Multi Label Node Classification for Predicting Brands in Tweets

2024 Spring Symposium

April 16, 2024

Department of Management Information System & Supply Chain Management

Taehyun Noh, Jimin Go, TackKyoon Woo, Seokwoo Song*

WEBER STATE UNIVERSITY

Table of Contents

Introduction

- In Social Network Analysis (SNA), social networks can be characterized by the sharing of information among users and the formation of connections based on common interests.
- As user interaction on social networks has become increasingly vibrant, monitoring user reactions on brands in social media has become essential for companies in formulating their marketing strategies.
- However, it can be challenging to determine the relevance of a post to a particular brand when the social media post doesn't indicate anything about the brand.
- This study attempts to compare multiple "graph-based" machine learning models for node classification.
- Further, we propose methods that combine Deep learning techniques with those existing models, to predict which brand a post is related to, even when the brand name is not mentioned in the post.

Related Work

No.	Title	Method	Prediction Type
1	The rendezvous algorithm:: Multiclass semi-supervised learning with Markov random walks (2007)	Heuristic	Single Label
2	A novel multi-label classification algorithm based on K-nearest neighbor and random walk (2020)	Heuristic	Multi Label
3	Semi-supervised classification with graph convolutional networks. (2016)	GCN	Single Label
4	Graph Attention networks (2017)	GAT	Multi Label
5	Inductive representation learning on large graphs (2017)	GraphSAGE	Single Label
6	Optimization of graph neural networks with natural gradient descent (2020)	SSP	Multi Label

Table 1. Related Work

- Current researches in node classification many of them focus on predicting a single label. In this study, we aim to conduct multi-label node classification predicting a relatively large number of classes.

