Sapienza Università di Roma

Department of Ingegneria Informatica, Automatica e Gestionale "Antonio Ruberti"

Enhancing Road Safety Modeling with Graph Neural Networks: The Impact of Slope Inclusion

Advanced Machine Learning

Professor:
Fabio Galasso

Students:
Claudio Giannini (2093898)
Arian Gharehmohammadzadeghashghaei (2103465)
Arash Bakhshaee Babaroud (2105709)
Ehsan Mokhtari (2108539)

Academic Year 2024/2025

1 Abstract

Recent studies have begun leveraging Graph Neural Networks (GNNs) to predict traffic accidents by uncovering spatiotemporal patterns in road networks (Nippani et al., 2017; Gao X. et al., 2024), facilitating improved urban planning and road safety. In this work, we aimed to replicate the findings of Nippani et al. (2017) and evaluate the impact of incorporating slope data into the model. Our experiments demonstrated an improvement of 60 basis points in AUROC after integrating slope information, highlighting its significance in enhancing predictive performance.

2 Introduction

In this study, we focused on the state of Montana, as it was one of the states analyzed in the original paper. Montana was chosen not only for consistency with prior work but also due to its smaller road network size, which reduced **computational demands** and **minimized API calls** required to retrieve elevation data for slope calculations. Our workflow was structured into two main phases: the **dataset creation** where we extended the original dataset by introducing a new tensor that encodes slope information for edges in the road networ and the **training and testing** where we trained and tested the model using the same Graph Convolutional Network (GCN) architecture recommended by the original paper, consisting of two layers.

3 Dataset and Benchmark

The base dataset used in this study was sourced from the Harvard Dataverse and covers traffic data for Montana, USA, from 2016 to 2020. Node-level features included latitude, longitude, node indegree/outdegree, betweenness centrality, and meteorological variables such as average, maximum, and minimum temperatures, precipitation, wind speed, and sea-level pressure. Edge-level features comprised binary indicators for one-way roads, multi-class road types, road lengths, and Annual Average Daily Traffic (AADT). Slope data was introduced as a key enhancement in this study, derived from the Open-Elevation API. The slope for each edge was calculated as the elevation difference between the target and source nodes divided by the edge length as in the formula below:

$$\text{slope} = \frac{E_{\text{target}} - E_{\text{source}}}{L} \tag{1}$$

This feature captures road inclines and declines, which are particularly relevant in mountainous terrains or areas with steep gradients. However, sharp elevation changes introduced noise into the slope data, which was mitigated by applying thresholds. Unrealistic slopes were filtered by capping extreme values above ± 0.3 to ± 0.05, and slopes exceeding ± 0.2 on roads shorter than 50 meters were set to 0. To further improve slope accuracy, higher-precision APIs, such as Google Maps, and alternative encodings (e.g., categorical representations of slope: flat, mild, or steep) are proposed for future studies. The processed slope data was ultimately incorporated into the edge features tensor to enable the GNN to leverage this information during training and prediction.

4 Model and Approach

We implemented a Graph Convolutional Network (GCN) with two graph convolutional layers to predict traffic accidents. The model utilized a hidden dimensionality of 256, a learning rate of 0.001, and was optimized using the Adam optimizer over 100 epochs. These hyperparameters and the GCN architecture were adopted in this project based on the recommendations from the original paper for comparison purposes. The model performed solidly across all states tested, and the chosen hyperparameters demonstrated the best performance in terms of predictive accuracy. Training and evaluation were conducted using a train/validation/test split, covering the years 2016–2017 for training (40,040 records), 2018 for validation (20,677 records), and 2019–2020 for testing (39,222 records).

5 Experimental Results

We anticipated that integrating slope data into the Graph Neural Networks (GNNs) for road safety modeling would lead to notable improvements in performance metrics. This expectation is supported by the findings of Kuşkapan, E., et al. (2024), who observed that slopes exceeding 6% combined with snow significantly impacted traffic accidents, emphasizing the critical role of slope data in understanding accident-prone road conditions. Specifically, the **AUROC** improved from 82.62% to 83.22%, marking an increase of 60 basis points. This enhancement demonstrates the model's improved ability to rank accident-prone road segments more effectively. Additionally, **Precision** increased from 5.26% to 8.13%, indicating a reduction in false positives and enhancing the reliability of the model's predictions. However, a slight decline in **Recall** was observed, dropping from 52.01% to 47.14%, which suggests that while the model became more precise, it slightly compromised its ability to identify true positives. This trade-off highlights the need for further optimization through feature engineering and hyperparameter tuning based on this new information. To summarize the performance metrics of both models, Table 1 provides a detailed comparison of the **AUROC**, **Precision**, and **Recall** for the Train, Validation, and Test datasets.

Metric	Model - Slope	Model - No Slope	Improvement (b.p)
Train AUROC	83.31 ± 0.35	82.99 ± 0.15	+32
Valid AUROC	83.03 ± 0.12	82.73 ± 0.04	+30
Test AUROC	83.22 ± 0.30	82.62 ± 0.21	+60
Train Precision	11.89 ± 4.32	8.29 ± 3.92	+360
Valid Precision	8.96 ± 2.93	6.46 ± 1.36	+250
Test Precision	8.13 ± 2.77	5.26 ± 1.25	+287
Train Recall	44.35 ± 5.56	55.46 ± 5.54	-1111
Valid Recall	46.79 ± 6.60	55.96 ± 3.90	-917
Test Recall	47.14 ± 8.47	52.01 ± 1.34	-487

Table 1: Performance comparison between models trained with and without slope data across AUROC, Precision, and Recall metrics for Train, Validation, and Test datasets. Improvements are expressed in basis points (b.p).

6 Conclusions and Future work

The findings from this study highlight the potential of integrating slope data into GNNs for improving road safety modeling. The enhanced AUROC and Precision reflect the model's better ranking of accident-prone areas and reduced false positives, contributing to more accurate predictions. However, the trade-off between Precision and Recall underscores the need for additional refinement of the model. Future work should focus on testing the model across diverse geographical areas, improve the quality of slope data using Google Maps API and optimizing the model parameters based on the new slope data.

References

- Nippani, A., Li, D., Ju, H., Koutsopoulos, H. N., Zhang, H. R. (2023). Graph Neural Networks for Road Safety Modeling: Datasets and Evaluations for Accident Analysis. Presented at NeurIPS 2023. Northeastern University, Boston. Retrieved from https://neurips.cc/virtual/2023/poster/73579
- Gao, X., Jiang, X., Haworth, J., Zhuang, D., Wang, S., Chen, H., Law, S. (2024). Uncertainty-aware probabilistic graph neural networks for road-level traffic crash prediction. SpaceTimeLab, University College London (UCL),Peking University (PKU), MIT, University of Florida, and The Bartlett Centre for Advanced Spatial Analysis, UCL.
- Kuşkapan, E.; Çodur, M.Y.; Sahraei, M.A. Investigation of the Effect of Slope and Road Surface Conditions on Traffic Accidents Occurring in Winter Months: Spatial and Machine Learning Approaches. Appl. Sci. 2024, 14, 11629. https://doi.org/10.3390/app142411629