ISTDM 2025

Navigating Data Heterogeneity: Integration, Preparation, and Embedding Challenges in on-street parking occupation forecasting

Ayman Agoube

Polytechnique Hauts-de-France, LAMIH, CNRS, UMR8201, Valenciennes, France
CIRRELT/Polytechnique Montréal, Department of Mathematics and Industrial Engineering, P.O. Box 6079, Station Centre-Ville, Montréal, QC H3C 3A7, Canada Ayman.Agoube@uphf.fr

Abdessamad Ait El Cadi

Univ. Polytechnique Hauts-de-France, LAMIH, CNRS, UMR8201, INSA Hauts-de-France, F-59313, Valenciennes, France Abdessamad.AitElCadi@uphf.fr

Thierry Delot

Polytechnique Hauts-de-France, LAMIH, CNRS, UMR8201, Valenciennes, France Thierry.Delot@uphf.fr

Martin Trépanier

CIRRELT/Polytechnique Montréal, Department of Mathematics and Industrial Engineering, P.O. Box 6079, Station Centre-Ville, Montréal, QC H3C 3A7, Canada mtrepanier@polymtl.ca

Abstract

The continuous increase in the number of vehicles in urban areas is putting growing pressure on on-street parking, making efficient parking management a key challenge for cities. To get a sense of how much land is dedicated to parking, consider this: in the United States, the total area occupied by parked vehicles is roughly equivalent to the size of Massachusetts. In Europe, it's estimated to cover about half the land area of Belgium [1]. Research has shown that the time drivers spend circling around to find an available parking spot doesn't just waste time, but it also slows down overall traffic flow, increases vehicle emissions, and contributes to unnecessary fuel consumption and congestion in urban areas. For example, a study conducted in Westwood Village [2], a commercial district near The University of California, Los Angeles (UCLA), estimated that each year, this search for parking results in the consumption of approximately 47,000 gallons of gasoline, the release of 730 tons of CO_2 into the atmosphere, and a staggering 95,000 hours, wish is the equivalent of about eleven years lost by drivers simply trying to park.

Addressing parking challenges in cities requires more than just expanding infrastructure or enforcing stricter regulations. These approaches are often expensive, time-consuming, and limited by financial and spatial constraints [3]. A more efficient and forward-looking solution lies in providing drivers with reliable, real-time, and predictive information about parking availability. By equipping drivers with accurate data on nearby available spots as they approach their destination, such systems can significantly reduce the time spent searching for parking, ease traffic congestion, and help urban planners optimize the use of existing parking spaces [4].

Thanks to advancements in sensor technology and the growing connectivity between urban infrastructures, cities now have access to vast amounts of data on on-street parking occupancy. Real-time

monitoring systems, often linked to online platforms, allow both drivers and urban planners to make more informed decisions, easing traffic flow and improving the efficiency of parking policies. In parallel, the collection and analysis of historical data from parking meters enable the development of predictive tools capable of estimating future parking occupation. By integrating these predictions into parking guidance applications, drivers can be directed toward available spots near their destination, increasing their chances of finding a space upon arrival while also helping cities better manage parking demand.

Several approaches have tackled parking occupancy prediction by treating it as a regression problem. Traditional statistical time series models have been used to forecast occupancy rates and have demonstrated solid performance for short-term predictions [5]. However, these conventional methods tend to fall short when it comes to long-term forecasting, largely due to the complex and dynamic nature of parking data. More recently, deep learning models have emerged as powerful alternatives for addressing this challenge. Among them, Long Short-Term Memory (LSTM), have proven particularly effective for capturing temporal patterns in parking occupancy data [6]. That said, parking occupation in urban areas isn't just influenced by time, but it also exhibits strong spatial dependencies, especially in densely populated areas where occupancy at one location can directly affect nearby zones. This is a significant limitation for RNN-based models like LSTM, which are not inherently designed to capture spatial relationships between parking zones.

An effective parking prediction model needs to capture both the spatial and temporal patterns inherent in parking data. One potential approach is combining convolutional neural networks (CNNs) with LSTM networks, allowing the model to extract spatial features from parking areas while also learning how occupancy evolves over time [7]. However, this type of hybrid model is best suited for data structured in regular grids, like images or videos. Since parking networks in urban areas follow irregular, non-Euclidean structures, with parking zones connected in complex ways that do not conform to a regular grid, traditional CNN-LSTM models struggle to accurately capture the spatial dependencies between these zones.

Effectively forecasting parking occupancy rates (POR) requires considering the irregular, non-Euclidean nature of urban parking networks. On-street parking areas can be represented as a graph, where each node corresponds to a parking zone and holds occupancy data that evolves over time. Graph Neural Networks (GNNs) have attracted growing attention for their ability to capture complex spatial relationships in graph-structured data. In recent years, they have been widely applied to both traffic and parking prediction tasks, thanks to their capacity to model spatial dependencies between different areas within a transportation network. For example, one study introduced Spatio-Temporal Graph Convolutional Networks (STGCN) [8] to capture spatial interactions between multiple locations and track how traffic conditions change over time using a combination of multi-layer graph neural networks, 1D convolution, and gated linear units (GLUs) to handle temporal dynamics. Building on this, another work applied a similar spatial-temporal graph convolutional approach specifically for parking occupancy forecasting [9]. Later research further refined this concept with Hybrid Spatial-Temporal Graph Convolutional Networks (HSTGCN) [10], which enhanced the modeling of spatial dependencies by integrating an attention mechanism into the graph learning process.

