International Journal on Applications in Engineering and Technology
Volume 9 : Issue 2 : April 2023, pp 1 – 7 www.aetsjournal.com ISSN (Online) : 2455 - 0523

DOCUMENTS REPRESENTING FINE GRAINED CO-OCCURRENCES FOR BEHAVIOR BASED FRAUD DETECTION

P. NATARAJAN , *M. ILAMATHI*

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]. Gen rally, its advantages can be summarized as follows: Firstly, behavior based methods adopt the non intrusion 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 behavior based 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 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.

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

P.Natarajan , Assistant Professor , Department of Computer Applications , Erode Sengunthar Engineering College (Autonomous), Perundurai , Erode.
(Email : palanisamynatarajan50@gmail.com)

M. Ilamathi, PG Scholar , Department of Computer Applications, Erode Sengunthar Engineering College (Autonomous), Perundurai , Erode.
(Email : *kodaisuku@gmail.com*)

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 and population-level models.

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.. 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.

As the main contribution of our work, we aim to effectively model the co-occurrences among transactional attributes for high-performance behavioral models. For this purpose, we propose to adopt the heterogeneous relation network, a special form of the knowledge graph [15], to represent the co-occurrences effectively. Here, a network node (or say an entity) corresponds to an attribute value in transactions, and an edge corresponds to a heterogeneous association between different attribute values. Although the relation network can express the data more appropriately, it cannot finally solve the data imperfection problem for behavioral modeling, that is, it has no effect on enhancing the original low-quality data.

An effective data representation preserving these comprehensive relationships can act as an important mean of relational data enhancement. To this end, we introduce network representation learning (NRL), which effectively capture deep relationships [16]. Deep relationships make up for low quality data in fraud detection and improve the performance of fraud detection models. By calculating the similarity between embedding vectors, more potential relationships could be inferred. It partly solves the data imperfection problem. In addition to data enhancement, NRL transforms the traditional network analysis from the artificially defined feature to the automatic learned feature, which extracts deep relationships from numerous transactions.

The final performance of behavioral modeling for online fraud detection directly depends on the harmonious cooperation of data enhancement and model enhancement. Different types of behavioral models need matching network embedding schemes to achieve excellent performance. This is one of the significant technical problems in our work. We aim to investigate the appropriate network embedding schemes for population-level models, individual-level models, and models with different generalized behavioral agents. More specifically, for population-level models, we design a label-free heterogeneous network to reconstruct online transactions and then feed the features generated in embedding spac into the state-of-the-art classifiers based on machine learning to predict fraud risks; while, for individual-level models, we turn to a label-aware heterogeneous network that distinguishes the relations between attributes of fraudulent transaction, and further design multiple naïve individual-level models that match the representations generated from the label-aware network. Furthermore, we combine the population-level and individual-level models to realize the complementary effects by overcoming each other's weaknesses.

The main contributions can be summarized as follows:

_ We propose a novel effective data enhancement scheme for behavioral modeling by representing and mining more fine-grained attribute-level co-occurrences. We adopt the heterogeneous relation networks to represent the attribute-level co-occurrences, and extract those relationships by heterogeneous network embedding algorithms in depth.

_ We devise a unified interface between network embedding algorithms and behavioral models by customizing the preserved relationship networks according to the classification of behavioral models.

_ We implement the proposed methods on a real world online banking payment service scenario.

It is validated that our methods significantly outperform the state-of-the-art classifiers in terms of a set of representative metrics in online fraud detection.

The rest of this paper is organized as follows. We provide a literature review in Section 2. Section 3 gives an overview of our solution. Then, we present out method in detail in Section 4 and make the validation in Section 5. Finally, we conclude the paper and envisage future work in Section 6.

II. OBJECTIVES

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.

III. LITERATURE SURVEY

In this paper, we present HITFRAUD that leverages heterogeneous information networks for collective fraud detection by exploring correlated and fast evolving fraudulent behaviors. First, a heterogeneous information network is designed to link entities of interest in the transaction database via different semantics.[1]

In this paper, we study the social behaviors of OSN users, i.e. their usage of OSN services, and the application of which in detecting compromised accounts. In particular, we propose a set of social behavioral features that can effectively characterize the user social activities on OSNs. We validate the efficacy of these behavioral features by collecting and analyzing real user clickstreams to an OSN website.[2].

