2. The specific process is as follows:

- Extract individual features and form underlying groups;
- Extract virtual group from multiple dimensions;
- Build classifier using the features extracted from the above steps.

We now discuss the details step-by-step.

Profile feature extraction. The model explores the different features of a payment user, including the frequency of transaction activity, the variations in the transaction time, average number of transactions over a certain period, the IP address variation over time for the transaction, etc. For business sensitivity, the detailed feature design cannot be disclosed in this paper, although we believe the above information can be useful to the community.

Physical information grouping. First, for each individual payment Apps user, the mobile devise ID, cell phone number and bank card (for payment) number associated with each of his/her transaction are studied to find its connection to others, and the aggregation number can be an indicative feature variable. More specifically, consider the user using the same cell phone (thus the same phone number) first uses card 'A' to make the payment, and then uses another bank card 'B' for payment again. Although, apart from the card number, no other information is unknown (e.g. the card owner) these two cards will be grouped with the same cluster according to the card number in order to denoting the profile of the user. In a more complex but more realistic example, the attackers may use the professional device i.e. MODEM Pool which can be equipped with multiple cell phone cards, to mimic the behavior of a collection of cell phones. As the device has a unique ID, the phone numbers integrated with this device can easily be detected and labeled in one group.

Virtual group. Based on the detected physical group, four versions of virtual groups are constructed by thresholding across four domains according to the attribute on the transaction network: 1) payment amount, 2) discount amount, 3) transaction count, 4) average time between transactions, respectively. Then these domain-specific groups are analyzed to extract the n-dimensional local features in each domain, which leads to a 4n-dimensional feature vector that reflects the structure of the local groups.

At the individual user level, we combine together the features of individuals, physical groups and virtual groups to form the

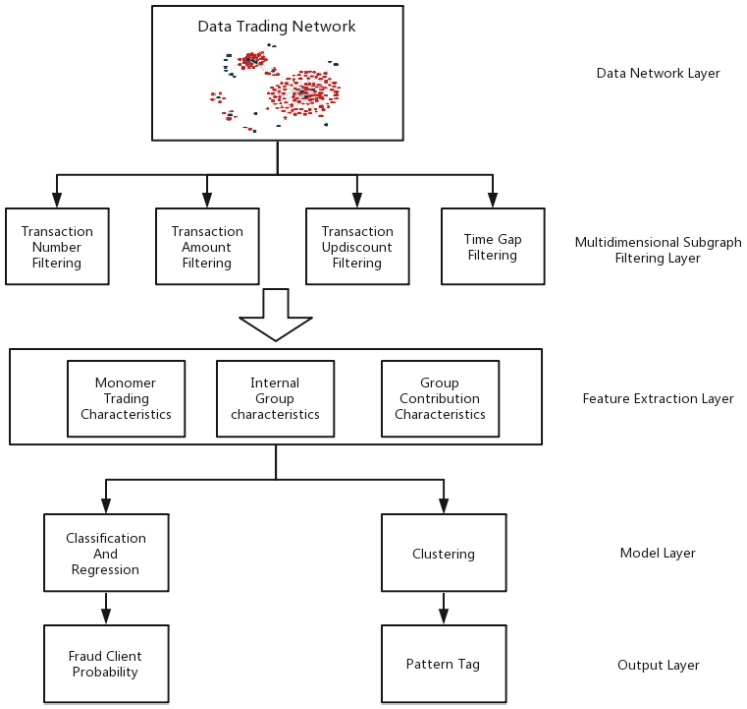

Figure 1 The overview of the individual level fraud detection pipeline. The pipeline adopts the raw transaction network as input and finally generates the fraud classification results and the probability for each client. Also, the clustered patterns for each individual client are also generated.

