Scientific Project on “Literature Review and Taxonomy on Reallocation Strategies for Bike-Sharing Platforms”

Submitted To

Jun.-Prof. Dr. Kai Heinrich
Faculty of Economics
Otto-von Guericke Universität Magdeburg

Submitted By

Name	Matriculation ID
Willem Franz Weiss	243418
Md. Ali Azzan Esha	234935
Zahed Hossain Fahim	235221

Table of Contents

List of Figures

Abstract

Bicycle-sharing systems have gained global attention as eco-friendly public transportation alternatives, resulting in a growing number of scientific works published related to this topic. We employ a systematic approach, from selecting the most recent and relevant papers to defining categories that can summarize the work of 30 articles effectively. Modern bike-sharing systems use a dockless or station-based approach, with different challenges. The most crucial aspect is maintaining the balance between supply and demand as it determines the system's success. Authors came up with various problems regarding the operational parts of bike-sharing systems and external effects like traffic and environment. Even the topic of social fairness is discussed. As the chosen methods for problem-solving are diverse, we provide seven categories in our taxonomy and differ between a predictive or prescriptive focus. The most widespread optimization methods were Clustering algorithms, Mixed and Linear Integer Programming, and Artificial Neural Networks for which the authors utilized real-time, simulated, weather, and spatial data. Further methods are Ordinary Least Squares, Birth-Death Processes, Long Short-Term Memory, Gradient-Boosted Regression Trees, and additional customized methods. Those methods generate results that can reveal high-demand zones with the help of density graphics and clustering. The results can help to solve bike imbalance problems while minimizing operation costs and maximizing customer satisfaction to give general guidance for an optimal bike reallocation strategy.

1. Introduction

Bike-sharing systems are nowadays used worldwide and provide a convenient transportation mode to citizens in many cities such as New York City, Berlin, Paris, and Beijing (Hulot et al., 2018; Y. Li et al., 2018). This system is so popular because of its usefulness and flexibility. Station-based and dockless are the main categories for bike-sharing systems. A station-based system is not flexible because someone who unlocks a bike from one station must park it at another station with a fixed location. On the other hand, the dockless bike-sharing system is a free-floating bike-sharing system that does not require fixed stations and allows customers to park their bikes anywhere (Su et al., 2022). A user can rent a bike from a station and return it to any other stations for a docked-based station system or can rent or return a bike at a random station via swiping their membership card, bank card, or mobile application for dockless bike-sharing system (Chiariotti et al., 2020; Y. Li et al., 2018).

In this paper, we will discuss the following problems - Imbalance with supply and demand, operational efficiency, unbalanced bike distribution, different traffic patterns, traffic relocation, social fairness and equality, and environmental impacts with possible methods to solve that we found in 30 research papers. There are different types of mathematical optimization methods – OLS (Ordinary Least Squares), Clustering Algorithms, Mixed and Linear Integer Programming Techniques, LSTM (Long Short-Term Memory), Neural Networks, GBRT (Gradient-Boosted Regression Trees), and a few customized methods used by the authors to solve the above problems.

Bike-sharing services provide greater flexibility than traditional public transportation, lowering vehicular traffic, reducing the city's environmental impact, and enhancing public health. Hence, they are a crucial part of the emerging intelligent city paradigm (Chiariotti et al., 2020). Bike-sharing systems provide numerous benefits for urban mobility, including reducing emissions, improving user’s health and lifestyle, alleviating traffic congestion, enhancing traffic systems, and integrating with public and multimodal transportation (Su et al., 2022).

This paper is structured as follows: First, we systematically overview our methodology. Then, we present our taxonomy as a form of conceptualization, details about the bike-sharing reallocation systems, and some limitations. This is followed by a summary of the author's results, split into three parts - problems, data scenarios, and optimization approaches. The last section provides a conclusion with an outlook.

2. Methodology and Systematic Overview

To give a broad overview of the literature on the topic “Reallocation strategies for bike-sharing platforms”, we utilized three large databases, namely “ACM (Association for Computing Machinery) Digital Library” (dl.acm.org), “IEEE (Institute of Electrical and Electronics Engineers)” (ieee.org) and “ScienceDirect” (sciencedirect.com). The research process is depicted in Figure 1 as a Prisma Flow Diagram.

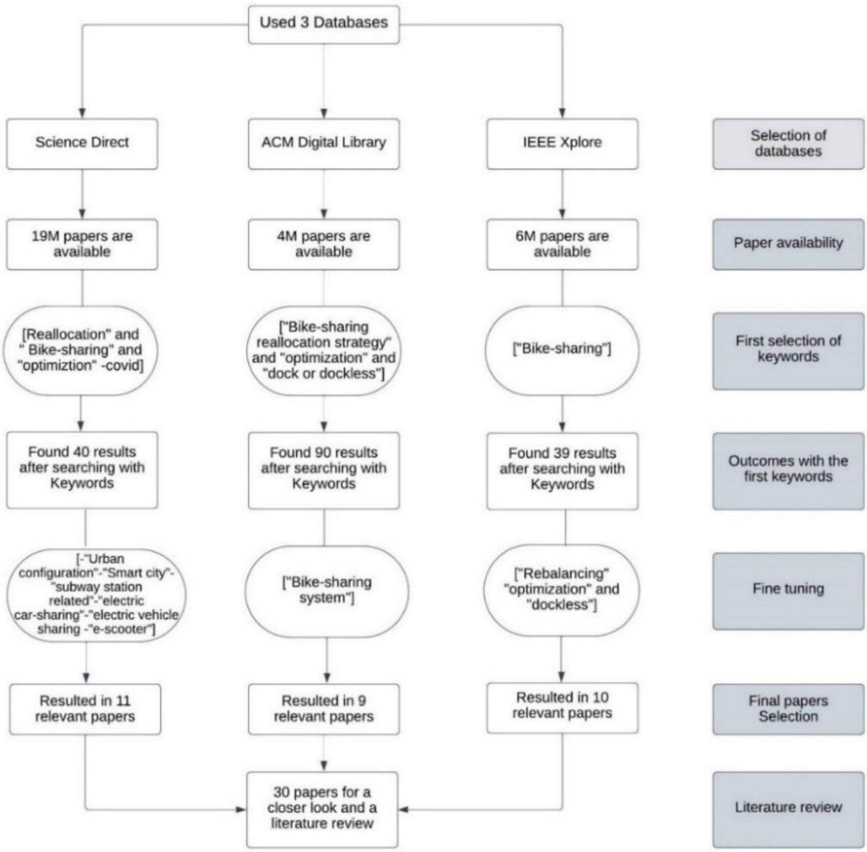

Figure 1: Prisma Flow Diagram of Methodology

ScienceDirect is a website that provides access to an extensive bibliographic database of scientific and medical publications of the Dutch publisher Elsevier, one of the most significant academic publishers. It thus offers a great variety of the most recent papers about bike-sharing systems. The search string ["Reallocation" and "bike sharing" and “optimization” -covid] revealed 40 results, 19 of which were from 2020 or newer. The keywords "Reallocation" and "bike sharing" will highlight the topic. At the same time, “-covid” filters out all papers related to the pandemic from 2020, in which bike-sharing had a vast demand explosion, but topics regarding the pandemic are not too relevant to our writing. As our Scientific Project focuses on comparable optimization strategies, the keyword “optimization” is required. Out of the 19 articles, we further refined our selection by eliminating eight papers found to be less pertinent through keyword fine-tuning. We utilized the following keywords [-"urban configuration" -"smart city" -"subway station related" -"electric car-sharing" -"electric vehicle sharing"-"e-scooter"] and ended up with 11 highly relevant papers.

