Received January 29, 2021, accepted February 28, 2021, date of publication March 4, 2021, date of current version March 12, 2021.

*Digital Object Identifier 10.1109/ACCESS.2021.3063881*

# Station Importance Evaluation in Dynamic Bike-Sharing Rebalancing Optimization Using an Entropy-Based TOPSIS Approach

**MINGJIA HE** [1,2], **XINWEI MA** [1,2], **AND YUCHUAN JIN** [3]
[1]School of Civil and Transportation Engineering, Hebei University of Technology, Tianjin 300401, China
[2]School of Transportation, Southeast University (SEU), Nanjing 210000, China
[3]School of Architecture and the Built Environment, KTH Royal Institute of Technology, 11428 Stockholm, Sweden

Corresponding author: Xinwei Ma (230169206@seu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 51678212.

**ABSTRACT** As an eco-friendly travel mode, bike-sharing has prevailed around the world. However, the systems are imbalanced due to the asymmetric spatial and temporal distribution of user demand. Station prioritization strategies are needed to rebalance more shared bikes for more important stations. This paper proposes an evaluation method of station importance in dynamic bike-sharing rebalancing. Firstly, a short-term demand prediction model is applied to capture the temporal and spatial characteristics of bike-sharing trip data and predict bike-sharing demand at the station level. Based on the prediction results, the method of determining rebalancing quantity is proposed with consideration of bike-sharing usage throughout the rebalancing period. Then, three criteria are employed to evaluate the importance of bike-sharing stations, including rebalancing quantity, closeness to inventory threshold, and distance from the key station. An entropy-based Technique for Order of Preference by Similarity to the Ideal Solution (TOPSIS) approach is proposed to weigh different criteria and evaluate station importance. Furthermore, the experiments on bike-sharing data from Nanjing City demonstrate the effectiveness of the proposed methods. This research is helpful for operators and managers to dynamically rebalance shared bikes with high efficiency and improve the service quality of bike-sharing systems.

**INDEX TERMS** Bike-sharing, short-term demand prediction, rebalancing demand, station importance, TOPSIS.

## I. INTRODUCTION

Bike-sharing systems bring new opportunities for urban development. They are able to help reduce motorized traffic, curtail pollutant emissions, and promote socially equitable transportation systems [1], [2]. Bike-sharing services also significantly improve the flexibility and accessibility of transportation resources for travelers and offer an attractive solution to the ''one/last mile problem'' in multimodal transportation. More than 2000 bike-sharing systems are operated globally, and more than 300 systems are in planning or under construction [3]. Bike-sharing systems can be categorized into docked bike-sharing and dockless bike-sharing systems [2], [4]. The docked bike-sharing systems have dedicated docking stations where bike rental and return occur. In contrast, the dockless bike-sharing systems allow users to rent and park shared bikes in the physical or geo-fencing designated parking areas via mobile phone applications [2]. Both types of bike-sharing systems have improved urban mobility. However, these systems are imbalanced due to the asymmetric spatial and temporal distribution of user demand [5]. During specific periods of a day, some stations might run out of bikes (deficit stations) while some have no space available to store the returned bikes (surplus stations) [6]. In light of this situation, effective measures are needed to rebalance bikes and optimize bike distribution.

The approaches to bike rebalancing can be divided into user-based rebalancing and operator-based rebalancing [7]. The user-based rebalancing approaches incentivize users to rent or return bikes in specific stations by offering monetary incentives [7]. However, the effectiveness of station rebalancing is affected by the willingness of user participation, and

The associate editor coordinating the review of this manuscript and approving it for publication was Baozhen Yao.

sometimes incentives do not bring desired economic benefits. Therefore, the operator-based rebalancing approaches are more widely adopted to rebalance shared bikes [8]. Performed by systems' service staff, the operator-based bike-sharing rebalancing is modeled as a many-to-many pickup-and-delivery vehicle routing problem (M-M PVRP), as introduced by Berbeglia *et al.* [9]. In the process of operator-based bike-sharing rebalancing, vehicles start from the depot with initial loads of bikes to pick up or drop off a certain number of bikes at surplus or deficit stations and finally return to their original depots [6].

In general, bike rebalancing optimization can be classified into static optimization and dynamic optimization [10]. For static optimization, the rebalancing operation is performed during the night, assuming that the demand for bikes is negligible. This optimization method is used to arrange bikes in the system for the next working day. The dynamic optimization method is used during the day, and the routes need to be updated regularly to handle varying demand. Effective rebalancing management is quite difficult due to the limitation of the operators' ability to redistribute bikes. Especially, in the dynamic rebalancing process, the demand varies in the daytime, and intervention resources such as operation time are limited. Therefore, station prioritization strategies are needed to service certain stations that are more important in the rebalancing process. Important stations are defined as stations that play a more important role than other stations in the network and affect the network structure and function [11]. Evaluating station importance can help the operators decide which station should be prioritized and implement effective rebalancing [12]. In order to improve the decision-making process in bike-sharing rebalancing, this research proposed a station importance evaluation method based on a multi-criteria decision-making (MCDM) approach. The contributions of this research are summarized as follows:

- A short-term demand prediction model is applied to capture the temporal and spatial characteristics of bike-sharing trip data and predict bike-sharing demand at the station level. Exogenous factors such as land use information, weather, and users' personal information, are included in the prediction model.
- Based on the bike-sharing demand prediction results, the method of determining rebalancing quantity is proposed. This method also considers bike-sharing usage throughout the rebalancing period.
- Three criteria are employed to evaluate the importance of bike-sharing stations, including rebalancing quantity, closeness to inventory threshold, and distance from the key station. The key station is defined as the station that is being served or to be served by rebalancing vehicles. The entropy-based technique for order of preference by similarity to the ideal solution (TOPSIS) approach is proposed to evaluate station importance.

The key contents of this paper are structured as follows. Section 2 reviews the literature on bike-sharing demand prediction and bike-sharing station prioritization. The methods to evaluate station importance are presented in detail in Section 3. Experiments and discussion results are conducted in Section 4. Finally, conclusions and future research directions are summarized in Section 5.

## II. LITERATURE REVIEW

### A. SHORT-TERM BIKE-SHARING DEMAND PREDICTION

Short-term demand prediction is important for managing transportation infrastructure and enhancing the reliability and accessibility of transportation systems. Accurate station-level bike-sharing demand prediction is necessary for the dynamic rebalancing process [2], [13], [14]. Generally, data-driven bike-sharing prediction can be categorized into parametric approaches and nonparametric approaches [15]. When applying statistical parametric techniques, bike-sharing demand prediction is usually defined as a time series prediction problem. The auto-regressive moving average model (ARMA) [16] and the auto-regressive integrated moving average (ARIMA) model [17] are well-known statistical parametric methods, as well as its diverse variants. For instance, using Dublin's bike-sharing data, Yoon *et al.* [18] put forward an improved ARIMA model, considering signals from neighboring stations and seasonal trends to estimate available bicycles at each station. Other statistic models have been widely used, such as the Bayesian network [19] and the Markov Chain model [20]–[22]. Although statistical models are easier to interpret, they may have limitations in prediction accuracy, data accessibility, and computing power [23], compared to nonparametric machine learning models. Unlike the statistical methods, machine learning automatically learns the relationship between the inputs and outputs [24] without making strong assumptions about the data structure. Various machine learning models have been applied for bike-sharing prediction, such as artificial neural networks (ANN) [25], support vector regression (SVR) [26] and regression trees (RT) [27]. Using a bike-sharing dataset from Washington D.C., Yin *et al.* [28] applied SVR, random forests (RF), and gradient boosted tree (GBT) to predict usage of bike-sharing system in an hour. They suggested that the RF method performs the best in terms of both prediction accuracy and training time.

