International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Identifying Fraud in Online Product Review Systems via Heterogeneous Graph Transformer

Yannam Divya Ramana Reddy, Student Member,M.Tech (CSE); Project guide:
G.ChennaKesava Reddy, M.Tech, Assistant Professor, Srinivasa Institute of Technology and Science, Kadapa.

ABSTRACT

Users can submit reviews about their purchased items or services in online product review systems. Fake reviews, on the other hand, frequently mislead consumers and cost businesses money. Traditional fraud detection algorithms rely primarily on rule-based methods, which are inadequate for rich user interactions and graph-structured data. In recent years, graph-based methods for dealing with this situation have been proposed, but few prior works have observed the camouflage fraudster's behaviour and inconsistency heterogeneous nature. Existing methods have either ignored or addressed these two issues only partially, resulting in poor performance. To address camouflage and inconsistency issues in a unified manner, we propose a new model called Fraud Aware Heterogeneous Graph Transformer (FAHGT).

FAHGT employs a type-aware feature mapping mechanism to handle heterogeneous graph data, followed by the implementation of various relation scoring methods to alleviate inconsistency and detect camouflage. Finally, the features of the neighbours are combined to form an informative representation. The experimental results on various types of real-world datasets show that FAHGT outperforms state-of-the-art baselines.

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 this complicated heterogeneous graph data well.The current approach is to model the data as a heterogeneous information network in order to discover similarities in fraudster characteristics and structure.Due to the effectiveness in learning the graph 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 issue has received a lot of attention, and many

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

solutions have been proposed.GraphConsisfound that there are three inconsistency problems in fraud detection and CAREGNN further proposed two camouflage behaviors. These issues can be summarised as follows:

Camouflage: Previous research demonstrated that crowd workers could adjust their behaviour to alleviate suspicion by connecting to benign entities such as highly reputable users, disguising fraudulent URLs with special characters, or generating domain-independent fake reviews using a generative language model to conceal their suspicious activities.

Inconsistency: Two users with different interests may be linked together by 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 commonactivity relation since fraudulent users tend to post many fraudulent reviews in the same short period.

EXISTING SYSTEM

Approximation is used in ChebNet [14] and GCN [15] to improve efficiency. GraphSAGE [16] samples a tree rooted at each node and computes the root's hidden representation by hierarchically aggregating hidden node representations from bottom to top for GNNs on spatial domain. GAT [17] goes on to propose learning in the spatial domain by calculating the relative importance of neighbour nodes using the masked selfattention mechanism. These methods are all intended for homogeneous graphs. They cannot be applied directly to a heterogeneous graph with various types of entities and relations.

Many heterogeneous GNN-based methods have been developed in recent years. HAN [18], HAHE [19], and Deep-HGNN [20] use handcrafted meta-paths to transform a heterogeneous graph into several homogeneous graphs, then apply GNN separately. GraphInception [21] constructs meta-paths between nodes with the same object type. HetGNN [22] 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 [23] extends transformer 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 [9] learns weighting parameters for different homogeneous subgraph. Player2Vec [7] and SemiGNN [8] 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

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

heterogeneous graph, GAS [3] 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 HeterogeneousGraph Transformer(FAHGT) to measure frauds exactly.
- This system is less performance due to lack of META RELATION SCORING.

PROPOSED SYSTEM

GraphConsist 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 neighbor 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.

Benefits

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.

LITERATURE SURVEY

1. X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu, "Heterogeneous graph attention network," in WWW, 2019, pp. 2022–2032.

