A Spatio-Temporal Graph Neural Network Approach for Bike-Sharing Demand Prediction

Alexander Julius Busch
University of Hamburg
Hamburg, Germany

Kaifeng Lu
University of Toronto
Toronto, Canada

Abstract—Bicycle sharing plays a significant role in the public transportation system. Accurate dynamic demand prediction is the key to support effective re-balancing and provide reliable service. In this study, we adapt and analyze the model spatio-temporal graph attention network (ST-GAT) architecture known from traffic prediction to predict bike-sharing windows and compare it with several baselines.

Index Terms—AI, Smart Mobility, GNN, Bike-Sharing

I. Introduction

Stations where bicycles are docked and taken from are a common way of implementing a bicycle sharing system. As bicycle sharing becomes one of the most popular ways to commute, it plays a significant role in the public transportation system. According to the Bike Share Toronto 2023 Business Review, the total number of rides in 2023 is estimated to be about 5.5 million, projected to 2025 to become more than 6.2 million [1]. The total number of stations deployed is planned to be more than $1,000$ with more than $10,000$ bikes available. However, the demand varies greatly between location, weekday, season and other factors, which leads to imbalances and congestions in the system. This results in customer dissatisfaction and unreliability of the system, jeopardizing the central role of bicycle in reaching emission-neutral transportation and providing convenience.

To properly implement dynamic solutions, such as adaptive dynamic pricing and terminal extensions, the demand needs to be reliably predicted. Since the demand fluctuates based on various aspects, we decided to train a machine learning model to investigate the relations between these aspects and the demand. With a sufficiently accurate model, bicycle sharing companies can adopt adjustments to provide more reliable service.

II. Literature Review

A. Scope and Overview

Due to the vastly growing number of bike sharing systems, bike sharing demand prediction has been investigated by a number of authors in the last 15 years. While classical machine learning approaches such as regression and boosting have been used in the early developments, in recent years, deep learning approaches have been found to give significant advantages in the forecasting domain [2]. Based on the literature, we compare the following machine learning approaches to bike sharing demand prediction and identify open problems.

B. Paper Summaries

a) *Bike-Sharing Demand Prediction at Community Level under COVID-19 Using Deep Learning [3]:*

This article investigates short-term bike-sharing demand forecasting in Montreal. It compares deep learning techniques for predicting bike pickups 15 minutes ahead in six communities identified within the city's bike-sharing network. The study compares the performance of LSTMs (Long Short Term Memory, a recurrent network type), CNN (Convolutional Neural Network)-LSTM hybrids, and ARIMA (Auto-Regressive Integrated Moving Average). The authors trained the models based on two main feature attributes: historical demand data in 15-minute intervals and weather conditions.

They found that the CNN-LSTM hybrid model outperforms the other models (MAE of 3.00 and RMSE of 4.77), compared to the LSTM model (MAE of 4.48 and RMSE of 6.86), and the ARIMA (MAE of 50.95 and RMSE of 61.17).

Pros:
- LSTM Models: These were found to capture temporal dependencies effectively with moderate complexity.
- CNN-LSTM Hybrid Models: These were found to combine spatial and temporal learning for improved accuracy in complex data.
- The ARIMA Model is fast to train and easy to interpret.

Cons:
- LSTM Models: Lacks spatial awareness, limiting prediction accuracy compared to hybrid models.
- CNN-LSTM Hybrid Models: These are computationally expensive and harder to interpret.
- The ARIMA Model was found to perform poorly for complex, nonlinear and disrupted data.

b) *Using graph structural information about flows to enhance short-term demand prediction in bike-sharing systems [2]:*

This paper investigates how graph theory can enhance short-term bike-sharing demand forecasting. It focuses on incorporating graph-based features derived from flow interactions such as Out strength, In-strength, Out-degree, In-degree, and PageRank to improve prediction accuracy. It compares three machine learning models: XGBoost, Multi-Layer Perceptron (MLP), and LSTM. This study found that including graph-based features significantly improves model performance compared to using traditional features like meteorological data alone. The authors here find that the XGBoost, MLP and LSTM model achieve MAPE values of 27.2%, 27.7%, and 27.0% as well as RMSE values of 6.78, 6.85, and 6.69 correspondingly. The LSTM model was here found to

be the most effective model for incorporating complex graph-based features, while the overall performance of the XGBoost is nearly comparable.

Pros:
- XGBoost excels at handling structured data and provides good performance with less tuning.
- The MLP architecture is simple and efficient with moderate performance.
- The LSTM is best at capturing temporal dependencies and sequential patterns, especially with time-lagged data.