ACM Digital Library is one of the largest collections of open-access research papers. This database currently offers nearly 4M open-access articles and a great variety of relevant documents regarding our topic, “Reallocation Strategies of Bike-Sharing Platforms”. To specify our search results, we used Boolean operators on the keywords. We searched for [Title: bike-sharing reallocation strategy] AND [All: optimization] AND [All: dock or dockless] AND [All: algorithm] AND [E-Publication Date: (01/01/2018 TO 12/31/2023)]. We again sorted the papers by relevance related to “bike-sharing systems”, only selected articles from 2018 to 2023, and ended up with nine research outputs.

We employed the IEEE Xplore database due to its extensive collection of papers on “Reallocation strategies for bike-sharing platforms”. The IEEE Xplore database is a powerful source of scientific and technical publications encompassing electrical engineering, computer science, electronics, and telecommunications fields. It is a widely recognized source of information for researchers. The database displayed relevant searches for the keyword [“Bike-sharing”]. We then filtered the results using criteria such as conferences, journals, magazines, early-access articles, and books. Then, we chose all open-access journal papers from 2018 to 2023. After conducting our search, we found 39 relevant documents. From these 39 papers, further keywords [“Rebalancing”], [“optimization”], and [“dockless”] lead to the selected final 10 papers.

In total, we will categorize 30 papers and analyze their different methodologies. The papers were organized based on occurring problems and their optimization models. The main problem with all the articles is reallocating the bikes. We can distinguish the issues into two categories - problems for station-based and dockless bike-sharing platforms. The common problems experienced with station-based and dockless bike-sharing platforms are an imbalance in supply and demand, operational efficiency, unbalanced bike distribution, different traffic patterns, social fairness and equality, and environmental impacts. In the discussed papers, the authors use different types of mathematical optimization methods, which are, in summary, OLS (Ordinary Least Squares), Clustering Algorithms, Mixed and Linear Integer Programming Techniques, LSTM (Long Short-Term Memory), Neural Networks, GBRT (Gradient-Boosted Regression Trees) and other customized methods.

3. Taxonomy Dimensions and Characteristics

The extracted data from our three databases with 30 selected papers were analyzed using a thematic analysis approach in the form of a table. The analysis identified common problems and the proposed reallocation strategies related to docked and free-floating bike-sharing systems across the literature.

Dimension	**Characteristics (frequency of occurrence)**		
Bike-Sharing System Type	*Dockless (23)*	*Docked (19)*	
Problem Type	*Imbalance with Supply and Demand (17)*	*Operational inefficiencies (8)*	*Traffic Patterns (15)*
	Social Fairness (2)	*Environmental effects (5)*	
Optimization Type	*Ordinary Least Squares (2)*	*Birth Death Process (1)*	*Clustering Algorithm (6)*
	Mixed/Linear Integer Programming (9)	*Long short-term memory (2)*	*Neural Networks (4)*
	Gradient boosted regression trees (2)	*Customized Methods (9)*	
Analytics	*Prediction (12)*	*Prescription (18)*	
Data Type	*Real Time (23)*	*Simulated (1)*	*Weather (1)*
	Spatial (5)		

Figure 2: Taxonomy of Reallocation Strategies for Bike-Sharing Platforms

Figure 2 gives an overview of the bike-sharing model type and the data scenarios used in the author's optimization of the selected papers. We distinguish between docked and dockless bike-sharing systems as their technical demands differ. As the optimization goal is optimal bike reallocation, using aggregated real-time and historic bike data is almost inevitable. As much bike-sharing data is available, only one paper used simulated data to forecast bike distribution. Spatial or weather datasets can be important for bike-sharing optimization as every place has its geographical characteristics, and the weather dramatically impacts the bike's demand.

We ended up with five problem categories: Imbalance with Supply and Demand, Operational Inefficiencies, Traffic Patterns and Relocation, Social Fairness, and Environmental Impacts. Among all the five categories, the "Imbalance with Supply and Demand"-problem has the highest number we found together. An imbalance in supply and demand indicates a situation where the quantity of the service available significantly differs from the quantity desired by consumers. In 8 papers among 30, we found an "Operational Inefficiency"-problem, which refers to the inability of a system, process, or organization to utilize resources effectively, resulting in wasted time, effort, and costs in the bike-sharing system, leading to increased operational costs. Fifty percent of the 30 papers discuss problems with "Traffic Patterns and Relocation". These problems indicate situations where shared bikes are not equally distributed across a city to fulfill customers' demands, causing some areas to have too few or too many bikes, often requiring relocation to ensure availability and usability. Finally, problems about Social Fairness and Environmental Impacts are less significant in our Literature Research.

As optimization types, we chose a selection of significant methods that serve a unique purpose within bike-sharing literature. Among them is Ordinary Least Squares (OLS). It is a fundamental statistical method used to estimate the relationships among variables. In bike-sharing, OLS might be employed to understand how various factors (weather, time of day, or day of the week) influence bike usage. The Birth-Death Process (BDP) optimization models events happen at specific intervals. For bike-sharing, this could be used to model the "birth" (arrival) and "death" (departure) of bikes at stations, helping to understand demand fluctuations and station dynamics. Clustering Algorithms can segment stations or users into groups based on similarities. In bike-sharing, this could identify clusters of stations with similar usage patterns or clusters of users with identical riding habits, which can be vital for targeted marketing or service provision. The Mixed and Linear Integer Programming Technique is an optimization technique that can help make decisions like where to place new bike stations or how to redistribute bikes for maximum efficiency, given certain constraints. Long Short-Term Memory (LSTM) is a type of recurrent neural network. LSTM is beneficial for time series forecasting. Given the temporal nature of bike-sharing data (hourly, daily usage), LSTM can predict future bike demand based on historical data, which is crucial for planning and operations. Neural Networks can be used for various predictive tasks in bike-sharing, such as forecasting demand or detecting anomalies. Neural networks can process complex datasets and uncover patterns that might be non-linear or intricate. Gradient-Boosted Regression Trees (GBRT) is a machine learning method that can be applied to classification and regression problems. In bike-sharing, it could help predict variables like future demand or classify usage patterns. We sorted the selected papers into predictive or prescriptive analytics. While the primary objective of predictive analytics is to forecast future events or outcomes based on historical and current data, prescriptive analysis is used to recommend specific actions to handle or benefit from a predicted future scenario.