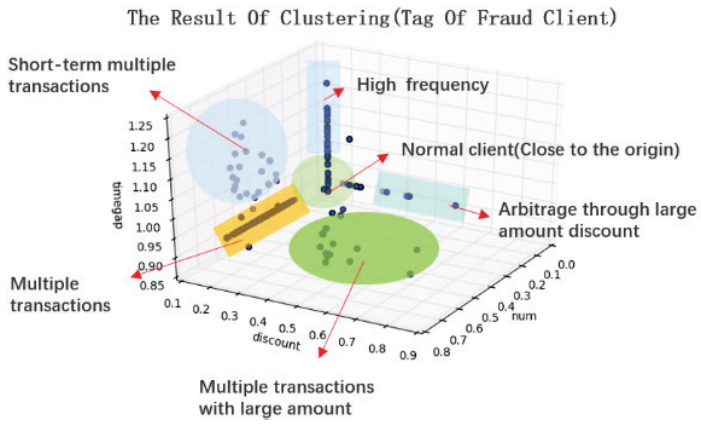

Figure 2 The clustered patterns from the individual clients. The three axes cover the time gap of transaction, the transaction fee discount, and the number of transactions in average over a certain period.

final input feature to the classifier for fraud detection. In particular, we use the GBDT model to score the potential individual attackers. It is worth noting that it is possible to use several graph matching techniques (in either a two-graph matching setting [14] or a multi-graph setting [15], [16] to fuse the information across the above networks), which will be addressed in our future work.

4. BUILDING LEARNING MODELS FOR MERCHANT FRAUD DETECTION

In this section, we discuss the machine-learning-based pipeline for fraud detection of merchants who have become the major channels facilitating fraud.

The main features considered include: 1) the merchant profile features such as the type of merchant (e.g. accommodation, entertainment, food service, retail, etc.), location of the merchant (usually at the regional level); 2) statistical features of merchant transactions, which can be categorized into main groups: general transaction statistics for the merchant including transaction count per hour (and median), and transaction amount, discount rate, the average time difference between two transactions. On the other hand, more fine-grained information from the payment user side is also collected and used, including the distribution of mobile App versions, the IP address distribution, the bar code scanning ratio, and payment type (online/offline), etc.

Using the aforementioned features, we apply a classification model. Specifically GBDT to estimate the daily status (i.e. normal or fraud) of each merchant. Then, the aggregated statistics for each merchant based on their daily status, for a certain period (e.g. a month), is computed to generate the risk score for each merchant over that period. Fig. 3 shows the relational overview of the merchant, user, and transaction

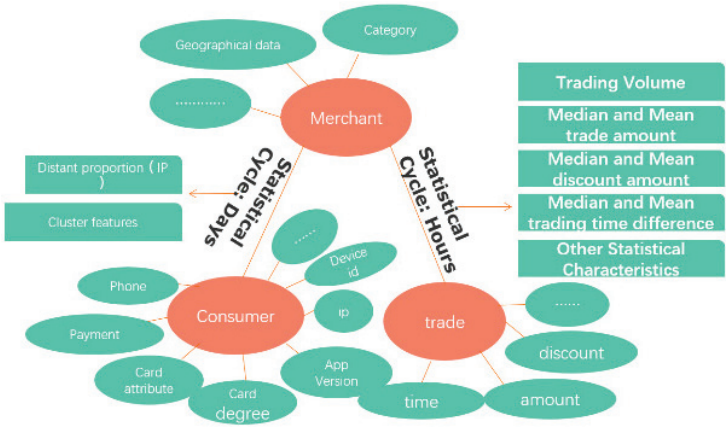

Figure 3 The relational overview of the merchant, user, and transaction as the three main entities analyzed in this paper. For each entity, different features can be extracted and used for further modeling and scoring

which are the three main entities involved in the payment process.

5. EXPERIMENTS AND CASE STUDY

The experiments are based on a case study using real-world payment data from the major payment channel in China. The users registered for this mobile payment channel now number over two billion, and we have extracted around three million transactions each day. The basic idea is to apply supervised learning methods to detect the high-risk transactions. Based on the result of the transaction level detection, individual clients and merchants suspected of fraud can further be discovered preparatory to taking further action.

One challenge, in reality, is that the collected transaction data has little profile information about the payment user. We also have little information about the Apps and the mobile phone hardware due to privacy. As a result, some simple and direct approaches cannot be applied. For instance, by analyzing the otherwise useful inertia sensors, malicious scripts that intentionally mimic the behavior of real users can be easily detected. However, this becomes very difficult given the privacy protection preventing the discolusre of physical information.

On the fraud detection side of the system, since massive numbers of transactions (in millions) accumulate each day which cannot be stored in a single machine, the distributed storage system is deployed for commercial use. Hence, Hadoop is used as the distributed file system which also involves the HIVE warehouse connector that transforms the data loading procedure as the MapReduce [17] task. The storage system is illustrated in Fig. 4.

