ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

MINING FRAUDSTERS AND FRAUDULENT STRATEGIES IN LARGE-SCALE MOBILE SOCIAL NETWORKS

Vinit Kumar Gunjan[1] , S SPANDANA[2] , K THANUJA[3] , D.SAI CHARAN[4]

ASSOCIATE PROFESSOR[1], UG SCHOLAR[2,3&4]

DEPARTMENT OF CSE, CMR INSTITUTE OF TECHNOLOGY, KANDLAKOYA VILLAGE, MEDCHAL RD, HYDERABAD, TELANGANA 501401

ABSTRACT: The rapid advancement of modern communication technologies, particularly mobile phone communications, has greatly facilitated human social interactions and information exchange, yet it has also led to the rise of telemarketing frauds, which can significantly deplete individual wealth and social resources, potentially causing broader economic setbacks. In this work, we aim to address the issue of telemarketing fraud by focusing on identifying the "precise fraud" phenomenon, where fraudsters selectively target victims based on specific criteria. To study this issue, we analyze a comprehensive dataset from Shanghai, which spans one month and includes telecommunication metadata for 54 million users and 698 million call logs. Our analysis reveals that personal user information is often leaked, providing fraudsters with the means to carefully select their victims based on factors such as the user's age and activity within the mobile network. This targeted approach by fraudsters raises significant concerns about privacy and the potential for widespread financial harm. To combat this issue, we propose a novel semi-supervised learning framework that effectively distinguishes fraudsters from non-fraudsters, using both labeled and unlabeled data to train the model. By leveraging this approach, we are able to improve the accuracy of fraud detection significantly. Our experimental results, which are based on real-world data, show that our proposed method outperforms several state-of-the-art fraud detection algorithms, achieving an average improvement of +0.278 in terms of the F1 score. This demonstrates the efficacy of our approach in identifying fraudulent activities with higher precision. The findings of this study are crucial for informing policy decisions by governments and mobile service providers, as they offer valuable insights into the strategies employed by fraudsters and the patterns that can be used to detect fraudulent behavior. Furthermore, our work highlights the importance of integrating advanced machine learning techniques in fraud detection systems to address the growing challenge of telemarketing fraud in the digital age. Through the application of these innovative methods, we aim to provide a more robust solution for preventing fraud and protecting consumers, ultimately contributing to the preservation of social and economic stability.

Index Terms—Social network, Fraud detection, Fraudulent strategy

I. INTRODUCTION Fraudulent activities, particularly phone fraud, have been rapidly increasing with the development of global communication technologies, causing significant financial losses and even life-threatening consequences for victims. In China, phone fraud has become a critical issue, with estimates showing over 500 million fraud cases in 2016 alone, leading to losses of approximately 16.4 billion USD, and less than 3% of these cases being resolved. This study focuses on identifying the behavior patterns of fraudsters using a large-scale

mobile social network dataset, consisting of 30 days of complete call logs from Shanghai in 2016, which includes 54 million users and 698 million call logs. The research explores the challenges in detecting fraud, particularly due to data sensitivity, as access to the content of call logs is restricted for privacy reasons. Instead, the study relies on meta-information, such as the timing and frequency of calls, to identify fraudulent patterns. The study reveals that fraudsters strategically target specific individuals based on factors such as age and activity in phone communications, rather than selecting victims randomly, showing that user information might have been leaked. Additionally, the study highlights the challenge of label imbalance in fraud detection, as the majority of users in the dataset are non-fraudsters, making fraud detection more difficult. To address these challenges, the researchers conduct an exploratory analysis of fraudster behavior and propose a novel factor-graph-based model, FFD, to distinguish fraudsters from non-fraudsters by incorporating fraudsters' structural information and targeting preferences. A semi-supervised learning framework is also introduced to handle the label sparsity issue by utilizing both labeled and unlabeled data. The experimental results show that this approach significantly outperforms several state-of-the-art fraud detection methods, achieving an average F1 score improvement of 0.278. This study not only uncovers key insights into fraudster strategies but also emphasizes the importance of protecting personal information to mitigate the risks of fraud. The proposed model offers a promising solution for detecting fraudsters in large-scale mobile networks, demonstrating its effectiveness in real-world applications. Through this work, the authors contribute to understanding fraudulent behavior in telecommunications and propose a powerful tool for fraud detection, potentially informing policies for governments and mobile service providers to combat telemarketing fraud more effectively.

