FRAUD DETECTION IN ONLINE PAYMENT SYSTEMS USING CO-OCCURRENCE PATTERNS

#1 J KUMARI, #2 M SIVA KRISHNA
#1 ASSISTANT PROFESSOR
#2 MCA SCHOLAR
DEPARTMENT OF MASTER OF COMPUTER APPLICATIONS,
QIS COLLEGE OF ENGINEERING AND TECHNOLOGY
VENGAMUKKAPALEM (V), ONGOLE, PRAKASAM DIST.,ANDHRA PRADESH- 523272.

ABSTRACT

The robust expansion of e-commerce fosters cybercrime. The detection of online payment fraud, a difficulty encountered by online services, is crucial in the fast expanding e-commerce landscape. Behavioral approaches are acknowledged as a viable approach for detecting online payment fraud. Nonetheless, constructing high-resolution behavioral models with low-quality behavioral data presents a significant problem. This study primarily tackles the issue through data augmentation for behavioral modeling. We derive detailed co-occurrence correlations of transactional attributes through the utilization of a knowledge graph. Additionally, we employ heterogeneous network embedding to enhance the representation of complex relationships. We specifically investigate tailored network embedding methodologies for various behavioral models, including population-level models, individual-level models, and generalized agent-based models. The efficacy of our strategy is substantiated by experiments conducted on a genuine dataset from a commercial bank. It can substantially enhance the efficacy of representative behavioral models in detecting online banking payment fraud. This is, to our knowledge, the first study to achieve data improvement for diverse behavioral models by the application of network embedding methods on attribute-level co-occurrence relationships.

1. INTRODUCTION

Online payment services have infiltrated individuals' lives. The improved convenience, however, is accompanied with intrinsic security issues Cybercrime related to online payment services typically exhibits features of diversification, specialization, industrialization, concealment, scenario-based tactics, and cross-regional operations, rendering the security prevention and control of online payments exceedingly difficult. An urgent necessity exists for the implementation of robust and comprehensive online payment fraud detection. The behavior-based approach is acknowledged as an effective framework for detecting online payment fraud The advantages can be succinctly expressed as follows:Initially,behavior-based techniques employ a nonintrusive detection scheme to ensure user experience without requiring user intervention during implementation. Secondly, it transforms the fraud detection paradigm from a singular occurrence to a continuous process, enabling verification of each transaction. Thirdly, although the fraudster may replicate the victim's everyday operational patterns, they must diverge from the user's behavior to exploit the victim's advantages. Deviation can be

identified by behavior-based strategies. This behavior-based approach can serve as a supplementary security measure, rather than supplanting other detection methods. The efficacy of behavior-based techniques is largely contingent upon the adequacy of user behavioral data User behavioral data utilized for online payment fraud detection is frequently of low quality or limited due to challenges in data gathering and user privacy regulations The primary problem is to construct a high-performance behavioral model with low-quality behavioral data. This complex issue can be effectively addressed using two approaches: data augmentation and model refinement. To strengthen behavioral models, a well-established method involves constructing models from many perspectives and integrating them effectively. One classification of models is predicated on the behavioral agent, as it is a pivotal element of behavioral models. Behavioral models can be categorized into individual-level models and population-level models based on agent granularity. This study emphasizes behavioral data enhancement. A fundamental principle is to thoroughly investigate the relationships inherent in the transaction data. More nuanced correlations may yield enhanced semantic information for the development of high-performance behavioral models. Current research in data augmentation for behavioral modeling primarily on the extraction and modeling of correlations, including co-occurrences, between behavioral attributes and labels To boost data augmentation, it is prudent to explore and leverage the more nuanced correlations within behavioral data, such as those within behavioral variables. Our primary contribution is the effective modeling of co-occurrences of transactional characteristics to enhance the performance of behavioral models. To do this, we propose utilizing the heterogeneous relation network, a specific variant of the knowledge graph to effectively describe co-occurrences. A network node represents an attribute value in transactions, while an edge signifies a heterogeneous link between distinct attribute values. While the relation network can represent the data more effectively, it ultimately fails to address the issue of data imperfection in behavioral modeling, as it cannot improve the inherent low-quality data. A proficient data representation that maintains these intricate relationships can serve as a crucial means of enhancing relational data. Consequently, we present network representation learning (NRL), which adeptly captures profound linkages [16]. Robust associations compensate for inferior data quality in fraud detection and enhance the efficacy of fraud detection algorithms. Calculating the similarity between embedding vectors may reveal more potential linkages. It somewhat addresses the issue of data inaccuracy. Besides data augmentation, NRL revolutionizes traditional network analysis by shifting from artificially defined features to automatically learnt features, thereby extracting profound linkages from extensive transactions. The ultimate efficacy of behavioral modeling for online fraud detection is contingent upon the synergistic collaboration of data augmentation and model refinement. Diverse behavioral models require corresponding network embedding strategies to attain optimal performance. This is a substantial technical challenge in our endeavors. We intend to examine suitable network embedding

