International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Special Issue: | Apr 2023 www.irjet.net p-ISSN: 2395-0072

International Conference on Trends in Engineering & Technology- 2023 (ICRTET)
Organised by: VSM College of Engineering, Ramachandrapuram

A HETEROGENEOUS GRAPH TRANSFORMER APPROACH TOFRAUD DETECTION IN ONLINE PRODUCT REVIEW SYSTEMS

1. N.S.C. MOHAN RAO, 2. S. DEVI CHANDRIKA, 3. K. RAKESH REDDY, 4. G. ANUSHA, 5. K.SAI LAKSHMAN, 6. R. DEVI PRAVALLIKA

[1].Associate Professor, [2-6]. Students B.Tech. Computer Science Engineering, V.S.M. College Of Engineering, Ramachandrapuram, A.P, India

---

ABSTRACT:

In online product review systems, users are allowed to submit reviews about their purchased items or services. However, fake reviews posted by fraudulent users often mislead consumers and bring losses to enterprises. Traditional fraud detection algorithm mainly utilizes rule-based methods, which is insufficient for the rich user interactions and graph-structured data. In recent years, graph-based methods have been proposed to handle this situation, but few prior works have noticed the camouflage fraudster's behavior and inconsistency heterogeneous nature. Existing methods have either not addressed these two problems or only partially, which results in poor performance. Alternatively, we propose a new model named Fraud Aware Heterogeneous Graph Transformer (FAHGT), to address camouflages and inconsistency problems in a unified manner. FAHGT adopts a type-aware feature mapping mechanism to handle heterogeneous graph data, then implementing various relation scoring methods to alleviate inconsistency and discover camouflage. Finally, the neighbors features are aggregated together to build an informative representation. Experimental results on different types of real-world datasets demonstrate that FAHGT outperforms the state-of-the-art baselines.

1. INTRODUCTION

Internet services have brought human beings with ecommerce, social networking, and entertainment platforms, which not only facilitate information exchange but also provide chances to fraudsters. Fraudsters disguise themselves as ordinary users to publish spam information or collect user privacy, compromising the interest of both platforms and users. In addition, multiple entities on the Internet are connected with multiple relationships. Traditional machine learning algorithms cannot handle thiscomplicated heterogeneous graph data well. The current approach is to model the data as a heterogeneous information network so that similarities in characteristics and structure of fraudsters can be discovered. Due to the effectiveness in learning the grap representation, graph neural networks (GNNs) have already been introduced into fraud detection areas including product review, mobile application distribution, cyber crime identification and financial services. However, most existing GNN based solutions just directly apply homogeneous GNNs, ignoring the underlying heterogeneous graph nature and camouflage node behaviors. This problem has drawn great attention with many solutions proposed . Graph Consist found that there are three inconsistency problems in fraud detection and CAREGNN further proposed two camouflage behaviors. These problems could be summarized as follows:

Camouflage:

Previous work showed that crowd workers could adjust their behavior to alleviate their suspicion via connectingto benign entities like connecting to highly reputable users, disguise fraudulent URLs with special characters or generate domain-independent fake reviews via generative language model to conceal their suspicious activities.

Inconsistency:

Two users with distinct interests could be connected via reviewing a common product such as food or movies. Direct aggregation makes GNNs hardly distinguish the unique semantic user pattern. Also, if a User is suspicious, then the other one should be more likely to be distrustful if they are connected by common activity relation since fraudulent users tend to post many fraudulent reviews in the same short period.

2. EXISTING SYSTEM

ChebNet and GCN are proposed to improve efficiency by using approximation. For GNNs on spatial domain, GraphSAGE samples a tree rooted at each node and computes the root's hidden representation by hierarchically aggregating hidden node representations from the bottom to top. GAT further proposes to learn in the spatial domain by computing different importance of neighbor nodes via the masked self attention mechanism. All these methods are designed for homogeneous graphs.

They cannot be directly applied to a heterogeneous graph with multiple types of entities and relations.

In recent years, lots of heterogeneous GNN based methods have been developed. HAN , HAHE and Deep-HGNN transforms a heterogeneous graph into several homogeneous graphs based on handcrafted meta-paths, applies GNN separately on each graph, and aggregates the output representations by attention mechanism. Graph Inception constructs meta-paths between nodes with the same object type. Het GNN first samples a fixed number of neighbors via random walk strategy. Then it applies a hierarchical aggregation mechanism for intra-type and intertype aggregation. HGT extendstransformer architecture to heterogeneous graphs. They directly calculate attention scores for all the neighbors of a target node and perform aggregation accordingly without considering domain knowledge.

For relation-aware graph fraud detectors, their main solution is to build multiple homogeneous graphs based on edge type information of the original graph then perform type independent node level aggregation and graph level concatenation. GEM learns weighting parameters for different homogeneous sub graph. Player2Vec and SemiGNN both adopt attention mechanism in feature aggregation and SemiGNN further leverages a structure loss to guarantee the node embeddings homophily. Some works directly aggregate heterogeneous information in the graph. For instance, under a user-review-item heterogeneous graph, GAS learns a unique set of aggregators for different node types and updates the embeddings of each node type iteratively.

Disadvantages:

- In the existing work, the system did not implement Fraud Aware Heterogeneous Graph Transformer (FAHGT) to measure frauds exactly.
- This system is less performance due to lack of META RELATION SCORING.

