JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Representing Fine-Grained Co-Occurrences for Behavior-Based Fraud Detection in Online Payment Services

[1]Prof. M.R.Rajput

[2]NEHA Vilas Patil [3]Prajakta Durgadas Pachpor [4]Sakshi Shrikrushna Tayade [5]Vrushali Mohan Kolte
[1]Professor Computer Science Engineering Department, Padm. Dr. VBKCOE, Malkapur, Maharashtra, India.
[2345]Student Computer Science And Engineering Department Padm Dr V.B.Kolte, Malakpur, Maharashtra, India.

ABSTRACT –

Cybercrime is fostered by e-commerce's rapid growth. The problem of online payment fraud detection, which online services must overcome, is crucial to the quickly developing e-commerce industry. It is acknowledged that behavior-based approaches have promise in the fight against online payment fraud. Nevertheless, using low-quality behavioural data to construct high-resolution behavioural models is quite difficult.

We primarily tackle this issue from data enhancement for behavioural modelling in our paper. Using a knowledge graph, we are able to extract transactional attribute co-occurrence correlations at a finer level. Additionally, in order to learn and get better at portraying comprehensive relationships, we use heterogeneous network embedding. In particular, we investigate tailored network embedding strategies for several behavioural model types, including generalised agent-based, population-level, and individual-level models.

Cybercrime is fostered by e-commerce's rapid growth. The problem of online payment fraud detection, which online services must overcome, is crucial to the quickly developing e-commerce industry. It is acknowledged that behavior-based approaches have promise in the fight against online payment fraud. Nevertheless, using low-quality behavioural data to construct high-resolution behavioural models is quite difficult.

We primarily tackle this issue from data enhancement for behavioural modelling in our paper. Using a knowledge graph, we are able to extract transactional attribute co-occurrence correlations at a finer level. Additionally, in order to learn and get better at portraying comprehensive relationships, we use heterogeneous network embedding. In particular, we investigate tailored network embedding strategies for several behavioural model types, including generalised agent-based, population-level, and individual-level models.

KEYWORDS –

Online payment services, fraud detection, network embedding, user behavioral modeling.

INTRODUCTION –

Online payment services have significantly impacted people's lives, but the enhanced convenience they offer comes with inherent security vulnerabilities [1]. The prevention and control of online payment security pose significant challenges, given the diversification, specialization, industrialization, concealment, scenario, and cross-regional features of cybercrime associated with these services [2]. Effectively addressing online payment fraud detection is of utmost importance.

The behavior-based approach has proven successful in detecting online payment fraud, offering several advantages [3]. Firstly, it employs a non-intrusive detection system to ensure a seamless user experience without requiring user intervention during installation. Secondly, it validates every transaction, transforming the fraud detection pattern from one-time to continuous. Thirdly, by diverging from user behavior, fraudsters attempting to exploit victims can be identified. Lastly, rather than replacing other detection techniques, the behavior-based approach functions as an additional layer of protection.

However, the effectiveness of behavior-based approaches relies heavily on the availability of sufficient user behavioral data [4]. Unfortunately, due to user privacy laws and data gathering challenges, the user behavioral data accessible for online payment fraud detection is often limited or of suboptimal quality [5]. Consequently, the key challenge lies in constructing a high-performance behavioral model using low-quality behavioral data. Data enhancement and model augmentation emerge as two evident strategies to address this formidable challenge.

An established method for enhancing behavioral models involves constructing them from various perspectives and integrating them effectively. Given the crucial role of the behavioral agent in these models, one classification categorizes behavioral models into individual-level models [6, 7, 8, 9, 10] and population-level models [11], [12], [13], based on the granularity of agents.

PROBLEM FORMULATION -

The challenge addressed in this study lies in the complex landscape of online payment fraud detection. Online payment services, while enhancing convenience, expose users to security vulnerabilities. The behavior-based approach has proven effective, relying on non-intrusive detection and continuous validation. However, the efficacy of these approaches is hindered by the limited availability and poor quality of user behavioral data due to privacy laws and data gathering challenges. The primary problem is formulating a robust behavioral model using low-quality data. To address this, strategies like data enhancement and model augmentation are considered. Additionally, the study explores the classification of behavioral models based on the granularity of agents, aiming to optimize fraud detection in online payment services.

PROPOSE SYSTEM METHODOLOGY -

The proposed system methodology involves a multi-faceted approach to enhance online payment fraud detection using behavior-based techniques. Firstly, data enhancement strategies will be implemented to address the limitations associated with low-quality user behavioral data. This may include the use of synthetic data generation techniques and feature engineering to augment the dataset.

Secondly, a model augmentation process will be employed to construct behavioral models from diverse perspectives and integrate them effectively. Individual-level models and population-level models, categorized based on the granularity of agents, will be explored and integrated for a comprehensive fraud detection system.

Furthermore, the system will leverage advanced machine learning algorithms and anomaly detection techniques to identify patterns indicative of fraudulent behavior. Continuous validation of transactions will be ensured, transforming fraud detection from a one-time event to a continuous monitoring process.