Limitations of this work include its exclusive reliance on real-time datasets neglecting simulation methodologies in investigating reallocation strategies for bike-sharing systems. Additionally, while the taxonomy is focused on emerging concerns and the corresponding optimization techniques, it needs an extensive analysis of the outcomes. Therefore, further dimensions with more details about the literature could be added, ultimately spanning the arc from problem approaches to results.

4. Summary of Results

4.1. Problems

Imbalance between supply and demand

Maintaining a supply and demand balance is crucial for a successful bike-sharing system (Y. Zhao et al., 2019). It becomes challenging to ensure the sustainability of bike-sharing without appropriately addressing the imbalance between supply and demand, especially during rush hours because of unbalanced traffic patterns in the cities (Chiariotti et al., 2020; Ji et al., 2020; Shan et al., 2020; Su et al., 2022). This issue can significantly impact urban traffic systems, as customers may encounter empty stations requiring a bike (Chiariotti et al., 2020; Y. Zhao et al., 2019). Without outside bike-sharing system interventions, the unequal demand distribution usually produces an imbalance between supply and demand for such services (Ji et al., 2020). According to Ji et al. (2020), during the morning rush hour, there would be an increase in passengers using shared bikes to go to subway stations. It might be difficult and frustrating to come across an empty dock while using the docked bike-sharing system. With a dockless bike-sharing system, the rental process is given more attention, and the imbalance problem is more visible because of the limited return locations (Ji et al., 2020; Wu, 2020). Also, overcrowded stations may negatively impact customer's experiences since they have to search for another station to return or rent a bike, which affects user interest in using the bike-sharing system (Zhu et al., 2023).

Accurately predicting destinations in a bike-sharing system can assist cities in establishing environmentally friendly transportation solutions for short distances (Dai et al., 2018; Hulot et al., 2018). However, predicting demand is often complicated because of weather conditions, traffic patterns, periods, and other dynamic factors (Chiariotti et al., 2020; Qiao et al., 2021). The authors examined how user demand for both docked and dockless bike-sharing systems is affected by factors such as bike-sharing fleets, socio-demographic characteristics, and land use, considering variations across different spatial locations and time periods (Ma et al., 2020). A separate study empirically explored riding behavior within the dimensions of both time and space (Y. Wang et al., 2022). Even the aspect of fairness in the distribution of bikes is accessed. Duran-Rodas et al. (2021) show that people need to be able to pay for the bike that they want to drive in the first place, and the authors confirm flaws in the spatial distribution of cycles depending on the prosperity of the specific region (Duran-Rodas et al., 2021).

Operational inefficiency

Bike-sharing system operators need help allocating resources effectively to ensure the continuous operation of a flexible and fully functional system. Fixing optimal resource allocation management is helpful (Drosouli et al., 2021). Improper or thoughtless parking of free-floating bicycle sharing (FFBS) can result in disputes between FFBS companies, who strive to maintain a sustainable business model, and local government authorities, who are concerned about the negative impact on the city's aesthetics and public perception (D. Zhao & Ong, 2021). Placing electric scooters or free-floating bikes on narrow sidewalks poses accessibility challenges for other individuals utilizing the same infrastructure. Moreover, accommodating these trips in designated parking facilities presents another obstacle, as finding suitable locations to accommodate a specific number of parking facilities can be problematic

(Sandoval et al., 2021). Broken bikes sometimes create an imbalance in bike availability. A recent study shows that broken bikes in China are estimated to reach 20 million annually (Xu & Zou, 2022). Because of the unavailability of bicycles at docking stations, it could be difficult for users to rent and return bikes as needed, and operational inefficiencies are caused by the availability imbalance and flawed rebalancing process (Afonso et al., 2022). Another study investigated that traditional methods, like truck-based relocation, need help keeping up with the changing demand for bikes and have major operating expenses (Zhu et al., 2023).

Traffic patterns and unbalanced bike distribution

As dockless bike-sharing traffic mode has increased rapidly, the bike-sharing's unbalanced spatial distribution and low utilization rate have rapidly expanded (Y. Zhao et al., 2019). There are so many reasons why users consider (not) using a shared bike. Influencing factors include distance, time, environment, activity type proportion, and service capacity of points of interests (POIs) as well as areas of interests (AOIs) (S. Li et al., 2021). Different traffic patterns in specific geographical regions can lead to empty or crowded docking stations during rush hours due to an imbalance in reallocation and an inaccurate demand prediction (Chiariotti et al., 2020; Hulot et al., 2018). Dockless bike-sharing faces challenges in reallocation and management due to unbalanced distribution over time and space (Liu & Xu, 2018). The asymmetrical spatial and temporal distribution of user demand leads to unstable bike distribution in bike-sharing systems (He et al., 2021). Due to spatial and temporal distribution, overflow or underflow could happen. Unbalanced bike distribution in bike-sharing systems highlights the need for reallocation to address empty or full stations (Afonso et al., 2022).

To solve the problem of unbalanced demand, the efficiency of the reallocation system needs to be improved (Y. Wang et al., 2022). Focusing on hourly demand prediction for demand rentals and returns at each system station could be an excellent solution to this problem (Hulot et al., 2018). The user's travel patterns differ between docked and dockless bike-sharing systems. Also, bike-sharing fleets, socio-demographic characteristics, and land use factors impact their demand within these systems (Ma et al., 2020). As bike-sharing systems encompass a wide range of bike availability and dynamic bike flows across different locations, operators face the challenge of redistributing bikes throughout the day within such expansive and evolving networks (Y.-J. Wang et al., 2022). Not only should bike stations get clustered by geographical location to improve the efficiency of bike-repositioning operations, a demand-driven approach to clustering bike stations should also be considered (Y.-J. Wang et al., 2022). At times of overflow, stations may carry more return bikes than they can accommodate, or underflow when vacant stations have no available bikes to rent (Wu, 2020).

Social fairness and equality

The successful implementation of a bike-sharing network relies on the assistance of public authorities and compliance with safety standards for bike lanes (Caggiani et al., 2020). Additionally, key considerations include the time required for implementation and the costs associated with developing the necessary infrastructure for the bike network. The prices vary based on the presence or absence of stations in the bike scheme, with docking stations and their used space being the most expensive components (Caggiani et al., 2020). Although metrics like the attributes of bike users, bike station features, and station accessibility are typically taken into account, achieving equal accessibility across diverse demographic groups in the entire urban area is often neglected (Caggiani et al., 2020). When considering spatial efficiency, there is a risk of excluding underprivileged areas from accessing the system (Duran-Rodas et al., 2021).