Recently, deep learning has attracted significant research interest due to its ability to extract latent features and model nonlinear relationships in the raw data. Deep learning has achieved great success in many areas, such as language processing and image recognition [29], [30]. In the bike-sharing research area, some researchers have applied deep learning to forecast short-term travel demand. Lu and Lin [31] input the rental records of the past time into Recurrent Neural Network (RNN) to predict the bicycle rental in the coming day. Pan *et al.* [32] used the deep long short-term memory (LSTM) sequence learning model to predict the rentals and returns at a single station based on historical trip data, weather data, and time data. Xu *et al.* [33] developed an LSTM model to predict the dockless bike-sharing demand at the traffic analysis zone (TAZ) level in different prediction horizons. Additionally, Ke *et al.* [34] employed convolutional neural networks (CNN) to predict bike-sharing inflows and outflows in each region based on historical trajectory data, weather and events of each area in a city. After splitting the whole city into grids with a predefined grid size and calculating the bike-sharing demand for each grid, Zhang *et al.* [35] applied CNN for large-scale spatio-temporal bike-sharing prediction. To capture spatial and temporal dependencies, researchers began to combine both networks and proposed a convolutional LSTM network (Conv-LSTM). Specifically, Du *et al.* [36] integrated irregular CNN and LSTM units to capture the features of spatial-temporal traffic flow and predict passenger traffic flows in different hours of a day based on the historical passenger flows and external factors such as weather, traffic control, sports events, and vocal concert. Ai *et al.* [37] proposed a Conv-LSTM model to predict the short-term distribution of the dockless bike-sharing systems by considering spatial-temporal variables and time-series variables. Ljubenkov [38] applied a convolutional neural network (CNN) to identify the spatial structures of bike flows and RNN-LSTM to find and predict its dynamic patterns.

However, when used to predict bike-sharing at the station level, CNN can only reflect inter-station relationship by geographical distance [39]. Some researchers attempted to apply deep learning architecture to graph data structure [39]–[42]. Taking bike stations as nodes, the bike-sharing network can be represented in a graph. Some researchers describe bike riding relationships in the complex heterogeneous spatial-temporal graph and used graph convolutional neural network (GCN) to capture non-Euclidean structures. Kim *et al.* [39] constructed a GCN prediction model to predict hourly bike-sharing demand at the station level by incorporating spatial characteristics, temporal patterns, and global variables (weather and weekday/weekend). Yoshida *et al.* [40] proposed a relational graph convolutional network-based method to predict the demand at the station level. Guo *et al.* [41] built a spatial-temporal graph neural network (ST-GNN) to model and predicted citywide bike-sharing demand. They used GCN to capture the spatial correlation and gated recurrent units (GRU) to capture the temporal dependency. Xiao *et al.* [42] developed the ST-GCN model to predict the picking up and returning demand of 186 stations for the Wenling public bike-sharing program. In their study, the ST-GCN model outperformed the RNN-based models in prediction accuracy and computation efficiency.

Previous studies have provided valuable insights for bike-sharing demand prediction. The combination of GCN and LSTM enables to capture the temporal characteristics and the bike-sharing network structure characteristics simultaneously. However, related researches have been insufficient. Few pieces of research make full use of multi-source heterogeneous data such as weather conditions, point of interest (POI) distribution and users' personal attributes [39].

### B. STATION PRIORITIZATION IN BIKE-SHARING REBALANCING

Many studies have explored bike-sharing rebalancing problems and proposed solutions. However, very few studies dealt with station prioritization during the rebalancing process. As shown in Table 1, rebalancing quantity, closeness to inventory threshold, and distance from the key station are the features commonly used to evaluate stations. Some studies made priority choices according to a single feature of stations. Xu *et al.* [43] proposed a dynamic scheduling model based on the multi-similarity inference model and adopted an enhanced genetic algorithm. In this algorithm, the hybrid crossover strategy is improved to give priority to stations that have larger user demand. To produce rebalancing plans for large-scale real-world bike-sharing systems, Mellou and Jaillet [44] gave priority to stations that need rebalancing to become leader stations. After all the leader stations were selected, mini-groups of stations were created by assigning each station to its closest leader, each of which is represented by its leader. Federico *et al.* [45] defined survival time as the shortest period of time after which the inventory state of a station exceeds a predefined threshold. They calculated survival time using historical usage data and assigned priority to the stations with the shortest survival time in the dynamic short-term relocation strategy. As indicated by Kadri *et al.* [46], only considering one feature of stations might mislead vehicle sequences' computation in some instances. Kadri *et al.* [46] compared two rebalancing prioritization strategies for the greedy search algorithm. The first strategy considered the state of unbalance and required travel time while the second considered just the required travel time. The results showed that the algorithm using the strategy that considers both aspects outperformed the one using the second strategy. Based on the previous research, Brinkmann *et al.* [47] used a safety buffer to decide whether a station is imbalanced. After serving an imbalanced station, its nearest imbalanced station would be served first. Benjamin [12] used a one-step policy improvement method to decide whether a specific station should be reset. The numerical experiments showed that their policy outperformed those built on simple prioritization rules. It indicated that a proper policy of station prioritization should be conducted based on multi-criteria.

As shown in Table 1, few researches provided insight into station importance evaluation. When more than one features are adopted, it is common to see that station importance is decided simply or considered separately. Moreover, the weights for each feature have not been explored. Shen *et al.* [16] illustrated that farther rebalancing distance brings higher rebalancing costs. As the basis for route optimization, the weighted path between two stations was calculated by dividing the station's imbalance by rebalancing distance. Jan *et al.* [48] proposed lookahead policies and defined three different target fill levels for stations according to inventory. A station is assumed to be imbalanced if its

**TABLE 1.** Summary of the characteristics of station prioritization for bike-sharing rebalancing in the existing studies.

| Study | Problem type | Bike type | Feature: Rebalancing quantity | Feature: Closeness to inventory threshold | Feature: Distance from the key station | Method | Research aim |
|---|---|---|---|---|---|---|---|
| Xu et al. (2018) [43] | Dynamic | Docked | √ | | | Qualitative | To enhance the genetic algorithm by adopting demand priority service crossover to optimize the scheduling route. |
| Mellou and Jaillet (2019) [44] | Dynamic | Docked | √ | | | Qualitative | To reduce the size of the large-scale network by developing a decomposition method based on proper station grouping. |
| Federico et al. (2018) [45] | Dynamic | Docked | | √ | | Qualitative | To assign a priority level to each station based on its urgency of being rebalanced in the dynamic short-term relocation strategy. |
| Kadri et al. (2018) [46] | Static | Docked | √ | | √ | Qualitative | To obtain vehicle sequences serving more imbalanced stations to minimize the total weighted time. |
| Brinkmann et al. (2020) [47] | Dynamic | Docked | √ | | √ | Qualitative | To introduce the short-term relocation policy that serving more imbalanced stations with lower travel costs. |
| Shen et al. (2020) [16] | Static | Dockless | √ | | √ | Qualitative | To determine the optimal rebalancing path by setting the weight to stations related to distance and imbalance. |
| Brinkmann et al. (2019) [48] | Dynamic | Docked | | √ | √ | Qualitative | To analyze the impacts of lookahead policies by defining two benchmark policies. |
| Zu et al. (2016) [49] | Static | Docked | √ | | √ | Qualitative | To generate the candidate initial solutions with the minimal total operation and penalty costs. |
| This study (2021) | Dynamic | Docked | √ | √ | √ | An entropy-based TOPSIS approach. | To improve the decision-making process in dynamic bike-sharing rebalancing by proposing a multi-criteria-based station importance evaluation method. |

Note: Qualitative means using basic statistical and comparative methods to determine and rank station importance.

fill level violates safety buffers. Considering the distance, the vehicle will serve the nearest imbalanced station in every decision state. Zu [49] proposed two priority selection methods to generate candidate initial solutions. The first one prioritized the station with the most considerable penalty cost reduction and the second one chose the station with the lowest traveling cost.