To improve the accuracy and interpretability of community discovery, we propose to infer users' social communities by in- corporating their spatiotemporal data and semantic information. Technically, we propose a unified probabilistic generative model, User-Community-Geo-Topic (UCGT), to simulate the generative process of communities as a result of network proximities, spatiotemporal co-occurrences and semantic similarity. With a well-designed multi-component model structure and a parallel inference implementation to leverage the power of multicores and clusters, our UCGT model is expressive while remaining efficient and scalable to growing large-scale geo-social networking data.[3].

IV. EXISTING SYSTEM

Vedran et al. [19] explored the complex interaction between social and geospatial behavior and demonstrated that social behavior could be predicted with high precision. Yin et al. [4] proposed a probabilistic generative model combining use spatiotemporal data and semantic information to predict user behavior. Naini et al. [7] studied the task of identifying the users by matching the histograms of their data in the anonymous dataset with the histograms from the original dataset. Egele etal. [8] proposed a behavior-based method to identify compromises of high-profile accounts. Ruan et al. [3] conducted a study on online user behavior by collecting and analyzing user clickstreams of a well known OSN.

Rzecki et al. [20] designed a data acquisition system to analyze the execution of single-finger gestures on a mobile device screen and indicated the best classification method for person recognition based on proposed surveys. Alzubaidi et al. [9] investigated the representative methods for user authentication on smartphone devices in smartphone authentication including seven types of behavioral biometrics, which are handwaving, gait, touchscreen, keystroke, voice, signature and general profiling.

Lee and Kim [21] proposed a suspicious URL detection system to recognize user anomalous behaviors on Twitter. Cao et al. [11] designed and implemented a malicious account detection system for detecting both fake and compromised real user accounts. Zhou et al. [12] proposed an FRUI algorithm to match users among multiple OSNs. Stringhini et al. [22] designed a system named EVILCOHORT, which can detect malicious accounts on any online service with the mapping between an online account and an IP address. Meng et al. [23] presented a static sentence-level attention model for text-based speaker change detection by formulating it as a matching problem of utterances before and after a certain decision point. Rawat et al. [24] proposed three methodologies to cope up with suspicious and anomalous activities, such as continuous creation of fake user accounts, hacking of accounts and other illegitimate acts in social networks.

VanDam et al. [25] focused on studying compromised accounts in Twitter to understand who were hackers, what type of content did hackers tweet, and what features could help distinguish between compromised tweets and normal tweets. They also showed that extra meta-information could help improve the detection of compromised accounts.

Zhao et al. [26] proposed a semi-supervised network embedding model by adopting graph convolutional network that is capable of capturing both local and global structure of protein-protein interactions network even there is no any information associated with each vertex. Li et al. [27] incorporated word semantic relations in the

latent topic learning by the word embedding method to solve that the Dirichlet Multinomial Mixture model does not have access to background knowledge when modeling short texts.

Baqueri et al. [28] presented a framework to model residents travel and activities outside the study area as part of the complete activity-travel schedule by introducing the external travel to address the distorted travel patterns. Chen et al. [29] proposed a collaborative and adversarial network (CAN), which explicitly models the common features between two sentences for enhancing sentence similarity modeling. Catolino et al. [30] devised and evaluated the performance of a new change prediction model that further exploit developerrelated factors (e.g., number of developers working on a class) as predictors of change-proneness of classes. Liu et al. [31] proposed a novel method for disaggregating the coarse-scale values of the group-level features in the nested data to overcome the limitation in terms of their predictive performance, especially the difficulty in identifying potential cross-scale interactions between the local and group-level features when applied to datasets with limited training examples.

Disadvantages

- ❖ The system is not implemented Composite Behavioral Models in which security is more for fraud detection.
- ❖ The system is not implemented Single-Agent Behavioral Model to detect the fraud.

V. PROPOSED SYSTEM

- ➢ The system proposes a novel effective data enhancement scheme for behavioral modeling by representing and mining more fine-grained attribute-level co-occurrences. We adopt the heterogeneous relation networks to represent the attribute-level co-occurrences, and extract those relationships by heterogeneous network embedding algorithms in depth.
- ➢ The system devises a unified interface between network embedding algorithms and behavioral models by customizing the preserved relationship networks according to the classification of behavioral models.
- ➢ The system implements the proposed methods on a realworld online banking payment service scenario. It is validated that our methods significantly outperform the state of-the-art classifiers in terms of a set of representative metrics in online fraud detection.

ADVANTAGES

The population-level models identify the fraud by detecting the population-level behavioral anomalies, e.g., behavioral outlier detection and misuse detection

At the population-level and individual-level, we utilize the intersection to integrate judgments. That is, the fraud is determined only if the judgments of both models are fraudulent. Our fraud detection model consists of two levels of models and plays a complementary role.