Proposed Method

- Basic Mechanism of Proposed Method

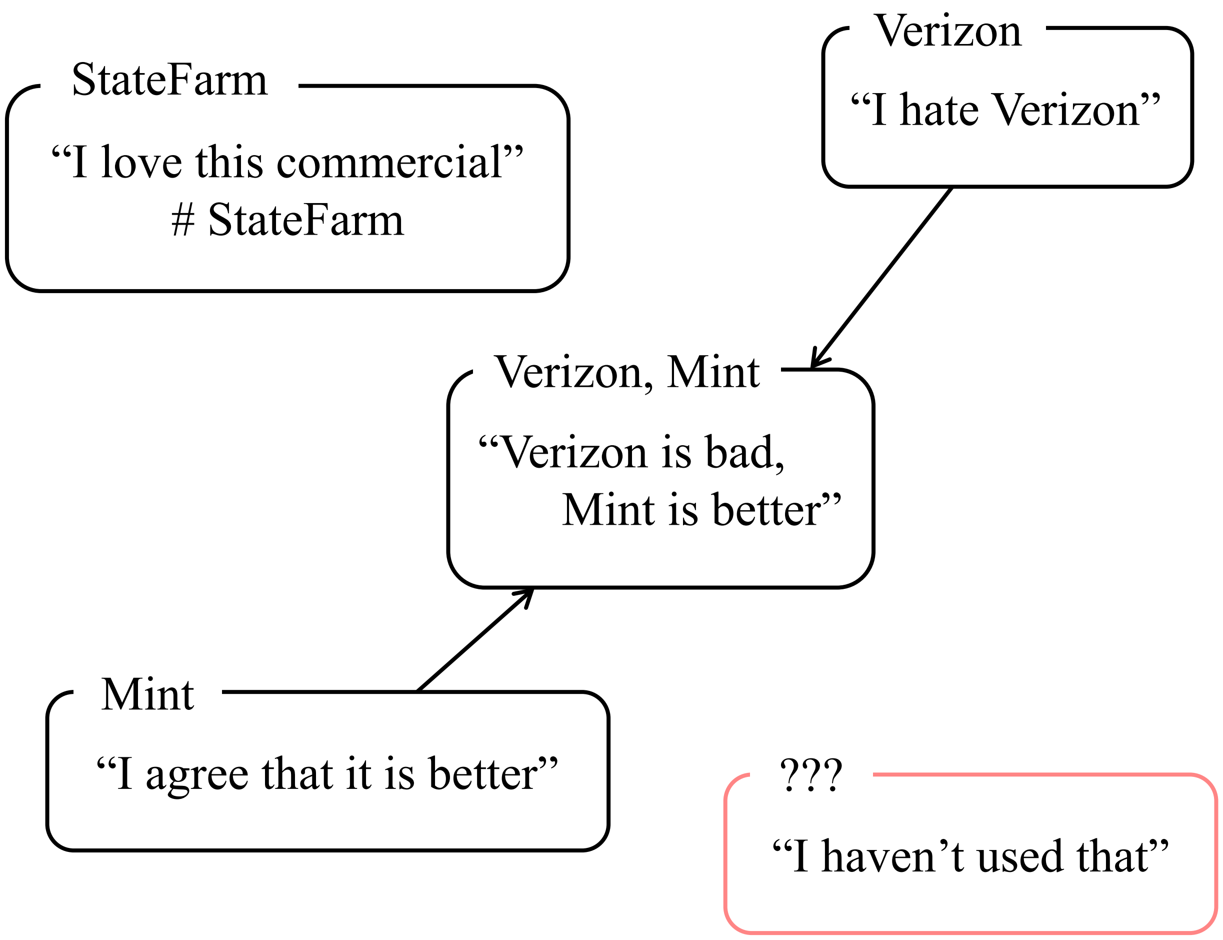

Fig 1. Example of twitter network

- If there are no mentions of a brand in the tweet, it’s difficult to figure out which brand is related with corresponding tweet just by looking at text features.
- However, if we consider network features, we could identify links between tweets.
- Therefore, by considering both features, it is possible to predict which tweets are related to a specific brand.