However, none of the aforementioned research combined all three features and made a comprehensive evaluation of station importance. Moreover, multi-criteria methods have not been used to evaluate bike-sharing station importance. To fill these research gaps, this paper employs rebalancing quantity, closeness to inventory threshold, and the distance from the key station as three criteria to evaluate bike-sharing stations. The entropy-based TOPSIS approach is applied. TOPSIS is among the most popular MCDM methods, where alternatives are evaluated based on Euclidean distances from an ideal and a nonideal solution [50]. This approach is helpful for operators to weigh various criteria and rank different alternatives [51].

## III. METHODOLOGY

### A. STUDY AREA AND DATA SOURCES

As the capital of Jiangsu province, Nanjing has long been ranked as the second-largest commercial center in the East China region. In order to ease traffic pressure and bring citizens great convenience, Nanjing launched docked bike-sharing programs in 2013. As of 2017, Nanjing had approximately 60,000 docked shared bikes and 450,000 dockless shared bikes [52]. In Fig. 1, the area circled by the Inner-ring Road in Nanjing is the main urban area, taking Xinjiekou Metro Station as the center. Previous research on docked bike-sharing in Nanjing has indicated a significant imbalance of temporal and spatial demand for bike-sharing trips in the main urban area [53]. Therefore, considerable efforts are needed to redistribute bike-sharing to keep a high level of service quality in this area. In this research, 49 stations within 13 TAZs around the city center (Xinjiekou Metro Station) are selected. Fig. 1 shows the bike-sharing station distribution in the study area.

The information of bike-sharing trip records, users' personal attributes, POI distribution and weather conditions are used for bike-sharing demand prediction. The bike-sharing dataset is provided by Nanjing Public Bicycle Company, involving bike-sharing trip records, bike-sharing station information and users' personal attributes. There are 350,701 bike-sharing trips from 1st Sep. 2017 to 30th Sep. 2017. As shown in Table 2, each bike-sharing trip records user ID, bike ID, rental time, rental station, return time and return station. When pre-processing data, trips with the following properties have been removed: (1) travel distance shorter than 100 m or longer than 5 km [54]; (2) travel duration less than 30 seconds or longer than 2 hours [55]; (3) incomplete information. Table 3 shows the information of docked bike-sharing stations, including station ID, longitude and latitude. Bike-sharing users' personal attributes, include age, gender, residency status. In addition, land use data contains POI and road density of each bike-sharing station with a radius

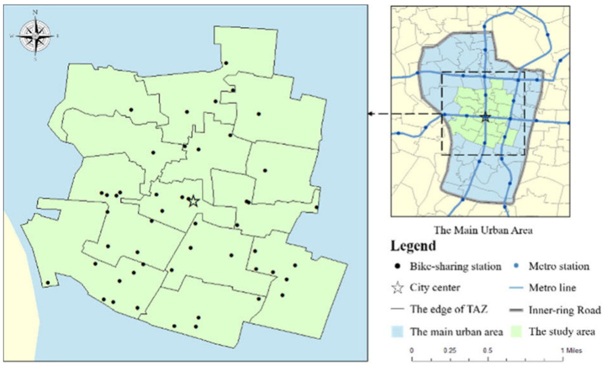



**FIGURE 1.** The selected bike-sharing station distribution map.

**TABLE 2.** A sequence of docked bike-sharing transaction records.

| User ID | Bike ID | Rental Time | Rental Station | Return Time | Return Station |
|---|---|---|---|---|---|
| NJ1...813 | 02203 | 2017/9/5 23:47:53 | 24 | 2017/9/5 23:58:21 | 25 |
| NJ1...343 | 35274 | 2017/9/5 23:47:46 | 05 | 2017/9/5 23:54:18 | 08 |
| NJH...204 | 35292 | 2017/9/5 23:38:14 | 21 | 2017/9/5 23:45:57 | 44 |
| NJ1...027 | 44092 | 2017/9/5 23:38:11 | 44 | 2017/9/5 23:46:26 | 41 |
| … | … | … | … | … | … |
| NJ1...577 | 22849 | 2017/9/5 23:37:42 | 12 | 2017/9/5 23:43:06 | 15 |

**TABLE 3.** The information of docked bike-sharing stations.

| Station ID | Longitude | Latitude |
|---|---|---|
| 01 | 118.776 | 32.043 |
| 02 | 118.777 | 32.043 |
| 03 | 118.781 | 32.043 |
| 04 | 118.783 | 32.042 |
| … | … | … |
| 49 | 118.775 | 32.036 |

of 300 m [56]. POIs are classified into working POIs, residential POIs, transport POIs, and other POIs [52] collected by using the Amap API. The road shapefile is downloaded from OpenStreetMap and the road density is calculated by using ArcGIS. Information of historical weather is obtained from Weather Underground [57].

To evaluate station importance, three criteria, including distance from the key station, rebalancing quantity and closeness to inventory threshold, need to be determined. Specifically, the longitude and latitude of the key station and bike-sharing station are used to calculate the distance from the key station. Bike-sharing demand prediction results, station capacity, and station inventory will be used to calculate rebalancing quantity. Station capacity and station inventory will be used to calculate closeness to inventory threshold. The information of station capacity and station inventory is extracted from historical trip records.

### B. THE FRAMEWORK OF STATION IMPORTANCE EVALUATION

The process of the dynamic rebalancing operation resolves variations of bike-sharing usage demand from time to time. Previous research has applied a rolling horizon to deal with changing demand [10]. As shown in Fig. 2, through a rolling horizon, the rebalancing operation is divided into several periods, transforming the dynamic rebalancing problem into several stages of static rebalancing problems [10]. Firstly, the initial conditions of the bike-sharing system are established, and the durations of the prediction and control periods are defined. Then, the first prediction period is solved and rebalancing plans are operated during the first control period. At this time, user demand and the state of stations are updated. The end of the control period is the start of next prediction period. Using information from the previous period, the rebalancing problem is solved again. Notably, the station that is being served or to be served by rebalancing vehicle at the end of the last operation period, is the key station of the next operation period. At the first stage of rebalancing, the depot is the key station, which is the starting point of rebalancing vehicles. It is assumed that the depot is the city center (Xinjiekou Metro Station). When the new rebalancing comes to the final period, the whole rebalancing procedure is completed.

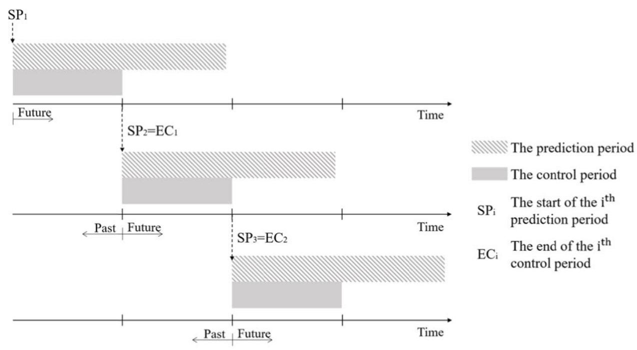



**FIGURE 2.** The dynamic rebalancing process based on the rolling horizon.

In the rolling horizon, the bike-sharing demand is predicted, then the station importance is evaluated at every rebalancing stage. The updates of station inventory and key station provide accurate information for the rebalancing operation, as shown in Fig. 3.