methodologies for population-level models, individual-level models, and models incorporating various generalized behavioral agents. Specifically, for population-level models, we develop a label-free heterogeneous network to reconstruct online transactions, subsequently inputting the features generated in the embedding space into advanced machine learning classifiers to predict fraud risks. Conversely, for individual-level models, we utilize a label-aware heterogeneous network that differentiates the relationships between attributes of fraudulent transactions, and further devise multiple naive individual-level models that align with the representations produced by the label-aware network. Moreover, we integrate the population-level and individual-level models to achieve complimentary impacts by mitigating each other's limitations. The key contributions can be encapsulated as follows: We offer an innovative and efficient data improvement strategy for behavioral modeling by representing and analyzing more granular attribute-level co-occurrences. We utilize heterogeneous relation networks to depict attribute-level co-occurrences and extract these links through in-depth heterogeneous network embedding methods. We create a cohesive interface between network embedding methods and behavioral models by tailoring the conserved relationship networks based on the classification of behavioral models. • We apply the offered methodologies to a practical online banking payment service context. Our approaches have been validated to significantly surpass state-of-the-art classifiers based on a range of representative parameters in online fraud detection.

2. RELATED WORKS

1. **Jurgovsky et al. (2018)** explored Recurrent Neural Network (RNN) models for fraud detection in online transactions, demonstrating how sequential behavior patterns can help in identifying fraudulent activities, laying the foundation for behavior-based approaches.
2. **Zhang et al. (2020)** proposed a fine-grained behavior modeling technique that captures co-occurrence patterns among transaction attributes (such as time, location, device type) to improve fraud detection accuracy in payment services.
3. **Wang et al. (2021)** introduced a graph-based representation method for user behavior in online payment platforms, where the co-occurrence of transaction attributes was modeled as graph edges, significantly enhancing fraud detection capabilities.
4. **Liu et al. (2020)** emphasized the importance of capturing subtle behavioral co-occurrences that fraudsters may exploit, proposing a fine-grained temporal-spatial analysis method to improve early fraud detection.
5. **Zhou et al. (2021)** implemented a deep learning-based framework incorporating feature co-occurrences and user behavior profiles, demonstrating significant improvements in detecting complex, evolving fraud patterns in digital payment services.
6. **Shao and Yang (2020)** discussed how fine-grained behavior analytics, when combined with machine learning models, can reveal hidden relationships between transaction attributes, making it

harder for fraudsters to mimic legitimate user patterns.

7. Recent studies indicate that modeling fine-grained co-occurrences in user behavior provides a powerful layer of defense for online payment services, outperforming traditional rule-based or isolated attribute detection systems.

3. SYSTEM ANALYSIS

EXISTING SYSTEM

Bahnsen et al. improve the detection performance by calibrating probabilities before establishing Bayes model. HMM model is used to model the customers' credit card shopping patterns for detection of credit card fraud. The shopping items indicate the hidden state and the corresponding prices from certain ranges are the observation. LR(Logistic Regression), Support Vector Machines(SVMs) and Random Forest(RF) are evaluated for credit card detection. The detection models are built on primary features and derived features from transaction.

Whitrow et al. proposed a new preprocessing strategy for better fraud detection with SVMs and KNN classification. Transactions aggregated in term of time window, then data with new features is used to model the pattern.

Wei etal.addressed the problem of unbalanced financial data and employed cost-sensitive neural network to punish the misclassification of fraud transaction. Sahinetal.incorporate cost function into decision tree to boost performance on unbalanced data. Following the general procedure of classification, feature selection is proceed to boost the detection performance of credit card fraud.

Perols [35] performed a systematic analysis of financial fraud detection with popular statistical and machine learning models. The evaluation is under the supervised manner. All these methods rely on accurate identification of fraud patterns from data set and these methods also suffer from the problem of unbalanced data. Bolton and David perform fraud detection with clustering methods. This unsupervised manner is under the assumption that small cluster indicates the anomaly in data.

