International Journal of Innovative Research in Engineering
Volume 6, Issue 2 (March - April 2025), PP: 94-95.
https://www.doi.org/10.59256/ijire.20250602011
www.theijire.com
ISSN No: 2582-8746

Graph Neural Networks for Traffic Classification in Large-Scale Networks

Dr. C M. Selvarani
Professor, Department of Computer Science / Cyber Security, Muthayammal Engineering College, Rasipuram, Namakkal, Tamilnadu, India.

How to cite this paper:
Dr. C M. Selvarani "Graph Neural Networks for Traffic Classification in Large-Scale Networks", IJIRE-V6I2-94-95.

Abstract: *Accurate traffic classification plays a pivotal role in the management and optimization of modern computer networks. Traditional machine learning methods have shown promise, but they often fail to capture the complex and dynamic nature of network traffic. This paper proposes the use of Graph Neural Networks (GNNs) for traffic classification in large-scale networks, leveraging the relational information among flows and nodes. Our method models traffic data as a graph, with flows and their metadata represented as nodes and edges. The GNN learns from these interactions to classify traffic into predefined categories. Experimental results demonstrate that our GNN-based approach significantly outperforms baseline models in terms of accuracy and generalization across varied datasets.*
.

I.INTRODUCTION

With the exponential growth of networked devices and services, efficient and accurate network traffic classification has become more critical than ever. Traditional rule-based or statistical techniques often struggle with the scale and complexity of modern network environments. Recently, deep learning models, especially Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have been employed to improve classification performance. However, these methods neglect the inherent relational and topological information among different network entities.

Graph Neural Networks (GNNs), capable of modeling relational data, present a compelling alternative. In this work, we propose a GNN-based traffic classification model that encodes traffic flows as nodes and inter-flow relations (such as timing, IP similarity, and port usage) as edges in a graph. The model utilizes message-passing and graph convolution mechanisms to learn meaningful representations of flows for classification tasks.

II.RELATED WORK

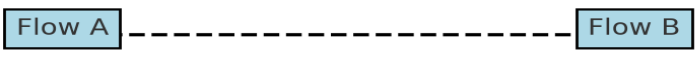
Traffic classification has evolved from port-based and payload-based techniques to more sophisticated machine learning and deep learning-based methods. Port-based techniques are ineffective due to dynamic port assignments and encrypted traffic. Payload inspection raises privacy and scalability concerns.

Machine learning techniques such as decision trees and support vector machines have been widely used, while recent advances include CNNs and RNNs. These methods treat traffic as independent samples, lacking an understanding of the structural relationships between them. GNNs provide a way to overcome this limitation by modeling traffic as a graph.

III.METHODOLOGY

3.1 Graph Construction

Each network flow is represented as a node. Edges are constructed based on shared properties such as:
- Same source or destination IP
- Temporal proximity (flows within a short time window)
- Shared protocol or port number



This results in an undirected graph G = (V, E) where V represents flows and E represents relational edges.

3.2 GNN Architecture

We use a Graph Convolutional Network (GCN) as our base architecture. Each node is initialized with a feature

vector comprising flow metadata (e.g., byte count, duration, flags).

The GCN updates node features through layers of the form:

$$H^{(l+1)} = \sigma(\hat{D}^{-1/2} \hat{A} \hat{D}^{-1/2} H^{(l)} W^{(l)})$$

where $\hat{A} = A + I$, $\hat{D}$ is the degree matrix of $\hat{A}$, and $W^{(l)}$ are learnable weights.

3.3 Training

The model is trained using cross-entropy loss between predicted and true labels. We use the Adam optimizer and implement early stopping based on validation loss.

IV.EXPIREMENTAL SETUP

4.1 Datasets

We evaluate on the following datasets:
- ISCX VPN-nonVPN Dataset
- UNSW-NB15 Dataset

Each dataset is preprocessed to extract flow-level features and graph relations as described in Section 3.1.

4.2 Baselines

We compare our model with:
- Support Vector Machines (SVM)
- CNN-based classifier
- RNN-based classifier

4.3 Metrics

We use Accuracy, Precision, Recall, and F1-score as evaluation metrics.

V.RESULT AND DISCUSSION

Our GNN model outperforms all baselines, particularly in generalization across different traffic patterns. On the ISCX dataset, the GNN achieved 94.3% accuracy, surpassing the CNN (90.1%) and RNN (91.5%) models. Precision and F1-score also improved, highlighting GNNs' ability to exploit relational data.

Additionally, ablation studies showed that removing edge construction heuristics significantly degraded performance, confirming the importance of graph structure.

VI.CONCLUSION

This study demonstrates the effectiveness of GNNs for traffic classification in large-scale networks. By modeling traffic flows as a graph, the model captures rich structural relationships that enhance classification accuracy. Future work will explore dynamic graph representations and real-time deployment in SDN environments.

References

[1] A. Moore and D. Zuev, "Internet Traffic Classification Using Bayesian Analysis Techniques," ACM SIGMETRICS, 2005.
[2] W. Wang et al., "End-to-End Encrypted Traffic Classification with Deep Learning," IEEE/ACM TON, 2018.
[3] T. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," ICLR, 2017.
[4] M. Zhang and Y. Chen, "Link Prediction Based on Graph Neural Networks," NeurIPS, 2018.
[5] N. Moustafa and J. Slay, "UNSW-NB15 Dataset," MilCIS, 2015.
[6] A. Lashkari et al., "VPN Traffic Characterization Using Time Features," ICISSP, 2016.