POSTER: Actively Detecting Implicit Fraudulent Transactions

Shaosheng Cao
AI Department (Hangzhou)
Ant Financial Services Group
shaosheng.css@antfin.com

XinXing Yang
AI Department (Beijing)
Ant Financial Services Group
xinxing.yangxx@antfin.com

Jun Zhou
AI Department (Beijing)
Ant Financial Services Group
jun.zhoujun@antfin.com

Xiaolong Li
AI Department (Seattle)
Ant Financial Services Group
xl.li@antfin.com

Yuan (Alan) Qi
AI Department (Hangzhou)
Ant Financial Services Group
yuan.qi@antfin.com

Kai Xiao
Security Department (Shanghai)
Ant Financial Services Group
xiaokai.xk@antfin.com

ABSTRACT

In this work, we propose to actively detect implicit fraudulent transactions. A novel machine learning method is introduced to distinguish anomalous electronic transactions based on the historical records. The transferor will be alerted during the on-going payment when the fraud probability is recognized as large enough. Compared with elaborative rule-based approaches, our model is much more effective in fraud detection.

CCS CONCEPTS

• **Security and privacy** → Intrusion/anomaly detection and malware mitigation; • **Computing methodologies** → Machine learning;

KEYWORDS

fraudulent transaction detection; transaction network; machine learning

ACM Reference Format:
Shaosheng Cao, XinXing Yang, Jun Zhou, Xiaolong Li, Yuan (Alan) Qi, and Kai Xiao. 2017. POSTER: Actively Detecting Implicit Fraudulent Transactions. In *Proceedings of CCS '17*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3133956.3138822

1 INTRODUCTION

Alipay, currently known as Ant Financial Services Group, is the largest mobile and online platform and money-market fund all over the world. In terms of the statistics information at the middle of 2016, there are more than 100 million daily active users and about 450 million annual active users[1]. However, there exist more than ten thousands fraudulent transactions per day that cause great losses. It is therefore a core component for the payment security of monitoring suspicious transactions. In general, fraudulent transactions can be summarized into two different situations, i.e., judgement afterwards and active detection on the fly.

In the first scenario, if a user has been aware of being cheated and has accused to us, it is necessary to judge whether he (or she) is in a fraud or not, in terms of the supporting proofs, transaction details and the profile. The above is an **explicit** fraud judgement after the incident has happened. In a secondary and more important situation, we desire to actively detect such an **implicit** fraudulent transaction before a user finishes transfering money into suspicious persons.

We emphasize on active identification of the potentially implicit fraudulent fraud at risk control strategy, in order to alert the transferor as soon as the anomalous transaction is identified. Based on our observations, we found that 1) fraudulent manners are changed rapidly with new patterns and attacks, and 2) the number of fraudulent transactions is much less than normal ones. To address the problems, we propose a novel machine learning method, which is automatically adaptive with constantly changing means of frauds as time goes by.

2 RELATED WORK

Fraudulent transaction detection has been widely investigated in the literature, e.g., credit card fraud, telecommunication fraud, and etc. [3, 16]. Rule-based approaches are introduced to produce assertion statement of `IF` {conditions} and `THEN` {a consequent} by [9, 15]. Brause et al. propose to generalize association rules by comparing fraud and normal records [5], and [2] shows a way that generates decision variables to identify potentially fraudulent calls. Supervised learning methods are presented in many literature, which yield a fraud probability for the judgement of a new record. Linear discriminative models are employed in [11], and neural networks are utilized later [1, 10, 14].

As for extremely unbalanced data, several unsupervised methods have been applied. Nigrini shows the effectiveness of the Benford's law in accounting fraud [13], and Bolton et al. describe unsupervised profiling methods for the fraud detection of credit cards [4]. Aggregation strategies and clustering methods are used, instead of the analysis targeting on a single transaction record. Vadoodparast et al. combine three traditional clustering methods to achieve better performance [17], and Casas et al. leverage k-means to group network security data, whose centroids and labels are fed into a classifier [7]. Detailed aggregation strategies are shown for detecting credit card fraud in [12, 18].