Cons:
- The XGBoost is less effective at handling complex time dependencies compared to the LSTM.
- The MLP struggles with sequential data and lacks the advanced handling of time dependencies.
- The LSTM is complex to train and requires significant computational resources.

c) *Modeling bike-sharing demand using a regression model with spatially varying coefficients [4]:*

This article focuses on investigating how various factors such as land use, socio-demographic attributes, and transportation infrastructure influence bike-sharing demand at different stations. Notably, they define a graph model, where the data for each bike-sharing cluster is accumulated according to its catchment region. The catchment region is calculated using Thiessen polygons, and ensured to be non-overlapping. The authors propose a spatially varying coefficients (SVC) regression model that accounts for local spatial effects, unlike previous regression models that assume the factors are spatially homogeneous.

Pros:
- The graph model includes connectivity between station clusters to be used.
- The more sophisticated modeling of station clusters defined by catchment areas instead of single stations notably allows to predict demand in regions where no stations are present and thus allows planning.

Cons:
- As the model does not directly regress on immediate historical data as e.g. the LSTM models, it fundamentally does not to allow accurate future predictions, but instead better fits general analysis purposes.

d) *Modeling Bike Availability in a Bike-Sharing System Using Machine Learning [5]:*

This paper explores predicting bike availability at San Francisco Bay Area Bike Share stations using machine learning algorithms. The authors apply three methods: Random Forest (RF) and Least Squares Boosting (LSBoost) for univariate regression and Partial Least-Squares Regression (PLSR) for multivariate regression. They found that factors like station neighbors, time of prediction, and weather conditions are significant in predicting bike availability. The RF, LSBoost and PLSR models achieve MAE of 0.37, 0.58 and 0.6 bikes per station correspondingly, i.e. the Random Forest (RF) model here is found to give the best prediction accuracy.

Pros:
- The RF model is highly accurate with low prediction error and robust to overfitting.

- The LSBoost model is effective for regression tasks with manageable computational complexity.
- The PLSR captures spatial correlations, useful for large networks with interdependent stations.

Cons:
- RF: Requires independent observations. Performance decreases as prediction horizon increases.
- LSBoost: Higher prediction error. Requires proper regularization against overfitting.
- PLSR: Yields the highest prediction error. Less accurate for smaller networks.

e) *Cross-Mode Knowledge Adaptation for Bike Sharing Demand Prediction Using DomainAdversarial Graph Neural Networks [6]:*

This recent paper focuses on improving the state of the art in bike sharing demand prediction by integrating additional features from other transport modes, here subway and ride-hailing traffic data, all from New York City. In order to integrate this heterogeneous graph structured data, recurrent CNNs are combined with graph convolutions to yield transport embeddings. The paper also introduces an adversarial training principle to learn these embeddings, such that the embeddings learned are optimized to be indistinguishable. These embeddings are then fed into a multiple GNNs, combined into a single representation, and passed to a final prediction layer.

Pros:
- The architecture developed seems to give a significant improvement to the compared approaches. However, the dataset itself is one of the most often used in bike-share prediction [2], allowing some degree of comparison of results.
- The architecture allows extension to integrate very heterogeneous other modes of transportation significant to bike-sharing that would otherwise be very difficult to integrate.

Cons:
- The architecture and total approach is complicated, requiring extensive effort to adapt to different formats of data from other systems.
- The data, especially from ride-hailing, is likely more noisy than the bike-sharing trip data (due to e.g. only several providers contributing to the ride-hailing data, and the different companies providing data using their own standards)
- It is unclear how much improvement the architecture would give without the additional features.
- Due to the station structure used for the graph, the model cannot predict demand for only places themselves.

C. Taxonomic Table and Tabular Comparison

We organize the used models into a tree based on the dominant design features for a brief overview of the common investigated models for bikesharing demand prediction. The fundamental division is between two lines of work between "classical" approaches and deep approaches, which split into different lines of works. Further, because the more advanced deep models build on LSTM-approaches to model immediate