5.1 Results on Individual User Fraud Detection

Note that Fig. 5(a) shows a typical malicious payment user who pays the same merchant over consecutive days. Moreover, the transaction amount and discount by the payment channel remain unchanged. Our model can easily detect this user. In another example, as shown in Fig.5(b), the individual user pays several times within three days and all the transactions involve only two merchants. This also suggests the possible risk of fraudulent payment around the person and these two merchants.

More quantitative results are given in Fig. 6, whereby the detection precision drops steadily as K increases for the top-K scoring suspected individuals. It can be easily seen that our model, which integrates both individual transaction profiles and group level features surpasses the performance of the baseline model that uses only the individual transaction profiles.

5.2 Results for Merchant Fraud Detection

The statistics of the used dataset for merchant study are shown in Table 1, and Table 2.

From Table 1, one can see that Fraudulent transactions do not occur very often but are more frequent during the promotion period. Table 2 shows that the consumption provinces of fraudulent merchants are more dispersed, while the geographical location of consumers is not concentrated.

For merchant fraud, we focus on detecting the merchants who intentionally conspire with individual clients in order to engage in fraud activities. The important features are discovered by GBDT classifier and are shown in Table 3.

We also compare our detection results with those produced by traditional handcrafted rules; the numbers are shown in Table 4. From the table, it can be seen that our computational method based on machine learning can identify more fraud merchants than can the traditional non-learning-based method (1907 vs. 1415). Meanwhile, they also have a considerable common part in common (1104 out of 1415 for handcrafted rule-based method).

5.3 Further Discussion

In this section, we conduct a more detailed comparison of merchant detection and individual fraud detection.

For merchant level detection, there is greater focus on the general behaviors of the payments. For instance, the concentration of the payment location (for offline) or IP address can indicate abnormal payment behavior. In particular, if many transactions are paid with an IP address in a region that is different from the merchant's location, it is highly likely that the

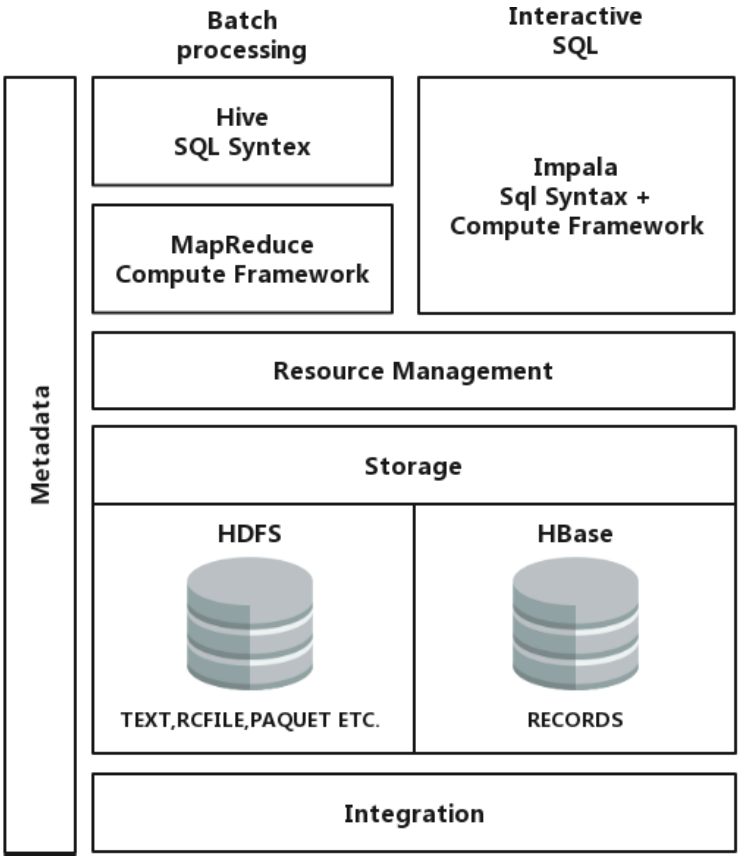

Figure 4 Overview of the adopted distributed file system and processing system

mchnt_cd	term_id	phone_no	trans_tm	trans_at	up_discount
QRC241000010088	00003528	182****8885	20181212095853	6000	3000
QRC241000010088	00003528	182****8885	20181213102817	6000	3000
QRC241000010088	00003528	182****8885	20181214102855	6000	3000
QRC241000010088	00003528	182****8885	20181216103006	6000	3000
QRC241000010088	00003528	182****8885	20181216110205	6000	3000