Proposed Method

- Basic Mechanism of Proposed Method

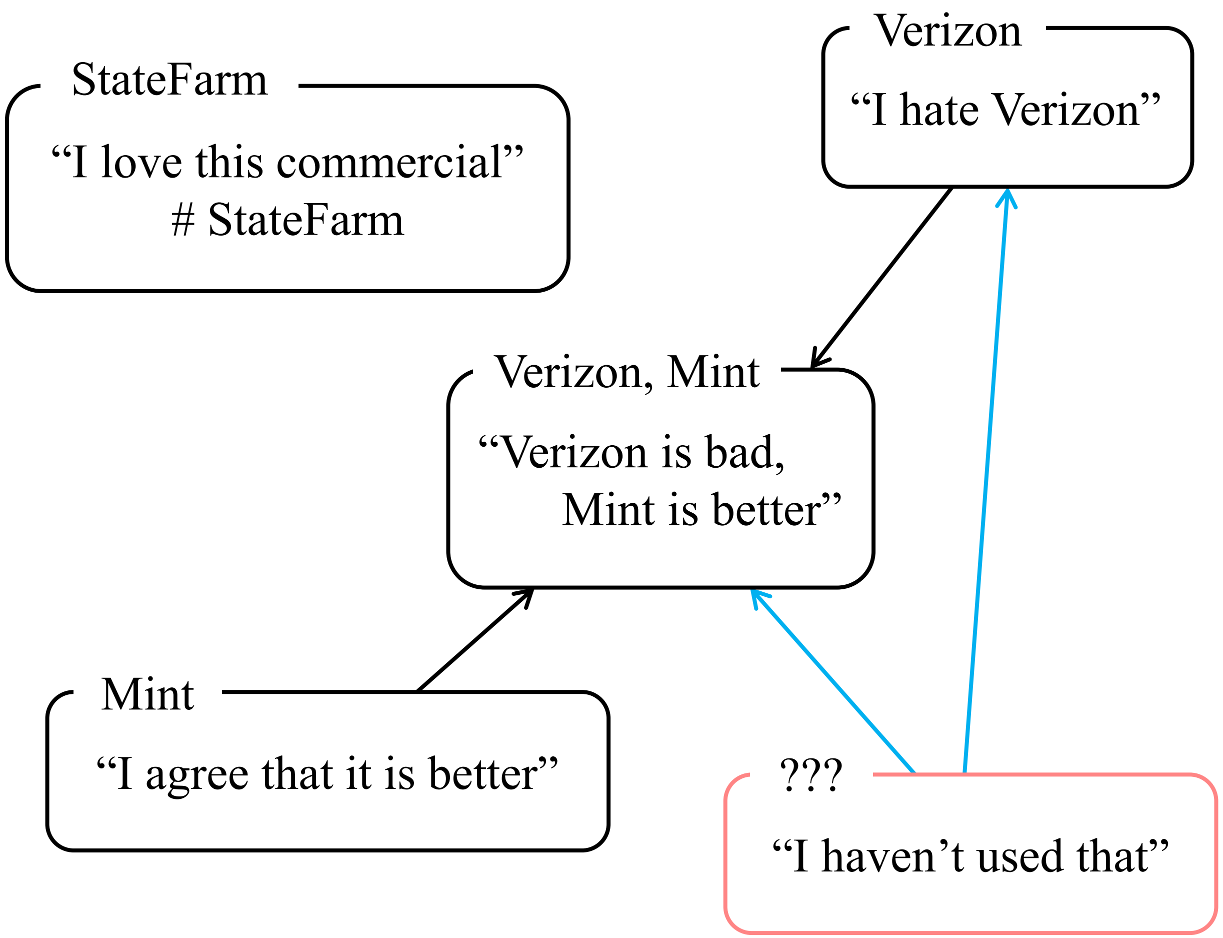

Fig 1. Example of twitter network

- If there are no mentions of a brand in the tweet, it's difficult to figure out which brand is related with corresponding tweet just by looking at text features.
- However, if we consider network features, we could identify links between tweets.
- Therefore, by considering both features, it is possible to predict which tweets are related to a specific brand.

