Learning Demand Functions for Bike-Sharing Using Spatio-Temporal Graph Neural Networks

Alexander Julius Busch
University of Hamburg
Hamburg, Germany

Kaifeng Lu
University of Toronto
Toronto, Canada

Abstract—Bike-sharing usage prediction has been implemented and analyzed using a variety of models, ranging from linear and logistic regression models with extensive spatial features [cit], decision features, ARIMA [cit] to deep learning models with convolutional [cit] and graph features [cit]. However, modeling and prediction of the underlying demand function that drives the usage has not been rigorously attempted to the best of our knowledge, possibly due to the ill-posed nature of the problem. We propose several defining properties of a demand function and extend the Spatio-Temporal Graph Neural Network (STGAT) model from traffic prediction to jointly predict both the actual usage rate and a suitable demand function matching the data, which we take from the City of Toronto's open data portal [cit]. We analyze accuracy and behavior of the model in normal usage prediction and illustrate the behavior of the learned demand function in various scenarios.

Index Terms—Bike-Sharing, GNN, STGAT, Demand Prediction

I. Introduction

Docked bike-sharing, i.e. a system of 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. Considered for the city of Toronto, 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 bike-sharing usage fluctuates nontrivially based on a multitude of different influences, such as weekday, daytime, events, weather and more features, machine learning and especially neural network approaches have been successfully employed to predict general usage [2], [3], [4]. However, only predicting the number of bikes taken in or out is not sufficient for implementing a dynamic pricing system: For example, when a small station is predicted to be completely full at a time, how much should we incentivize taking out bikes? The demand in this case

could range from zero (e.g. in the night) to almost arbitrarily high (e.g. during rush-hours in the day), but this quantity can not be directly computed from a purely predictive model.

In this article, our goal is to learn a demand function from data that quantitatively encodes user demand. For this, we extend the STGAT [5] prediction model to simultaneously predict usage and an infered demand function and define an adapted loss function to capture the notion of the demand. Notably, it is not our goal to outperform other purely predictive approaches, but improve the learned demand function via simultaneous learning of the pure prediction task and verify the model capability.

The structure of this article is as follows:
- We give an overview on other approaches on bike-sharing usage prediction and the STGAT architecture.
- We then mathematize the notion of a demand function and formalize the bike-sharing joint prediction task.
- In the following section, we translate the notion of a demand function into a proper regularizer and describe our variation of the STGAT architecture and the data processing.
- We then compare the predictive performance of our model to a linear baseline. We investigate its quality over the prediction horizon and show qualitative prediction results. We qualitatively compare different modeling choices in the resulting demand model and evaluate quantitative aspects of the demand model.

A. Motivation

For an illustration of the difference between demand and usage, consider a bike-sharing system with docks. Bikes can be taken out at a dock and have to be docked in back at any dock, with the price of the ride being proportional to the time between the dockings. Consider a user that wants to go from A to B at some day, possibly to reach their workplace or possibly to spontaneously take a ride to a park as leisure activity. In the case that there are bikes available, the user takes one and the demand is equal to the number of bikes taken out at that station. However, if the station is empty, the user might either walk to a nearby station, take another transportation medium or cancel the planned ride all together. In this case, the demand can not be inferred from the rate, but probably exhibits other regularities: The station might not be full at similar dates and the local demand is greatly defined by people commuting regularly to work, or other complex regularities. However, manually extracting the components of the demand is infeasible: From

the author's own experience with the Toronto bike-sharing system, what people do when a station is full depends on weather, close stations, personal and time considerations, daytime and other factors. Thus we propose to use a deep learning approach to learn a suitable demand function from usage data using proper regularization.

For those familiar with the notion of counterfactual reasoning, the demand can be seen as counterfactual quantity: It is the rate of bikes taken in or out if the station would not have been empty at that point.

II. Literature Review

III. Problem Formulation and Modeling

A. Axioms of Demand

From the described characteristics of the demand in Section I.A, we postulate the following properties of an ideal demand function, which we translate into a loss minimization problem for machine learning.

In the following, $\text{IR}(s,t), \text{OR}(s,t)$ denote the rates of bikes docked in or taken out at station $s \in \mathbb{N}$ and time $t \in \mathbb{R}$ and ID, OD the respective demand functions. The predicates $\text{atmax}(s,t), \text{atmin}(s,t)$ are defined to be 1 if the station is at its maximum capacity or empty respectively, and 0 otherwise.

