International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 1772-1775 www.ijaem.net ISSN: 2395-5252

Fraud Identification in Online Product Review Systems

A. Swarna latha[1], P.V.R.R.K.Sri Charan[2], P.Satish[3], G.Prudhvi sai[4], T.Sai Charishma[5],

Department of Computer Science and Engineering, Raghu Institute of Technology, Visakhapatnam, Andhra Pradesh, India.

Submitted: 05-06-2022 Revised: 17-06-2022 Accepted: 20-06-2022

ABSTRACT: When looking at internet product reviews, there are many that are fraudulent and utilized to lower the product's value. Users have the ability to leave feedback on products or services they have purchased. Fake reviews made by unscrupulous people, on the other hand, may mislead consumers and cause businesses to lose money. The rule-based technique used by most fraud detection algorithms is insufficient for large user interactions and graph-structured data. In recent years, graph-based strategies or approaches have been presented to deal with situations like this, but some previous research has noted the camouflage fraudster's inconsistent behavior. Existing approaches may not solve both difficulties or may only address one of them, resulting in poor performance. To solve the camouflages and inconsistent difficulties properly, we present a novel model called Fraud Aware Heterogeneous Graph Transformer (FAHGT). To handle the graph data, FAHGT uses a type-aware feature mapping procedure, followed by several relation scoring methods to overcome inconsistencies and uncover fraudulent users. Finally, the features of the neighbors are combined to provide an informative representation.

I. INTRODUCTION:

Humans are forced to use e-commerce, social networking apps, and entertainment platforms as a result of internet services, which not only reduce the chances of information transmission but also give opportunities for impostors. These fraudsters pose as ordinary users in order to publish spam [1] or get user privacy, as well as to give and take the interests of both platform and end users. Furthermore, many Internet units will be associated with multiple relationships. This convoluted heterogeneous graph data is too difficult for typical machine learning techniques to manage. The present technique is to describe the data as a heterogeneous information network in order to detect similarities in fraudsters' traits and basic structure. Graph neural networks (GNNs) have already been used in fraud detection domains such as product review systems [2]–[5], mobile application distribution [6], cyber crime detection [7], and financial services [8],[9] due to their effectiveness in learning the graph representation. However, existing GNN-based solutions will only use homogeneous GNNs, ignoring the underlying heterogeneous graph nature and hide node features. This problem has gotten a lot of attention, and there have been a lot of solutions presented [4],[5],[10]. We discovered three inconsistency problems in fraud identification in the graphs consis problem [4], and CAREGNN [5] proposed two camouflage behaviors. These issues can be summed up as follows:

i. Camouflage: Previous research has shown that crowd workers can adjust their behavior to reduce their suspicion by connecting to tender entities such as highly esteemed users, disguising fraudulent links with special characters or symbols [3],[6], or creating domain-independent fake reviews using a generative model [11].

ii. Inconsistency: Reviewing a similar product or service, such as movies or gadgets, could unite multiple users with well-defined preferences. Direct assembling makes it difficult for GNNs to recognize a single meaningful user pattern. Furthermore, if one person appears suspicious, the other should be wary if the two are linked by a common action, as fraudulent users are more likely to write numerous fraud reviews in a short period of time.

II. LITERATURE SURVEY:

1) ChebNet [14] and GCN [15] are two more approaches that use approximation to improve efficiency. GraphSAGE [16] examines a tree rooted at each node for GNNs on a contiguous

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 1772-1775 www.ijaem.net ISSN: 2395-5252

domain and computes the root's hidden representation by hierarchically aggregating hidden node representations from the bottom to the top. GAT [17] also suggests learning in the spatial domain by using the masked self attention method to compute varied relevance of neighbor nodes. All of these strategies are intended for graphs that are homogeneous. They can't be immediately applied to a graph with several sorts of entities and relations because they're heterogeneous.

2) Several heterogeneous GNN-based approaches have been developed in recent years. HAN [18], HAHE [19], and Deep-HGNN [20] use constructed meta-paths to split a heterogeneous graph into numerous homogeneous graphs, perform GNN independently on each graph, and aggregate the output representations using an attention method.
3) Graph Inception [21] creates meta-paths between nodes that share an object type. HetGNN [22] uses a random walk approach to sample a fixed number of neighbors. Then, for intra-type and intertype aggregation, it uses a hierarchical aggregation approach. HGT [23] adds heterogeneous graphs to the transformer design. They calculate attention scores for all of a target node's neighbors and aggregate them without taking domain knowledge into account.

III. PROBLEM STATEMENT:

Users can leave reviews on things or services they've purchased using online product review systems. Fake reviews, on the other hand, frequently mislead consumers and cause businesses to lose money. After reading the reviews, the customer may conclude the product isn't very good. This causes the product to be less popular on the market, resulting in a loss for the manufacturer. Graph-based approaches have been presented in recent years to deal with this problem, however few previous studies have taken into account the camouflage fraudster's behavior and inconsistent diverse character.