VI. IMPLEMENTATION AND EXECUTION

A. Modules:

1)User module:

In this module the user will search the Server that are available and buy the service that the user is interested the user will redirect to the online transaction page after selecting the service then if the user have the transactional id and password they can login and pay the amount to the service provider account by online transfer method.

2)Service module:

In this module the Service Provider will add the new services or products to view for the user .Then the user will search the services that are available and select the required service and then move to the transaction page to pay the money for activation of the service

3)Bank Transaction module:

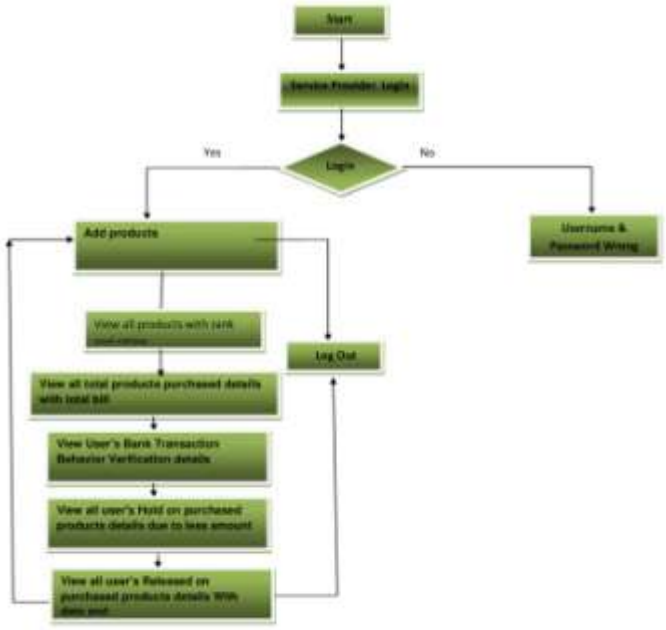
In transaction module the user will get redirect to this page after the selection of the service that they want for online payment. In this transaction module first the user will apply for the loan for buying the service and the bank admin should approve the loan after successfully verifying the details of the user once the bank approved the loan the amount will automatically credit to the user account in his account number. Then the user can directly transfer the amount to the service provider account.

4)Admin module:

In admin module the admin will authorize the users and add the services that are launched newly to view in the user side. Admin can check directly his bank account details and purchase transaction behaviors after login. In admin module we will have all the user details and their transactions behavior.

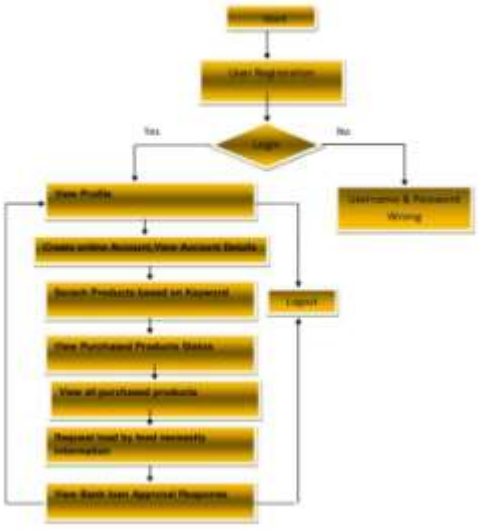
VII. DESIGN AND ALL DIAGRAMS

1)Flow Chat : Service Provideer



Flow Chat : Bank Transsaction Behaviour verify(BTBV)

2)Flow Chat : User



VIII. FUTURE ENHANCEMENT

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.

IX. 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 co occurrence 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 state of- the-art classifiers with lightweight feature engineering methods. 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.

X. 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. W¨ uchner, 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. P´erez-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 alicious accounts on online services," in Proc. USENIX Security 2015,Washington, D.C., USA, August 12-14, 2015, pp. 563–578.
[23] Z. Meng, L. Mou, and Z. Jin, "Hierarchical RNN with static sentence-level attention for text-based speaker change detection," in Proc. ACM CIKM 2017, Singapore, November 06 - 10, 2017, pp. 2203–2206.
[24] A. Rawat, G. Gugnani, M. Shastri, and P. Kumar, "Anomaly
recognition in online social networks," International Journal of
[22] Security and Its Applications, vol. 9, no. 7, pp. 109–118, 2015.
[23]] C. VanDam, J. Tang, and P. Tan, "Understanding compromised
[24] accounts on twitter," in Proc. ACM WI 2017 , Leipzig, Germany,August 23-26, 2017, pp. 737–74.