II. LITERATURE SURVEY

A) L. Peel and A. Clauset, "Detecting change points in the large-scale structure of evolving networks." in AAAI, 2015, pp. 2914–2920.

The challenge of identifying change points in the structure of dynamic, large-scale networks. As networks evolve over time, understanding when and how significant changes occur is critical for analyzing various complex systems such as social networks, communication networks, and biological networks. The authors propose a novel method for detecting these change points, focusing on the large-scale structural shifts that may indicate important events or transitions within the network. Their approach combines statistical models with efficient computational techniques to track and identify structural changes, offering a robust framework for detecting when a network's properties—such as connectivity, centrality, or community structure—undergo significant alterations. Through the use of a series of real-world network data sets, including social networks and communication networks, Peel and Clauset demonstrate the effectiveness of their method in identifying meaningful change points that are not easily detectable by traditional approaches. The proposed method accounts for the challenges posed by the size and complexity of evolving networks and provides a means for detecting subtle changes that could have major implications for understanding the dynamics of network growth and evolution. The experimental results show that their method outperforms existing techniques in terms of both accuracy and scalability, making it a valuable tool for researchers and practitioners seeking to study the temporal evolution of large networks. Ultimately, this work contributes to the growing field of network science by providing a scalable, statistically grounded approach for

change point detection in evolving networks, which has broad applications in monitoring, analyzing, and intervening in complex systems.

B) M. A. Peabody, "Finding groups of graphs in databases," Ph.D. dissertation, Drexel University, 2002.
It explores the problem of detecting groups or clusters of similar graphs within large graph databases. As graphs are commonly used to represent relationships or structures in various fields such as computer science, biology, and social networks, the ability to identify similar graph structures is crucial for pattern recognition, data mining, and knowledge discovery. Peabody proposes novel algorithms and methods for efficiently finding groups of graphs that share certain properties or structural similarities, even within massive datasets. The dissertation presents a comprehensive framework that combines graph theory, database indexing, and clustering techniques to group graphs based on structural features such as connectivity, subgraphs, and patterns of edges and nodes. Peabody also addresses the challenges posed by the complexity and size of graph data, providing solutions to efficiently manage and search through large-scale graph databases while maintaining accuracy in detecting groups. Through the development of specialized algorithms, the work introduces a methodology for optimizing graph comparison and clustering, thereby enhancing the ability to detect meaningful relationships between graphs that may not be immediately obvious through traditional methods. The dissertation includes experimental results demonstrating the effectiveness of the proposed methods, showing their ability to identify graph groups in a variety of domains, including social networks, chemical structures, and computer networks. Peabody's work contributes to the field of graph mining by providing new techniques for clustering graph data, with implications for a wide range of applications, from understanding complex networks to improving data retrieval systems. Ultimately, this research lays the foundation for more efficient and effective methods for graph-based data analysis, which has applications in numerous disciplines dealing with structured data.

C) P. Shoubridge, M. Kraetzl, W. Wallis, and H. Bunke, "Detection of abnormal change in a time series of graphs," Journal of Interconnection Networks, vol. 3, no. 01n02, pp. 85–101, 2002.
It explore methods for detecting abnormal changes in the structure of graphs over time. With the growing importance of dynamic networks in various fields, such as communication, social networks, and biological systems, identifying sudden or anomalous changes in graph structures becomes crucial for understanding underlying shifts or disruptions in the system being studied. The authors propose a framework for analyzing time series data of graphs, where each graph in the series represents a snapshot of a network at a specific time. Their approach focuses on detecting significant changes or outliers in the evolution of graph structures, which may indicate critical events such as system failures, attacks, or structural changes. To achieve this, they apply statistical methods to model the normal behavior of graphs over time and then identify deviations from this behavior, using tools like graph similarity measures and change detection algorithms. The paper presents a detailed analysis of various techniques for detecting abnormal changes, including the use of graph edit distance and spectral methods, and compares their effectiveness in different scenarios. Through experimental results on several real-world datasets, the authors demonstrate the utility of their proposed methods in identifying abnormal changes in time series of graphs, such as detecting shifts in social networks or abnormalities in network traffic patterns. This work contributes to the field of network analysis by providing a novel approach to detecting and understanding temporal

changes in graph-based data, offering valuable insights for applications in network monitoring, anomaly detection, and predictive modeling. The proposed methods enable more robust and adaptive analysis of dynamic systems, helping researchers and practitioners better respond to unexpected changes in evolving networks.

