Enhancing Graph Neural Network-based Fraud Detectors: A Redesign and Evaluation

Seyyed Masoud RezvaniNejad

Abstract

One of the most important applications of machine learning is fraud detection, where these scammers camouflage their behavior in order to avoid certain detection methods. The Camouflage-Resistant Graph Neural Network, invented for identifying camouflaged fraudsters from a multi-relation graph, is revisited with further improvements in this paper. We implemented and evaluated a variant of the model within computational constraints, inspired by the proposed framework of CARE-GNN. The key components of CARE-GNN, the adjustments in our redesign, and the challenges faced in its implementation are discussed in this work. These findings highlight future areas of development and demonstrate the potential of GNNs to combat smart fraud.

1 Introduction

The increasing prevalence of fraud in online platforms necessitates robust detection methods. Fraudsters often camouflage their behavior by mimicking legitimate users, making detection challenging. Graph Neural Networks (GNNs) have shown promise in tackling such issues due to their ability to aggregate information from graph-structured data.

The CARE-GNN framework addresses fraudster camouflage through:

- **Label-aware Similarity Measure**: Identifying informative neighbors using supervised learning.
- **Reinforcement Learning-based Neighbor Selection**: Dynamically filtering irrelevant nodes.
- **Relation-aware Aggregation**: Aggregating information across multiple relations.

In this project, we redesigned the CARE-GNN framework to explore its efficacy under limited computational resources. Our implementation includes the adaptation of core components, testing on real-world datasets, and an evaluation of the model's performance.

2 Model

2.1 CARE-GNN Framework

The CARE-GNN model introduces the following innovations to detect camouflaged fraudsters:

- **Label-aware Similarity Measure:** Computes the similarity between nodes based on domain-specific labels and features, helping to distinguish fraudsters from benign entities.
- **Reinforcement Learning (RL) for Neighbor Selection:** Learns the optimal number of neighbors to consider during aggregation using an RL-based thresholding mechanism.
- **Relation-aware Aggregation:** Aggregates information from different node relations to enhance detection accuracy.

2.2 Redesign and Implementation

Our redesign focused on implementing a simplified version of CARE-GNN due to computational constraints:

- Adapted the similarity measure using lightweight neural modules.
- Simplified the RL module to reduce computational overhead while maintaining the dynamic selection of neighbors.
- Aggregated information using standard techniques without introducing additional complexity.

The implementation was carried out in a Jupyter Notebook ('GNN_Fraud_Detection.ipynb') and followed the workflow described in the accompanying 'README.md'.

2.3 Datasets and Experiments

We tested the model on two real-world datasets:

- **Yelp Dataset:** Includes hotel and restaurant reviews with labeled fraudulent and legitimate reviews.
- **Amazon Dataset:** Contains product reviews, where fraudulent users are labeled based on feedback patterns.

The experiments were limited by GPU resources, which restricted the size of datasets and model complexity. Metrics such as AUC and Recall were used to evaluate the model's performance.

3 Conclusion

This work successfully implemented a redesigned CARE-GNN model for fraud detection. While our results demonstrate the potential of GNNs in addressing camouflaged fraud, the limitations imposed by computational resources impacted the scale and robustness of our experiments. Future work could focus on:

- Leveraging cloud-based GPUs for large-scale experiments.
- Exploring lightweight GNN architectures to reduce computational demands.
- Integrating additional domain-specific features for improved fraud detection.

4 References

- Yingtong Dou, Zhiwei Liu, Li Sun, Yutong Deng, Hao Peng, Philip S. Yu. Enhancing Graph Neural Network-based Fraud Detectors against Camouflaged Fraudsters. *ACM CIKM 2020.* `https://github.com/YingtongDou/CARE-GNN`.
- CARE-GNN GitHub Repository: `https://github.com/YingtongDou/CARE-GNN`.
- The implementation file: 'GNN_Fraud_Detection.ipynb' : `https://github.com/smasoudrezvani/smasoudrezvani.github.io/blob/main/assets/Powerpoint/GNN_Fraud_Detection.ipynb`