Multi-Channel Spatio-Temporal Graph Neural Network for Bike-sharing Demand Prediction: Integrating public transport and weather data

L. Belkessa[a,b,*], M. Ameli[a], M. Ramezani[b] and M. Zargayouna[a]

[a] COSYS-GRETTIA, Univ. Gustave Eiffel, Champs-sur-Marne, France
linda.belkessa2@univ-eiffel.fr, mostafa.ameli@univ-eiffel.fr, mahdi.zargayouna@univ-eiffel.fr
[b] School of Civil Engineering, The University of Sydney, Sydney, Australia
mohsen.ramezani@sydney.edu.au
[*] Corresponding author

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)
September 02-03, 2024, Crete, Greece

April 29, 2024

Keywords: Micromobility, Public transport integration, Demand prediction, Graph Neural Network, Spatio-temporal dynamics

1 INTRODUCTION

Increasing urban density and environmental concerns have led to the adoption of shared micromobility (MM) solutions such as bicycles and e-scooters as sustainable transport options. These solutions address challenges such as traffic congestion and pollution and are crucial for first— and last-mile connectivity with public transport (PT) systems. Integrating MM with PT is essential for improving urban mobility (Martens, 2007). In this context, utilizing PT data helps predict MM demand, optimize bike fleet management, and enhance service efficiency and user satisfaction by ensuring timely and reliable bike availability.

Recent research uses both traditional statistical methods and advanced deep learning techniques to integrate PT data for MM demand prediction. While traditional methods like Poisson regression and SARIMA are known for their interpretability, they often fall short in handling complex urban dynamics (Yu *et al.*, 2023). On the other hand, machine learning methods like Random Forests and Gradient Boosting Machines and deep learning approaches like Graph Convolutional Networks (GCNs) and Spatial-Temporal Graph Convolutional Networks (STGCNs) have improved addressing non-linear data relationships and spatial dependencies Goh *et al.* (2019), Lin *et al.* (2018), Xiao *et al.* (2021). However, these methods typically overlook PT checkout patterns influencing short-term and long-term bike demand, which could significantly improve demand forecasting Lin *et al.* (2018). This study aims to address the research gap of underutilization of PT checkout data, which provides crucial insights into MM demand at the station level, where PT and MM interactions are most significant Martens (2007). This paper proposes a novel approach integrating PT checkouts with MM demand data using a Multi-Channel Spatio-Temporal Graph Convolutional Network (MC-STGCN) to enhance prediction accuracy. In addition, external features such as weather are considered as an input to the framework. The proposed model investigates various strategies for generating adjacency matrices and evaluates different configurations to adapt to demand volumes.

2 METHODOLOGY

The proposed MC-STGCN is designed to integrate subway checkout data from public transport and external weather factors to capture both spatial and temporal dynamics for bike-sharing demand prediction. The proposed methodology employs a graph representation of the urban transport network. The nodes represent stations such as bike-sharing and subway stations. The edges capture the connectivity between these entities, distinguishing the intra-mode (bike-to-bike) and inter-mode (subway-to-bike) connections by feeding the adjacency matrices (calculated given spatial proximity) into different channels. We construct detailed adjacency matrices to model these relationships, enabling our network to address and leverage the complex interactions within the transport system. Figure 1 presents a global overview of the architecture of the MC-STGCN, which integrates PT checkouts data, bike pickups data, and weather data to forecast micromobility demand.

Data preprocessing involves normalizing various time-series data streams using Min-Max scaling ($x' = \frac{x-\min(x)}{\max(x)-\min(x)}$) and Z-score normalization ($x' = \frac{x-\mu}{\sigma}$) to standardize inputs where x is the original value and x' denotes the normalized value. The model constructs spatial graphs where nodes represent stations, using an adjacency matrix normalized as $A = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ to reflect station connectivity, where A denotes the adjacency matrix and D is the diagonal degree matrix.