#### 1) BIKE-SHARING DEMAND PREDICTION

Long short-term memory (LSTM) has been widely used in bike-sharing demand prediction to capture the temporal characteristics of time series data [58], [33]. However, the bike-sharing network structure characteristics and the impact of the states of station neighbors were not considered in these researches. In the previous research, the graph convolution network (GCN) was proposed to handle data with graph structures, such as social network data and meteorological station network measurements [59]. In this paper, the GCN

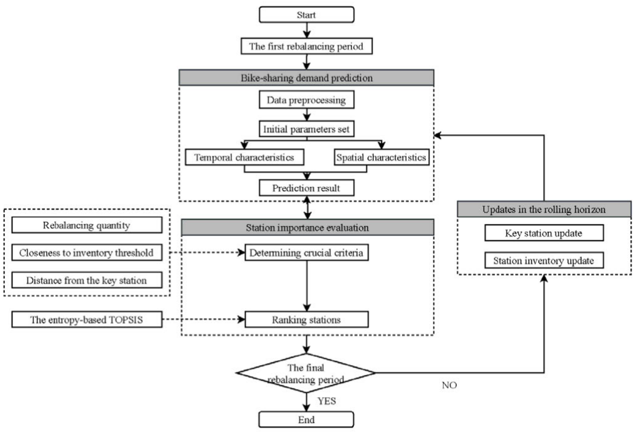



**FIGURE 3.** **Bike-sharing demand prediction and station importance evaluation based on the rolling horizon.**

model is embedded in the LSTM model to predict bike-sharing demand. In this GC-LSTM model, LSTM is applied to capture the temporal characteristics of bike-sharing trip data at the station level. GCN is used to extract the structural characteristics of the bike-sharing network by performing convolution operations on the hidden layer state and considering the influence of the neighbors' hidden state on the hidden state of the node. Fig. 4 shows the structure of GC-LSTM. The following equations explain the meaning of notations in Fig. 4.

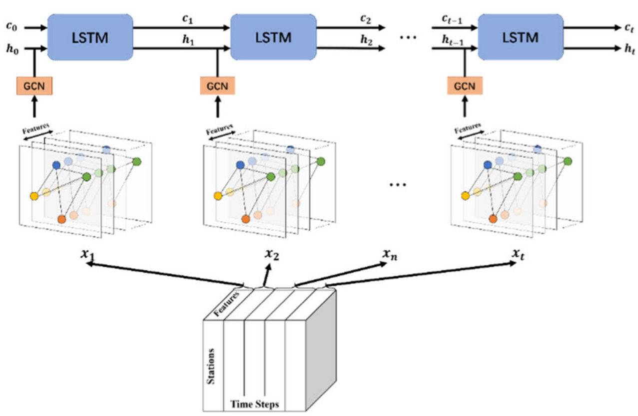



**FIGURE 4.** **The network structure of GC-LSTM model.**

This model has three gates: the forget gate $f_t$, the update gate $u_t$, and the output gate $o_t$. The forget gate $f_t$ is to decide what information will be dropped from the previous cell state. The value 0 indicates that the information will be dropped, and 1 means the information will be reserved. In equation (3), $\tilde{c}_t$ is generated by a tanh layer and contains the new candidate vector. As depicted in equation (4), the cell state can be updated using the update and the forget gates to calculate a weighted average of the candidate $\tilde{c}_t$ and $c_{t-1}$ from the last timestep. Then, the hidden layer vector $h_t$ is decided by the output gate $o_t$ and the updated cell $c_t$ in equation (6) [60]. The last hidden cell state will be passed to an activation function to get the final prediction normally.

$$f_t = \sigma(W_f A_t + \sum_{k=0}^{K} \theta_{hkf} T_k \tilde{L}_{t-1} h_{t-1} + b_f) \tag{1}$$

$$u_t = \sigma(W_u A_t + \sum_{k=0}^{K} \theta_{hku} T_k \tilde{L}_{t-1} h_{t-1} + b_u) \tag{2}$$

$$\tilde{c}_t = \tanh(W_c A_t + \sum_{k=0}^{K} \theta_{hkc} T_k \tilde{L}_{t-1} h_{t-1} + b_c) \tag{3}$$

$$c_t = f_t \odot c_{t-1} + u_t \odot \tilde{c}_t \tag{4}$$

$$o_t = \sigma(W_o A_t + \sum_{k=0}^{K} \theta_{hko} T_k \tilde{L}_{t-1} h_{t-1} + b_o) \tag{5}$$

$$h_t = o_t \odot \tanh(c_t) \tag{6}$$

where $A_t$ represents the adjacency matrix as the input data at the time $t$. $h_{t-1}$ represents the hidden state at the time $t-1$. $W_f$, $W_u$, $W_c$, $W_o$ and $b_f$, $b_u, b_c, b_o$ are the weight and bias matrix of three gates in GC-LSTM. $\sum_{k=0}^{K} \theta_{hkf}$, $\sum_{k=0}^{K} \theta_{hku}$, $\sum_{k=0}^{K} \theta_{hkc}$, $\sum_{k=0}^{K} \theta_{hko}$ denote the parameter vector of the graph convolution.

In this paper, we build a three-layer neural network with GC-LSTM layers to predict the rental and return demand of bike-sharing at the station level. This proposed model can automatically learn the temporal and spatial information among stations based on the combined structures of LSTM and GCN. To evaluate the performance of the proposed model, mean absolute error (MAE) and root mean square error (RMSE) are used as Lin *et al.* [61] did:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |x_i - \hat{x}_i| \tag{7}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2} \tag{8}$$

where $x_i$ is the observed value, $\hat{x}_i$ is the predicted value, and $n$ is the number of testing samples.

### 2) STATION IMPORTANCE EVALUATION

#### *a: CRITERIA DETERMINATION*

In this paper, station importance is evaluated based on multi-criteria, which are rebalancing quantity, closeness to inventory threshold, and distance from the key station. This section shows the definitions and calculation methods of these criteria.

*Rebalancing quantity:* In previous research, rebalancing quantity is entirely decided by the prediction result without consideration of inventory [43], or it should serve stations to a fixed fill level (50% in O'Mahony's research [62]) at the moment of rebalancing. However, in dynamic rebalancing, some stations that have been served may still have bike surpluses or bike deficits during the peak hour. This is because bike-sharing demand changes quickly and varies from time to time, especially during peak periods. To tackle with this problem and improve rebalancing efficiency, this research considers the demand variation between rebalancing stages. It aims to reduce unmet demand throughout the rebalancing process. Based on demand prediction, station inventory and station capacity, the calculation method of bike-sharing rebalancing quantity is introduced.