Importance of environmental impacts

Sustainability is a topic of public interest when evaluating dockless bike-sharing. The impact of dockless bike-sharing systems on the environment, public interest, and individual users involves assessing various factors such as efficient resource utilization through bike-sharing, reduction of greenhouse gas (GHG) emissions, enhancement of urban transport efficiency, and the exploitation of public space resources in terms of usage and occupation (Tao & Zhou, 2021). Further, the impact of bike-sharing systems on the environment can be divided into three categories: land use, transport facility, and accessibility (Chen & Ye, 2021). If not appropriately considered, the appearance of too many bikes on the streets can lead to conflicts. The convenience of parking bikes everywhere also brings unintended consequences: the surplus of freely roaming bicycles in the city, the problem of unauthorized bicycle parking, and broken bikes. A recent study shows that in China, broken bikes create environmental issues, affecting user experience and polluting the environment (Teng et al., 2020). The indiscriminate placement of shared bicycles obstructs the smooth traffic flow and hinders the movement of pedestrians and vehicles

(D. Zhao & Ong, 2021). Even people with disabilities can be prevented from moving freely on the sidewalks if bikes or scooters are wrongly parked (Sandoval et al., 2021).

4.2. Data Scenarios

Aggregated real-time dataset of bike-sharing systems

Anywhere in the world where bike-sharing systems are installed, data accrues and much data is collected in China, as some of the most extensive bike-sharing systems worldwide are located there (Chang & Ferreira, 2021; Ma et al., 2020). Recording bike trajectories enables real-time analysis of regional bike operations and travel patterns, facilitating a balance between bike supply and demand (Tao & Zhou, 2021). Every entry within the real-world dataset comprises user identification, bike identification, origin and destination coordinates, departure time, and arrival time (Chen & Ye, 2021; Y.-J. Wang et al., 2022). Additionally, the corresponding station IDs of docked bike-sharing systems are analyzed (Ma et al., 2020). Using a multisource dataset can help to enhance the accuracy (Y. Wang et al., 2022). Aggregated real-time data of bike-sharing systems consists of bike journeys, including the timing and positions of rental start and end points, station placements, and the service area boundaries (Duran-Rodas et al., 2021). Necessary for the usage of aggregated data is to prepare and clean the data sufficiently. That includes removing unusual bike trips shorter or longer than a fixed distance and distance and travels without a complete dataset (S. Li et al., 2021; Ma et al., 2020; Sandoval et al., 2021).

Data with an average of 1.9 shared bikes per person was used by the authors and gathered from Wuhan, Hubei province, China, on January 20, 2019 (Y. Zhao et al., 2019). Real data was collected from the New York Citi Bike system through 2017 to test these matters – distribution of interarrival times and the approval of their predictions (Chiariotti et al., 2020; Dai et al., 2018; Y. Li et al., 2018). The historical dataset of the Montreal bike-sharing system, BIXI Montreal, between April 2015 and November 2016 has been used for calculating the matrix of dimension (number of hours) × (number of stations) (Hulot et al., 2018). For a different study, historical information about the New York City bike-sharing system was gathered in the Jersey City neighborhood between January 2019 and January 2020. This information included the trip's duration, start and stop timestamps, start and end station names, station identifiers, station latitude and longitude, bike identifiers, user type (customer or subscriber), gender, and birth year. Fifty-three stations were open then, and 609 bikes were used (Drosouli et al., 2021). The authors used the average demand for pickups and the intermediate need for drop-offs per hour in one day at all stations in New York. Data was collected from publicly available sources from the website of Citi Bike, as of May 2019 for “A Multi-Objective Optimization Model for Bike-Sharing” (Shan et al., 2020).

The bike-sharing system's aggregated real-time and historical dataset shows how bicycles are distributed and moved over time. It uses past bike location records to improve its accuracy and usefulness (Liu & Xu, 2018). By analyzing historical data, it is possible to gain insights into the availability of bicycles at each station within the network and track the journeys taken throughout the year (Afonso et al., 2022). Historical data records, station details, timings, and other information are vital for understanding bike-sharing dynamics (Zhu et al., 2023). The article "How Does Dockless Bike-Sharing System Behave by Incentivizing Users to Participate in Rebalancing" used historical trip data from the Mobike bike-sharing company, explicitly focusing on five weekdays in September 2017. Based on average travel demand patterns, the data determines centroids for 1,134 of 1,185 regular hexagons in Nanjing's downtown study area (Ji et al., 2020). Nanjing Public Bicycle Company's datasets also provide historical records of bike-sharing journeys in Nanjing from September 1 to September 30, 2017. These datasets contain bike-sharing trip records, station specifics, and user attributes (He et al., 2021).

Simulated Dataset

The authors use a generic grid layout adaptable to various urban configurations and public transportation routes to simulate an urban environment. This simulation primarily depends on zone population and point-to-point travel duration (Caggiani et al., 2020). In our sample of articles, only one paper used simulated data. The reason could be the broad availability of bike-sharing data and, thus, no need for a simulation process.

Spatial Dataset

Spatial data analysis relies on datasets that provide comprehensive geographic insight. Spatial data categories can be described by the concentration of features within each spatial segment, the proportion of parts based on their class, or pedestrian accessibility, determined using an exponential cost calculation (Duran-Rodas et al., 2021). One such dataset is the Mobike bicycle dataset in Chengdu, China, which includes essential details like longitude and latitude. It is an excellent example of a spatial dataset enriched with geographic information (Zhang et al., 2019). Another similar dataset involves the dockless bicycles in Nanjing, which provides real-time insights into the spatial distribution and trajectories of these bikes through GPS coordinates. This dataset also counts as spatial data since it integrates geographical information in another paper where the study develops into the differences in bike rental and return patterns across various locations (Liu & Xu, 2018). This in-depth analysis relies on geographic information analysis, revealing valuable insights about the disparities in different parts of the city (Wu, 2020). The dataset in Zhao and Ong's (2021) paper consists of roughly 34.3 million distinct location data entries for Mobike shared bicycles situated on Xiamen Island, China (D. Zhao & Ong, 2021).

Weather Dataset

Bike-sharing systems will be affected by dynamic factors like the weather, which challenge shared bike scheduling (Qiao et al., 2021). Weather data is essential for bike-sharing systems as it offers valuable insights into environmental conditions that can significantly affect user behavior and system operations. Weather data includes different variables such as temperature, wind direction, wind scale, wind speed, air pressure, precipitation probability, and humidity with hourly temporal granularity (Afonso et al., 2022; Zhang et al., 2019). Analyzing weather data can assist in predicting bike demand trends with greater accuracy (Zhu et al., 2023). Weather data, along with the historical dataset of the Montreal bike-sharing system, BIXI Montreal, between April 2015 and November 2016, has been used for calculating the matrix of dimension (number of hours) × (number of stations) (Hulot et al., 2018). Weather data from New York City was used for experimenting with a principle system that only covers urban centers in Citi Bike’s accurate data, including station status data, dock demand at each station, and more (Y. Li et al., 2018).

4.3. Optimization approaches

4.3.1. Predictive types of optimizations

Ordinary Least Squares (OLS)

A standard method for investigating the determinants of bike-sharing user demand is through a regression model. However, the OLS approach has faced criticism for frequently overlooking bike-sharing usage's spatial and temporal fluctuations. A geographically weighted regression (GWR) or a geographically and temporally weighted regression (GTWR) can be a helpful expansion (Ma et al., 2020). Those analyses can determine the user's demand over space and time and create plots of density graphics that show the predicted bike-sharing system usage in certain regions (Ma et al., 2020).

