IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

INTERNET FINANCIAL FRAUD DETECTION BASED ON A DISTRIBUTED BIG DATA APPROACH WITH NODE2VEC

[1]A NAGARAJU, [2]S. AYESHA

[1](Assistant Professor), MSC, DNR college(A) PG courses Bhimavaram

[2]MSC, scholar, DNR college(A) PG courses Bhimavaram

ABSTRACT

The rapid development of information technologies like Internet of Things, Big Data, Artificial Intelligence, Blockchain, etc., has profoundly affected people's consumption behaviors and changed the development model of the financial industry. The financial services on Internet and IoT with new technologies has provided convenience and efficiency for consumers, but new hidden fraud risks are generated also. Fraud, arbitrage, vicious collection, etc., have caused bad effects and huge losses to the development of finance on Internet and IoT. However, as the scale of financial data continues to increase dramatically, it is more and more difficult for existing rule-based expert systems and traditional machine learning model systems to detect financial frauds from large-scale historical data. In the meantime, as the degree of specialization of financial fraud continues to increase, fraudsters can evade fraud detection by frequently changing their fraud methods. In this article, an intelligent and distributed Big Data approach for Internet financial fraud detections is proposed to implement graph embedding algorithm Node2Vec to learn and represent the topological features in the financial network graph into low-dimensional dense vectors, so as to intelligently and efficiently classify and predict the data samples of the large-scale dataset with the deep neural network. The approach is distributed performed on the clusters of Apache Spark Graph and Hadoop to process the large dataset in parallel. The groups of experimental results demonstrate that the proposed approach can improve the

ISSN 2347–3657

Volume 12, Issue 3, 2024

efficiency of Internet financial fraud detections with better precision rate, recall rate, F1-Score and F2-Score.

1. INTRODUCTION

With the rapid development of the information technologies like Internet of Things, Big Data, Artificial Intelligence, Blockchain, etc., the digital life led by financial technology has profoundly affected people's consumption behaviors and changed the development model of the traditional financial industry to a certain extent [1]. In particular, technical products such as mobile payment, IOT financial services and Internet financial wealth management have penetrated into lots of aspects of economic and social activities. From 2014 to the present, the development momentum of China's Internet consumer finance industry has been good, and various mobile e-commerce companies have entered the consumer finance field through installment payments and small loans, which has promoted the development of related industries. Internet financial services based on consumer credits in China, such as Huabei launched by Ant Financial and Alipay of Alibaba Group, JingdongBaitiao operated by JD.com, WeiLiDai launched by WeBank of Tencent, etc., have enabled consumers to enjoy the online shopping experience of "consumption first, pay later", and covered the e-commerce installment shopping, cash borrowing and other businesses. Especially in 2020, the COVID-19 pandemic [2] has caused a surge in online transaction volume and brought a large number of online customers to online service providers. It has cultivated the habit of more groups of users to make online purchases and payments through mobile phones and IOT devices, which brings continuous impetus to the development of the Internet financial industry.

The rapid development of mobile and IOT financial payment services has not only provided convenience and efficiency for consumers, but also brought more hidden fraud risks. Due to the concealment of the complex network, there could be a breeding ground for fraudulent activities by criminals. The control of fraud risks is becoming more and more difficult and fraud cases occur frequently, which causes the fraud losses to commercial banks and financial institutions are also increasing. The continuous happening of Internet financial fraudulent problems, such as the agreement cash-out incident of Huabei and Taobao merchants,

ISSN 2347–3657

Volume 12, Issue 3, 2024

and "Baitiao" multiple fraud incidents, have not only damaged the legitimate rights and interests of the service platform, but also caused consumers to question the company's account security and risk identification capabilities.

A large number of violations are beyond the scope of the industry's existing laws and regulations, and industry regulation has always lagged behind the innovative development of Internet consumer finance, which makes the regulatory laws and regulations are often in a state of absence so that it impossible to deal with industry violations in a timely manner. Fraud, arbitrage, vicious collection and other phenomena are becoming more and more rampant in online financial service platforms, which has caused bad effects and huge losses to the development of consumer finance on Internet and IOT. Fraud is an illegal or criminal deception aimed at obtaining financial or personal benefits.

Fraud generally has the attributes of abnormal or unfair transactions. Due to the inconsistency with previous fund operation rules or other normal behaviors, fraudulent behavior presents various abnormal characteristics, including abnormal transaction amount, abnormal transaction time, abnormal transaction account, abnormal transaction IP, or abnormal personal credit rating.