[1] https://en.wikipedia.org/wiki/Ant_Financial#cite_note-toknow-8

CCS '17, October 30-November 3, 2017, Dallas, TX, USA

ACM ISBN 978-1-4503-4946-8/17/10…$15.00
https://doi.org/10.1145/3133956.3138822

3 OUR MODEL

In this section, we will give the details of the proposed model.

3.1 The Main components

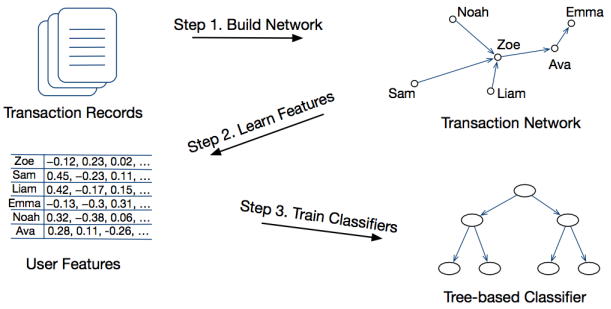


Figure 1: The trainer module of the proposed model.

Our model consists of two components, including trainer and predictor modules. As illustrated in Figure 1, we first build a transaction network from the historical records, and then learn user features using unsupervised learning. Finally, tree-based classifiers are trained by labeled transactions. Once user features and classifiers are ready, the predictor module of our model is able to yield a fraud probability score to alert the transferor if necessary, as described in Figure 2.

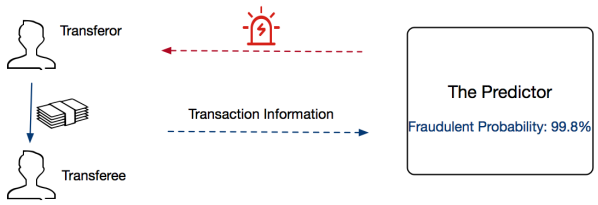


Figure 2: The predictor module of the proposed model.

The key of the task is to learn the topological feature information of each user node in the transaction network and train robust classifiers from unbalanced categories.

3.2 Learning User Features from Transaction Network

Since the fraud patterns vary over time, we aim to extract the useful features automatically. An electronic transaction involves two different roles of users, i.e, transferor and transferee, where a user is treated as a node and an edge exists if there is a transaction between them.

Let us define $G =< V, E >$ as the transaction network, where V is the collection of the nodes and E is the collection of the edges. Given the network G, our goal is to learn a representational matrix D ($D \in |V| \times d$), where the d-dimensional vector of the i-th row of learned matrix denotes the topological feature information of the i-th user node.

Intuitively, the most intimate nodes of a node are its (1-hop) neighbours, and next initimate ones are 2-hop neighbours, and so on. Motivated by [6], we define the loss function by measuring the topological similarity between a node and its neighbours:

$$\mathcal{L}_1 = -\sum_{c \in \tau(w)} \left(\log \sigma(\vec{w} \cdot \vec{c}) + \sum_{i=1}^{\lambda} \mathbb{E}_{c' \sim U} [\log \sigma(-\vec{w} \cdot \vec{c'})] \right) \quad (1)$$

where $\mathcal{L}_1$ is the loss function, and $\tau(w)$ is the collection of neighbours of w within a fixed number of hops. $\mathbb{E}_{c' \sim U}[\cdot]$ denotes the expectation, where c' follows the node distribution U. c' is a negative sample that does not occur in the neighbours but is randomly selected from the whole node collection V, and λ represents the number of negative samples. $\vec{w}$ and $\vec{c}$ are the low dimensional representational vectors of the node w and its neighbour node c. Besides, σ is sigmoid function.

$\vec{w}$ is randomly initialized at first and updated by the term of partial derivative of the loss:

$$\vec{w}_{t+1} = \vec{w}_t - \alpha \cdot \frac{\partial \mathcal{L}_1}{\partial w} \quad (2)$$

where α is a hyper-parameter by means of learning rate. After enough iterations of updates, we get the final d-dimensional feature vector of node w. The procedure is not influenced by unbalanced labels at all, since only transaction records are needed.