The Process Mining technique in bike-sharing systems, as seen in Figure 3, solves critical inquiries about how the system functions, such as insights information, identifying travel patterns, discovering user's demands, and revealing bike imbalance problems, which will assist the professionals adequately handling the bike distribution (Drosouli et al., 2021). It provides explainable visualizations based on frequency and time data to better understand and analyze the process (Drosouli et al., 2021).

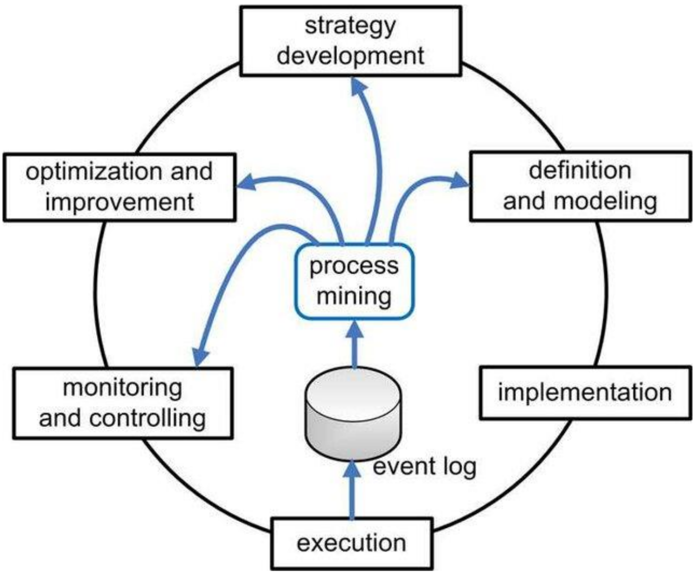

Figure 3: Process Mining (Houy et al., 2010)

Birth-Death Process (BDP)

The Birth-Death Process (BDP) in a bike-sharing system is a formal stochastic modeling framework employed to depict the system's dynamic behavior by capturing the arrivals and departures of bikes at different stations over a given period. This modeling approach can be applied to various fields, including population dynamics and analyzing transportation systems (Chiariotti et al., 2020). Within the BDP, “birth" signifies the arrival of a new bike at a station, while "death" represents the departure of a bike from a station. It is assumed that these arrivals and departures transpire independently and at random intervals (Chiariotti et al., 2020)

Clustering algorithms (Spatial and Temporal)

A clustering algorithm should help facilitate efficient bike repositioning, visualize the user’s demand, and enhance the user’s utility (Y.-J. Wang et al., 2022, p. 2). Wang et al. (2022) split large bike-sharing networks into smaller subnetworks for repositioning. They propose a method called geographical clustering, which finds communities within the data to reach their objective. The clusters give insight into bike-sharing systems' high-demand zones (Y.-J. Wang et al., 2022). Various researchers cluster bike-sharing systems into three categories: location-based clustering, a clustering technique according to their geographic coordinates, and pattern-based clustering, which recognizes usage patterns based on geographic locations. Additionally, count series clustering utilizes trip data to depict station usage and movement patterns (Dai et al., 2018; Shan et al., 2020). A Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and a classical k-means clustering algorithm has been used to determine optimal parking lots (D. Zhao & Ong, 2021). They introduce clusters characterized by parking demand to pinpoint potential sites and capacities for bicycle parking facilities during the planning phase (D. Zhao & Ong, 2021). DBSCAN and k-means (KM), as well as Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), as an extension to DBSCAN, were also utilized in Sandoval et al. 's (2021) analysis of placement selection for bike and e-scooter parking facilities (Sandoval et al., 2021). They aim to predict a series of specific spots where micro-mobility parking facilities could be placed to ensure that as many unique trips as possible can be used, as depicted in Figure 4 below.

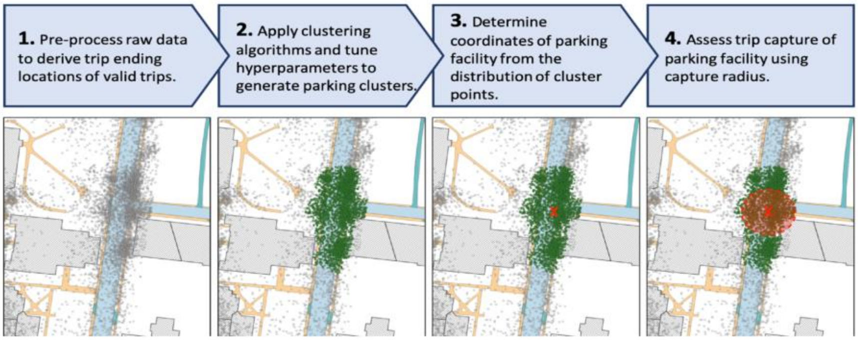

Figure 4: Step-by-step process of the clustering's algorithm utilization (Sandoval et al., 2021)

Clustering algorithms are used in unsupervised machine learning to analyze data and gather numerous data points into several categories by considering their inherent traits or patterns (Dai et al., 2018; Shan et al., 2020; Y. Zhao et al., 2019). Y. Zhao et al. (2019) employed four clustering algorithms, including KM (a distance-based clustering algorithm), Ant Colony clustering algorithm (ACO – which makes use of a positive feedback mechanism and distributed parallel computing), Fuzzy C-Means clustering algorithm (FCM – a genetic algorithm for pre-classifying rental records), and Mean Shift clustering algorithm (MS – a clustering algorithm that identifies high-density regions iteratively), to examine the connection between geographic location and the density of bike distribution (Y. Zhao et al., 2019).

Based on demand prediction, X. Zhang et al. (2019) formulated the "Zone Based Two-Stage Rebalancing Model" clustering model. The model's objective is to optimize the level of synchronization between bike demand and the quantity of shared bikes allocated to each zone (Zhang et al., 2019). The researched geographical area is split into two categories in this proposed model – one zone with deficient bikes (ZDB) and a second zone with sufficient bikes (ZSB). Given the accompanying dispatching expenses and the various levels of bike shortages across different zones, certain zones can achieve balance by only moving bikes from nearby zones. Keeping reasonable cycles in each zone is essential for the best system functioning. In the first stage, rebalancing range size depends on bike shortage level, with one ZDB receiving bikes from other ZSB within range. In the second stage, ZDB bikes can be acquired by progressively expanding the rebalancing range until balanced (Zhang et al., 2019).