The rebalancing quantities are calculated in each stage of the rolling-horizon rebalancing operation. Let $t_0$ represent the starting time of rebalancing operation, $\tau$ represent the duration of rebalancing operation. The horizon divides the operation into $n$ stages, with each stage having a duration of $l$. Different stages are described as and the time intervals are $[t_0, t_0+l], [t_0+l, t_0+2l], \ldots, [t_0+(n-1)l, t_0+nl]$. Let $init_{k,i}$ represent the initial number of bikes available at the station $i$ at the stage $e_k$. The number of bikes available at the station $i$ at the stage $e_k$ can be described as:

$$avail_{i,k}(t) = \max\{0, init_{i,k} - v_{i,k} \cdot (t - t_0 - kl)\}$$
$$(t_0 + kl < t < t_0 + \tau, k = 0, 1, \ldots, n-1) \quad (9)$$

where $v_{i,k}$ represents the difference between the rental rate and return rate at the stage $e_k$. The equation is as follows:

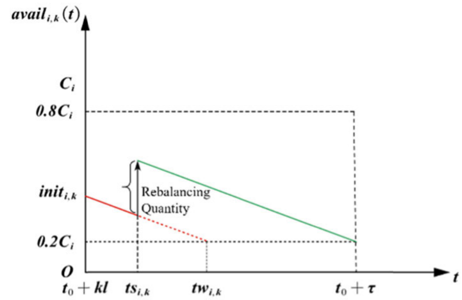
$$v_{i,k} = \frac{N_{i,k}^{pred,rent} - N_{i,k}^{pred,return}}{\tau} \quad (10)$$

where $N_{i,k}^{pred,rent}$ and $N_{i,k}^{pred,return}$ respectively represent predicted rental quantities and return quantities at the station $i$ during the period $[t_0 + kl, t_0 + \tau]$.

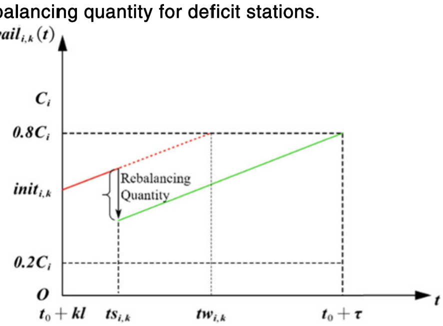
Considering the uncertainties in users' bike renting and returning, the upper and lower limits of the rebalancing threshold should be set for each station. Let $C_i$ represents the maximum station capacity. Generally, the upper and lower limits of the rebalancing threshold of station $i$ are set to $0.2C_i$ and $0.8C_i$ respectively, according to previous research and the operational experience [63], [64]. Let $tw_{i,k}$ represent the station warning time in the current rolling period. This parameter means that at the moment $tw_{i,k}$, the number of bikes available at the station $i$ reaches the upper/lower limit of the rebalancing threshold. $tw_{i,k}$ can be calculated as:

$$tw_{i,k} = \begin{cases} t_0 + kl + \frac{init_{i,k} - 0.2C_i}{v_{i,k}}, & v_{i,k} > 0, 0.2C_i < init_i < 0.8C_i \\ t_0 + kl + \frac{init_{i,k} - 0.8C_i}{v_{i,k}}, & v_{i,k} < 0, 0.2C_i < init_i < 0.8C_i \\ t_0, & otherwise \end{cases} \quad (11)$$

For stations that need to be reset in the rolling horizon, the rebalancing quantity of bikes at stations can be depicted in Fig. 5. The red line represents the number of bikes available at the station in each period without rebalancing operation. The green line represents the number of bikes available at the station in each period with rebalancing operation. Specifically, this research aims to keep station inventory not exceeding the upper or lower limit of the rebalancing threshold throughout rebalancing stages.




(a) Rebalancing quantity for deficit stations.




(b) Rebalancing quantity for surplus stations.

**FIGURE 5. Rebalancing quantity of bikes at stations. (a) Rebalancing quantity for deficit stations. (b) Rebalancing quantity for surplus stations.**

To ensure that the inventory of a deficit station would not fall below the lower limit of the rebalancing threshold ($0.2C_i$) before the end of the rolling horizon, and ensure that the inventory of a surplus station would not exceed the upper limit of the rebalancing threshold ($0.8C_i$). The rebalancing quantity at deficit station ($\gamma = 0.2$) and surplus station ($\gamma = 0.8$) can be calculated by:

$$Q_{i,k} = v_{i,k}(\tau - kl) - init_{i,k} + \gamma C_i \quad (12)$$

*Closeness to inventory threshold:* A higher level of the closeness to inventory threshold means that it is easier to reach the inventory threshold for these stations and the lack of bikes or docks is more likely to occur. Therefore, more importance would be attached to stations with higher closeness to inventory threshold. Closeness to inventory threshold can be calculated in equation (13).

$$c_{i,k} = \left| \frac{|avail_{i,k,t} - 0.2C_i| - |avail_{i,k,t} - 0.8C_i|}{C_i} \right|$$
$$(k = 0, 1, \ldots, n-1) \quad (13)$$

where $avail_{i,k,t}$ represents the number of bikes available at station $i$ and at the stage of $e_k$.

*Distance from the key station:* The distances from the key station represents the distances between the unserved stations and the key station. It is notable that the key station is updated in the process of rebalancing operation. In this research, it is assumed that the depot is the city center at the first rebalancing stage. Haversine formulation is used to calculate distances from the unserved stations to the key station [65].

$$D_i = 2R \cdot \arcsin \sqrt{\sin^2(\frac{lat_k, lat_i}{2}) + \cos(lat_k) \cdot \cos(lat_i) \cdot \sin^2(\frac{lon_k - lon_i}{2})} \quad (14)$$

where $lat_k, lon_k$ represent the latitude and longitude of the key station and $lat_i, lon_i$ represent the latitude and longitude of the unserved station $i$; $R$ represents the earth radius.

#### b: THE ENTROPY-BASED TOPSIS APPROACH

The TOPSIS method developed by Hwang and Yoon [66] is an MCDM technique based upon the concept that the chosen alternative should have the shortest distance to an ideal solution and the farthest distance from a nonideal solution simultaneously. In this research, an entropy-based TOPSIS approach is applied to evaluate station importance based on three criteria. The process to evaluate station importance can be described in the following steps.

*Step 1:* is to construct the normalized decision matrix:

$$n_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}}, \quad i = 1, 2, \ldots, m; \; j = 1, 2, \ldots, n \quad (15)$$

where $x_{ij}$ is outcome of $i^{th}$ station with respect to the $j^{th}$ criterion; $m$ is the number of stations to be rebalanced, $n$ is the number of criteria ($n = 3$).

*Step 2:* is to calculate the weighted normalized decision matrix. An entropy method is used to determine the weight $w_j$.

$$E_j = -\frac{1}{\ln(m)} \sum_{i=1}^{m} P_{ij} \ln(P_{ij}), \quad j = 1, 2, \ldots, n \quad (16)$$

$$w_j = \frac{|1 - E_j|}{\sum_{j=1}^{n} |1 - E_j|}, \quad j = 1, 2, \ldots, n \quad (17)$$

$$v_{i,j} = w_j n_{i,j}, \quad i = 1, 2, \ldots, m; \; j = 1, 2, \ldots, n \quad (18)$$

*Step 3:* is to determine the ideal solution $A^+$ and nonideal solution $A^-$:

$$\begin{aligned} A^+ &= \{v_1^+, v_2^+, \ldots, v_n^+\} \\ &= \{(\max v_{ij} | j \in J), (\min v_{ij} | j \in J')\} \end{aligned} \quad (19)$$

$$\begin{aligned} A^- &= \{v_1^-, v_2^-, \ldots, v_n^-\} \\ &= \{(\min v_{ij} | j \in J), (\max v_{ij} | j \in J')\} \end{aligned} \quad (20)$$

where $J$ is related to benefit criteria, $J = \{j = 1, 2, \ldots, n | j\}$; $J'$ is related to cost criteria, $J' = \{j = 1, 2, \ldots, n | j\}$.