In the MC-STGCN architecture, each data channel processes its respective channel, employing graph convolution layers ($H^{(l+1)} = \sigma(\Theta^{(l)} \cdot (D^{-\frac{1}{2}}AD^{-\frac{1}{2}}) \cdot H^{(l)})$) to capture local and global spatial structures, where $H^{(l+1)}$ is the output features of the next layer, σ is the activation function, $\Theta^{(1)}$ is the weight matrix for layer 1, and $H^{(l)}$ is the input features from the current layer. Temporal dynamics are incorporated through convolutional layers, synthesizing information across time to predict demand more accurately. Integration at the fusion layer combines these processed features, which are then utilized in the output layer to generate predictions ($y = W_{\text{out}} \cdot x + b_{\text{out}}$), where y is the predicted output, W_{out} is the weight matrix of the output layer, x is the input to the output layer, and b_{out} is the bias from the output layer, effectively capturing the interplay between different urban mobility modes and environmental factors.

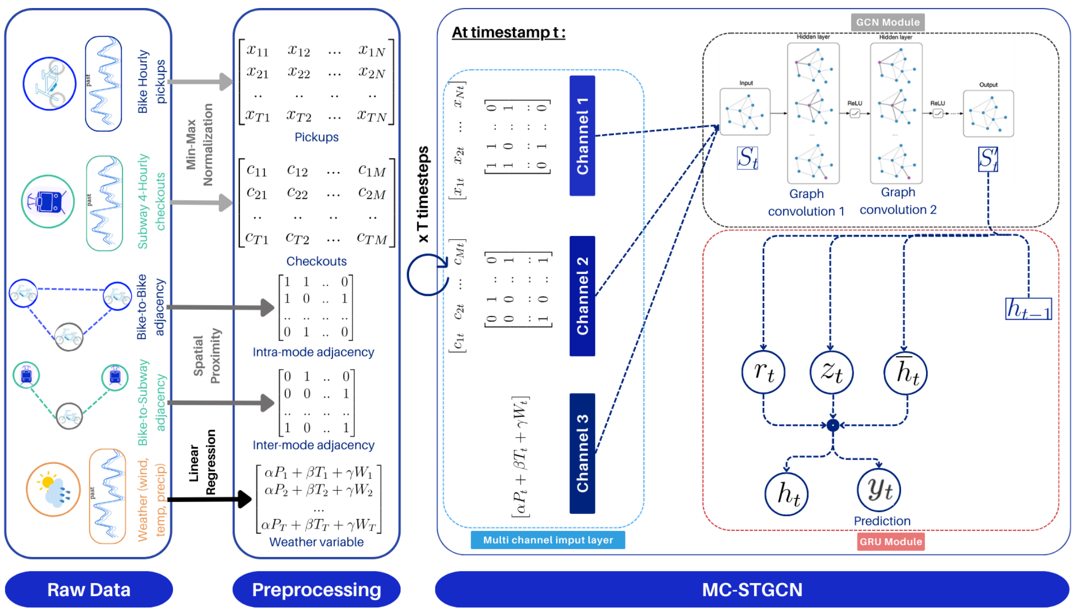

Figure 1 – *Overview of the proposed MC-STGCN architecture, where r_t is the reset gate, z_t is the update gate, and $\tilde{h}_t$ is the candidate activation vector and σ denotes the sigmoid activation function used in the gating mechanisms.*

3 RESULTS & DISCUSSION

This section presents the results of applying MC-STGCN to real data set of Manhattan's Citibike and subway system data. The study utilizes comprehensive datasets from 2022, including hourly weather data, Citibike data aggregated to four-hour intervals and focuses on Manhattan, equipped with 151 MTA (Metropolitan Transportation Authority) subway stations and over 1,800 Citibike stations, providing a framework for analyzing micromobility and public transport usage within this urban network.