To maintain user experience, a non-intrusive detection system will be implemented, minimizing the need for user intervention during installation. The system will also consider the evolving nature of cyber threats by incorporating adaptive mechanisms to detect new patterns of fraudulent behavior.

In summary, the proposed system methodology encompasses data enhancement, model augmentation, diverse model perspectives, continuous validation, non-intrusive detection, and adaptability to effectively combat online payment fraud while preserving user experience.

WORKING ON LANGUAGES -

In the project focused on language processing using Python (ES13), the professional environment chosen is Anaconda with Jupyter as the development platform. The database for storing relevant information is in CSV file format, providing a Long-Term Support (LTS) solution for data storage. Additionally, the server-side implementation involves the creation of a web page to facilitate interaction and presentation of results.

The Python language, in its ES13 version, provides a robust foundation for natural language processing tasks. Anaconda, with its comprehensive package management system, ensures easy installation and management of libraries essential for language processing within the Jupyter environment, fostering a seamless development experience.

Data is stored and managed in CSV files, offering a simple and accessible solution for handling structured data. This LTS approach ensures the stability and longevity of the data storage solution. The utilization of a web page as a server-side component enhances user interaction, allowing for a more dynamic and user-friendly experience.

In summary, the chosen technology stack of Python (ES13), Anaconda with Jupyter, CSV files, and a web page server creates an efficient and effective environment for developing language processing solutions, emphasizing ease of use, data management, and user interaction.

LITERATURE SURVEY-

focus on behavior-based fraud detection in online payment services. Pay attention to papers that specifically mention fine-grained co-occurrences or similar techniques

.Pay attention to the methods and techniques used in the papers you find. Note any common approaches or innovative methods for representing fine-grained co-occurrences in fraud detection system. Look for influential authors in the field and check the papers they've authored or cited. This can lead you to additional relevant research.

Explore conference proceedings from relevant conferences such as IEEE Symposium on Security and Privacy, ACM CCS, and journals like IEEE Transactions on Dependable and Secure Computing, Journal of Cybersecurity, and Journal of Financial Crime.

Look for industry reports from cybersecurity firms, financial institutions, and consulting companies. These reports often contain valuable insights and case studies on fraud detection in online payment services. Summarize the findings from the papers you've reviewed, noting common methodologies, challenges, and emerging trends in representing fine-grained co-occurrences for behavior-based fraud detection. Highlight any gaps or areas where further research is needed. This could include areas with limited coverage in the literature or emerging challenges in online payment fraud detection.

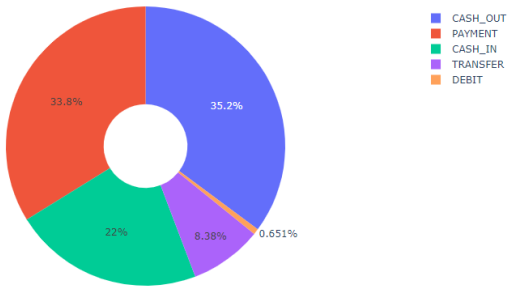

Fig. Distribution of Transaction Type

	step	type	amount	nameOrig	oldbalanceOrg	newbalanceOrig	nameDest	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud
0	1	PAYMENT	9839.64	C1231006815	170136.00	160296.36	M1979787155	0.00	0.00	0	0
1	1	PAYMENT	1864.28	C1666544295	21249.00	19384.72	M2044282225	0.00	0.00	0	0
2	1	TRANSFER	181.00	C1305486145	181.00	0.00	C553264065	0.00	0.00	1	0
3	1	CASH_OUT	181.00	C840083671	181.00	0.00	C38997010	21182.00	0.00	1	0
4	1	PAYMENT	11668.14	C2048537720	41554.00	29885.86	M1230701703	0.00	0.00	0	0
...	...	...	...	...	...	...	...	...	...	...	...
6362615	743	CASH_OUT	339682.13	C786484425	339682.13	0.00	C776919290	0.00	339682.13	1	0
6362616	743	TRANSFER	6311409.28	C1529008245	6311409.28	0.00	C1881841831	0.00	0.00	1	0
6362617	743	CASH_OUT	6311409.28	C1162922333	6311409.28	0.00	C1365125890	68488.84	6379898.11	1	0
6362618	743	TRANSFER	850002.52	C1685995037	850002.52	0.00	C2080388513	0.00	0.00	1	0
6362619	743	CASH_OUT	850002.52	C1280323807	850002.52	0.00	C873221189	6510099.11	7360101.63	1	0