*Step 4:* is to determine the Euclidean distance for the ideal and nonideal solutions of each alternative:

$$d_i^+ = \sqrt{\sum_{j=1}^{n} (v_{ij} - v_j^+)^2}, \quad i = 1, 2, \ldots, m \quad (21)$$

$$d_i^- = \sqrt{\sum_{j=1}^{n} (v_{ij} - v_j^-)^2}, \quad i = 1, 2, \ldots, m \quad (22)$$

*Step 5:* is to calculate the TOPSIS score:

$$Score_i = \frac{d_i^-}{d_i^+ + d_i^-}, \quad i = 1, 2, \ldots, m; 0 < Score_i < 1 \quad (23)$$

where $Score_i$ is the relative distance of the station $i$ from a nonideal solution, which implies the importance of the station $i$. Stations with higher TOPSIS scores should be given service priorities [11].

## IV. RESULTS

In this paper, the rebalancing operation duration is 60 minutes, starting at 8 a.m. and ending at 9 a.m. Due to the variance of real-time demand, the rebalancing operation should be appropriately adjusted to provide better service for users. Therefore, station importance needs to be frequent updated [67], [68]. As suggested by Marte and Kristine [68], Yu [69] and Feng [70], the horizon is divided into four stages in this paper, with each stage having a duration of 15 minutes. The duration of the prediction horizon is set to be 30 minutes. The next prediction and operation period are activated by the end of the last operation period, as shown in Fig. 2. Taking the first stage of the rebalancing operation as an example, experiments in this paper illustrate how to predict bike-sharing demand based on GC-LSTM models and evaluate station importance based on the entropy-based TOPSIS method. With updates of the key station and the station inventory, the process of applying prediction and evaluation at following stage will be the same as the first stage.

### A. BIKE-SHARING DEMAND PREDICTION

Experiments are implemented using bike-sharing trips on weekdays, taking data generated in the first twelve days as the training dataset, data in the following four days as validation dataset, and data in the last five days as testing dataset. In each model, two output layers are produced to predict rental demand and return demand, respectively. To deal with dimensional problems, the min-and-max normalization is applied. The hyper-parameters, learning rate, batch size, and the number of hidden units are searched using a grid search. Specifically, the learning rate starts from 0.001 and ends at 0.01 with a step of 0.001; Both the batch size and the number of hidden units start from 32 and end at 128 with a step of 32. To prevent overfitting, the criteria are set for the early stop process. If the MAE on validation does not decrease by over 0.00001 for more than 10 training epochs, the model will stop training. The prediction models are coded by Python. All experiments are implemented on a Windows 10 computer with CPU (Intel(R) Core (TM) i7-8700 CPU @ 3.20GHz), 16 GB random-access memory (RAM) and one NVIDIA RTX 2060 GPU with 6 GB memory.

To testify the performance of GC-LSTM model for bike-sharing demand prediction in the dynamic rebalancing, the proposed GC-LSTM model is compared with the baseline models on the real bike-sharing datasets under the same parameter settings and coding environment, which are introduced as follows.

- Historical average (HA): It uses the historical average demand in the same prediction period for prediction.
- Support vector regression (SVR): It transforms the feature vectors into a linear high-dimensional space using kernel-based radial basis function. This method is very effective to replace the complex calculation in high-dimensional space.
- Extreme gradient boosting (XGBoost): It is an implementation of gradient boosting decision trees. All features are placed into a one-dimensional vector and used for prediction. It optimizes the loss function through the second derivative and applies a regularized model to prevent overfitting.
- Artificial neural network (ANN): It is originated from the biological neural network (BNN). This method is able to learn from a set of input-output parameter space and to approximate functions with high-dimensional data requiring no detailed information about the system.
- Graph convolutional network (GCN): It is a multi-layer neural networks operating on graph-structured data. The convolutional layer enables the network to construct node embeddings by fusing both features of nodes and relationships between nodes.
- Long-short term memory (LSTM): It is a long short-term memory neural network. As an improved version of RNN, LSTM can deal with vanishing gradients and handle complex temporal dependencies.

Table 4 shows the RMSE and MAE of each model in different prediction intervals based on 10 times of training results. For rental demand and return demand, the RMSE and MAE become smaller when the prediction interval becomes shorter. The GC-LSTM model with the 15-min interval has the best performance among the four prediction intervals. Moreover, the results of model comparison show that the bike-sharing demand prediction model based on GC-LSTM outperforms its counterparts. Compared to the LSTM structure, the RMSE and MAE drop 20.5%, 27.91%, respectively, with an interval of 15min for rental demand prediction, and 24.69%, 31.82% for return demand prediction. Notably, the performance of the GC-LSTM model improves quickly than the LSTM model. This may because, compared with the LSTM model, the GC-LSTM model captures more fluctuation characteristics in the bike-sharing graph structure. Similar results can be found in the research of Zhang *et al.* [71], the performance of GCN-based models improved more obvious than LSTM-based models, as the time interval decreased.

**TABLE 4. Comparison among prediction models.**

| Prediction model | Rental demand prediction | | Return demand prediction | |
|---|---|---|---|---|
| | RMSE | MAE | RMSE | MAE |
| *Prediction interval:15min* | | | | |
| HA | 3.18 | 1.81 | 3.12 | 1.80 |
| SVR | 1.43 | 0.78 | 1.41 | 0.77 |
| XGBoost | 1.44 | 0.78 | 1.39 | 0.76 |
| ANN | 1.40 | 0.77 | 1.37 | 0.76 |
| GCN | 1.38 | 0.74 | 1.34 | 0.73 |
| LSTM | 0.78 | 0.43 | 0.81 | 0.44 |
| GC-LSTM | 0.62 | 0.31 | 0.61 | 0.30 |
| *Prediction interval:30min* | | | | |
| HA | 6.07 | 3.50 | 5.95 | 3.49 |
| SVR | 2.21 | 1.26 | 2.13 | 1.22 |
| XGBoost | 2.22 | 1.27 | 2.21 | 1.23 |
| ANN | 2.12 | 1.22 | 2.06 | 1.19 |
| GCN | 2.11 | 1.20 | 2.02 | 1.16 |
| LSTM | 0.88 | 0.52 | 0.85 | 0.50 |
| GC-LSTM | 0.73 | 0.40 | 0.71 | 0.39 |
| *Prediction interval:45min* | | | | |
| HA | 8.88 | 5.18 | 8.72 | 5.18 |
| SVR | 2.86 | 1.66 | 2.82 | 1.63 |
| XGBoost | 2.86 | 1.66 | 2.83 | 1.64 |
| ANN | 2.72 | 1.59 | 2.67 | 1.55 |
| GCN | 2.71 | 1.57 | 2.67 | 1.55 |
| LSTM | 0.85 | 0.51 | 0.84 | 0.50 |
| GC-LSTM | 0.81 | 0.47 | 0.82 | 0.48 |
| *Prediction interval:60min* | | | | |
| HA | 11.68 | 6.89 | 11.40 | 6.87 |
| SVR | 3.53 | 2.06 | 3.45 | 2.02 |
| XGBoost | 3.53 | 2.07 | 3.45 | 2.01 |
| ANN | 3.37 | 1.99 | 3.28 | 1.93 |
| GCN | 3.37 | 1.97 | 3.27 | 1.91 |
| LSTM | 0.92 | 0.56 | 0.88 | 0.53 |
| GC-LSTM | 0.89 | 0.54 | 0.88 | 0.53 |

### B. STATION IMPORTANCE EVALUATION

The initial inventory and station capacity can be obtained based on the historical bike-sharing data. Bike-sharing demand in the next 30 minutes is predicted by the GC-LSTM model. Rebalancing quantity is decided by bike-sharing demand, initial inventory and station capacity. Closeness to inventory threshold is determined by initial inventory and station capacity. Besides, the distance from the key station is calculated using haversine formulation. Table 5 shows the results of three criteria for station importance evaluation.