IV. METHODOLGY & IMPLEMENTATION:

The inconsistency problem is addressed by GraphConsis, which computes the similarity score between node embeddings, which cannot discriminate nodes of different types. CAREGNN uses a reinforcement learning-based neighbor selector and relation aware aggregator to improve GNN-based fraud detectors against disguised fraudsters. Despite this, the graph's heterogeneity limits its performance.

We propose heterogeneous mutual attention to resolve the inconsistency problem and create a label-aware neighbor selector to overcome the camouflage problem in the Fraud Aware Heterogeneous Graph Transformer (FAHGT). Both are combined into a single device known as the "score head mechanism." On a variety of real-world datasets, we demonstrate the efficacy and efficiency of FAHGT. Experimental results show that FAHGT outperforms state-of-the-art GNNs and GNN-based fraud detectors in terms of KS and AUC.

In this study, we used four different types of classifiers. They are

1) Naïve Bayes
2) Logistic Regression
3) SVM (Support Vector Machine)
4) Decision Trees

Following a review of the data, the classifier that delivers the highest accuracy among the others is chosen as the main classifier for our project. Following this, it will mostly employ the classifier with the highest accuracy and test the review entered by the end user.

This will assist others in immediately identifying the bogus reviews.

The training of the data set is the most important step in this implementation. We'll train the data set once the classifier is fixed. We may also download the data set after we've trained it. The product name, product text (review), the rating that unique user gave to that product, and the time when he delivered the review are the major components that were important for the review classification.

The data mining method of feature extraction from the raw data of the data set will be based on these parameters.

We can now test the review by inserting it in a specific text field after we've trained the data set.

The key variables will determine whether or not the review is fake. If we modify the data set, we must also change the variables in the files and match them to the new data set. When a user enters a review, it will be separated into different segments containing only words, which will aid the algorithm in swiftly classifying the review.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 6 June 2022, pp: 1772-1775 www.ijaem.net ISSN: 2395-5252

4.1. Architecture:

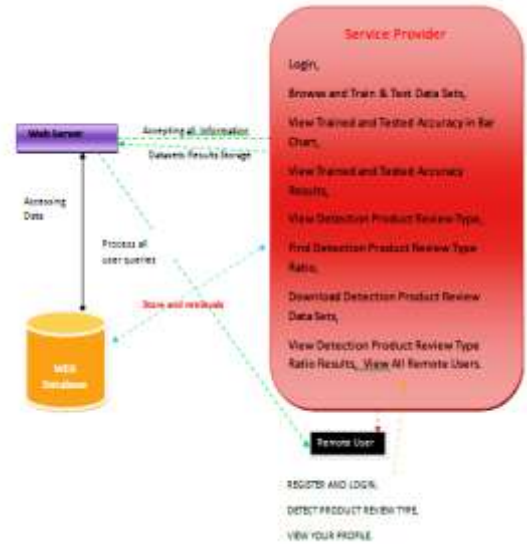
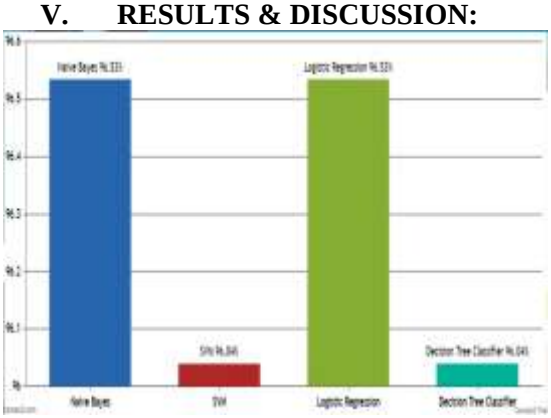
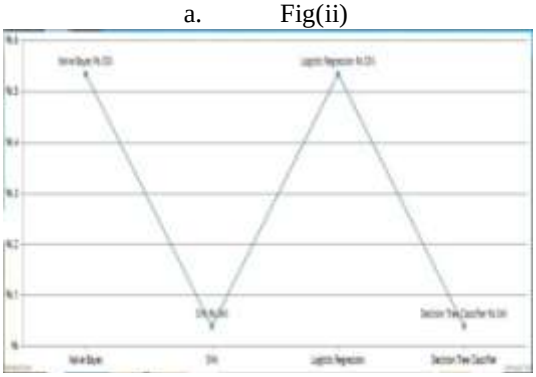


Fig (i)