Mixed and Linear Integer Programming Techniques

Mixed Integer Programming (MIP) is an extension of linear programming (LP) that allows integer variables in addition to continuous variables and determines the optimal route for one or more vehicles visiting predetermined stations, more specifically bicycle or bike routing problems (Chiariotti et al., 2020). Mixed-Integer Linear Programming (MILP) is often utilized for system analysis and optimization because it provides a flexible and efficient solution to address significant and complex problems, such as the transportation problem (Kantor et al., 2020). This model combines multiple vehicle routing and news vendor sub-issues (Teng et al., 2020). They can provide constructive suggestions for managing dockless bikes, including reasonably setting demand and supply (Liu & Xu, 2018) and creating station rebalancing tasks with the shortest cycling distance (Zhu et al., 2023). In a few cases, the authors utilized Linear Integer Programming Techniques, a matter of Mixed Integer Programming where all variables are restricted to integer values only, without allowing continuous variables and predicting the algorithm (Hulot et al., 2018).

The primary purpose of using Integer Programming Techniques is to solve bike allocation difficulties, bike routing and scheduling problems, effective rebalancing strategies, fill the gap in traffic prediction and pricing as well as incentive design (Chiariotti et al., 2020; Hulot et al., 2018). Another objective of using MILP is to maximize revenue by considering bike demand, returns, recalled and replenished bikes, and optimal vehicle routes for delivery and pickup operations (Teng et al., 2020). Research has indicated that implementing the MILP approach to assess the recycling of damaged bikes in bike-sharing systems reduces costs and improves productivity (Xu & Zou, 2022).

Neural Networks

Demand predictors could be based on the interactions of variables like land use, points of interest (POIs), time, transportation infrastructure, and weather significantly impacting dockless bike-sharing usage (Y. Wang et al., 2022). However, if linear assumptions were made, they would miss the intricate connections between the influencing factors, producing contradictory results. The random forest (RF) model, known for its ability to handle nonlinear variables, is well suited to estimate dockless bike-sharing mobility patterns. Partial dependence plots (PDP) can be utilized to assess the nonlinear associations between factors and dockless bike-sharing usage. Consequently, variables such as the distance to specific points of interest, such as restaurants or bus stops, are quantitatively estimated and utilized as demand indicators (Y. Wang et al., 2022). PDP plots offer practical directions for repositioning in bike-sharing systems and urban planning because they show the nonlinear effects of the built environment on bike-sharing usage (Chen & Ye, 2021).

Dynamic Convolutional Neural Networks (DCNNs) are a type of neural multi-layer network structure, each layer consisting of multiple feature planes and can retrieve dynamic features from the static data. In contrast, the traditional shared bike demand forecasting algorithms only consider static factors based on experiences. They cannot meet the real-time demand forecasting for bikes at the stations (Qiao et al., 2021). It can learn and update the filters during the training process, and a dynamic matrix is employed to predict the bike demand at each station under different weather conditions in different periods (Qiao et al., 2021). The proposed shared-bike demand forecasting model based on DCNNs contains four working layers. The first is the data acquisition layer, which saves information regarding each state and station per minute. The second one is the data preprocessing layer, which analyzes the impact of weather and time factors on bike demand. The third is the DCNN layer, which predicts the bike demand; the fourth is the data visualization layer, which visualizes the forecasted results and compares them with the original dataset (Qiao et al., 2021).

4.3.2. Prescriptive types of optimizations

LSTM (Long Short-Term Memory)

LSTM, a widely used bike-sharing prescription technique, is extensively applied to capture the time-dependent features of time series data, and it is often complemented by Graph Convolutional Networks (GCN). GCNs are commonly employed to handle data structures like social network data and measurements from meteorological station networks (He et al., 2021). The authors have integrated the GCN model into their LSTM model to predict bike-sharing demand. In this so-called GC-LSTM model, LSTM captures the temporal characteristics of bike-sharing trips at the station level. At the same time, GCN extracts the structural elements of the bike-sharing network through convolution operations (He et al., 2021). A neural network with three layers incorporates GC-LSTM to forecast the bike-sharing rental and return demand on a station-by-station basis. The GC-LSTM model accurately predicts short-term bike-sharing requests, containing exogenous factors such as land use, weather, and user data to determine rebalancing quantity (He et al., 2021). Combining LSTM with

GCN makes capturing the bike-sharing network's temporal properties and structural aspects possible to give optimal advice (He et al., 2021).

Gradient-Boosted Regression Trees (GBRT)

GBRT is a data mining technique that offers significant advantages. It is a regression algorithm that combines the strengths of decision trees and gradient descent optimization used to predict the total number of bikes in a bike-sharing system (Qiao et al., 2021). It merges two critical algorithms: regression trees, which are a group of models employed for regression analysis, and boosting, an adaptive technique that combines multiple models to improve predictive precision. This integration of regression trees and expansion enables GBRT to provide advanced data mining capabilities and improve predictions' overall performance (Chen & Ye, 2021). With the modeling results of GBRT and the help of PDPs, the effective ranges and tipping points of built environment factors to encourage bike-sharing can be identified (Chen & Ye, 2021).

This model aims to produce highly accurate prescriptions and functions with various data types (Qiao et al., 2021). It creates multiple regression trees through several iterations, with each learner being trained using the residuals from the previous learner. To determine the final learner, it computes the weighted sum of numerous obtained learners (Qiao et al., 2021). The GBRT model is enhanced iteratively using gradient descent optimization. The following decision trees are constructed to forecast the negative gradients or residuals from the previous trees. Each new tree aims to minimize the loss function, typically mean squared error, by adjusting to the negative slopes (Qiao et al., 2021).

4.3.3. Customized Methods

With further algorithms and methodologies, custom-based optimizations are designed to improve systems or processes per particular restrictions and objectives while considering all relevant factors and limitations. Authors devised different custom-made optimization methods to solve bike-sharing relocation problems, presented below.

Life-Cycle Analysis

The Life-Cycle Analysis evaluates the resources required, greenhouse gas (GHG) emissions, user transport time and expenses, and the demand for road and roadside parking space associated with transporting one individual for a distance of one kilometer. It considers the environmental impacts of resource extraction, manufacturing, distribution, use, and end-of-life treatment. LCA helps identify potential environmental hotspots and guides decision-making for bike-sharing providers (Tao & Zhou, 2021).

Demand And/oR Equity (DARE)

DARE should help improve the planner's prioritized infrastructure distribution according to their bike-sharing system goals by transparently modifying the weights for spatial fairness and efficiency and communicating these priorities for a fair design (Duran-Rodas et al., 2021). A different approach for improving the equality (accessibility and coverage) in bike-sharing systems is an author-constructed network model with 693 nodes and 2616 arcs on a defined grid representing a specific region. It optimizes the bike distribution based on equalized demand while having budget restrictions (Caggiani et al., 2020).

Gravity Model

The gravity model explains the correlations between two regions in a geographical sense. The authors refer to a bike-sharing system and use the model to examine how the distribution of activity types and the service capacity of POIs affect the conclusions drawn regarding trip purposes (S. Li et al., 2021). According to this model, the level of interaction and influence between two regions is directly proportional to their respective populations. As the distance between them increases, the relationship becomes inversely proportional to the square of that distance (S. Li et al., 2021). The outcomes of this analysis yield spatial-temporal patterns, which can be visualized in the form of heat maps. These heat maps depict regions characterized by high demand, clearly representing areas where the need for a particular variable or phenomenon is notably elevated, especially POIs (S. Li et al., 2021).