Proposed Method

- Existing Graph Neural Network (GNN) Model Architecture

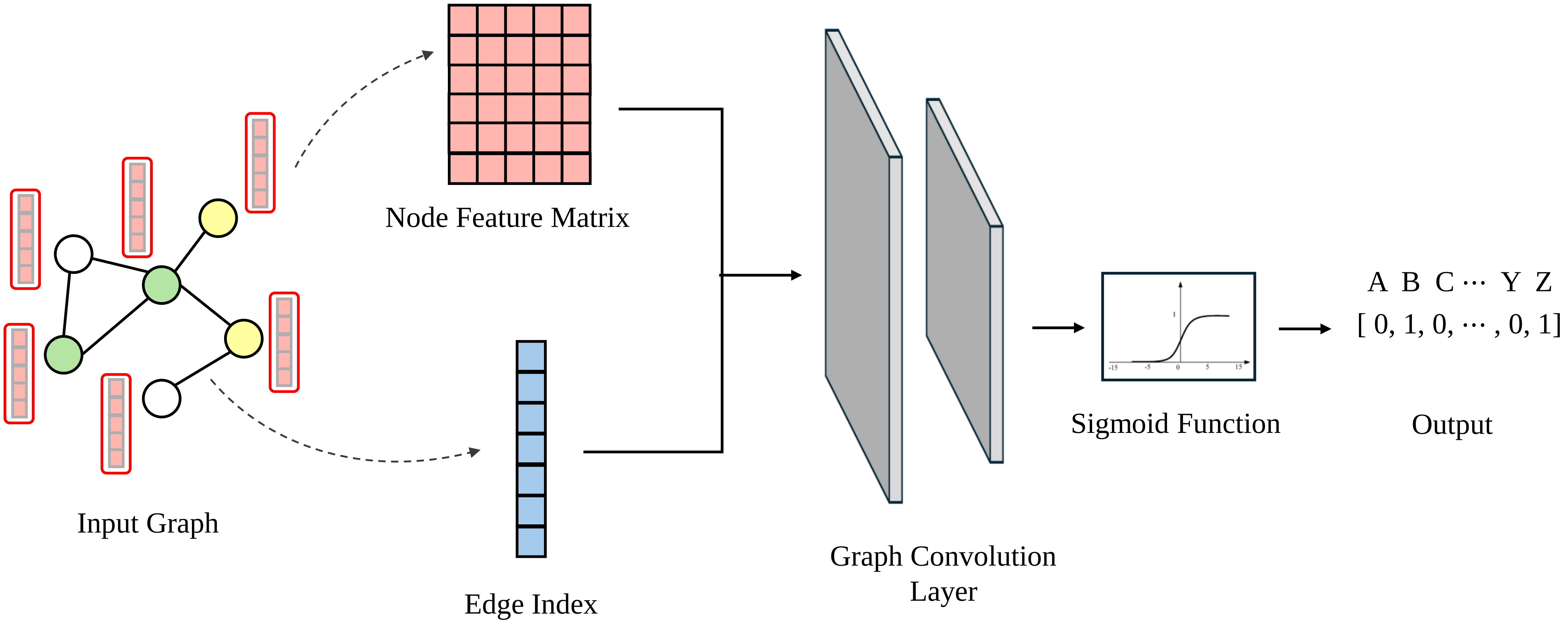

Fig 2. Existing GNN Model Architecture

Proposed Method

- Proposed Model Architecture

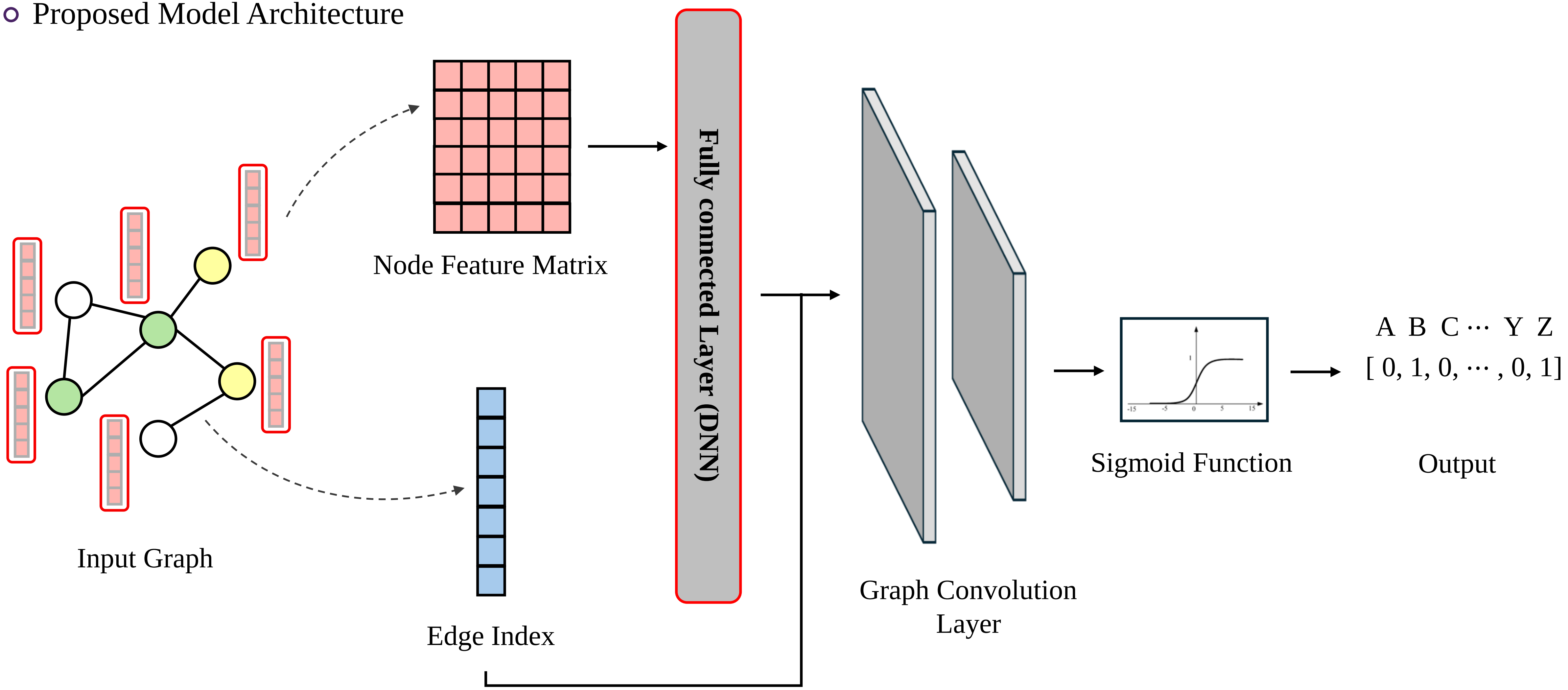

Fig 3. Proposed Model Architecture

Experiments

- Dataset
  - We have 14,180 tweets mentioning brands that advertised during the 2024 Super Bowl game.
  - We conducted experiments with the top 31 brands, excluding the top 2 brands, as NFL is not a brand and Temu has heavily skewed information.
  - The dataset contains a total of 6,096 tweets, with 4.8% of them having multi-labels.

