Optimizing Bike Sharing Station Placement through Urban Data and Climate-Informed Deep Learning Models

Surya Prakash Meesala
Florida State University
Tallahassee, Florida, USA
sm23w@fsu.edu

Venkata Vaibhav Parasa
Florida State University
Tallahassee, Florida, USA
vvp23@fsu.edu

Ajay Jagini
Florida State University
Tallahassee, Florida, USA
aj23p@fsu.edu

ABSTRACT

The proliferation of bike-sharing systems necessitates the identification of ideal sites for bike stations to enhance urban mobility and promote sustainable city design. This approach utilises diverse urban data—including land use, street networks, points of interest (POIs), demographics, and climatic data—to forecast bike-sharing demand within a metropolitan region. Our suggested model incorporates a convolutional neural network (CNN) for spatial data, a long short-term memory (LSTM) network for temporal usage patterns, and dense layers for category data, providing a comprehensive method to detect high-demand locations. This research seeks to test the model's efficacy using assessment measures such as accuracy and demand prediction precision, with prospective uses for urban planners.

KEYWORDS

Urban data, bike-sharing demand, CNN, LSTM, climate data, urban planning, mobility management

ACM Reference Format:
Surya Prakash Meesala, Venkata Vaibhav Parasa, and Ajay Jagini. 2024. Optimizing Bike Sharing Station Placement through Urban Data and Climate-Informed Deep Learning Models. In . ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Urban bike-sharing systems have gained popularity as an eco-friendly transportation alternative in cities globally, enhancing urban mobility, alleviating traffic congestion, and mitigating greenhouse gas emissions. Efficient bike-sharing initiatives necessitate precise demand forecasts to enhance station locations, bike distributions, and availability, thereby ensuring supply aligns with demand. Predicting bike-sharing demand is intricate due to its dependence on numerous elements, such as spatial configuration, demographic attributes, transit accessibility, and weather conditions.

Conventional techniques for station placement predominantly depend on surveys and observational studies, both of which are labour-intensive and expensive. These methods often involve urban planners performing on-site surveys to assess demand, analyse mobility patterns, and pinpoint areas with potential for significant utilisation. This method, while effective, becomes unfeasible when applied to extensive urban regions or swiftly expanding cities.

The growing accessibility of diverse urban data offers a chance to enhance the accuracy of demand forecasting for bike-sharing systems. Open data portals now offer comprehensive datasets encompassing urban land use, transportation networks, demographic data, and more city-specific metrics, like points of interest (POIs) and transit stops. Utilising these facts, researchers may create predictive models that evaluate demand in various urban regions, allowing planners to determine ideal sites for bike stations based on current and anticipated demand.

Notwithstanding recent progress, numerous obstacles persist in establishing a comprehensive demand forecast framework for the placement of bike-sharing stations. Bike-sharing demand is fundamentally spatial and temporal, varying by geographic location, time of day, day of the week, and season. Spatial attributes, including nearness to transportation stations or prominent locations, affect demand in manners necessitating intricate modelling. Likewise, temporal patterns, such commuting hours or seasonal variations, must be recorded to ensure precise projections.

To tackle these issues, we offer a multi-input deep learning model architecture that integrates convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and dense layers. This model amalgamates data from diverse urban datasets, encompassing land use, street networks, points of interest, demographics, and climatic data, to forecast bike-sharing demand with elevated spatial and temporal precision. Through the amalgamation of these specialised layers, the model proficiently synthesises geographical, temporal, and categorical data, resulting in enhanced accuracy in predicting bike-sharing demand. Our methodology signifies a progressive advancement in urban planning, wherein demand forecast models are both data-driven and attuned to local and temporal conditions.

2 RELATED WORK

The placement of bike-sharing stations is a well examined subject in urban planning, transportation modelling, and data-informed decision-making. A multitude of studies has examined the determinants of bike-sharing demand, employing methodologies that range from survey-based analysis to machine learning models utilising extensive metropolitan data.

- Traditional Demand Assessment Methods
- Data-Driven Urban Mobility Models

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/XXXXXXX.XXXXXXX

- Machine Learning and Deep Learning Approaches for Demand Prediction

2.1 Traditional Demand Assessment Methods

Garcia-Palomares et al (**2012**) employed Geographic Information Systems (GIS) to examine spatial characteristics and population density, amalgamating this information with survey findings to assess demand and enhance station positioning[1]. Despite the incorporation of GIS-based methodologies providing spatial insights, these techniques remained largely reliant on manual data collection and frequently failed to adequately capture temporal variations in demand, including seasonal or diurnal swings.

2.2 Data-Driven Urban Mobility Models

Zhang et al (**2011**) utilised social media check-ins and Points of Interest (POI) data to forecast bike-sharing demand through the analysis of human activity patterns[2]. Proximity to certain points of interest, including retail establishments and transit stops, is a significant predictor of bike station use.

Shaheen et al (**2010**) examined the demographic characteristics influencing bike-sharing adoption, observing that younger and less affluent demographics utilise bike-sharing systems more frequently[3].