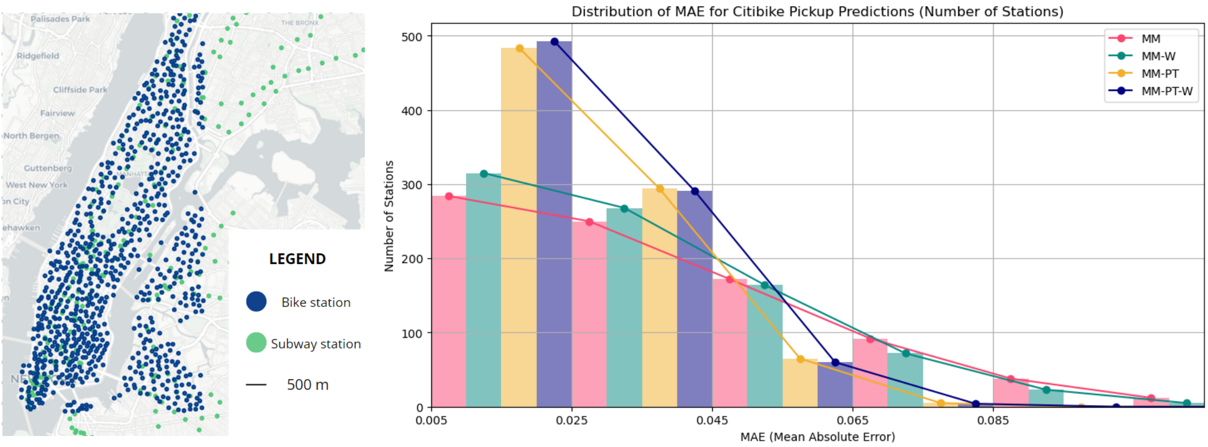

Figure 2 – *(a) Distribution of New York Citibike and subway stations; (b) Distribution of MAE.*

The geospatial distribution of bike and metro stations across Manhattan is shown in Figure 2(a), showing potential correlations with station density and usage patterns, impacting model performance variably across regions. Four configurations of the model are considered as follows: **MM**: Micromobility only; **MM-W**: Micromobility + Weather; **MM-PT**: Micromobility + Public transport; **MM-PT-W**: Micromobility + Public transport + Weather. The Mean Absolute Error (MAE) is measured to evaluate the methods' performance. Figure 2(b) presents the distribution of MAE, indicating that incorporating weather data generally enhances the model's predictive accuracy. Further improvements are observed when public transportation data in both MM-PT and MM-PT-W is integrated, indicating that multimodal data integration is beneficial.

Table 1 – *MC-STGCN Performance results*

Metric	Short-term prediction (4 hours)				Daily prediction				Weekly prediction			
	MM	MM-W	MM-PT	MM-PT-W	MM	MM-W	MM-PT	MM-PT-W	MM	MM-W	MM-PT	MM-PT-W
MSE	0.0189	0.0187	0.0184	0.0182	0.0132	0.0128	0.0124	0.0122	0.0140	0.0138	0.0134	0.0132
MAE	0.1023	0.1015	0.1008	0.1005	0.0790	0.0782	0.0774	0.0770	0.083	0.0820	0.0812	0.0810
R^2	0.2200	0.2250	0.2300	0.2320	0.4300	0.4600	0.4800	0.4820	0.4000	0.4450	0.5300	0.5500

The model's performance at different horizon lengths demonstrates differences in accuracy for urban mobility forecasting (see Table 1). The integration of distinct data sources as separate input channels, contrary to a feature augmentation strategy, allows the model to develop specialized representations for each type of data, enhancing prediction robustness and accuracy by 15% compared to the MM scenario. Different temporal contexts within model configurations, such as daily and weekly aggregations, significantly improved predictive performance, showcasing the model's capability to harness temporal patterns effectively as the MM-PT-W increased the coefficient of determination (R^2) by 13% for the daily resampling and 22% for the weekly resembling compared to the same model configuration in 4-hourly frequency. This indicates that the model extracts and leverages the long term patterns.

Station-wise performance analysis over epochs (test prediction iterations) indicates improved model stability and convergence speed when multimodal data is used, as shown in Figure 3(a).

The results show that the inclusion of subway data reduces the MAE. The model configuration using PT data is able to reach lower MAE values; for instance, the MM-PT-W converges rapidly to the lowest error to below 0.01. Besides, this figure shows that the convergence of the model is enhanced by adding multimodal data. Models with subway checkouts exhibit a more stable and quicker convergence, suggesting that transit data provides a critical temporal context that enhances learning. In terms of geospatial distribution, the integration of subway data helps to lower the MAE across Manhattan, which is illustrated by the prevalence of green markers indicating lower MAE values in the heatmap shown in Figure 3(b). This underscores the spatial correlation of prediction accuracy with station connectivity and urban dynamics.