Fig. Fraud Detection History

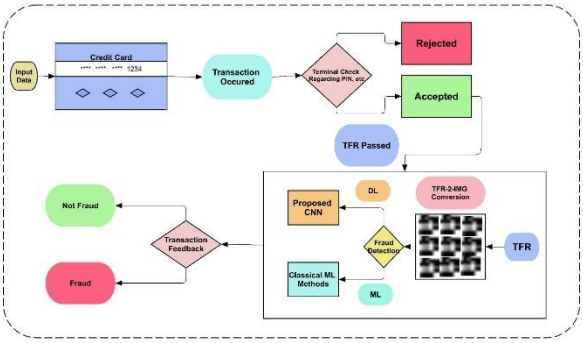

Fig. Block Diagram Fraud Detection

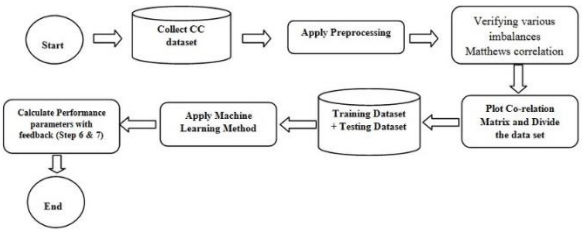

Fig: Flowchart of Online Fraud Detection

RELATED WORKING -

Certainly, the rapid evolution of online payment services has given rise to a continuous stream of fraudulent activities in online transactions. Addressing this challenge, the use of behavioral models for fraud detection has become a focal point of extensive research, capturing the attention of numerous researchers.

As online transactions proliferate, so do the tactics employed by fraudsters, necessitating innovative approaches for detection. Behavioral models offer a promising avenue, leveraging patterns in user behavior to identify anomalies indicative of fraudulent activity. The dynamic nature of online fraud requires continuous exploration and refinement of these models, prompting researchers to delve into various methodologies to enhance the effectiveness of fraud detection systems.

This vibrant and evolving field reflects the commitment of researchers to stay ahead of emerging threats in the online payment landscape, aiming to provide secure and reliable services for users. The exploration of behavioral models signifies a proactive response to the challenges posed by the dynamic and sophisticated nature of online transaction fraud.

FRONT-END TECHNOLOGY -

The front-end of a web application is constructed using HTML for structuring content, CSS for styling and layout, and JavaScript for interactivity. HTML provides the basic framework of the webpage, CSS enhances its visual presentation, and JavaScript adds dynamic and interactive features. On the back end, Python is employed, likely through web frameworks like Flask or Django, to handle server-side logic, manage data, and facilitate communication with databases. This combination of technologies allows for a comprehensive development approach, ensuring a visually appealing and interactive user interface coupled with robust server-side functionality.

BACK-END TECHNOLOGY -

In the back-end technology stack, Python serves as the primary programming language, facilitating server-side logic and overall application functionality. The professional environment is established through Anaconda with Jupyter, providing a robust development platform for Python. Jupyter notebooks within Anaconda offer an interactive and collaborative environment for coding, analysis, and documentation. For data storage and retrieval, a CSV file format is employed as the database, offering a lightweight and easily manageable solution with Long-Term Support (LTS) for data persistence. This back-end technology configuration, featuring Python with Anaconda and CSV files, ensures efficient server-side processing, seamless development, and a straightforward approach to data management in a reliable and sustainable manner.

CONCLUSION –

In conclusion, this paper introduces a pioneering approach to enhance online payment fraud detection through the representation of fine-grained co-occurrences in transactional data. The meticulously designed co-occurrence relation networks, coupled with heterogeneous network embedding techniques, demonstrate remarkable efficacy in capturing complex relationships among transactional attributes.

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 Int. Conf. Data Mining, 2017,pp. 769–774.

[2] M. A. Ali, B. Arief, M. Emms, and A. P. A. van Moorsel, "Does theonline card payment landscape unwittingly facilitate

fraud?"IEEE Security Privacy, vol. 15, no. 2, pp. 78–86, Mar./Apr. 2017.

[3] X. Ruan, Z. Wu, H. Wang, and S. Jajodia, "Profiling online socialbehaviors for compromised account detection," IEEE Trans. Inf.Forensics Security, vol. 11, no. 1, pp. 176–187, Jan. 2016.

[4] H. Yin et al., "Discovering interpretable geo-social communities for user behavior prediction," in Proc. IEEE 32nd Int. Conf. Data Eng., 2016, pp. 942–953.

[5] Y.-A. De Montjoye 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. Knowl. Discov. 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. Inf. Forensics Security, vol. 11, no. 2, pp. 358–372, Feb. 2016.

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

[9] A. Alzubaidi and J. Kalita, "Authentication of smartphone users using behavioral biometrics," IEEE Commun. Surveys Tuts., vol. 18, no. 3, pp. 1998–2026, Third Quarter 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 33rd Int. Conf. Data Eng., 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 Conf. Comput. Commun. Secur., 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. Knowl. Data Eng., vol. 28, no. 2, pp. 411–424, Feb. 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 Comput., vol. 16, no. 1, pp. 99–112, Jan./Feb. 2019.

[14] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 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. 21st Int. Conf. Inf. Fusion, 2018, pp. 2392–2399.

[16] P. Cui, X. Wang, J. Pei, and W. Zhu, "A survey on network embedding," IEEE Trans. Knowl. Data Eng., vol. 31, no. 5, pp. 833–852, May 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. Inf. Forensics Security, vol. 12, no. 5, pp. 1042–1055, May 2017.