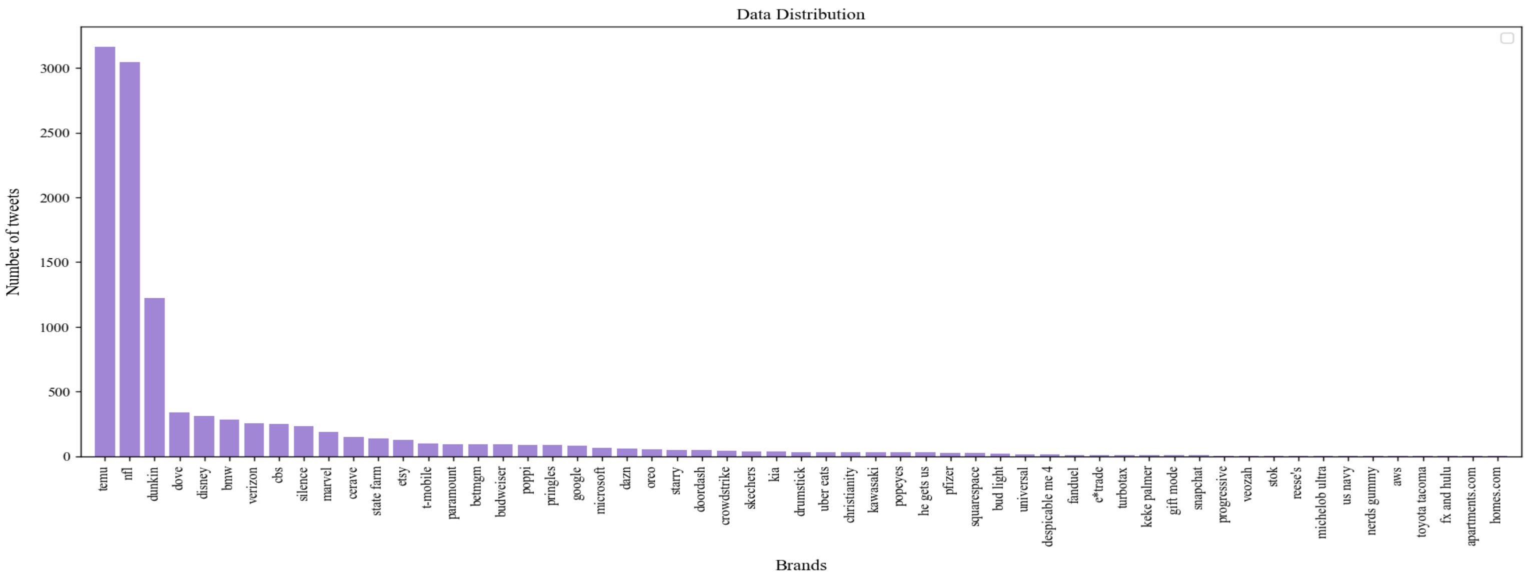

Fig 4. Entire Data Distribution

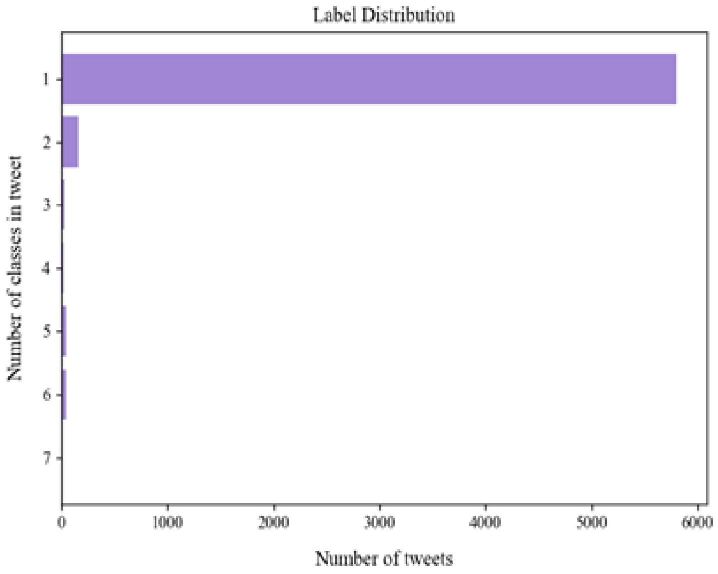

Fig 5. Label Distribution

Experiments

- Experiment Settings
  - In this study, nodes represents each tweets and edges are derived by mentions and retweets between nodes.
  - In graph-based learning, it is crucial to determine what information to embed for the features of nodes.
  - We conducted two different experiments to investigate the impact of retweets as node features :

    - For the first experiment, we only considered information of hashtag, author and text of tweet as node features, without the text of retweeted tweet.

    - Second experiment, we included the text of retweeted tweets into node features, with hashtag, author and text of tweet.

Experiments

- Evaluation Metrics
  - When evaluating classifier model, metrics such as accuracy, precision, recall and F1 score are commonly measured based on the confusion matrix.
  - However, since these metrics are suitable for binary classifier model, we utilized the micro and macro values of each metric to evaluate the performance of multi classifier model.

    - Micro: calculate metrics globally by counting the total true positives, false negatives, and false positives.

    - Macro: calculate metrics for each label and find their unweighted mean.

Experiments

- Without RT Experiment Results

	Micro-F1	Macro-F1	Macro-ROC	Average Precision
SSP	0.4712	0.1294	0.7611	0.1952
GCN	0.4332	0.1381	0.7263	0.1687
SAGE	0.4066	0.1271	0.7548	0.1710
SSP+DNN	0.5258	0.2519	0.8009	0.3061
GCN+DNN	0.4346	0.1157	0.7244	0.1516
SAGE+DNN	0.4280	0.1527	0.7932	0.2068