2.3 Machine Learning and Deep Learning Approaches for Demand Prediction

Chen et al (**2015**) employed a hybrid approach using artificial neural networks (ANNs) and regression models to forecast bike-sharing demand by integrating points of interest (POI), demographic information, and street network data. This model surpassed conventional regression methods by more effectively capturing the non-linear correlations between metropolitan characteristics and bike-sharing demand[4]. Nonetheless, in the absence of specialised layers to separately manage spatial and temporal data types, ANN-based models frequently fail to accurately grasp intricate urban patterns.

3 DATA ANALYSIS

In our framework, accurate bike-sharing demand forecasting depends on the efficient use of varied urban datasets that encompass several geographical, temporal, demographic, and environmental attributes of a city. This study will use datasets from two major urban areas: New York and California and the research primarily utilises land use, street and transit networks, places of interest (POIs), demographic information, and climate data. Each dataset offers distinct insights on future bicycle usage trends, collectively enhancing the comprehensive understanding of urban mobility demand.

- **Land Use Data**: Provides information on residential, commercial, and mixed-use zones, facilitating the identification of high-demand locations based on urban configuration.
- **Street and Transit Networks**: Illustrates streets, bicycle lanes, and transit stops, offering spatial insights into the accessibility of prospective bicycle stations.
- **Points of Interest (POIs)**: Encompasses amenities such as parks, educational institutions, and commercial centres, facilitating demand forecasting based on proximity.
- **Demographic Data**: Analyses population density, age demographics, and socioeconomic variables to predict bike-sharing uptake based on user attributes.
- **Climate Data**: Offers historical meteorological patterns, encompassing temperature, precipitation, and seasonal variations, which are known to substantially influence bicycle utilisation. Every dataset is subjected to meticulous pre-processing to guarantee compatibility and enhance predictive accuracy inside the multi-input model framework:
- **Spatial Alignment**: Land use, street networks, and points of interest data are spatially organised onto a uniform grid pattern to standardise input for convolutional neural network processing.
- **Normalisation**: Demographic and climatic data are standardised to address scale discrepancies, enabling the model to process these variables efficiently and impartially.
- **Categorical Encoding**: Demographic variables, including age groups and economic status, are encoded to ensure compatibility with dense layer inputs.
- **Temporal Aggregation**: Climate data is consolidated by day or season, depending on the required level of temporal granularity, hence improving LSTM processing efficiency.

4 FRAMEWORK DESIGN

The proposed framework utilises a multi-input neural network architecture that amalgamates spatial, temporal, demographic, and climate data, aimed at encapsulating the intricacy and fluctuation of bike-sharing demand in urban settings. This design consists of three primary neural network components: Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) network, and Dense Layers.

- CNN Layer for Spatial Data
- LSTM Layer for Temporal Data
- Dense Layers for Categorical Data
- Combining Layers in a Multi-Input Architecture

4.1 CNN Layer for Spatial Data

The Convolutional Neural Network (CNN) layer analyses spatial information, concentrating on land utilisation, street networks, and point of interest (POIs). Convolutional Neural Networks (CNNs) excel at identifying spatial associations in grid-based data, rendering them suitable for analysing urban layout patterns that affect bike-sharing demand.

- **Input Preparation**: The spatial datasets, including land use, roadway networks, and points of interest (POIs), are depicted as grid maps, with each cell denoting a geographic area inside the city. Each grid cell encompasses data indicative of spatial characteristics, including the existence of residential zones, nearness to transit stations, or concentration of commercial points of interest.
- **Convolutional Layers**: The CNN layers apply filters over the spatial grid, learning important spatial features that indicate potential high-demand areas.

- **Feature Extraction and Pooling**: After convolution, pooling layers reduce the spatial resolution, allowing the model to focus on the most relevant spatial features while reducing computational complexity.

4.2 LSTM Layer for Temporal Data

The Long Short-Term Memory (LSTM) network is engineered to analyse temporal and climatic data, which reflect cyclical and seasonal fluctuations in bike-sharing demand. LSTM layers are optimal for processing sequential data, as they can learn temporal dependencies, rendering them appropriate for discerning patterns affected by time of day, weather, and seasonality.

- **Temporal Input**: Historical usage patterns and climate data, including temperature, precipitation, and humidity, are fed into the LSTM layers.
- **Sequential Processing**: The LSTM layers process this temporal data sequentially, allowing the model to understand how bike demand changes over time.
- **Climate Sensitivity**: Climate data is crucial for cities experiencing unusual weather patterns, as the demand for bike-sharing systems may significantly decline under specific conditions. By integrating meteorological data directly into the LSTM layer.

4.3 Dense Layers for Categorical Data

The Dense Layers process demographic and land use data, enabling the model to analyse non-sequential, categorical variables that affect bike-sharing demand. Dense layers excel at learning weighted correlations among variables, rendering them appropriate for analysing data where each feature is independent, such as demographic information and category land use classifications.

4.4 Combining Layers in a Multi-Input Architecture

The outputs from the CNN, LSTM, and dense layers are amalgamated to create a cohesive feature vector, integrating spatial, temporal, and category information. This multi-input architecture allows the model to easily incorporate various data types, encapsulating the complete intricacy of urban bike-sharing demand.