After obtaining the values of three criteria at the station level, the entropy-based TOPSIS method is employed to evaluate bike-sharing station importance. The entropy is employed to weigh the rebalancing quantity, closeness to inventory threshold, and distance from the key station that influence station importance.

The decision matrix that combines rebalancing quantity, closeness to inventory threshold, and distance from the key station can be obtained from Table 5. The calculation results are shown in Table 6. The first step in the TOPSIS approach is to compute the normalized decision matrix using equation (15). Then, the weighted decision matrix is calculated by equation (16-18). It is noteworthy that the weighted decision matrix is obtained by multiplying a normalized decision matrix by the weight of the three criteria. The ideal and nonideal solutions are computed by equation (19) and equation (20). And then, the distances from ideal and nonideal solutions are calculated by equation (21) and equation (22). Finally, the ranks of all alternatives (bike-sharing stations) are computed according to the relative closeness to ideal solution by equation (23). Table 7 shows the information of the top 10 stations of importance ranking, including Station ID, distance from ideal and nonideal solutions, TOPSIS score, and importance ranking.

**TABLE 5. The determination of criteria for station importance evaluation.**

| Station ID | Initial inventory | Station capacity | Rebalancing quantity | Closeness to inventory threshold | Distance from the key station (m) |
|---|---|---|---|---|---|
| 01 | 11 | 42 | 21 | 0.4762 | 796 |
| 02 | 19 | 30 | 8 | 0.2667 | 635 |
| 03 | 22 | 37 | 6 | 0.1892 | 322 |
| 04 | 0 | 29 | 6 | 0.6000 | 103 |
| 05 | 19 | 41 | 6 | 0.0732 | 1069 |
| … | … | … | … | … | … |
| 49 | 8 | 35 | 10 | 0.5429 | 1058 |

**TABLE 6. The calculation process of the station importance evaluation.**

| | Rebalancing quantity | Closeness to inventory threshold | Distance from the key station |
|---|---|---|---|
| *Weight* | | | |
| $w_j$ | 0.2415 | 0.2363 | 0.5222 |
| *Station ID* | | | |
| 01 | 0.0056 | 0.0020 | 0.0068 |
| 02 | 0.0021 | 0.0035 | 0.0085 |
| 03 | 0.0016 | 0.0049 | 0.0167 |
| 04 | 0.0016 | 0.0016 | 0.0522 |
| 05 | 0.0016 | 0.0127 | 0.0050 |
| … | … | … | … |
| 49 | 0.0027 | 0.0017 | 0.0051 |
| *Ideal and nonideal solutions* | | | |
| $A^+$ | 0.0150 | 0.01862 | 0.1168 |
| $A^-$ | 0.0003 | 0.0016 | 0.0036 |

As shown in Table 6, the weights of the three criteria are computed using the entropy approach. Distance from the key station has a weight of 52.22%. This consideration plays a central role in station importance evaluation. Rebalancing quantity play an important role in the bike-sharing with 24.15% weight, while the lowest importance is associated with closeness to inventory threshold (23.63%).

**TABLE 7. The results of the top 10 stations of importance ranking.**

| Station ID | Distance from ideal solution | Distance from nonideal | TOPSIS score | Importance ranking |
|---|---|---|---|---|
| 19 | 0.0167 | 0.1135 | 0.8721 | 1 |
| 4 | 0.0682 | 0.0486 | 0.4163 | 2 |
| 24 | 0.0949 | 0.0230 | 0.1953 | 3 |
| 46 | 0.1047 | 0.0195 | 0.1572 | 4 |
| 23 | 0.0994 | 0.0177 | 0.1509 | 5 |
| 18 | 0.1094 | 0.0162 | 0.1292 | 6 |
| 29 | 0.1089 | 0.0157 | 0.1257 | 7 |
| 3 | 0.1019 | 0.0136 | 0.1178 | 8 |
| 41 | 0.1126 | 0.0139 | 0.1102 | 9 |
| 40 | 0.1129 | 0.0137 | 0.1081 | 10 |

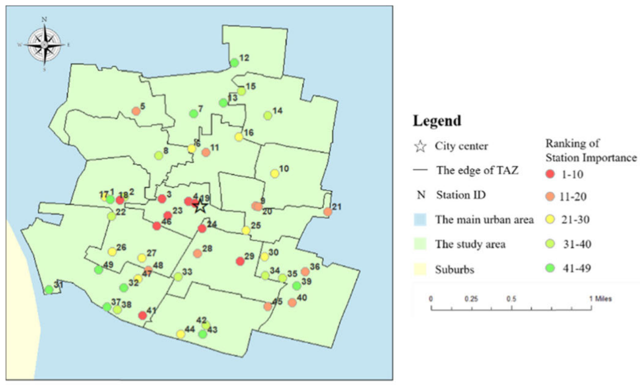



**FIGURE 6. The spatial distribution of station importance evaluation result.**

The results of the TOPSIS approach in Table 7 showed that station No.19 is ranked first as it obtained the highest amount of TOPSIS score (0.8721). This station also had the shortest distance from an ideal solution (0.0167). The results also showed that the TOPSIS score for station No. 4 is 0.4163 — the second-highest among all stations. Fig. 6 showed the spatial distribution of station importance evaluation. Highly ranked bike-sharing stations should be prioritized in vehicle routing plan. As Fig. 6 showed, most stations with a high level of importance aggregated around the key station. One of the main reasons is that distance from the key station is the most significant factor for station importance according to the results of weighting aspects for each criterion in Table 6. In Fig. 6, some stations obtained a relatively high ranking although they are distributed a bit far from the key station in the south. Rebalancing operation should also be implemented for these stations to improve service quality better.

## V. CONCLUSION

This research proposed the method of station importance evaluation for more effective dynamic bike-sharing rebalancing. The rebalancing quantity, closeness to inventory threshold, and distance from the key station were three criteria for evaluating bike-sharing stations. The GC-LSTM model was adopted for short-term bike-sharing demand prediction to determine rebalancing quantity with higher precision, considering exogenous factors such as land use information, weather, and users' personal information. The combination of LSTM and GCN enable to captured the temporal characteristics and the bike-sharing network structure characteristics. The experiments showed that GC-LSTM models outperformed baseline models using Nanjing docked bike-sharing dataset. Then, determination of rebalancing quantity was determined by considering bike-sharing usage throughout the rebalancing period. The entropy method was used to weigh the influences of different criteria. Then, the TOPSIS approach was employed to evaluate station importance and rank stations.

Numerical examples were set up to illustrate the proposed methods. Smart card data of docked bike-sharing in Nanjing City was obtained, and 49 stations within 13 TAZs around the city center were extracted. GC-LSTM models with different prediction intervals were established, and the RMSE and MAE were calculated to illustrate the performance of models. The results showed that the GC-LSTM model with a 15-min interval has the best performance. After calculating the rebalancing quantity, closeness to inventory threshold, and distance from the key station, the entropy-based TOPSIS method was employed to evaluate bike-sharing station importance. The entropy method results showed that distance from the key station had the most significant influence on station importance with a weight of 52.22%. Besides, based on the TOPSIS approach, the importance of each station was calculated, and the ranking results and spatial distribution were analyzed. It was indicated that many stations close to the key station gained a high level of importance. Meanwhile, there were also stations distributed far from the key station with higher rankings due to higher levels of other criteria. The average computational time of the GC-LSTM model is 5.9 seconds. For the entropy-based TOPSIS method, it is 1.4 seconds. The running time is acceptable in real-world bike-sharing rebalancing operations.