Table 2. Without RT Experiment Results

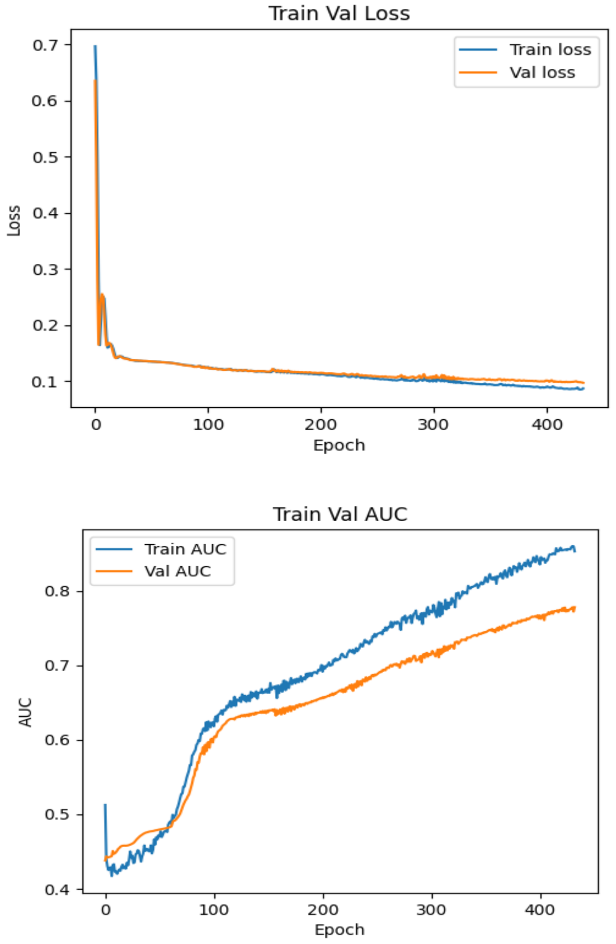

Fig 6. Best Model Learning Curve

Experiments

- With RT Experiment Results

	Micro-F1	Macro-F1	Macro-ROC	Average Precision
SSP	0.7603	0.2814	0.9142	0.3927
GCN	0.5829	0.1532	0.8476	0.2242
SAGE	0.5856	0.1766	0.8705	0.2600
SSP+DNN	0.7858	0.3329	0.9215	0.4572
GCN+DNN	0.7275	0.2928	0.9409	0.4012
SAGE+DNN	0.7797	0.3541	0.9520	0.4564

Table 3. With RT Experiment Results

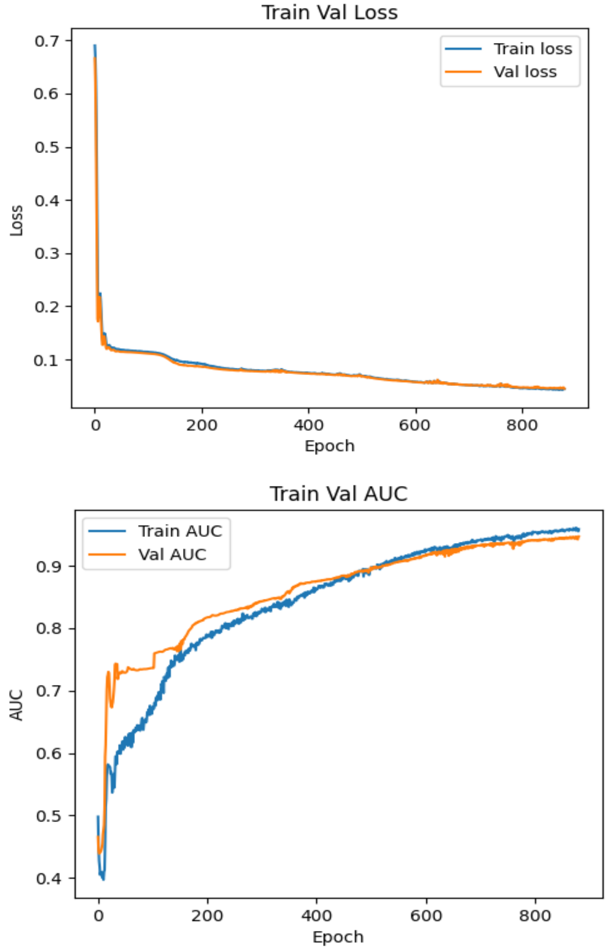

Fig 7. Best Model Learning Curve

Experiments

- The Comparison of Model Performances on Multi-labeled Nodes
  - Due to the presence of many nodes in the dataset with only one label, we excluded those nodes and compared the classification results of each model for nodes with multiple labels.

