REPRESENTING FINE GRINED CO-OCCURANCES FOR BEHAVIOR BASED FRAUD DETECTION IN ONLINE SYSTEM

[1]GEETHA PRATHIBA, [2]MENDE MANEESHA, [3]G.SHARANYA, [4]K.SONYSANDHYA

[1]Assistant Professor, Department of Information Technology, **MALLA REDDY ENGINEERING COLLEGE FOR WOMEN**, Maisammaguda, Dhulapally Kompally, Medchal Rd, M, Secunderabad, Telangana.

[2, 3,4] Student, Department of Information Technology, **MALLA REDDY ENGINEERING COLLEGE FOR WOMEN**, Maisammaguda, Dhulapally Kompally, Medchal Rd, M, Secunderabad, Telangana.

ABSTRACT

The vigorous development of e-commerce breeds cybercrime. Online payment fraud detection, a challenge faced by online service, plays an important role in rapidly evolving e-commerce. Behavior-based methods are recognized as a promising method for online payment fraud detection. However, it is a big challenge to build high-resolution behavioral models by using low-quality behavioral data. In this work, we mainly address this problem from data enhancement for behavioral modeling. We extract fine-grained co-occurrence relationships of transactional attributes by using a knowledge graph. Furthermore, we adopt the heterogeneous network embedding to learn and improve representing comprehensive relationships. Particularly, we explore customized network embedding schemes for different types of behavioral models, such as the population-level models, individual-level models, and generalized-agent-based models. The performance gain of our method is validated by the experiments over the real dataset from a commercial bank. It can help representative behavioral models improve significantly the performance of online banking payment fraud detection. To the best of our knowledge, this is the first work to realize data enhancement for diversified behavior models by implementing network embedding algorithms on attribute-level co-occurrence relationships.

I.INTRODUCTION

Online payment services have penetrated into people's lives. The increased convenience, though, comes with inherent security risks [1]. The cybercrime involving online payment services often has the characteristics of diversification, specialization, industrialization, concealment, scenario, and cross-region, which makes the security prevention and control of online

payment extremely challenging [2]. There is an urgent need for realizing effective and comprehensive online payment fraud detection. The behavior-based method is recognized as an effective paradigm for online payment fraud detection [3]. Generally, its advantages can be summarized as follows: Firstly, behavior-based methods adopt the nonintrusion detection scheme to guarantee the user experience without user operation in the implementation process. Secondly, it changes the fraud detection pattern from one-time to continuous and can verify each transaction. Thirdly, even if the fraudster imitates the daily operation habits of the victim, the fraudster must deviate from the user behavior to gain the benefit of the victim. The deviation can be detected by behaviorbased methods. Finally, this behavior-based method can be used cooperatively as a second security line, rather than replacing with other types of detection methods.

However, the effectiveness of behavior-based methods often depends heavily on the sufficiency of user behavioral data [4]. As a matter of fact, user behavioral data that can be used for online payment fraud detection are often low-quality or restricted due to the difficulty of data collection and user privacy requirements [5]. In a word, the main challenge here is to build a highperformance behavioral model by using low-quality behavioral data. Then, this challenging problem can naturally be solved in two ways: data enhancement and model enhancement. For behavioral model enhancement, a widely recognized way is to build models from different aspects and integrate them appropriately. For model classifications, one type is based on the behavioral agent since it is a critical factor of behavioral models. According to the granularity of agents, behavioral models can be further divided into the individual-level models [6], [7], [8], [9] and population-level models [10], [11], [12], [13]. In this work, we focus on the other way, i.e., behavioral data enhancement. As for this way, a basic principle is to deeply explore relationships underlying the transaction data. The more fine-grained correlations can possibly provide richer semantic information for generating high performance behavioral models. Existing studies in data enhancement for behavioral modeling mainly focus on mining and modeling the correlations (including co occurrences) between behavioral features and labels [14]. To further improve data enhancement, a natural idea is to investigate and utilize the more fine-grained correlations in

behavioral data, e.g., ones among behavioral attributes.

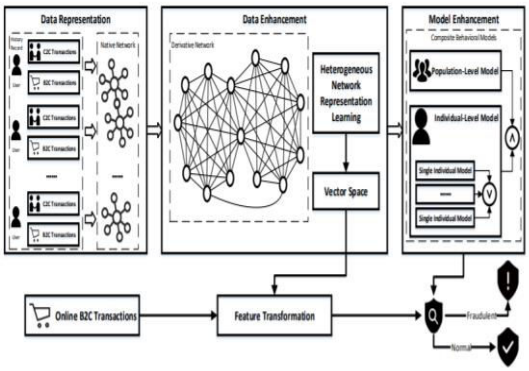

fig 1 : system architecture

II.EXISTING SYSTEM:

The effectiveness of behavior-based methods often depends heavily on the sufficiency of user behavioral data. As a matter of fact, user behavioral data that can be used for online payment fraud detection are often low-quality or restricted due to the difficulty of data collection and user privacy requirements. In a word, the main challenge here is to build a high performance behavioral model by using low-quality behavioral data. Then, this challenging problem can naturally be solved in two ways: data enhancement and model enhancement.