CoDetect is an unsupervised model which is based on matrices co factorization. The matrices from graph represent the genuine proprieties (features and connections) of financial data. The detection results give a better understanding of fraud patterns and furthermore, help to trace the originate of fraud groups.

Disadvantages

1. There is no Evaluation with Subspace Clustering Methods.

2. There is no SVM Classification in Credit Card Fraud Detections.

PROPOSED SYSTEM

1. In the proposed system, the system would like to develop a novel framework for fraud detection by considering the special detecting and tracing demanding of fraud entities and behaviors. Specifically, we investigate: (1) how to utilize both graph matrix and feature matrix for fraud detection and

fraud tracing; (2) how to mathematically model both graph matrix and feature matrix so as to simultaneously achieve the tasks of fraud detection and tracing. In an attempt to solve these challenges.

2. The system proposed a novel detection framework Co Detect for financial data, especially for money laundering data. The system incorporates fraud entities detection and anomaly feature detection in the same framework to find fraud patterns and correspondingfeaturessimultaneously.Combining entities detection and feature detection enables us to build a novel fraud detection framework for noisy and sparse financial data: relevant fraud patterns help the identification of fraud identities, and relevant features in turn help revealing of the nature of fraud activities.

Advantages

- Provide an approach to establish weighted graph from financial network, incorporating properties of nodes and links.
- Demonstrate different scenarios of financial fraud and formulate the patterns of fraud in term of graph and sparse matrix.
- Propose a novel unsupervised framework, CoDetect, for the problem of complex patterns discovery and anomaly features identification, employing two matrices residual analysis on graph-based financial network.

4. IMPLEMENTATION

Modules:

Bank Admin

In this module, the Admin has to login by using valid user name and password. After login successful he can do some operations such as Bank Admin's Profile ,View Users and Authorize ,View Ecommerce Website Users and Authorize, Add Bank ,View Bank Details ,View Credit Card Requests, View all Products with rank ,View all Financial Frauds ,View all Financial Frauds with Random Forest Tree With wrong CVV ,View all Financial Frauds with Random Forest Tree with Expired Date Usage ,List Of all Users with Majority of Financial Fraud ,Show Product Rank In Chart ,Show Majority Voting With Wrong CVV Fraud in chart,Show Majority Voting with Expiry date Usage in chart.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

View Chart Results

Show Product Rank In Chart, Show Majority Voting With Wrong CVV Fraud in chart, Show Majority Voting with Expiry date Usage in chart.

Ecommerce User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the

database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like, Add Category, Add Products, View all Products with rank, and View all Purchased Products with total bill, View All Financial Frauds.

End User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like, View My Profile, Manage Bank Account, Request Credit Card, View Credit Card Details, Transfer Money to Your Credit Card Account, Search for Products by Keyword, View all Purchased Products with Total Bill.

Methodology:

The proposed methodology focuses on modeling fine-grained co-occurrence patterns among transaction attributes and user behavior to improve the accuracy and robustness of fraud detection in online payment platforms.

1. Data Collection and Preprocessing

- Collect real-time transactional data from online payment services, including:
  - Transaction amount, timestamp, device type, location, IP address, payment method, and user profile.
- Clean and preprocess the dataset by:
  - Handling missing values
  - Normalizing continuous attributes
  - Encoding categorical variables
- Label historical transactions as legitimate or fraudulent based on ground truth.

2. Fine-Grained Feature Extraction

- Extract key behavioral features such as:
- Time-based features (e.g., transaction time intervals, frequency patterns)
- Location-based features (e.g., IP geolocation consistency)
- Device-based features (e.g., browser fingerprint, device type)
- Identify co-occurrence patterns among these features that indicate suspicious behavior.
- Represent the co-occurrences as structured feature vectors, graphs, or matrices.

3. Co-Occurrence Representation Modeling

- Model fine-grained co-occurrences using one or more of the following techniques:
- **Graph-based models:** Represent users and transaction attributes as nodes, with edges capturing co-occurrence relationships.
- **Matrix-based models:** Use co-occurrence matrices to quantify the frequency and correlation of attribute pairs.
- **Sequence-based models:** Capture temporal co-occurrence using sequences or event streams.

4. Behavior-Based Fraud Detection Model Development

- Train machine learning or deep learning models to learn from the co-occurrence representations, such as:
  - Random Forest or Gradient Boosted Trees for structured features
  - Graph Neural Networks (GNN) for graph-based representations
  - Recurrent Neural Networks (RNN) or Long Short-Term Memory (LSTM) for sequence-based data
- Optimize model hyperparameters to balance detection accuracy and false positive rates.