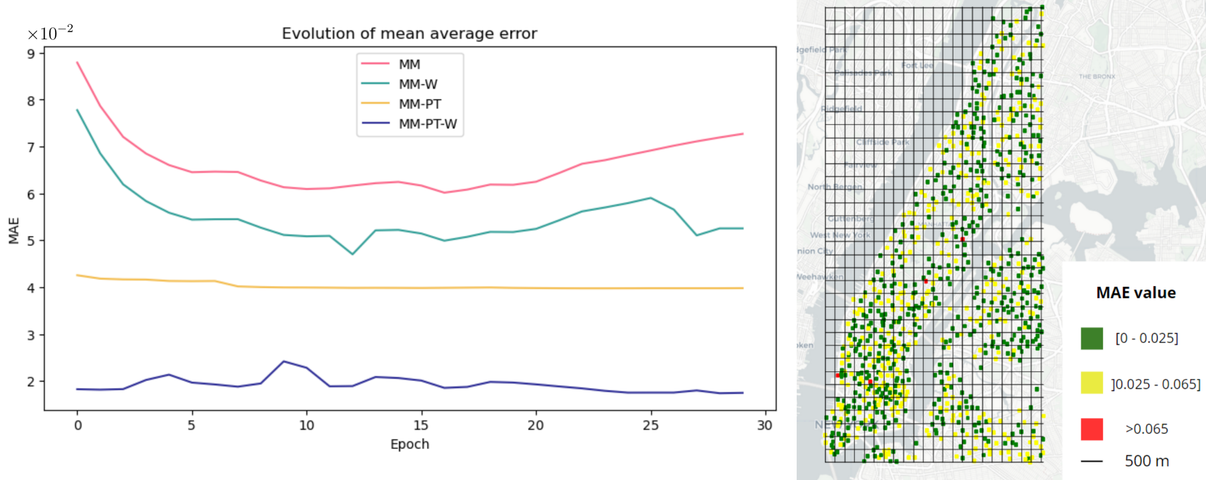

Figure 3 – *(a) MAE evolution over test set; (b) Heatmap of MAE values in Manhattan stations*

4 CONCLUSION

This research investigates the integration of public transport (PT) checkout data with micro-mobility (MM) demand prediction using MC-STGCN. By leveraging diverse data streams, the developed methodologies enhance prediction accuracy and provide deeper insights into urban mobility dynamics, facilitating effective MM fleet management. To complete our study, we explored various predefined adjacency matrices. These matrices could be based on GPS timings from subway to bike stations, enhancing our refinement of spatial relationship modeling. Additionally, assessing our model's performance across different temporal configurations—such as weekdays, weekends, and peak hours, is crucial to understanding demand variability more comprehensively. We are integrating real-time data feeds, including weather prediction and PT schedules, which will be essential to enhance our model's responsiveness. We are currently conducting multiple experiments in this direction, and the primary results are promising.

References

Goh, Chong Yang, Yan, Chiwei, & Jaillet, Patrick. 2019. Estimating primary demand in bike-sharing systems. *SSRN Electronic Journal*.

Lin, Lei, He, Zhengbing, & Peeta, Srinivas. 2018. Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. *Transportation Research Part C: Emerging Technologies*, **97**, 258–276.

Martens, Karel. 2007. Promoting bike-and-ride: The Dutch experience. *Transportation Research Part A: Policy and Practice*, **41**(4), 326–338.

Xiao, Guangnian, *et al.* 2021. Demand prediction for a public bike sharing program based on spatio-temporal graph convolutional networks. *Multimedia Tools and Applications*, **80**.

Yu, Liang, *et al.* 2023. Demand prediction and optimal allocation of shared bikes around urban rail transit stations. *Urban Rail Transit*, **9**(1), 57–71.