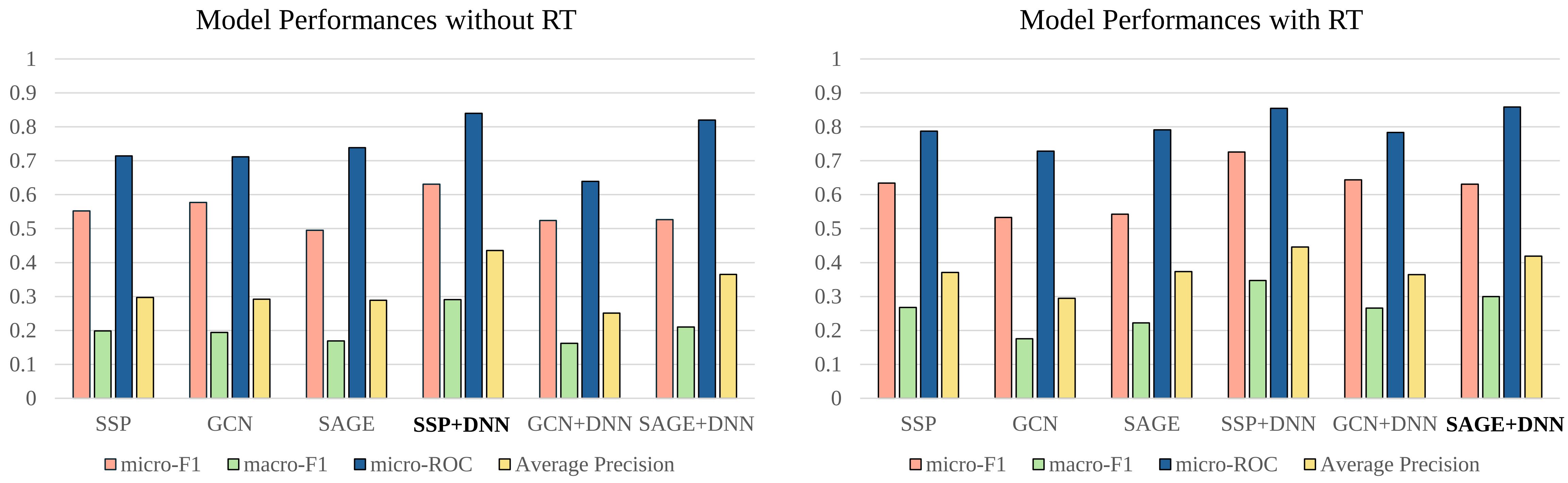

Fig 8. The Comparison of Model Performances on Multi-labeled Nodes

Conclusion

- In this study, we proposed a method of adding a Deep Neural Network (DNN) layer to the existing Graph Neural Network model to more effectively reflect text features within Twitter data.

- We confirmed that model performs better when using retweets as node features, so we were able to figure out that retweets serve as important features for determining the label of nodes.

- Furthermore, It was observed that adding a DNN layer in both experiments, where node features were differently reflected, resulted in a significant performance improvement.

- Moreover, the superiority of the proposed model was evident in the results regarding how well it predicted multi-labels.

- This model is expected to contribute companies to analyze social network for their marketing strategies by utilizing social media posts that were previously not used in analysis or required classification.

Discussion

- Limitation
  - The dataset may not be suitable for accurately evaluating the performance of the proposed model due to the low proportion of multi-labels, low network density, and insufficient diversity of text features.

- Future Work
  - To more accurately validate the model performance, we will use datasets with appropriate proportions of multi-labels and density of networks.
  - In addition to combining a DNN layer, we plan to explore and apply various methods to effectively reflect text features.

Reference

[1] Wang, Z. W., Wang, S. K., Wan, B. T., & Song, W. W. (2020). A novel multi-label classification algorithm based on K-nearest neighbor and random walk. *International Journal of Distributed Sensor Networks*, 16(3), 1550147720911892.

[2] Azran, Arik. "The rendezvous algorithm: Multiclass semi-supervised learning with markov random walks." *Proceedings of the 24th international conference on Machine learning*. 2007.

[3] Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

[4] Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. *stat*, 1050(20), 10-48550.

[5] Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30.

[6] Izadi, Mohammad Rasool, et al. "Optimization of graph neural networks with natural gradient descent." *2020 IEEE international conference on big data (big data)*. IEEE, 2020.

Thank you for listening

Department of Management Information System & Supply Chain Management

Taehyun Noh, Jimin Go, TackKyoon Woo, Seokwoo Song*

WEBER STATE UNIVERSITY

Q & A