5.Real-TimeFraudDetection Framework

- Integrate the trained model into a real-time fraud detection pipeline.
- For each new transaction:
- Extract fine-grained features and co-occurrence patterns
- Generatecorresponding representations
- Predict the likelihood of fraud based on learned behavior patterns
- Flag suspicious transactions for further investigation or automated blocking.

6. Evaluation and Performance Analysis

- Evaluate the system using standard metrics such as:
- Accuracy, Precision, Recall,
- Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC)
- Compare performance against baseline models that do not incorporate co-occurrence patterns.
- Conduct stress tests to evaluate system performance under high transaction volumes.

5. RESULTS AND DISCUSSION

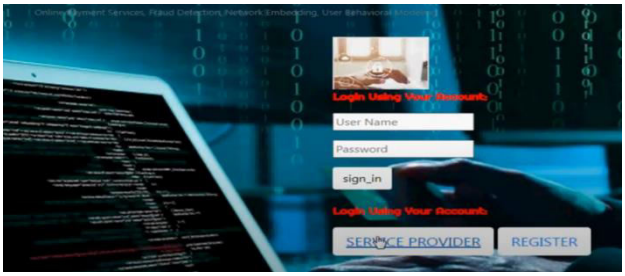

Fig 1

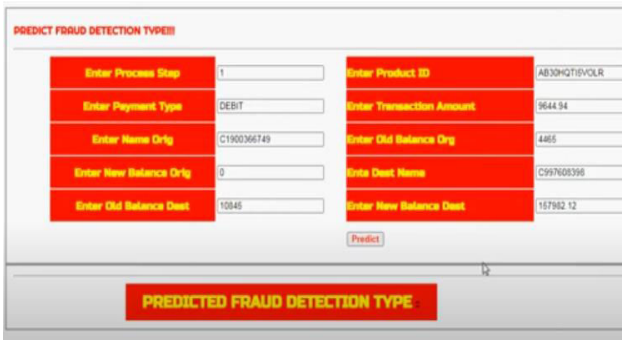

Fig 2

FUTURE SCOPE AND CONCLUSION

We offer an effective data enhancement approach for behavioral models in online payment fraud detection by modeling the co-occurrence relationships of transactional characteristics. We develop tailored co-occurrence relation networks and employ heterogeneous network embedding techniques to represent online transaction data for various behavioral models, such as individual-level and population-level models. The methodologies are corroborated through application on a real-world dataset. They

surpass the leading classifiers through efficient feature engineering techniques. Consequently, our methods can likewise function as a viable model for autonomous feature engineering. Several intriguing topics remain for exploration: Future research could involve expanding the data augmentation strategy to encompass additional behavioral models, such as group-level models and generalized-agent-based models, beyond the population-level and individual-level models examined in this study. It would be intriguing to examine the specialized enhancement strategies for more sophisticated individual-level models, as the employed naive individual-level model fails to fully leverage the benefits of the proposed data representation scheme utilizing heterogeneous network embedding techniques. It is expected to illustrate the applicability of the proposed method by utilizing it in various real-world application scenarios.

REFERENCES

[1] B. Cao, M. Mao, S. Viidu, and P. S. Yu, “Hitfraud: A broad learning approach for collective fraud detection in heterogeneous information networks,” in Proc. IEEE ICDM 2017, New Orleans, LA, USA, November 18-21, 2017, pp. 769–774.

[2] M. A. Ali, B. Arief, M. Emms, and A. P. A. van Moorsel, “Does the online card payment landscape unwittingly facilitate fraud?” IEEE Security & Privacy, vol. 15, no. 2, pp. 78–86, 2017.

[3] X. Ruan, Z. Wu, H. Wang, and S. Jajodia, “Profiling online social behaviors for compromised account detection,” IEEE Trans. Information Forensics and Security, vol. 11, no. 1, pp. 176–187, 2016.

[4] H. Yin, Z. Hu, X. Zhou, H. Wang, K. Zheng, N. Q. V. Hung, and S. W. Sadiq, “Discovering interpretable geo-social communities for user behavior prediction,” in Proc. IEEE ICDE 2016, Helsinki, Finland, May 16-20, 2016, pp. 942–953.

[5] Y.-A. De Montjoye, L. Radaelli, V. K. Singh et al., “Unique in the shopping mall: On the reidentifiability of credit card metadata,” Science, vol. 347, no. 6221, pp. 536–539, 2015.