3.3 Training Classifiers from Transaction Labels

We define $\mathcal{D} = \{(x_i, y_i)\}$ as the collection of labeled dataset, where x_i is the feature vector of the i-th instance and y_i is the label. The features of an instance are made up of the learned features of the involved transferor and transferee, as well as the basic information in the transaction situation. $y_i = 1$ when it is a fraudulent case; otherwise, $y_i = 0$. $\hat{y}_i$ denotes the predictive fraud score of the i-th instance by our model, and $l(y_i, \hat{y}_i)$ is a differentiable convex function of decision tree between y_i and $\hat{y}_i$. Besides, a regularization term $\Omega(\cdot)$ is also added. Inspired by [8], we show the loss as follows:

$$\mathcal{L}_2 = \sum_{i=1}^{|\mathcal{D}|} \left(g_i f(x_i) + \frac{1}{2} h_i f^2(x_i) \right) + \Omega(f) \quad (3)$$

where $\mathcal{L}_2$ is the loss function, g_i and h_i is the first order and second order partial derivative of $l(y_i, \hat{y}_i)$:

$$g_i = \frac{\partial l(y_i, \hat{y}_i)}{\partial \hat{y}_i} \qquad h_i = \frac{\partial^2 l(y_i, \hat{y}_i)}{\partial \hat{y}_i \partial \hat{y}_i} \quad (4)$$

Although linear classifiers like logistic regression are also widely applied in supervised learning, we choose the above gradient boosting based models instead as for its high accuracy in the case. When the training module is finished, the predictor module has the ability of active detection.

4 EXPERIMENTS

In this section, we will show the effectiveness of the proposed model versus rule-based approach in the real electronic transactions.

4.1 Benchmark, Baseline and Evaluation Metrics

We collect the transaction records from December 1, 2016 to February 20, 2017 in Alipay, from which about 57 million records are

Table 1: The performance comparison between baseline and our model.

Methods	F1	KS	AUC	REC@100	REC@500	REC@1000
Baseline	61.09%	86.18%	98.23%	73.04%	51.77%	41.93%
Our Model	**65.22%**	**88.75%**	**98.79%**	**78.00%**	**57.48%**	**48.26%**

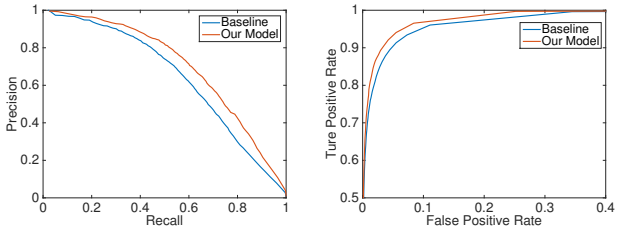
sampled, so as to build the transaction network and learn user features. We also randomly select 2 million records from February 24, 2017 to April 9, 2017 for training the classifier, and adopt 0.8 million records from April 10, 2017 to April 20, 2017 as the test dataset.

In order to test the performance of our model, we compare the experimental results with the rule-based baseline. Inspired by several guidelines[2], dozens of the rules are summarized from the current transfer environments. For example, if the IP address or telephone are from a same city, if the transferees has been complained in the past and so on.

To make a fair comparison, we evaluate using different evaluation metrics. Receiver Operating Characteristic (ROC) curve reflects the diagnostic capacity of a binary classifier, which is decided by drawing true positive rate against the false positive rate, while Area Under a Curve (AUC), as described by its name, is the value of the area under (the ROC) curve[3]. Another important metric is Precision-Recall (PR) plotting curve as the discrimination threshold varies, and the F1 Score is interpreted as harmonic mean of precision and recall[4]. Besides, Kolmogorov-Smirnov (KS) test is a nonparametric test of probability distribution between predictive results and golden standard[5].

4.2 Empirical Results

As described in Table 1, our proposed method can consistently outperform baseline over different testing metrics. In practice, we pay more attention on recall at k predictive samples. Specifically, the value of "REC@100" equals 73.04% means recall value is 73.04% if we alert only 1 time in 100 transaction records. So the higher value of "REC@k" is, the more accurate it is. In addition, both AUC values are close to 100%, as for extremely few fraudulent transactions in statistics.