Process Mining Technique

The process mining method offers crucial details regarding the structure and operation of the bike-sharing system. It can provide insightful information, such as travel patterns, user demand, and bike imbalance issues at each station, assisting domain experts in resolving the bike distribution issue and achieving optimal resource allocation (Drosouli et al., 2021). It finds patterns, trends, and specifics in event logs. To better understand and analyze processes, explainable visualizations based on frequency and time data are also provided (Drosouli et al., 2021).

Spatio-Temporal Reinforcement Learning model (STRL)

The STRL model is an approach that combines reinforcement learning techniques with spatial and temporal information to learn optimal actions in dynamic environments over time and also targets minimizing customer loss over a long period (Y. Li et al., 2018). This model is carefully constructed to capture the system dynamics and real-time uncertainties using a deep neural network to estimate the best long-term value function for each STRL, from which the best repositioning strategy can be deduced (Y. Li et al., 2018).

User-based Model

Bicycle-sharing systems face asymmetric demand-supply relationships, particularly during rush hours, making dynamic rebalancing challenging for operators due to remote pick-up and drop-off processes (Ji et al., 2020). To enhance system patronage, these authors developed a user-based model for rebalancing or addressing the imbalance issue of bike-sharing systems. It is specially designed for a dockless bike-sharing system (Ji et al., 2020). The fundamental concept behind this model is to use monetary incentives to encourage bike return to bike-deficient regions rather than surplus regions within an acceptable maximum walking distance, and they are expected to respond to the system in a way that maximizes their utility (Ji et al., 2020). The principle is shown in Figure 5 below.

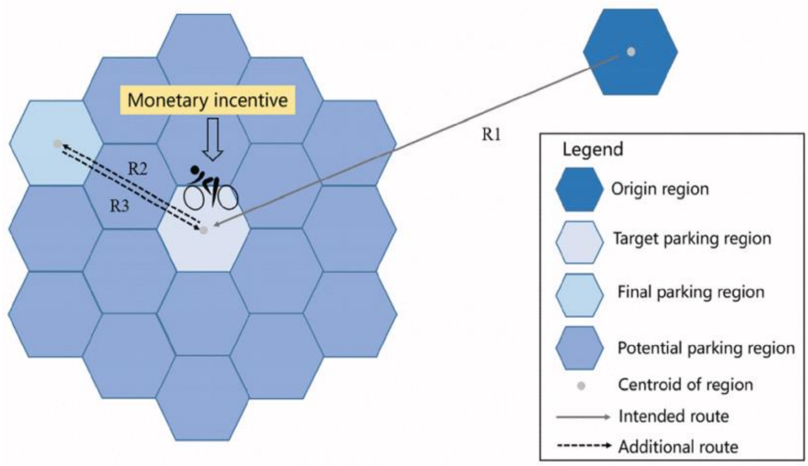

Figure 5: Simple illustration of the User-based Model (Ji et al., 2020)

5. Conclusion & Outlook

In conclusion, this paper explores the vast field of bicycle-sharing systems, providing insight. We have systematically examined and categorized the findings from 30 recent and relevant research papers, primarily focusing on addressing the challenges faced by modern bike-sharing systems, whether they employ a dockless or station-based approach. One of the central themes that emerged from our analysis is the critical importance of maintaining a delicate balance between supply and demand within these systems, as this equilibrium largely dictates their success. Authors have highlighted various operational issues and external effects, ranging from traffic patterns to environmental considerations and even concerns related to social fairness.

Our taxonomy has revealed diverse problem-solving methods researchers employ, including clustering algorithms, mixed and linear integer programming, and artificial neural networks. These methods, often driven by real-time, simulated, weather, and spatial data, have generated results that assist in identifying high-demand zones, mitigating bike imbalance problems, and optimizing operational costs, all while maximizing customer satisfaction. The findings of this research underscore the significance of bicycle-sharing systems in the broader context of smart cities, offering flexible, environmentally friendly transportation solutions that contribute to reduced emissions, improved public health, and alleviated traffic congestion. These systems have the potential to enhance urban mobility, integrate seamlessly with other modes of public transportation, and foster sustainable urban development.

However, it is essential to acknowledge the limitations and gaps identified in the literature. Despite the wealth of research available, specific geographic contexts, operational challenges, and emerging trends still need to be explored. Further empirical studies and cross-disciplinary collaborations are required to address these gaps, enabling the development of context-specific reallocation solutions that cater to the unique characteristics of different cities and user populations. This review shows that no one-size-fits-all reallocation strategy exists, and the design choice should be context-specific, considering the unique aspects of each bike-sharing system.

6. Bibliography

Afonso, A. S., Pires, J. M., Datia, N., & Birra, F. (2022). Bicycle Demand Prediction to Optimize the Rebalancing of a Bike Sharing System in Lisbon. 2022 26th International Conference Information Visualisation (IV), 366–372. https://doi.org/10.1109/IV56949.2022.00067

Caggiani, L., Colovic, A., & Ottomanelli, M. (2020). An equality-based model for bike-sharing stations location in bicycle-public transport multimodal mobility. Transportation Research Part A: Policy and Practice, 140, 251–265. https://doi.org/10.1016/j.tra.2020.08.015

Chang, S. K. J., & Ferreira, A. F. (2021). Bike-Sharing System: Uncovering the "Success Factors." In R. Vickerman (Ed.), International Encyclopedia of Transportation (pp. 355–362). Elsevier. https://doi.org/10.1016/B978-0-08-102671-7.10348-3

Chen, E., & Ye, Z. (2021). Identifying the nonlinear relationship between free-floating bike sharing usage and built environment. Journal of Cleaner Production, 280, 124281. https://doi.org/10.1016/j.jclepro.2020.124281

Chiariotti, F., Pielli, C., Zanella, A., & Zorzi, M. (2020). A Bike-sharing Optimization Framework Combining Dynamic Rebalancing and User Incentives. ACM Transactions on Autonomous and Adaptive Systems, 14(3), 11:1-11:30. https://doi.org/10.1145/3376923

Dai, P., Song, C., Lin, H., Jia, P., & Xu, Z. (2018). Cluster-Based Destination Prediction in Bike Sharing System. Proceedings of the 2018 Artificial Intelligence and Cloud Computing Conference, 1–8. https://doi.org/10.1145/3299819.3299826

Drosouli, I., Theodoropoulou, G., Miaoulis, G., & Voulodimos, A. (2021). A Process Mining Approach for Resource Allocation Management in a Bike Sharing System. Proceedings of the 24th Pan-Hellenic Conference on Informatics, 327–333. https://doi.org/10.1145/3437120.3437334

Duran-Rodas, D., Wright, B., Pereira, F. C., & Wulfhorst, G. (2021). Demand And/oR Equity (DARE) method for planning bike-sharing. Transportation Research Part D: Transport and Environment, 97, 102914. https://doi.org/10.1016/j.trd.2021.102914