[6] A. Khodadadi, S. A. Hosseini, E. Tavakoli, and H. R. Rabiee, “Continuous-time user modeling in presence of badges: A probabilistic approach,” ACM Trans. Knowledge Discovery from Data, vol. 12, no. 3, pp. 37:1–37:30, 2018.

[7] F. M. Naini, J. Unnikrishnan, P. Thiran, and M. Vetterli, “Where you are is who you are: User identification by matching statistics,” IEEE Trans. Information Forensics and Security, vol. 11, no. 2, pp. 358–372, 2016.

[8] M. Egele, G. Stringhini, C. Kruegel, and G. Vigna, “Towards detecting compromised accounts on social networks,” IEEE Trans. Dependable and Secure Computing, vol. 14, no. 4, pp. 447–460, 2017.

[9] A. Alzubaidi and J. Kalita, “Authentication of smartphone users using behavioral biometrics,” IEEE Communications Surveys and Tutorials, vol. 18, no. 3, pp. 1998–2026, 2016.

[10] H. Mazzawi, G. Dalaly, D. Rozenblatz, L. Ein-Dor, M. Ninio, and O. Lavi, “Anomaly detection in large databases using behavioral patterning,” in

Proc. IEEE ICDE 2017, pp. 1140–1149.

[11] Q. Cao, X. Yang, J. Yu, and C. Palow, “Uncovering large groups of active malicious accounts in online social networks,” in Proc. ACM SIGSAC 2014, pp. 477–488.

[12] X. Zhou, X. Liang, H. Zhang, and Y. Ma, “Cross-platform identification of anonymous identical users in multiple social media networks,” IEEE Trans. Knowledge and Data Engineering, vol. 28, no. 2, pp. 411–424, 2016.

[13] T. Wuchner, A. Cislak, M. Ochoa, and A. Pretschner, “Leveraging ¨ compression-based graph mining for behavior-based malware detection,” IEEE Trans. Dependable Secure Computing, vol. 16, no. 1, pp. 99–112, 2019.

[14] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in Proc. ACM SIGKDD 2016, CA, USA, August 13-17, 2016, pp. 785–794.

[15] B. Jia, C. Dong, Z. Chen, K. Chang, N. Sullivan, and G. Chen, “Pattern discovery and anomaly detection via knowledge graph,” in Proc. FUSION 2018, Cambridge, UK, July 10-13, 2018, pp. 2392– 2399.

[16] P. Cui, X. Wang, J. Pei, and W. Zhu, “A survey on network embedding,” IEEE Trans. Knowledge and Data Engineering, vol. 31, no. 5, pp. 833–852, 2019.

[17] M. Abouelenien, V. Perez-Rosas, R. Mihalcea, and M. Burzo, ´ “Detecting deceptive behavior via integration of discriminative features from multiple modalities,” IEEE Trans. Information Forensics and Security, vol. 12, no. 5, pp. 1042–1055, 2017.

[18] W. Youyou, M. Kosinski, and D. Stillwell, “Computer-based personality judgments are more accurate than those made by humans,” PNAS, vol. 112, no. 4, pp. 1036–1040, 2015.

[19] V. Sekara, A. Stopczynski, and S. Lehmann, “Fundamental structures of dynamic social networks,” PNAS, vol. 113, no. 36, pp. 9977–9982, 2016.

[20] K. Rzecki, P. Plawiak, M. Niedzwiecki, T. Sosnicki, J. Leskow, and M. Ciesielski, “Person recognition based on touch screen gestures using computational intelligence methods,” Information Science, vol. 415, pp. 70–84, 2017.

[21] S. Lee and J. Kim, “Warningbird: Detecting suspicious urls in twitter stream,” in Proc. NDSS 2012, San Diego, California, USA, February 5-8, 2012, vol. 12, pp. 1–13.

[22] G. Stringhini, P. Mourlanne, G. Jacob, M. Egele, C. Kruegel, and G. Vigna, “EVILCOHORT: detecting communities of malicious accounts on online services,” in Proc. USENIX Security 2015, Washington, D.C., USA, August 12-14, 2015, pp. 563–578

.

GUIDE PROFILE

Mrs. J.KUMARI is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. She earned MCA from Osmania University, Hyderabad, and her M.Tech in Computer Science and Engineering (CSE) from Jawaharlal Nehru Technological University, Kakinada (JNTUK). Her research interests include ML, programming language. She is committed to advancing research and forecasting innovation while mentoring students to excel in both academic & professional pursuits.

Mr. M.SIVA KRISHNA, currently pursuing Master of Computer Applications at QIS College of Engineering and Technology (Autonomous), Ongole, Andhra Pradesh.

,