- If the station is empty at a time, the out-demand should be greater or equal than the bikes taken out: $\text{OD}(s,t) \geq \text{OR}(s,t)$ if $\text{atmin}(s,t)$
- If the station is completely full, the in-demand should be lower than the rate of bikes docked. $\text{OD}(s,t) \geq \text{OR}(s,t)$ if $\text{atmax}(s,t)$
- Otherwise, the demand should be equal to the respective in or out rates:
  ‣ $\text{ID}(s,t) = \text{IR}(s,t)$ if not $\text{atmax}(s,t)$
  ‣ $\text{OD}(s,t) = \text{OR}(s,t)$ if not $\text{atmin}(s,t)$
- Both in-demand and out-demand functions should be smooth, i.e. its intuitively it should not change abruptly. This can be modeled in several different ways. We propose to demand that the third derivative is small, i.e. $\frac{d^3}{dt^3}(\text{ID}(s,t) + \text{OD}(s,t)) \ll 1$.

  The motivation for this is that the third derivative is the smallest derivative providing a natural shape for demand. Notably, the functions with minimal third derivatives are quadratic functions, while functions minimizing the first or second derivative are piecewise constant or affine linear functions respectively, which would be an unnatural way to extend a demand function. The minimization of the third derivative can be seen as analogous to motion planning in robotics, where one minimizes jerk (the third derivative of the position) in order to obtain a smooth trajectory with small changes in acceleration.

B. Relaxation to Loss Function

From the ideal properties, we construct an analogous relaxed loss function, which allows to model the problem as mini-mization problem. We define the loss in terms of three components:

- the rate error $\left(\widehat{\text{IR}} - \text{IR}\right)^2 + \left(\widehat{\text{OR}} - \text{OR}\right)^2$
- the demand constraint violation

$$\left(\text{IR} - \widehat{\text{ID}}\right)^2 \cdot (\text{atmax} \cdot [\text{IR} \geq \text{ID}] + 1 - \text{atmax})$$
$$+ \left(\text{OR} - \widehat{\text{OD}}\right)^2 \cdot (\text{atmin} \cdot [\text{OR} \geq \text{OD}] + 1 - \text{atmin}) \quad (1)$$

- the smoothness violation $S_3\left(\widehat{\text{ID}}\right)^2 + S_3\left(\widehat{\text{OD}}\right)^2$, where S_3 is a numerical finite difference approximation of the third derivative.

The final loss is now formed as mean over stations and times of the previous terms:

$$\frac{1}{N_s \cdot N_t} \sum_{s,t} \text{RateError} + \text{DemViol} + \alpha \cdot \text{SmoothViol}, \quad (2)$$

where $\alpha \in \mathbb{R}$ is a regularization constant and we used abbreviations to denote the error components.

IV. Proposed Solution

In order to suitably solve the minimization problem, we adapt the spatio-temporal graph attention neural network architecture (STGAT) from [6]. We base our adaptation on the open implementation from [7]. Generally, a graph neural network (GNN) is a biologically inspired approach to machine learning, which operates on graph-structured data. The fundamental layer employed in a graph neural network is a graph convolutional layer (GCN), which computes output node features by computing and then aggregating features from each incoming node. The graph attention layer (GAT) employed in the STGAT is an extension of this layer, which weights the computed features by a learned attention score in order to compute a more refined representation. Notably, the fundamental principle of an attention mechanism, itself biologically inspired, has proven to be successful in many machine learning domains.

The STGAT architecture in [6] is constructed to predict car traffic velocities at measurement points, given historical velocities over all measurement points as graph. We adapt the final linear layer to give bike in-, out-rate and in-, out-demand predictions. We choose to investigate the STGAT architecture because of its performance in predicting roughly the next 45 minutes, given the last hour of information, which is a time horizon that would be useful for demand prediction for a dynamic pricing problem, and the similarity of the problems.

Additionally, because bike-sharing depends significantly on time feature information [[8]],

The base network consists of the following layers:

- A graph attention layer with 8 heads, followed by dropout.
- An LSTM layer with hidden size 32, which takes the reshaped historic data over all nodes.
- A second LSTM layer with hidden size 128.
- A final linear layer operating on the last LSTM prediction, outputting information for all 9 timepoints.

For the evaluation of our demand prediction task, we choose to compare three reference models:

- The base STGAT model, with number of nodes and final output size adapted to the problem, without dropout.
- An upscaled and regularized variant, with LSTM sizes $(128, 256)$, dropout of 0.95 and weight decay 0.4.
- A variation, where we replace the LSTMs by a decoder-only transformer with 4 layers, 8 attention heads, and an embedding size of 32.