Currently, fraud detection schemes in the industry mainly include rule-based expert systems and machine learning-based model systems. The rule-based expert system requires anti-fraud experts to manually analyze a large amount of normal and abnormal transaction data, accurately identify the behavior of fraudsters, find important features that can effectively distinguish fraud, and write expert rules for fraud detection. Therefore, the rule-based expert system strongly relies on the professional knowledge and business knowledge of the anti-fraud experts. If the experts cannot detect increasingly complex fraud patterns in a timely and keen manner, it will cause huge losses. With the continuous increase of machine computing power, model systems based on machine learning have emerged. The machine learning-based model system is generally divided into four modules: data preprocessing, feature engineering, model training and model prediction [3]. Data preprocessing includes missing value processing, sampling and other steps. After

ISSN 2347–3657

Volume 12, Issue 3, 2024

the processing is completed, cumulative calculations are usually performed based on historical transaction data to convert the original data into characteristic data. After that, models such as machine learning regression or classification are used for training and evaluation on the data set. Finally, the model goes online for fraud detection. However, as the scale of financial transaction data continues to increase dramatically, it is more and more difficult for rule-based expert systems and traditional machine learning model systems to detect transaction frauds or fraudulent behavior patterns from large-scale historical data when faced with massive data levels. In the meantime, as the degree of specialization of financial fraud continues to increase, fraudsters can evade fraud detection by frequently changing their own fraud methods. Nevertheless, it is difficult for fraudsters to change all their associated relationships. When the associated network graph can cover a large area, even if a fraudster or fraudulent behavior is careful, it may unwittingly reveal clues. Therefore, in the context of large-scale financial data, how to effectively mine the topological structure characteristics of the association network graph in real time and improve the effect of models for financial fraud detection is a new direction for researchers to explore. In this article, an intelligent and distributed Big Data approach for Internet financial fraud detection is proposed to implement graph embedding algorithm Node2Vec to learn and represent the topological features in the financial network graph into low-dimensional dense vectors, so as to intelligently and efficiently classify and predict the data samples of the large-scale dataset with the deep neural network. The approach is distributedly performed on the clusters of Apache Spark GraphX and Hadoop to process the large dataset in parallel. The groups of experimental results demonstrate that the proposed approach can improve the efficiency of Internet financial fraud detections with better precision rate, recall rate, F1-Score and F2-Score. The rest of the article is organized as follows. Literature of related works is described in Section 2. Section 3 demonstrates the graph embedding algorithm of Node2Vec representation learning. An intelligent and distributed Big Data approach for Internet financial fraud detection is proposed in Section 4. In Section 5, groups of experiments are implemented to evaluate the efficiency of the proposed approach.

ISSN 2347–3657

Volume 12, Issue 3, 2024

Conclusions and future works are summarized in Section 6.

2.EXISTINGSYSTEM

Allen et al. find that there are many credit channels in the United States and based on the research of American household credit models, and that household consumption, household income, credit banks and credit scale are obviously related [7]. Kregel studies the development trend of consumer finance and finds that the development of Internet consumer finance companies must fully consider the current market legal environment, financial market and consumer behavior factors, etc. Internet consumer finance is directly related to the current development of the national financial system [8]. Momparler et al. take the American Internet consumer finance company as the research object, study the risks and advantages of the Internet consumer finance platform, and design a related risk management model [9].

Allen et al. find that there are many credit channels in the United States and based on the research of American household credit models, and that household consumption, household income, credit banks and credit scale are obviously related [7]. Kregel studies the development trend of consumer finance and finds that the development of Internet consumer finance companies must fully consider the current market legal environment, financial market and consumer behavior factors, etc. Internet consumer finance is directly related to the current development of the national financial system [8]. Momparler et al. take the American Internet consumer finance company as the research object, study the risks and advantages of the Internet consumer finance platform, and design a related risk management model [9].

Ficawoyi et al. analyze the positive relationship between Internet exposure levels and credit card default through surveys on consumer finance and income nodes [14]. The research points out that Internet access, low income, and male families are more likely to cause credit card defaults. Giudici et al. propose how to improve credit risk accuracy of P2P Internet financial platforms and of those who lend to small and medium enterprises [15]. The augment traditional credit scoring methods are put forward with “alternative data” that consist of centrality measures derived from similarity networks among borrowers and

ISSN 2347–3657

Volume 12, Issue 3, 2024

deduced from their financial ratios. The experimental findings suggest that the proposed approach improves predictive accuracy as well as model explainability.