(a) Normal merchant trend.

phone_no	trans_tm	trans_at	up_discount	mchnt_nm
135******93	20181211232317	3000	3000	mchnt1
135******93	20181211232329	200	135	mchnt1
135******93	20181212064047	800	799	mchnt1
135******93	20181212064125	1000	500	mchnt1
135******93	20181212064137	3000	1000	mchnt1
135******93	20181212064158	200	100	mchnt1
135******93	20181212065259	3000	1000	mchnt1
135******93	20181212065323	500	237	mchnt1
135******93	20181212065333	200	112	mchnt1
135******93	20181212073158	20	19	mchnt1
135******93	20181212102157	3000	1500	mchnt1
135******93	20181212104323	3000	1500	mchnt2
135******93	20181212104405	3000	1500	mchnt2
135******93	20181212234604	3000	1500	mchnt2
135******93	20181212234657	3000	1500	mchnt2
135******93	20181213074142	1000	500	mchnt1
135******93	20181213080126	3000	1000	mchnt1
135******93	20181213080145	500	128	mchnt1

(b) Fraud merchant trend.

Figure 5 Two typical cases of individuals who make fraudulent payments to one or more merchants. Perhaps there is a conspiracy between the individual user and the merchants.

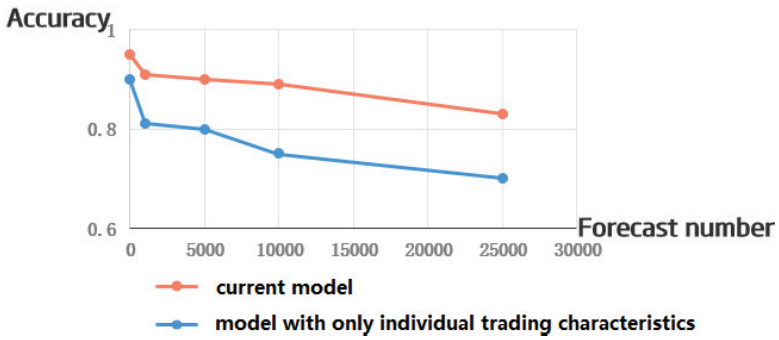

Figure 6 Performance comparison between handcrafted rule-based method and our proposed GBDT classifier-based supervised learning method for the individual fraud detection experiment.

Table 1 The ratio of payments with discount against the total transaction count comparison between the fraud merchants and normal merchants.

merchant	normal	fraud
merchant count	1148957	91814
discount ratio (count) whole period	43.1%	1.4%
discount ratio (amount) whole period	83.9%	29.4%
discount ratio (count) in promotion	50%	78%
discount ratio (amount) in promotion	1.75%	3.31%

Table 2 Payment occurrence location comparison between the fraud merchants and normal merchants. The mean client province count denotes the average province number from which a merchant's payment occurrence location.

merchant	normal	fraud
mean client province count	1.8	1.3

Table 3 A partial list of features and their importance by the GBDT classification model on the merchant fraud detection task.

feature	importance
IP address concentration	0.062
Apps version distribution	0.04
discount rate	0.03
median of the transaction amount	0.026
maximum period of consecutive transactions	0.025
total transaction amount	0.021
payment mode	0.015

Table 4 Fraud merchant numbers detected by our GBDT model and by handcrafted rules.

detected by GBDT model only	detected by rule only	detected by both
803	1104	311

merchant is involved in some fraud. From another perspective, the malicious merchants often have a low rate during 'normal' periods, but when promotions are launched, the number of transactions multiplies considerably.

On the other hand, for individual fraud detection, the analysis focuses more on the transaction information itself, usually in a series of payments, which are concentrated in one or a few merchants, and within a short time period. This phenomenon is clearly shown in Fig. 7.

6. CONCLUSION

We have presented a comprehensive description of the techniques and behind the motivation for devising the machine learning-based methods and tools for mobile payment fraud detection. Here the fraud scenario is relatively new as closely involved with the fast development of the mobile payment in China, whereby the attackers maliciously explore the potential arbitrage opportunities in the process of marketing promotion for the mobile payment tools. Supervised methods are developed to help.