V. RESULTS & DISCUSSION:

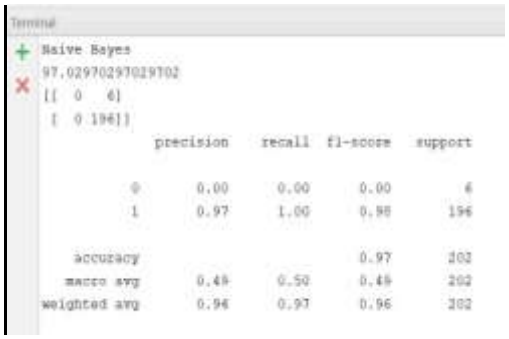


a. Fig(ii)



b. Fig(iii)

The nave Bayes algorithm provided the best accuracy for the data set, as shown in the graphs I and (ii). This will assist us in comprehending and employing the nave Bayes method for determining whether or not the reviews are fraudulent.



```
Naive Bayes
97.02970297029702
[[  0   6]
 [  0 196]]
              precision    recall  f1-score   support

           0       0.00      0.00      0.00         6
           1       0.97      1.00      0.98       196

    accuracy                           0.97       202
   macro avg       0.49      0.50      0.49       202
weighted avg       0.96      0.97      0.96       202
```

c. Fig(iii)

This is the nave bayes algorithm's confusion matrix, which gave the review system a score of 97.03 percent.

VI. CONCLUSION & FUTURE WORK:

We propose FAHGT, an unique heterogeneous graph neural network for detecting fraudulent users in online review systems, in this paper. For automatic Meta path generation, we use heterogeneous mutual attention to address inconsistent characteristics. We created label aware scoring to filter out loud neighbors in order to discover camouflage activities. In the final feature aggregation, two neural modules are merged in a unified fashion called the "score head mechanism," and both contribute to edge weight computation. The good efficacy of FAHGT on fraud detection has been validated by experiment findings on real-world company datasets. Our model's stability and efficiency are further demonstrated by the hyper-parameter sensitivity and visual examination. In summary, FAHGT is capable of removing inconsistency and detecting camouflage, resulting in state-of-the-art performance in the majority of circumstances. We intend to expand our model's ability to handle dynamic graph data and incorporate fraud detection into other domains, such as robust item suggestion in e-commerce or loan default prediction in financial services, in the following days.

REFERENCES:

[1]. V. S. Tseng, J. Ying, C. Huang, Y. Kao, and K. Chen, “Fraud detector: 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.
[7]. Y. Zhang, Y. Fan, Y. Ye, L. Zhao, and C. Shi, "Key player identification in underground forums over attributed heterogeneous information network embedding framework," in CIKM, 2019.
[8]. D. Wang, J. Lin, P. Cui, Q. Jia, Z. Wang, Y. Fang, Q. Yu, and J. Zhou, "A semi-supervised graph attentive network for fraud detection," in ICDM,2019.
[9]. Z. Liu, C. Chen, X. Yang, J. Zhou, X. Li, and L. Song, "Heterogeneous graph 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.
[11]. P. Kaghazgaran, M. Alfifi, and J. Caverlee, "Wide-ranging review manipulation attacks: Model, empirical study, and countermeasures," in CIKM, 2019.
[12]. Z. Zhang, P. Cui, and W. Zhu, "Deep learning on graphs: A survey," TKDE, 2020.
[13]. J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," arXiv preprint arXiv: 1312.6203, 2013.
[14]. M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering" in NeurIPS, 2016, pp. 3844–3852.
[15]. T. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks" in ICLR, 2017.
[16]. W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs" in NeurIPS, 2017.
[17]. P. Veli˘ckovi´c, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks" in ICLR, 2017.
[18]. 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.
[19]. S. Zhou, J. Bu, X. Wang, J. Chen, and C. Wang, "Hahe: Hierarchical attentive heterogeneous information network embedding" arXiv preprint arXiv: 1902.01475, 2019.
[20]. S. Wang, Z. Chen, D. Li, Z. Li, L.A. Tang, J. Ni, J. Rhee, H. Chen, and P. S. Yu, "Attentional heterogeneous graph neural network: Application to program reidentification" in Proceedings of the 2019 SIAM International Conference on Data Mining. SIAM, 2019, pp. 693–701.
[21]. Y. Zhang, Y. Xiong, X. Kong, S. Li, J. Mi, and Y. Zhu, "Deep collective classification in heterogeneous information networks," in Proceedings of the 2018 World Wide Web Conference, 2018, pp. 399–408.
[22]. C. Zhang, D. Song, C. Huang, A. Swami, and N. V. Chawla, "Heterogeneous graph neural network" in KDD, 2019, pp. 793–803.
[23]. Z. Hu, Y. Dong, K.Wang, and Y. Sun, "Heterogeneous graph transformer," in WWW, 2020, pp. 2704–2710.