(a) PR Curve (b) ROC Curve

Figure 3: PR Curve and ROC Curve

Figure 3 shows the PR curve and the ROC curve of our model against baseline. Obviously, in the PR curve, the performance of our model outperforms baseline consistently. For the other comparison of the ROC curve, it is easy to observe that the true positive rate is 93.56% of our model when false positive rate is 5%. By contrast, the value of baseline is only 90.58% with same false positive rate.

[2]https://www.bluefin.com/merchant-support/identifying-fraudulent-transactions
[3]https://en.wikipedia.org/wiki/Receiver_operating_characteristic#Area_under_the_curve
[4]https://en.wikipedia.org/wiki/F1_score
[5]https://en.wikipedia.org/wiki/Kolmogorov-Smirnov_test

5 CONCLUSION

We propose a novel method for actively detecting implicit fraudulent transactions. From the empirical results, the proposed model significantly outperforms rule-based baseline. In our future work, we will investigate more possible solutions to reduce the fraud cases further.

ACKNOWLEDGEMENT

The authors thank the anonymous reviewers for their valuable suggestions.

REFERENCES

[1] Emin Aleskerov, Bernd Freisleben, and Bharat Rao. 1997. Cardwatch: A neural network based database mining system for credit card fraud detection. In *CIFEr*. IEEE, 220–226.
[2] Gerald Donald Baulier, Michael H Cahill, Virginia Kay Ferrara, and Diane Lambert. 2000. Automated fraud management in transaction-based networks. (Dec. 19 2000). US Patent 6,163,604.
[3] Richard J Bolton and David J Hand. 2002. Statistical fraud detection: A review. *Statistical science* (2002), 235–249.
[4] Richard J Bolton, David J Hand, et al. 2001. Unsupervised profiling methods for fraud detection. *Credit Scoring and Credit Control VII* (2001), 235–255.
[5] R Brause, T Langsdorf, and Michael Hepp. 1999. Neural data mining for credit card fraud detection. In *ICTAI*. IEEE, 103–106.
[6] Shaosheng Cao, Wei Lu, and Qiongkai Xu. 2015. Grarep: Learning graph representations with global structural information. In *CIKM*. ACM, 891–900.
[7] Pedro Casas, Alessandro D'Alconzo, Giuseppe Settanni, Pierdomenico Fiadino, and Florian Skopik. 2016. POSTER:(Semi)-Supervised Machine Learning Approaches for Network Security in High-Dimensional Network Data. In *CCS*. ACM, 1805–1807.
[8] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *SIGKDD*. ACM, 785–794.
[9] William W Cohen. 1995. Fast effective rule induction. In *ICML*. 115–123.
[10] Sushmito Ghosh and Douglas L Reilly. 1994. Credit card fraud detection with a neural-network. In *System Sciences*, Vol. 3. IEEE, 621–630.
[11] David J Hand. 1981. Discrimination and classification. *Wiley Series in Probability and Mathematical Statistics, Chichester: Wiley, 1981* (1981).
[12] Sanjeev Jha, Montserrat Guillen, and J Christopher Westland. 2012. Employing transaction aggregation strategy to detect credit card fraud. *Expert systems with applications* 39, 16 (2012), 12650–12657.
[13] Mark J Nigrini. 1999. I've got your number. *Journal of accountancy* 187, 5 (1999), 79.
[14] Raghavendra Patidar, Lokesh Sharma, et al. 2011. Credit card fraud detection using neural network. *IJSCE* 1, 32-38 (2011).
[15] J Ross Quinlan. 1990. Learning logical definitions from relations. *Machine learning* 5, 3 (1990), 239–266.
[16] Donald Tetro, Edward Lipton, and Andrew Sackheim. 2000. System and method for enhanced fraud detection in automated electronic credit card processing. (Aug. 1 2000). US Patent 6,095,413.
[17] Massoud Vadoodparast, Abdul Razak Hamdan, et al. 2015. Fraudulent Electronic Transaction Detection Using Dynamic KDA Model. *IJCSIS* 13, 3 (2015), 90.
[18] Christopher Whitrow, David J Hand, Piotr Juszczak, D Weston, and Niall M Adams. 2009. Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge Discovery* 18, 1 (2009), 30–55.