In future work, we intend to improve the design of the machine learning component, as well as the whole system for effective and efficient fraud detection in an online real-time fashion. In particular, more volatile features will be explored and discovered via a trial-and-error procedure which can lead to a better understanding of fraudulent activities in the context of mobile payment. We also intend to transfer the learned models and relevant features to other companies and settings, in the

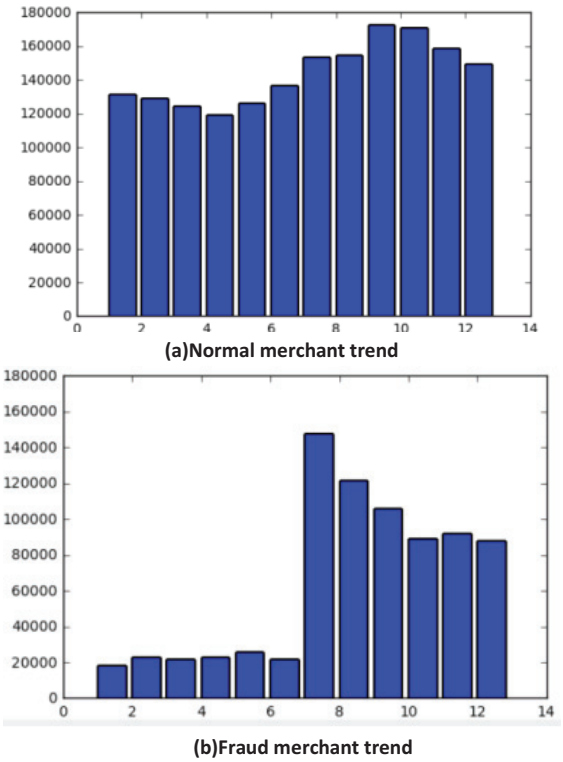

Figure 7 Comparison of the normal and fraud merchants in terms of their payment amount (in 10K RMB) over the 12 weeks before (the first 6 weeks) and during (the second 6 weeks) the launch of the promotion.

sense of transfer learning, which is an important task in machine learning research.

In addition, the framework based on temporal point process [18] will be a promising direction as a systematic way of modeling the dynamic behavior of users. Specifically, we have noted a series of new methods ranging from nonparametric learning of temporal point process [19], learning with missing attributes [20], to the deep network-based methods for point process, including both recurrent neural network methods [21], [22] and generative and adversarial learning methods [23], [24]. On the other hand, we may also adopt time-series-based methods in the discrete time space for behavior modeling and forecasting. In such a setting, interpretable deep time series methods [25] will be our first choice.

Last, we will pay more attention to the network embedding-based approaches [26]–[28] which are an expressive means of modeling the relational networks. Currently, we have not adopted these techniques mainly due to scalability issues for billions of users to model is hard to solve. In the long term, we will explore the scalable models and systems for network representation and embedding. It is hoped that by transforming the node with complex local and global structures into vectorized feature forms, traditional machine learning methods such as SVM, decision tree, logistic regression, etc. can be readily used in the Euclidean space. This also ensures the scalability of network analytics for large and dynamic networks. In particular, we are also interested in fusing two or multiple networks for more in-depth analysis across networks, although this also suffers from the scalability issue in the traditional graph matching setting [15], [16]. However, it has been noted that there are recent works [29] offering more scalability and flexibility.

ACKNOWLEDGMENT

The work is partially supported by NSFC (U19B2035, 61972250), National Key Research and Development Program of China (2018AAA0100704, 2016YFB1001003).