- **Feature Concatenation**: Following the processing of each component's designated data type, the extracted features are merged into a singular vector. This unified feature vector encompasses spatial relationships derived from the CNN, temporal and climatic dependencies from the LSTM, and categorical insights from the dense layers.
- **Final Prediction Layer**: The aggregated features are input into additional dense layers that generate a final demand forecast. The final layers assign weights to each feature, enabling the model to generate a comprehensive prediction based on all contributing elements.

5 EVALUATION

The evaluation phase is essential for assessing the efficacy and dependability of the proposed model in forecasting bike-sharing demand and pinpointing ideal station placements. This section delineates the experimental configuration, assessment measures, and baseline comparisons employed to evaluate the model's performance.

- Experimental Setup
- Evaluation Metrics

5.1 Experimental Setup

To ensure that the evaluation captures both spatial and temporal variations in bike-sharing demand, the experimental setup includes:

- Training and Testing Split
- Cross-Validation
- Hyperparameter Tuning

5.2 Training and Testing Split

The datasets are partitioned into training and testing sets, generally following an 80:20 ratio. Temporal data, including climate and history bicycle riding, are meticulously segregated to avert data leakage, wherein future data could unintentionally affect training.

5.3 Cross Validation

A K-fold cross-validation method is utilised to evaluate the model's robustness and generalisation abilities across various data subsets. This approach partitions the dataset into K subsets, utilising K-1 subsets for training and the remaining subset for testing in each cycle.

5.4 Hyperparameter Tuning

Essential hyperparameters, including learning rate, batch size, the quantity of filters in the CNN, and the count of LSTM units, are optimised by grid search or random search techniques.

5.5 Evaluation Metrics

The model's performance is assessed through several key metrics that capture its accuracy, relevance in recommendations, and overall ranking quality.

- Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE)
- Normalized Discounted Cumulative Gain (nDCG@K)
- F1 Score for High-Demand Zones

5.6 Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE)

Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are utilised to assess the overall accuracy of demand forecasts by computing the average and squared discrepancies between expected and actual demand values. Reduced MAE and RMSE values signify enhanced predictive accuracy, especially beneficial for pinpointing demand hotspots.

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

5.7 Normalized Discounted Cumulative Gain (nDCG@K)

nDCG evaluates the quality of the ranking produced by the model, comparing it to the ideal ranking. It's particularly useful for top-K recommendations (e.g., top 10 locations) by penalizing incorrect placements more heavily when they appear higher on the list.

$$\text{nDCG@K} = \frac{\text{DCG@K}}{\text{IDCG@K}}$$

$$\text{DCG@K} = \sum_{i=1}^{K} \frac{2^{\text{rel}_i} - 1}{\log_2(i+1)}$$

5.8 F1 Score for High-Demand Zones

The F1 Score for High-Demand Zones assesses the model's precision and recall specifically in areas of high demand, where precise station placement is paramount. The F1 Score facilitates the equilibrium between pinpointing critical locations for station placement and reducing false positives in areas with low demand.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

6 CONCLUSION

This research presents a novel, data-driven architecture that utilises recent advancements in deep learning to enhance the positioning of bike-sharing stations. By integrating geographical, temporal, and demographic data into a CNN-LSTM-Dense architecture, the model effectively tackles the intricacies of urban configurations and the variability of user demand patterns. The suggested system improves forecast accuracy and offers a scalable, customisable tool for city planners, facilitating data-driven decision-making that matches with sustainable and efficient urban mobility objectives.

7 CONTRIBUTION

The work on this project was a collaborative effort, with each team member contributing equally to the research, development, and documentation. Each team member's contribution is quantified as follows:

- ***Venkata Vaibhav Parasa (VVP23)*** : 33.3%
- ***Ajay Jagini (AJ23P)*** : 33.3%
- ***Surya Prakash Meesala (SM23W)*** : 33.3%

Each member was actively involved in all phases of the project, including data analysis, framework design, model implementation, and evaluation.

REFERENCES

[1] Longbiao Chen, Daqing Zhang, Gang Pan, Xiaojuan Ma, Dingqi Yang, Kostadin Kushlev, Wangsheng Zhang, and Shijian Li. 2015. Bike sharing station placement leveraging heterogeneous urban open data. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 571–575. https://doi.org/10.1145/2750858.2804291

[2] Juan García-Palomares, Javier Gutiérrez, and Marta Latorre. 2012. Optimizing the location of stations in bike-sharing programs: A GIS approach. *Applied Geography* 35 (11 2012), 235–246. https://doi.org/10.1016/j.apgeog.2012.07.002

[3] Susan A. Shaheen, Stacey Guzman, and Hua Zhang. 2010. Bikesharing in Europe, the Americas, and Asia. *Transportation Research Record* 2143 (2010), 159 – 167. https://api.semanticscholar.org/CorpusID:40770008

[4] Daqing Zhang, Bin Guo, and Zhiwen Yu. 2011. The Emergence of Social and Community Intelligence. *Computer* 44, 7 (July 2011), 21–28. https://doi.org/10.1109/MC.2011.65

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009