history, this marks another division in the tree. In order to have concise names for the approaches, we denote every approach by the fundamental architectural characteristic and its paper (e.g. LSTM [3] denotes the LSTM approach in [3]), although this does neglect other important differences in the approaches.
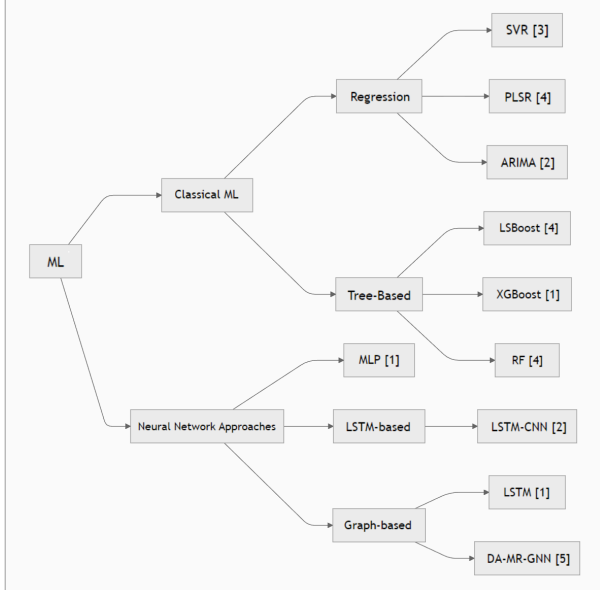

Fig. 1: A classification tree of the employed ML approaches

We summarize the main advantages and disadvantages in the following table. We classify how well each approach integrates spatial, recent historical and connectivity information. For example, the pure LSTM approach handles and generalizes recent historical information well, while the regression does not have explicit integration of these features, it only incorporates historical information via engineered features.

TABLE I: Comparison of the Employed ML Approaches

Approach	Spatial Handling	Temporal Handling	Connectivity Modeling	Model Complexity	Model Performance
PLSR [5]	Decent	Rudimentary	Rudimentary	Rudimentary	Low
LSBoost in [5]	Rudimentary	Rudimentary	Rudimentary	Low	Decent - Good
RF in [5]	Rudimentary	Rudimentary	Rudimentary	Low	Good
XGBoost in [3]	Rudimentary	Rudimentary	Rudimentary	Low	Good
ARIMA in [3]	Rudimentary	Short Term	Rudimentary	Low	Decent
MLP in [3]	Rudimentary	Short Term	Rudimentary	Rudimentary	Decent
SVR [4]	Good	Rudimentary	Decent - Good	Medium (Complex Feature Engineering)	Good for spatial modeling
DA-MR-GNN [6]	Decent	Very good	Very Good	Very high	Very good

D. Directions for Future Research

As noted in [2], there is no single standard benchmarking dataset for the specific objective of bike sharing demand prediction, making it difficult to obtain a precise ranking of the models. Notably, the New York Citibike data has been used in multiple approaches [2], [6], but the precise data used still changes. As given in [2], several studies have suggested that XGBoost performing as well most of the state of the art approaches, having also won the 2014 Kaggle competition [7], however the best approach was found to depend strongly on the dataset and modeling. Notably [2] suggests, that well-performing deep model architectures in traffic prediction are likely to show good performance in bike-sharing demand prediction. In this line of work, [6] combines well performing spatio-temporal GNN approach, known from traffic prediction [8] and [9], with a domain-adversarial network. However, little work has been done on the extent that performance of traffic prediction carries over to bike-sharing performance and adaptation of successful models from other adjacent fields, (e.g. models in [8], [9]), appears to be a promising research direction.

III. Problem Formulation and Modeling

A. Formulation

Our goal is to predict the demand of bicycle sharing using a machine learning approach given historical data and additional features, such as the day of the week and the daytime.

We consider a bicycle sharing system consisting of N_S stations with fixed capacities Cap_s, i.e. a maximum of Cap_s bicycles can be docked at station $s \in \mathbb{N}$. We denote the current number of bicycles at station s by $B_s(t)$ We distinguish two types of demands:

- $\mathrm{OutDem}_s(t)$: the number of bicycles that people take out the dock at (t, s) if there are enough bicycles present.
- $\mathrm{InDem}_s(t)$: the number of bicycles that people dock at (t, s) if there is be enough space present.

In order for the number of bicycles to be conserved, a pair of suitable demand functions has to satisfy

$$\sum_{t,s} \mathrm{InDem}_s(t) - \mathrm{OutDem}_s(t) = 0, \tag{1}$$

where the sum runs over all times and states.

These functions encode the psychological, social and logistical aspects of bicycle sharing demand. They define the behavior of the system (i.e. the number of bicycles at station s) via the difference equation

$$B_s(t + 1) = B_s(t) + \min(\mathrm{InDem}_s(t), \mathrm{Cap}_s(t)) \\ - \min(\mathrm{OutDem}_s(t), B_s(t)). \tag{2}$$

In order to formulate a proper machine learning problem, we now approximate the discrete quantities by real numbers, i.e. we introduce

- the hourly rates of In-Demand $\mathrm{ID}_s(t) \in \mathbb{R}$ and out demand $\mathrm{OD}_s(t) \in \mathbb{R}$
- the actual hourly rates of bicycles docked $\mathrm{In} \in \mathbb{R}$ and bicycles taken out $\mathrm{Out} \in \mathbb{R}$.