REFERENCES

1. M. Artís, M. Ayuso, and M. Guillén, "Detection of automobile insurance fraud with discrete choice models and misclassified claims", *Journal of Risk and Insurance*, vol. 69, no. 3, pp. 325–340, 2002.
2. D. Xing and M. Girolami, "Employing latent dirichlet allocation for fraud detection in telecommunications", *Pattern Recognition Letters*, vol. 28, no. 13, pp. 1727–1734, 2007.
3. A. C. Bahnsen, A. Stojanovic, D. Aouada, and B. Ottersten, "Improving credit card fraud detection with calibrated probabilities", in *Proceedings of the 2014 SIAM International Conference on Data Mining*. SIAM, 2014, pp. 677–685.
4. C. Xie, H. Cai, Y. Yang, L. Jiang, and P. Yang, "User profiling in elderly healthcare services in china: Scalper detection", *IEEE Journal of Biomedical and Health Informatics,* vol. 22, no. 6, pp. 1796– 1806, 2018.
5. J. H. Friedman, "Greedy function approximation: a gradient boosting machine", *Annals of Statistics*, pp. 1189–1232, 2001.
6. Y. Liu and X. Yao, "Ensemble learning via negative correlation", *Neural Networks*, vol. 12, no. 10, pp. 1399–1404, 1999.
7. S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.
8. G. Rätsch, T. Onoda, and K.-R. Müller, "Soft margins for adaboost", *Machine Learning,* vol. 42, no. 3, pp. 287–320, 2001.

9. C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines", *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.
10. Y.-J. Zheng, X.-H. Zhou, W.-G. Sheng, Y. Xue, and S.-Y. Chen, "Generative adversarial network-based telecom fraud detection at the receiving bank", *Neural Networks*, vol. 102, pp. 78–86, 2018.
11. A. Bekirev, V. Klimov, M. Kuzin, and B. Shchukin, "Payment card fraud detection using neural network committee and clustering", *Optical Memory and Neural Networks*, vol. 24, no. 3, pp. 193–200, 2015.
12. C. S. Hilas, P. A. Mastorocostas, and I. T. Rekanos, "Clustering of telecommunications user profiles for fraud detection and security enhancement in large corporate networks: a case study", *Applied Mathematics & Information Sciences*, vol. 9, no. 4, p. 1709, 2015.
13. J. Yan, Y. Wang, K. Zhou, J. Huang, C. Tian, H. Zha, and W. Dong, "Towards effective prioritizing water pipe replacement and rehabilitation", in *Twenty-Third International Joint Conferenceon Artificial Intelligence*, 2013.
14. J. Yan, C. Li, Y. Li, and G. Cao, "Adaptive discrete hypergraph matching", *IEEE Transactions on Cybernetics*, vol. 48, no. 2, pp. 765– 779, 2018.
15. J. Yan, J. Wang, H. Zha, X. Yang, and S. Chu, "Consistency-driven alternating optimization for multigraph matching: A unified approach", *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 994–1009, 2015.
16. J. Yan, M. Cho, H. Zha, X. Yang, and S. Chu, "Multi-graph matching via affinity optimization with graduated consistency regularization", *TPAMI*, 2016.
17. J. Dean and S. Ghemawat, "Mapreduce: simplified data processing on large clusters", *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.
18. D. Daley and V.-J. David, "An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure", *Springer Science & Business Media*, 2007.
19. E. Lewis and G. Mohler, "A nonparametric em algorithm for multiscale Hawkes processes" *Journal of Nonparametric Statistics*, 2011.
20. L. Li and H. Zha, "Dyadic event attribution in social networks with mixtures of Hawkes processes" in *CIKM*. ACM, 2013, pp. 1667–1672.
21. N. Du, H. Dai, R. Trivedi, U. Upadhyay, M. Gomez-Rodriguez, and L. Song, "Recurrent marked temporal point processes: Embedding event history to vectore," in *KDD*, 2016.
22. S. Xiao, J. Yan, X. Yang, H. Zha, and S. Chu, "Modeling the intensity function of point process via recurrent neural networks", in *AAAI,* 2017.
23. S. Xiao, M. Farajtabar, X. Ye, J. Yan, L. Song, and H. Zha, "Wasserstein learning of deep generative point process models", in *NIPS*, 2017.
24. S. Xiao, H. Xu, J. Yan, M. Farajtabar, X. Yang, L. Song, and H. Zha, "Learning conditional generative models for temporal point processes", in *AAAI*, 2018.
25. L. Li, J. Yan, X. Yang, and Y. Jin, "Learning interpretable deep state space model for probabilistic time series forecasting", in *IJCAI*, 2019.
26. B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2014, pp. 701–710.
27. J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "Line: Large-scale information network embedding", in *WWW*, 2015.
28. A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks", in *SIGKDD*, 2016.
29. T. Yu, J. Yan, W. Liu, and B. Li, "Incremental multi-graph matching via diversity and randomness-based graph clustering", in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 139–154.