As a powerful deep learning-based graph representation technique, graph neural networks have demonstrated superior performance and sparked significant

research interest. It has not, however, been fully considered in graph neural networks for heterogeneous graphs with different types of nodes and links. The heterogeneity and wealth of semantic information present significant challenges when designing a graph neural network for a heterogeneous graph. One of the most exciting recent advances in deep learning is the attention mechanism, which has shown great promise in a variety of fields.We first propose a novel heterogeneous graph neural network based on hierarchical attention, which includes node-level and semantic-level attentions, in this paper. Specifically, node-level attention seeks to learn the significance of a node and its meta-path-based neighbours, whereas semantic-level attention can learn the significance of various meta-paths. The importance of node and meta-path can be fully considered with the learned importance from both node-level and semantic-level attention. The proposed model can then generate node embedding by hierarchically aggregating features from meta-path basedneighbours.Extensive experimental results on three real-world heterogeneous graphs not only demonstrate our proposed model's superior performance over the state-of-the-arts, but also its potentially good interpretability for graph analysis

2. S. Zhou, J. Bu, X. Wang, J. Chen, and C. Wang, "Hahe: Hierarchical attentive heterogeneous information network embedding," arXiv preprint arXiv:1902.01475, 2019.

Heterogeneous information network (HIN) embedding has recently received a lot of attention due to its effectiveness in dealing with complex heterogeneous data. Meta path, which connects different object types with different semantic meanings, is widely used in existing HIN embedding works. However, several issues have yet to be addressed. First, different meta paths convey different semantic meanings, whereas existing works assume that all nodes share the same meta path weights and ignore the personalised preferences of different nodes on different meta paths. Second, given a meta path, nodes in HIN are connected by path instances, whereas existing works fail to fully explore the differences between path instances that reflect nodes' preferences in the semantic space.r To address the aforementioned issues, we propose a Hierarchical Attentive Heterogeneous Information Network Embedding (HAHE) model that captures personalised preferences on meta paths and path instances in each semantic space. Because path instances are based on a specific meta path, a hierarchical attention mechanism is used to model personalised preferences on meta paths and path instances. Extensive experiments on a variety of real-world datasets show that our proposed model model outperforms state-of-the-art methods in a variety of data mining tasks.

3. S. Wang, Z. Chen, D. Li, Z. Li, L.-A. Tang, J. Ni, J. Rhee, H. Chen, andP. S. Yu, "Attentional heterogeneous graph neural network: Application toprogram reidentification," in Proceedings of the 2019 SIAM InternationalConference on Data Mining. SIAM, 2019, pp. 693–701.

A programme or process is an essential component of almost every IT/OT system. Can we rely on the program's identity/ID (for example, executable name)? To avoid detection, malware may disguise itself with the ID of a legitimate programme, and a system tool (e.g., PowerShell) used by the attackers may have the fake ID of another common software, which is less sensitive. However, existing intrusion detection techniques frequently overlook this critical programme reidentification problem (i.e., checking the program's identity). In this paper, we propose an attentional heterogeneous graph neural network model (DeepHGNN) to verify the program's identity based on system behaviours. The central idea is to use the representation learning of the heterogeneous programmebehaviour graph to guide the reidentification process. We create the programme reidentification.

We propose an attentional heterogeneous graph neural network model (DeepHGNN) in this paper to verify the program's identity based on system behaviours. The central idea is to guide the reidentification process by leveraging the representation learning of the heterogeneous programmebehaviour graph. To solve the programme reidentification problem, we formulate it as a graph classification problem and devise an efficient attentional heterogeneous graph embedding algorithm. Extensive experiments using real-world enterprise monitoring data and real attacks demonstrate DeepHGNN's effectiveness across multiple popular metrics as well as its robustness to normal dynamic changes such as programme version upgrades.

4. Z. Hu, Y. Dong, K.Wang, and Y. Sun, "Heterogeneous graph transformer,"in WWW, 2020, pp. 2704–2710.

Graph neural networks (GNNs) have seen increasing success in modelling structured data in recent years. However, most GNNs are designed for homogeneous graphs with identical nodes and edges, making it impossible to represent heterogeneous structures. We present the Heterogeneous Graph Transformer (HGT) architecture for modelling Web-scale heterogeneous graphs in this paper. We create node- and edge-type dependent parameters to characterise the heterogeneous attention over each edge, allowing HGT to maintain dedicated representations for different types of nodes and edges. We designed the heterogeneous mini-batch graph sampling algorithm—HGSampling—for efficient and scalable training on Web-scale graph data. Extensive tests on the Open Academic Graph, which has 179 million nodes and 2 billion edges, show that the proposed method works.The dataset and source code of HGT are publicly available at https://github.com/acbull/pyHGT.