He, M., Ma, X., & Jin, Y. (2021). Station Importance Evaluation in Dynamic Bike-Sharing Rebalancing Optimization Using an Entropy-Based TOPSIS Approach. IEEE Access, 9, 38119–38131. https://doi.org/10.1109/ACCESS.2021.3063881

Houy, C., Fettke, P., Loos, P., Aalst, W., & Krogstie, J. (2010). BPM-in-the-Large – Towards a Higher Level of Abstraction in Business Process Management (Vol. 334, p. 244). https://doi.org/10.1007/978-3-642-15346-4_19

Hulot, P., Aloise, D., & Jena, S. D. (2018). Towards Station-Level Demand Prediction for Effective Rebalancing in Bike-Sharing Systems. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 378–386. https://doi.org/10.1145/3219819.3219873

Ji, Y., Jin, X., Ma, X., & Zhang, S. (2020). How Does Dockless Bike-Sharing System Behave by Incentivizing Users to Participate in Rebalancing? IEEE Access, 8, 58889–58897. https://doi.org/10.1109/ACCESS.2020.2982686

Kantor, I., Robineau, J.-L., Bütün, H., & Maréchal, F. (2020). A Mixed-Integer Linear Programming Formulation for Optimizing Multi-Scale Material and Energy Integration. Frontiers in Energy Research, 8. https://www.frontiersin.org/articles/10.3389/fenrg.2020.00049

Li, S., Zhuang, C., Tan, Z., Gao, F., Lai, Z., & Wu, Z. (2021). Inferring the trip purposes and uncovering spatio-temporal activity patterns from dockless shared bike dataset in Shenzhen, China. Journal of Transport Geography, 91, 102974. https://doi.org/10.1016/j.jtrangeo.2021.102974

Li, Y., Zheng, Y., & Yang, Q. (2018). Dynamic Bike Reposition: A Spatio-Temporal Reinforcement Learning Approach. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 1724–1733. https://doi.org/10.1145/3219819.3220110

Liu, M., & Xu, X. (2018). Dockless Bike-Sharing Reallocation Based on Data Analysis: Solving Complex Problem with Simple Method. 2018 IEEE

Third International Conference on Data Science in Cyberspace (DSC), 445–450. https://doi.org/10.1109/DSC.2018.00072

Ma, X., Ji, Y., Yuan, Y., Van Oort, N., Jin, Y., & Hoogendoorn, S. (2020). A comparison in travel patterns and determinants of user demand between docked and dockless bike-sharing systems using multi-sourced data. Transportation Research Part A: Policy and Practice, 139, 148–173. https://doi.org/10.1016/j.tra.2020.06.022

Qiao, S., Han, N., Huang, J., Yue, K., Mao, R., Shu, H., He, Q., & Wu, X. (2021). A Dynamic Convolutional Neural Network Based Shared-Bike Demand Forecasting Model. ACM Transactions on Intelligent Systems and Technology, 12(6), 70:1-70:24. https://doi.org/10.1145/3447988

Sandoval, R., Van Geffen, C., Wilbur, M., Hall, B., Dubey, A., Barbour, W., & Work, D. B. (2021). Data driven methods for effective micromobility parking. Transportation Research Interdisciplinary Perspectives, 10, 100368. https://doi.org/10.1016/j.trip.2021.100368

Shan, Y., Xie, D., & Zhang, R. (2020). A Multi-Objective Optimization Model for Bike-Sharing. Proceedings of the 2019 7th International Conference on Information Technology: IoT and Smart City, 383–387. https://doi.org/10.1145/3377170.3377175

Su, S., Xiong, D., & Yu, H. (2022). A Multiobjective Evolutionary Algorithm with Variable Neighborhood Search for the Dynamic Rebalancing of a Bike-sharing System. Proceedings of the 4th International Conference on Management Science and Industrial Engineering, 421–428. https://doi.org/10.1145/3535782.3535837

Tao, J., & Zhou, Z. (2021). Evaluation of Potential Contribution of Dockless Bike-sharing Service to Sustainable and Efficient Urban Mobility in China. Sustainable Production and Consumption, 27, 921–932. https://doi.org/10.1016/j.spc.2021.02.008

Teng, Y., Zhang, H., Li, X., & Liang, X. (2020). Optimization Model and Algorithm for Dockless Bike-Sharing Systems Considering Unusable Bikes in China. IEEE Access, 8, 42948–42959. https://doi.org/10.1109/ACCESS.2020.2967398

Wang, Y., Zhan, Z., Mi, Y., Sobhani, A., & Zhou, H. (2022). Nonlinear effects of factors on dockless bike-sharing usage considering grid-based spatiotemporal heterogeneity. Transportation Research Part D: Transport and Environment, 104, 103194. https://doi.org/10.1016/j.trd.2022.103194

Wang, Y.-J., Kuo, Y.-H., Huang, G. Q., Gu, W., & Hu, Y. (2022). Dynamic demand-driven bike station clustering. Transportation Research Part E:

Logistics and Transportation Review, 160, 102656. https://doi.org/10.1016/j.tre.2022.102656

Wu, J. (2020). Challenges and opportunities in algorithmic solutions for re-balancing in bike sharing systems. Tsinghua Science and Technology, 25(6), 721–733. https://doi.org/10.26599/TST.2020.9010002

Xu, G., & Zou, A. (2022). A Recycling Routing Problem of Broken Bikes With Incentives in Bike Sharing Systems. IEEE Access, 10, 106191–106201. https://doi.org/10.1109/ACCESS.2022.3211945

Zhang, X., Yang, H., Zheng, R., Jin, Z., & Zhou, B. (2019). A Dynamic Shared Bikes Rebalancing Method Based on Demand Prediction. 2019 IEEE Intelligent Transportation Systems Conference (ITSC), 238–244. https://doi.org/10.1109/ITSC.2019.8917099

Zhao, D., & Ong, G. P. (2021). Geo-fenced parking spaces identification for free-floating bicycle sharing system. Transportation Research Part A: Policy and Practice, 148, 49–63. https://doi.org/10.1016/j.tra.2021.03.007

Zhao, Y., Dai, L., Peng, L., Song, Y., & Zhou, Z. (2019). Analysis of Spatial Distribution of China's station-free bike-sharing by Clustering Algorithms. Proceedings of the 2019 4th International Conference on Mathematics and Artificial Intelligence, 15–19. https://doi.org/10.1145/3325730.3325748

Zhu, H., Shou, T., Guo, R., Jiang, Z., Wang, Z., Wang, Z., Yu, Z., Zhang, W., Wang, C., & Chen, L. (2023). RedPacketBike: A Graph-Based Demand Modeling and Crowd-Driven Station Rebalancing Framework for Bike Sharing Systems. IEEE Transactions on Mobile Computing, 22(7), 4236–4252. https://doi.org/10.1109/TMC.2022.3145979