A. Modeling Details

We use the historic bike-sharing data provided from the city of Toronto's open data portal. For all computations, we use the data for the full month May 2024. For the station geographical locations, we use the current live information provided. Notably, because historic station capacities are not given, we have to estimate when a station is full or empty from the bikes taken in and out. For an exploratory spatial data analysis of Toronto's bike-sharing system, see our Medium article.

a) *Station Occupancy Estimation:* The behavior at most stations follows a clear daily pattern, where bikes show wave-like patterns, however, the cumulative number of bikes either increases or declines over the month. This is probably due to bikes taken out for repairing or relocation, which are not logged in the ridership data. In order to estimate when a station is empty or full, we thus take daily minima and maxima and consider the station full if the cumulative number of bikes at that timepoint is within 2 bikes close to the maximum or minimum, respectively, as marked in Fig. 1. Notably, we aim to err on the side of overestimation, because if a station is almost full at a time, usually users remember this behavior and refrain from using this station during that time, although the demand exists.
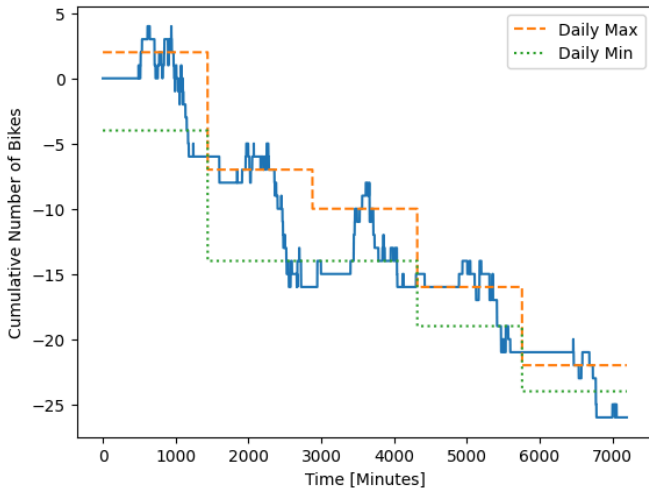

Fig. 1: Cumulative number of bikes in a station over several days with capacity bounds marked.

b) *Temporal Horizon and Rate Calculation:* In modeling bike sharing demand, the prediction time horizon and the averaging horizon fundamental parameters. In [3], the number of pickups in 15 minutes was found to be reliably predictable, which is equivalent to predicting a uniform average over 15 minutes. In our approach, we follow [6] and predict 9 timepoints in 5 minute intervals over the next 45 minutes, given 12 timepoints in the past, for all stations at the same time. Due to our averaging approach, which averages over partial timepoints in the future, we choose the standard evaluation horizon to be 20 minutes, which has in [2] also been found to be the most effective horizon for prediction using standard ML models. Because we are interested in finding a suitable demand function over time, which is related to rate data, a way of averaging the bike pickups has to be chosen. Notably, if one chooses to predict the exact number of bikes taken out or in each minute, one finds almost random behavior, because whether one arrives a minute later or not depends on many other factors, which are insignificant to the demand. Thus, we choose to average with a gaussian filter with $\sigma = 10$ min to calculate rate information. For this standard deviation of 10 minutes, thus roughly 63% of the information accumulated lies in the interval ± 10 min around each datapoint and $\approx 95\%$ in the ± 20 min interval. Empirically, this interval yields nontrivial prediction results. (Notably, smoothing with $\sigma = 60$ min renders the prediction task trivial, yielding similar accuracies for both linear and more complex models.)

V. Performance Evaluation

VI. Conclusions and Recommendations

VII. Code

References

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

[2] 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.

[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] 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.

[5] 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

[6] X. Kong, W. Xing, X. Wei, P. Bao, J. Zhang, and W. Lu, "STGAT: Spatial-Temporal Graph Attention Networks for Traffic Flow Forecasting," *IEEE Access*, vol. 8, no. , pp. 134363–134372, 2020, doi: 10.1109/ACCESS.2020.3011186.

[7] T. C. Julie Wang Amelia Woodward, "Predicting Los Angeles Traffic with Graph Neural Networks." [Online]. Available: https://medium.com/stanford-cs224w/predicting-los-angeles-traffic-with-graph-neural-networks-52652bc643b1

[8] E. Eren and V. E. Uz, "A review on bike-sharing: The factors affecting bike-sharing demand," *Sustainable Cities and Society*, vol. 54, p. 101882, 2020, doi: https://doi.org/10.1016/j.scs.2019.101882.