Disadvantages

1. Not effectively model the co-occurrences among transactional attributes for high-performance behavioral models

III.PROPOSED SYSTEM:

we mainly address this problem from data enhancement for behavioral modeling. We extract fine-grained co-occurrence relationships of transactional attributes by using a knowledge graph. Furthermore, we adopt the heterogeneous network embedding to learn and improve representing comprehensive relationships. Particularly, we explore customized network embedding schemes for different types of behavioral models, such as the population-level models, individual-level models, and generalized-agentbased models. The performance gain of our method is validated by the experiments over the real dataset from a commercial bank. It can help representative behavioral models improve significantly the performance of online banking payment fraud detection. To the best of our knowledge, this is the first work to realize data enhancement for diversified behavior models by implementing network embedding algorithms on attribute-level co-occurrence relationships.

Advantages

1. effectively model the co-occurrences among transactional attributes for high-performance behavioral models

IV.METHODOLOGY

The methodology for this project begins with defining the primary objective, which is to enhance fraud detection capabilities in online systems through fine-grained co-occurrence analysis of user behaviors. This approach seeks to identify fraudulent activities by analyzing detailed patterns and relationships among various user actions and system events.

Data collection is the first step, involving the identification and gathering of data from diverse sources such as user logs, transaction records, and system interaction histories. This data should comprehensively cover user behaviors and system states. Following collection, data preprocessing is crucial to clean and normalize the data, removing noise and handling any missing values to ensure consistency and facilitate effective analysis.

Feature extraction involves deriving fine-grained behavioral features from user interactions, such as click patterns and time intervals between actions, as well as transaction details. Additionally, co-occurrence features are extracted by analyzing relationships between different behaviors, like simultaneous logins and transactions or repeated failed login attempts followed by successes. These features represent the detailed relationships among user actions.

In the co-occurrence representation phase, frequency analysis is performed to compute how often various behaviors co-occur, using statistical methods to identify significant patterns. A graph-based representation is also constructed, where nodes represent individual behaviors or events and edges represent their co-occurrence relationships. Graph analytics are used to identify clusters and patterns that may indicate fraudulent activities.

For model development, appropriate machine learning algorithms are selected based on their suitability for detecting anomalies in fine-grained co-occurrence patterns. Techniques such as Random Forest, Gradient Boosting Machines, or Neural Networks are employed. The model is trained on labeled datasets containing known instances of fraud and non-fraud, with cross-validation used to assess its performance and robustness.

Behavioral pattern analysis involves using the trained model to identify unusual patterns and deviations from normal behavior. Anomaly detection techniques are applied to flag behaviors

that significantly deviate from established norms, with a focus on investigating these anomalies to determine their relevance to fraud.

The integration and deployment phase requires embedding the fraud detection model into the existing online system to enable real-time monitoring of user behaviors. It is essential to ensure that the system can handle the scale and complexity of online transactions effectively. Continuous monitoring mechanisms are implemented to update the model based on new data and emerging fraud patterns, with regular reviews and refinements to adapt to evolving fraudulent tactics.

Finally, the evaluation and improvement stage involves assessing the system's performance using metrics such as precision, recall, F1-score, and ROC-AUC. Analyzing the effectiveness of the system in reducing false positives and negatives is key. A feedback loop is established to review detected fraud cases, refine the system's accuracy, and incorporate feedback for ongoing improvements. This comprehensive methodology aims to enhance fraud detection in online systems through a detailed analysis of user behavior co-occurrences.

V.CONCLUSION

For behavioral models in online payment fraud detection, we propose an effective data enhancement scheme by modeling co-occurrence relationships of transactional attributes. Accordingly, we design customized cooccurrence relation networks, and introduce the technique of heterogeneous network embedding to represent online transaction data for different types of behavioral models, e.g., the individual-level and population-level models. The methods are validated by the implementation on a real-world dataset. They outperform the stateof-the-art classifiers with lightweight feature engineering methods.

VI.FUTURE ENHANCEMENT

Therefore, our methods can also serve as a feasible paradigm of automatic feature engineering. There are some interesting issues left to study: (1) An interesting future work is to extend the data enhancement scheme into other types of behavioral models, e.g., the group-level models and generalized-agent-based models, except the population-level and individual-level models studied in this work. (2) It would be interesting to investigate the dedicated enhancement schemes for more advanced individual-

level models, since the adopted naive individual-level model does not fully capture the advantages of the proposed data representation scheme based on the techniques of heterogeneous network embedding. (3) It is anticipated to demonstrate the generality of the proposed method by applying it to different real-life application scenarios.

VII. BIBLIOGRAPHY

[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.