5. G. Wang, S. Xie, B. Liu, and S. Y. Philip, "Review graph based onlinestore review spammer detection," in ICDM. IEEE, 2011, pp. 1242–1247.

Consumers can learn a lot about products and services by reading online reviews. Spammers, on the other hand, are joining the community in an attempt to mislead readers by writing fake reviews.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

Previous spammer detection attempts relied on reviewer behaviours, text similarity, linguistic features, and rating patterns. These studies can identify specific types of spammers, such as those who post numerous similar reviews about a single target entity. However, there are other types of spammers who can manipulate their behaviours to appear to be genuine reviewers and thus are undetectable by the available techniques. We propose a novel concept of a heterogeneous review graph in this paper to capture the relationships between reviewers, reviews, and stores that the reviewers have reviewed.We explore how interactions between nodes in this graph can reveal the cause of spam and propose an iterative model to identify suspicious reviewers. This is the first time such intricate relationships have been identified for review spam detection. We also develop an effective computation method to quantify the trustiness of reviewers, the honesty of reviews, and the reliability of stores. Different from existing approaches, we don't use review text information. Our model is thus complementary to existing approaches and able to find more difficult and subtle spamming activities, which are agreed upon by human judges after they evaluate our results.

We explore how interactions between nodes in this graph can reveal the cause of spam and propose an iterative model to identify suspicious reviewers. This is the first time such intricate relationships have been identified for review spam detection. We also develop an effective computation method to quantify the trustiness of reviewers, the honesty of reviews, and the reliability of stores. Different from existing approaches, we don't use review text information. Our model is thus complementary to existing approaches and able to find more difficult and subtle spamming activities, which are agreed upon by human judges after they evaluate our results.

MODULES OF PROJECT

3.4.1. Admin

Admin must login to this module using a valid username and password. He can perform some operations after successfully logging in, such as View and Authorize Users. Create and View Categories, Create and View Sub-Categories View User Search History, View Fraud Detection On Product Reviews, View Products with Ranks and Comments, View Products Ranks Results.

Viewing and Authorizing Users

In this module, the admin views all users details and authorize them for login permission. User Details such as User Name, Address, Email Id and Mobile Number.

Viewing and Authorizing Users

In this module, the admin views all users details and authorize them for login permission. User Details such as User Name, Address, Email Id and Mobile Number.

3.4.2 User

In this module, there are n numbers of users are present. User should register before doing some. After registration successful he can login by using valid user name and password. Login successful he will do some

operations Register and Login, My Profile, Search Products & Give Review,View Top Products,My Search History.

Viewing Profile Details

In this module, the user can see their own profile details, such as their address, email, mobile number, profile Image.

CONCLUSION

In this paper, we propose FAHGT, a 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.

BIBLIOGRAPHY

[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 Conferenceon 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 withgraph convolutional networks," in CIKM, 2019.

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

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

[7] Y. Zhang, Y. Fan, Y. Ye, L. Zhao, and C. Shi, "Key player identificationin underground forums over attributed heterogeneous information networkembedding framework," in CIKM, 2019.

[8] D. Wang, J. Lin, P. Cui, Q. Jia, Z. Wang, Y. Fang, Q. Yu, and J. Zhou, "Asemi-supervised graph attentive network for fraud detection," in ICDM,2019.

[9] Z. Liu, C. Chen, X. Yang, J. Zhou, X. Li, and L. Song, "Heterogeneousgraph neural

networks for malicious account detection," in CIKM, 2018.
[10] Y. Dou, G. Ma, P. S. Yu, and S. Xie, "Robust spammer detection by nash reinforcement learning," in KDD, 2020.