Our given given data consists of a list of all rides including start, end stations and start and and times in minutes. The discrete behavior previously stated now translates to the partially continuous loss

$$
\text{loss}(t) = \begin{cases}
(\text{ID}(t) - \text{In})^2 + (\text{OD}(t) - \text{Out})^2 \\
\text{if } 0 < B(t) < \text{Cap}(t) \\[1.5em]
(\text{OD} - \text{Out})^2 + \max(0, \text{ID} - \text{In})^2 \\
\text{if } B(t) = \text{Cap}(t) \\[1.5em]
(\text{ID} - \text{In})^2 + \max(0, \text{OD} - \text{Out})^2 \\
\text{if } B(t) = 0
\end{cases}
\tag{3}
$$

i.e. the demands should normally be identically to the rates of the bicycles taken in or out, only if the station is either full or empty, the demands may be higher, but not lower, than the input or output bicycles.

B. Modeling

Modeling factors:

- Controllable Parameters: These include the sharpness of the hourly rate approximation, which depends on the window. Choosing a large window to compute the rates leads to loss of precision in the time domain, while too small window might not capture the continuity of the demand well. Other control parameters of the historic data collection include the accuracy of time resolution which in the given dataset is minutes.
- Signals (i.e. input features): To predict the future hourly demand, we input
  – the historic input and output rates before the prediction time
  – the daytime
  – the day of the week.
- Error States (i.e. failure modes of the model):
  – Possible failure modes include cases if the future prediction are not accurate enough to be useful, the predictions become unstable or the in-demand does not match the out demand.
- Noise Factors:
  – Capacity changes of the stations during the month, which appears due to e.g. repositioning of stationing.These changes are not included in the published data.
  – Population density changes, irregular road and facility closures which are not present in the training data but have a significant influence on bicycle sharing behavior.

IV. Proposed Solution

V. Performance Evaluation

VI. Conclusions & Recommendations

VII. Code

References

[1] J. Hanna, *Bike share toronto 2023 Business Review*, 2023, [Online]. Available: https://www.toronto.ca/legdocs/mmis/2023/pa/bgrd/backgroundfile-240804.pdf

[2] Y. Yang, A. Heppenstall, A. Turner, and A. Comber, "Using graph structural information about flows to enhance short-term demand prediction in bike-sharing systems," *Computers, Environment and Urban Systems*, vol. 83, p. 101521, 2020, doi: https://doi.org/10.1016/j.compenvurbsys.2020.101521.

[3] A. Mehdizadeh Dastjerdi and C. Morency, "Bike-Sharing Demand Prediction at Community Level under COVID-19 Using Deep Learning," *Sensors*, vol. 22, no. 3, 2022, doi: 10.3390/s22031060.

[4] X. Wang, Z. Cheng, M. Trépanier, and L. Sun, "Modeling bike-sharing demand using a regression model with spatially varying coefficients," *Journal of Transport Geography*, vol. 93, p. 103059, 2021, doi: https://doi.org/10.1016/j.jtrangeo.2021.103059.

[5] H. I. Ashqar, M. Elhenawy, M. H. Almannaa, A. Ghanem, H. A. Rakha, and L. House, "Modeling bike availability in a bike-sharing system using machine learning," in *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 2017, pp. 374–378. doi: 10.1109/MTITS.2017.8005700.

[6] Y. Liang, G. Huang, and Z. Zhao, "Cross-Mode Knowledge Adaptation for Bike Sharing Demand Prediction Using Domain-Adversarial Graph Neural Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 5, pp. 3642–3653, 2024, doi: 10.1109/TITS.2023.3322717.

[7] W. Cukierski, "Bike Sharing Demand." Kaggle, 2014.

[8] T. Yu, H. Yin, and Z. Zhu, "Spatio-temporal Graph Convolutional Neural Network: A Deep Learning Framework for Traffic Forecasting," *ArXiv*, 2017, [Online]. Available: https://api.semanticscholar.org/CorpusID:4972291

[9] S. Zhang, Y. Guo, P. Zhao, C. Zheng, and X. Chen, "A Graph-Based Temporal Attention Framework for Multi-Sensor Traffic Flow Forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, pp. 7743–7758, 2022, [Online]. Available: https://api.semanticscholar.org/CorpusID:234906657