Disadvantages

1) The system doesn't support Resilient Distributed Datasets.
2) There is no Directed Acyclic Graph method to find fraud accurately.
3)

3.PROPOSED SYSTEM

Through studying a large number of Internet financial fraud cases, two important characteristics are found: (1) The pattern of Internet financial fraud continues to evolve and develop over time, not just repeating the existing individual behavior patterns appeared in historical cases; (2) With the advancement of anti-fraud technology, it is getting harder for individuals to commit Internet financial fraud. It needs to be organized and conducted through related and connected groups. A graph is an abstract graph formed by a number of nodes and the edges connecting each node [31], [32]. It is usually used to describe a specific relationship between things. A relational network graph refers to a graph-based data structure composed of nodes and edges. Each node represents an entity, and each edge is the relationship between an entity and the other connected entity. The relationship network graph connects different entities together according to their relationships, thus it could provide the ability to analyze problems from the perspective of "relationship"

In anti-fraud applications, entities in the network graph, such as people, equipment, mailboxes, card numbers, etc., can be represented by nodes, and the relationships between these nodes in the business can be represented by edges. Through continuous construction and reproduction of the associated relationships hidden covertly in Internet financial frauds, fraud characteristics can be detected and corresponding risk control strategies can be designed. The graph algorithms can characterize various high-risk features in the Internet finance, such as batch attacks, intermediary participation, etc., which is more effective to identify abnormal group frauds from normal behaviors.

Advantages

1) Node2Vec is a graph embedding algorithm that introduces two biased random walk methods---BFS

ISSN 2347–3657

Volume 12, Issue 3, 2024

(Breadth First Search) and DFS (Depth First Search) on the basis of Deep Walk.

2) AN INTELLIGENT AND DISTRIBUTED BIG DATA APPROACH FOR INTERNET FINANCIAL FRAUD DETECTION.

4. OUTOUT SCREENS

Login page

Registration page:

Admin page

Profile page:

Remote user page:

Predictive bar chart:

ISSN 2347–3657

Volume 12, Issue 3, 2024

5. CONCLUSION

The occurrences of Internet financial fraud cases have caused huge losses to commercial banks or financial institutions. In order to enhance the efficiency of financial fraud detections, an intelligent and distributed Big Data approach is proposed in this article. The approach mainly includes four modules: data preprocessing module, normal data feature module, graph embedding module, prediction module. The graph embedding algorithm Node2Vec is implemented on Spark GraphX and Hadoop to learn and represent the topological features of each vertex in the network graph into a low-dimensional dense vector, so as to improve the classification effectiveness of deep neural network and predict the fraudulent samples of the dataset. The experiments evaluatethe indicators of precision rate, recall rate, F1-Score and F2-Score, and the results show that due to the Node2Vec properties of structural equivalence and homophily, the features of samples can be better learned and represented and the proposed approach is better than the comparative methods. In future work, the inductive graph embedding network algorithms, such as GraphSage, PinSage, etc., would be improved and implemented to effectively learn the features of newly generated vertices in a dynamic network graph, so as to achieve the better effect of financial fraud detection.

6. REFERENCES

[1] U. Paschen, C. Pitt, and J. Kietzmann, “Artificial intelligence: building blocks and an innovation typology,” *Business Horizons*, vol. 63, no. 2, pp. 147-155, 2020.

[2] P. Yu, Z. Xia, J. Fei, and S. K. Jha, “An application review of artificial intelligence in prevention and cure of COVID-19 pandemic,” *CMC-Computers* Materials & Continua, vol. 65, no. 1, pp. 743-760, 2020.

ISSN 2347–3657

Volume 12, Issue 3, 2024

[3] L. Shen, X. Chen, Z. Pan, K. Fan, F. Li, and J. Lei, "No-reference stereoscopic image quality assessment based on global and local content characteristics," *Neurocomputing*, vol. 424, no. 2, pp. 132-142, 2021.

[4] H. Beck, "Banking is essential, banks are not, the future of financial intermediation in the age of the Internet," Netnomics, vol. 3, no. 1, pp. 7-22, 2001.

[5] G. N. Weiss, K. Pelger, and A. Horsch, "Mitigating adverse selection in p2p lending–Empirical evidence from prosper.com," *SSRN Electronic Journal*, vol. 19, no. 7, pp. 65-93, 2010.

[6] Y. Houston, C. Jongrong, J. H. Cliff, and H. Y. Chih, "E-commerce, R&D, and productivity: firm-level evidence from Taiwan," *Information Economics and Policy*, vol. 18, no. 5, pp. 561-569, 2013.

[7] F. Allen, J. Mcandrews, and P. Strahan, "E-finance: an introduction," *Center for Financial Institutions Working Papers*, vol. 22, no. 1, pp. 25-27, 2012.

[8] J. A. Kregel, "Margins of safety and weight of the argument in generating financial fragility," *Journal of Economics Issues*, vol. 6, no. 31, pp. 543-548, 2016.

[9] A. Momparler, C. Lassala, and D. Ribeiro, "Efficiency in banking services: a comparative analysis of Internet-primary and branching banks in the US," *Service Business*, vol. 7, no. 4, pp. 641-663, 2013.

[10] V. Jambulapati and J. Stavins, "Credit card act of 2009: what did banks do?," *Banking & Finance*, vol. 46, no. 9, pp. 21-30, 2014.