IMPLEMENTATION

Modules

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Browse Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View All Antifraud Model for Internet Loan Prediction, Find Internet Loan Prediction Type Ratio, View Primary Stage Diabetic Prediction Ratio Results, Download Predicted Data Sets, View All Remote Users.

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.

Remote 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 REGISTER AND LOGIN, PREDICT PRIMARY STAGE DIABETIC STATUS, VIEW YOUR PROFILE.

CONCLUSION

In this paper, we have addressed the problem of identifying fraudsters and uncovering fraudulent strategies in large-scale mobile networks by analyzing a one-month dataset of telecommunication metadata from Shanghai, which includes 698 million call logs from 54 million users. Our study reveals that fraudsters exhibit distinct behaviors compared to non-fraudsters when communicating with others, and that they tend to target users based on factors such as age and activity within the mobile network. Based on these findings, we proposed a novel semi-supervised model to differentiate fraudsters from non-fraudsters. Experimental results show that our model significantly outperforms several state-of-the-art methods in detecting fraudulent activities. For future work, we suggest exploring the detection of fraud groups rather than individual fraudsters, as these groups often consist of fraudsters with varying roles and responsibilities, which could lead to revealing collaboration patterns within fraudulent networks. Additionally, further research could incorporate geographical information to track fraudsters' offline behavior and analyze their movement patterns within the city. Although our work provides valuable insights into fraud detection, it is constrained by the available data. Despite Shanghai being a major global city and China Telecom being a leading service provider, the dataset's selection bias may limit the generalizability of our findings to other regions or networks. Nevertheless, our approach offers a promising direction for detecting

and understanding fraud in large mobile networks, with potential applications in improving security and privacy for users. Future research in this area could build on our work by addressing the limitations of the data and extending the framework to broader contexts, enhancing the robustness of fraud detection systems.

REFERENCES

[1] B. Hooi, H. A. Song, A. Beutel, N. Shah, K. Shin, and C. Faloutsos, "Fraudar: Bounding graph fraud in the face of camouflage," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 895–904.

[2] V. S. Tseng, J. Ying, C. Huang, Y. Kao, and K. Chen, "Fraudetector: A graph-mining-based framework for fraudulent phone call detection," in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015, pp. 2157– 2166.

[3] J. Xu, A. H. Sung, and Q. Liu, "Behaviour mining for fraud detection." Journal of Research and Practice in Information Technology, 2007.

[4] M. I. M. Yusoff, I. Mohamed, and M. R. A. Bakar, "Fraud detection in telecommunication industry using gaussian mixed model," in Research and Innovation in Information Systems (ICRIIS), 2013 International Conference on, 2013, pp. 27–32.

[5] P. Chan, W. Fan, A. L. Prodromidis, and S. J. Stolfo, "Distributed data mining in credit card fraud detection," IEEE Intelligent Systems & Their Applications, vol. 14, no. 6, pp. 67–74, 1999.

[6] T. Ormerod, N. Morley, L. Ball, C. Langley, and C. Spenser, "Using ethnography to design a mass detection tool (mdt) for the early discovery of insurance fraud," in CHI'03 Extended Abstracts on Human Factors in Computing Systems, 2003, pp. 650–651.

[7] S. Aral, L. Muchnik, and A. Sundararajan, "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks," Proceedings of the National Academy of Sciences, vol. 106, no. 51, pp. 21 544–21 549, 2009.

[8] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," Information Theory, IEEE Transactions on, vol. 47, no. 2, pp. 498–519, 2001.

[9] J. M. Hammersley and P. Clifford, "Markov fields on finite graphs and lattices," Unpublished manuscript, 1971.

[10] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in UAI'99, 1999, pp. 467–475.

ISSN 2454-9940
www.ijasem.org
Vol 18, Issue 2, 2024

[11] J. Kleinberg, “Hubs, authorities, and communities,” ACM Computing Surveys, vol. 31, p. 5, 1999.

[12] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” arXiv preprint arXiv:1609.02907, 2016.