3. PROPOSED SYSTEM

GraphConsis addresses the inconsistency problem by computing the similarity score between node embeddings, which cannot distinguish nodes with different types. CAREGNN enhances GNN-based fraud detectors against camouflaged fraudsters by reinforcement learning based neighbour selector and relation aware aggregator. Its performance still suffers from the heterogeneous graph.

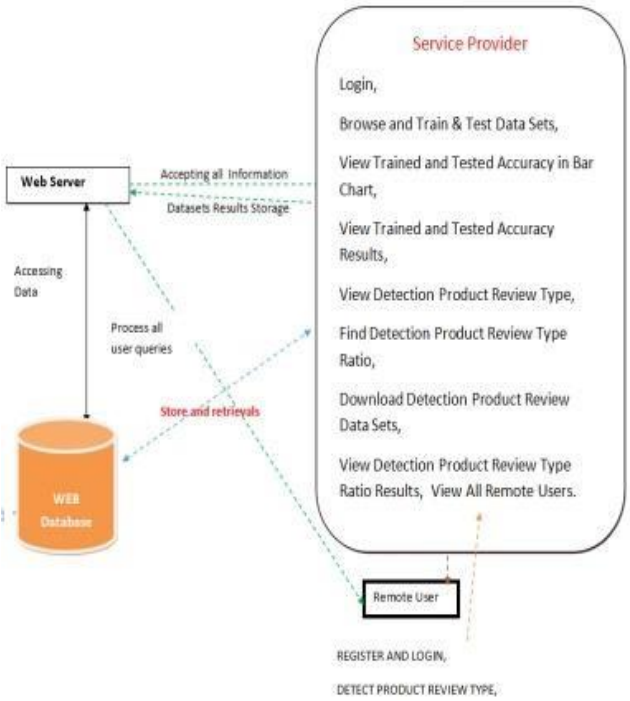
In this paper, the system introduces the Fraud Aware Heterogeneous Graph Transformer (FAHGT), where we propose heterogeneous mutual attention to address the inconsistency problem and design a label-aware neighbor selector to solve the camouflage problem. Both are implemented in a unified manner called the “score head mechanism”. We demonstrate the effectiveness and efficiency of FAHGT on many real world datasets. Experimental results suggest that FAHGT can significantly improve KS and AUC over state- of-the-art GNNs as well as GNN-based fraud detectors.

Advantages

The advantages of FAHGT can be summarized as follows.

- Heterogeneity: FAHGT is able to handle heterogeneous graphs with multi-relation and multi-node type without designing meta-path manually.
- Adaptability: FAHGT attentively selects neighbors given a noise graph from real-world data. The selected neighbors are either informative for feature aggregation or risky for fraud detection.
- Efficiency: FAHGT admits a low computational complexity via a parallelizable multi-head mechanism in relation scoring and feature aggregation.
- Flexibility: FAHGT injects domain knowledge by introducing a flexible relation scoring mechanism. The score of a relation connecting two nodes not only comes from direct feature interaction but is also constrained by domain knowledge.

4. SYSTEM ARCHITECTURE:



5.MODULES IMPLEMENTATION

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 and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Detection Product Review Type, Find Detection Product Review Type Ratio, Download Detection Product Review Data Sets, View Detection Product Review Type Ratio Results, View All Remote 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, DETECT PRODUCT REVIEW TYPE, VIEW YOUR PROFILE.

6. CONCLUSION

In this paper, we propose FAHGT, a 6. novel heterogeneous graph neural network for fraudulent user detection in online review systems. To handle inconsistent features, we adopt heterogeneous mutual attention for automatic meta path construction. To detect camouflage behaviors, we design the label aware scoring to filter noisy neighbors. Two neural modules are combined in a unified manner called "score head mechanism" and both contribute to edge weight computation in final feature aggregation. Experiment results on real-world business datasets validate the excellent effect on fraud detection of FAHGT. The hyper-parameter sensitivity and visual analysis further show the stability and efficiency of our model. In summary, FAHGT is capable of alleviating inconsistency and discover camouflage and thus achieves state-of-art performance in most scenarios. In the future, we plan to extend our model in handing dynamic graphs data and incorporate fraud detection into other areas, such as robust item recommendation in E-commerce or loan default prediction in financial services.

7. REFERENCES

[1] 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, Sydney, NSW, Australia, August 10-13, 2015, L. Cao, C. Zhang, T. Joachims, G. I. Webb, D. D. Margineantu, and G. Williams, Eds. ACM, 2015, pp. 2157–2166. [Online]. Available: https://doi.org/10.1145/2783258.2788623

[2] J. Wang, R. Wen, and C. Wu, "Fdgars: Fraudster detection via graph convolutional networks in online app review system," in WWW Workshops, 2019.

[3] A. Li, Z. Qin, R. Liu, Y. Yang, and D. Li, "Spam review detection with graph convolutional networks," in CIKM, 2019.

[4] Z. Liu, Y. Dou, P. S. Yu, Y. Deng, and H. Peng "Alleviating the inconsistency problem of applying graph neural network to fraud detection," in SIGIR, 2020.

[5] Y. Dou, Z. Liu, L. Sun, Y. Deng, H. Peng, and P. S. Yu, "Enhancing graph neural network-based fraud detectors against camouflaged fraudsters," in CIKM, 2020.

[6] R. Wen, J. Wang, C. Wu, and J. Xiong, "Asa: Adversary situation awareness via heterogeneous graph convolutional networks," in WWW Workshops, 2020.

International Conference on Trends in Engineering & Technology- 2023 (ICRTET)
Organised by: VSM College of Engineering, Ramachandrapuram