We developed and tested an on-street parking occupancy forecasting model based on the Simplified Spatio-Temporal Graph Neural Network (SST-GNN) architecture. The model calculates Parking

Occupancy Rates (POR) by analyzing instantaneous parking statuses and performing both spatial and temporal aggregation to capture inter-zone dependencies and evolving occupancy trends. By leveraging a spatio-temporal graph structure, the model effectively predicts occupancy across multiple parking zones. Using real-world data from Montreal, in collaboration with the Agence de Mobilité Durable (AMD), our model consistently outperformed traditional approaches (ARIMA, LSTM) and advanced baselines like STGCN, achieving significant reductions in RMSE for both short-term (15 minutes) and longer-term (45-60 minutes) forecasts. These results highlight the value of incorporating spatial and temporal dynamics through graph-based learning when forecasting POR.

While spatio-temporal Graph Neural Networks (GNNs) have shown promising results in capturing the spatial and temporal dependencies inherent to parking data, several key challenges persist due to the heterogeneous nature of the data and the limitations of current modeling approaches. This paper investigates three major dimensions of data heterogeneity in parking space forecasting.

First, existing GNN-based approaches typically construct spatial graphs using Euclidean distances between parking zones, which poorly reflect the reality of driver behavior. Drivers navigate the city based on road network distances, traffic conditions, and practical accessibility rather than straight-line proximity. We explore alternative graph construction techniques that incorporate road network distances and contextual driving patterns, aiming to align the graph topology with real-world mobility.

Second, we investigate the challenge of learning expressive and interpretable node embeddings within GNN models. Traditional graph convolution techniques rely heavily on feature aggregation from neighboring nodes, which often results in a known issue called over-smoothing, where nodes that are close in the graph structure end up with nearly identical embeddings. This loss of uniqueness hinders the model's ability to distinguish between zones with genuinely different parking behaviors [11]. This problem becomes even more pronounced in urban environments, where adjacent parking zones may operate under very different regulatory and operational rules. For example, some areas might enforce residential-only parking, while others apply dynamic pricing, limited-time parking, or special loading zone restrictions. These operational constraints directly shape both driver behavior and parking turnover, yet they are rarely embedded into the core representation learning process. Therefore, we emphasize the importance of embedding strategies that explicitly account for local regulations and operational constraints, ensuring that even geographically close zones retain distinct representations when their regulatory conditions differ. This approach allows the model to better reflect the diversity of operational contexts found in urban parking networks.

The third challenge involves effectively integrating external data sources and operational rules into the spatio-temporal forecasting framework. Parking occupation is shaped not only by spatial and temporal dependencies but also by external influences, including municipal regulations, nearby amenities, weather, and traffic. Parking zones located near commercial areas naturally experience different occupation profiles compared to those in residential neighborhoods. Moreover, regulatory factors such as seasonal parking bans or temporary restrictions during special events contribute to the complexity of forecasting occupancy. Existing models often treat these external factors as static inputs appended to the node features, limiting the model's ability to dynamically adapt to changing conditions. To address this, we propose a hybrid embedding framework that fuses spatial embeddings with embeddings derived from external data, enabling the model to dynamically adjust spatial relationships based on external influences.

Using real-world parking data from urban environments, we empirically demonstrate the importance of addressing these three dimensions. By aligning spatial graphs with realistic mobility patterns, enhancing the interpretability of learned embeddings, and incorporating multi-source external data into the forecasting pipeline, we show significant improvements in both predictive accuracy and model robustness. This work advances the state of spatio-temporal forecasting in smart cities, providing actionable insights for both researchers and urban planners working at the intersection of transportation, data science, and smart mobility.

REFERENCES

[1] Inci, E., A review of the economics of parking. Economics of Transportation, Vol. 4, No. 1, 2015, pp. 50–63, special Issue on Collective Contributions in the Honor of Richard Arnott.

[2] D. Shoup, Cruising for parking. Access Magazine 1, 30 (2007), 16–23, 2007.

[3] Cao, R., Choudhury, F., Winter, S., & Wang, D. Z. W. (2024). Artificial intelligence for parking forecasting: an extensive survey of machine learning techniques. *Transportmetrica A: Transport Science*, 1–39.

[4] Qian, Zhen (Sean), and Ram Rajagopal. 2015. "Optimal Dynamic Pricing for Morning Commute Park ing." TransportmetricaA: TransportScience11(4):291–316.

[5] Yu, F., J. Guo, X. Zhu, and G. Shi, Real time prediction of unoccupied parking space using time series model. In 2015 International Conference on Transportation Information and Safety (ICTIS), 2015, pp. 370–374.

[6] Shao, W., Y. Zhang, B. Guo, K. Qin, J. Chan, and F. D. Salim, Parking Availability Prediction with Long Short-Term Memory Model. In Grid and Pervasive Computing, 2018.

[7] Chandra, R., U. Bhattacharya, A. Bera, and D. Manocha, TraPHic: Trajectory Prediction in Dense and Heterogeneous Traffic Using Weighted Interactions. CoRR, Vol. abs/1812.04767, 2018.

[8] Yu, B., H. Yin, and Z. Zhu, Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875, 2017.

[9] Xiao, X., Z. Jin, Y. Hui, N. Cheng, and T. H. Luan, Spatial-Temporal Graph Convolutional Networks for Parking Space Prediction in Smart Cities. In 2021 IEEE 94th Vehicular Tech nology Conference (VTC2021-Fall), 2021, pp. 1–5.

[10] Xiao, X., Z. Jin, Y. Hui, Y. Xu, and W. Shao, Hybrid Spatial–Temporal Graph Convolu tional Networks for On-Street Parking Availability Prediction. Remote Sensing, Vol. 13, No. 16, 2021.

[11] Chen, D., Y. Lin, W. Li, P. Li, J. Zhou, and X. Sun, Measuring and relieving the over smoothing problem for graph neural networks from the topological view. In Proceedings of the AAAI conference on artificial intelligence, 2020, Vol. 34, pp. 3438–3445.