The findings of this study are helpful for operators to make a better decision on which stations should be prioritized in the rebalancing process. It also can be applied in user-based operation. Ji *et al.* [72] have suggested that higher monetary incentive will encourage users to participate into bike-sharing rebalancing and significantly alleviate bike-sharing imbalance issue. Station importance evaluation can be helpful to find which stations should be rebalanced first by users. By assigning higher monetary incentive to more important stations, users' engagement in rebalancing at those stations can be encouraged. Moreover, this framework is designed for docked bike-sharing, so it cannot be directly applied for dockless bike-sharing due to lack of fixed stations. However, similar to docked bike-sharing, the decision-making process of dockless bike-sharing rebalancing still needs to consider multi-criteria, such as travel distance of rebalancing vehicle, rebalancing quantity.

For different bike-sharing datasets in different geographic location, theoretically, GC-LSTM is able to capture temporal and spatial features of bike-sharing data. The results of station importance evaluation may not be the same as this case study, and the weights of criteria may be different. This is because the entropy-based TOPSIS can mine dataset characteristics, considering different key stations, station distribution and travel patterns in bike-sharing systems. This study has several limitations that need to be addressed in future research. Firstly, this research only took Nanjing docked bike-sharing as a case study. If other bike-sharing datasets are available, more experiments can be done using methods in this research for further verification and validation. Secondly, this framework can not directly be implemented on dockless bike-sharing. There are no fixed stations for dockless bike-sharing. Therefore, generating virtual stations and calculating loading and unloading time for decentralized dockless shared bikes are crucial questions in future research. Thirdly, this research only depicted the first stage of rebalancing in a rolling horizon to illustrate how to predict bike-sharing demand using GC-LSTM models and evaluate station importance based on multi-criteria. To complete the whole rebalancing process, the pickup-and-delivery vehicle routing problem should be solved to update the key station and bike-sharing station inventory. And then, the following rebalancing stages can be investigated and comparisons can be made to investigate how station prioritization improves vehicle routing plans. Fourthly, bike-sharing demand prediction is necessary for the dynamic rebalancing process. In this paper, the GC-LSTM models predict bike-sharing demand at the station level to propose station importance evaluation strategies. Although the performance of the GC-LSTM models on the Nanjing bike-sharing dataset is better than baseline models, state-of-art prediction models, such as temporal convolutional network (TCN) based methods or dual-stage attention-based recurrent neural network (DARNN) based methods, can be further explored to improve the accuracy of bike-sharing demand prediction. Finally, other MCDM methods could be used and compare the obtained results with the current study results. Future research can continue towards employing fuzzy MCDM methods to deal with the uncertainty involved in bike-sharing daily usage. It would assist operators and managers to dynamically rebalance shared bikes with high efficiency and improve the service quality of bike-sharing systems.

## REFERENCES

[1] Y. Cao and D. Shen, "Contribution of shared bikes to carbon dioxide emission reduction and the economy in Beijing," *Sustain. Cities Soc.*, vol. 51, Nov. 2019, Art. no. 101749.

[2] A. Pal and Y. Zhang, "Free-floating bike sharing: Solving real-life large-scale static rebalancing problems," *Transp. Res. C, Emerg. Technol.*, vol. 80, pp. 92–116, Jul. 2017.

[3] M. Dell'Amico, M. Iori, S. Novellani, and A. Subramanian, "The bike sharing rebalancing problem with stochastic demands," *Transp. Res. B, Methodol.*, vol. 118, pp. 362–380, Dec. 2018.

[4] T. Gu, I. Kim, and G. Currie, "To be or not to be dockless: Empirical analysis of dockless bikeshare development in China," *Transp. Res. A, Policy Pract.*, vol. 119, pp. 122–147, Jan. 2019.

[5] F. Soriguera and E. Jiménez-Meroño, "A continuous approximation model for the optimal design of public bike-sharing systems," *Sustain. Cities Soc.*, vol. 52, Jan. 2020, Art. no. 101826.

[6] Y. LuU Benlic and Q. Wu, "An effective memetic algorithm for the generalized bike-sharing rebalancing problem," *Eng. Appl. Artif. Intel.*, vol. 95, Dec. 2020, Art. no. 103890.

[7] L. Pan, Q. Cai, Z. Fang, P. Tang, and L. Huang, "Deep reinforcement learning framework for rebalancing dockless bike sharing systems," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 1393–1400.

[8] A. Singla, M. Santoni, G. Bart Ok, P. Mukerji, M. Meenen, and A. Krause, "Incentivizing users for balancing bike sharing systems," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 723–729.

[9] G. Berbeglia, J.-F. Cordeau, I. Gribkovskaia, and G. Laporte, "Static pickup and delivery problems: A classification scheme and survey," *TOP*, vol. 15, no. 1, pp. 1–31, May 2007.

[10] C. S. Shui and W. Y. Szeto, "Dynamic green bike repositioning problem—A hybrid rolling horizon artificial bee colony algorithm approach," *Transp. Res. D, Transp. Environ.*, vol. 60, pp. 119–136, May 2018.

[11] R. Salih-Elamin and H. Al-Deek, "Utilizing station importance to maximize locations coverage in a bike share system network," *Adv. Transp. Stud., Int. J.*, vol. 51, pp. 121–140, Dec. 2020.

[12] B. Legros, "Dynamic repositioning strategy in a bike-sharing system; how to prioritize and how to rebalance a bike station," *Eur. J. Oper. Res.*, vol. 272, no. 2, pp. 740–753, Jan. 2019.

[13] Y. Liu, W. Y. Szeto, and S. C. Ho, "A static free-floating bike repositioning problem with multiple heterogeneous vehicles, multiple depots, and multiple visits," *Transp. Res. C, Emerg. Technol.*, vol. 92, pp. 208–242, Jul. 2018.

[14] L. Caggiani, R. Camporeale, M. Ottomanelli, and W. Y. Szeto, "A modeling framework for the dynamic management of free-floating bike-sharing systems," *Transp. Res. C, Emerg. Technol.*, vol. 87, pp. 159–182, Feb. 2018.

[15] Y. Zhang, T. Cheng, and Y. Ren, "A graph deep learning method for short–term traffic forecasting on large road networks," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 34, no. 10, pp. 877–896, Oct. 2019.

[16] S. Shen, Z. Wei, L. Sun, K. S. Rao, and R. Wang, "A hybrid dispatch strategy based on the demand prediction of shared bicycles," *Appl. Sci.*, vol. 10, no. 8, p. 2778, Apr. 2020.

[17] A. Kaltenbrunner, R. Meza, J. Grivolla, J. Codina, and R. Banchs, "Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system," *Pervas. Mobile Comput.*, vol. 6, no. 4, pp. 455–466, Aug. 2010.

[18] J. W. Yoon, F. Pinelli, and F. Calabrese, "Cityride: A predictive bike sharing journey advisor," in *Proc. IEEE Int. Conf. Mobile Data Manage.*, Jul. 2012, pp. 306–311.

[19] J. E. Froehlich, J. Neumann, and N. Oliver, "Sensing and predicting the pulse of the city through shared bicycling," in *Proc. 21st Int. Joint Conf. Artif. Intell.*, Pasadena, CA, USA, 2009, pp. 1420–1426.

[20] Y. Zhou, L. Wang, R. Zhong, and Y. Tan, "A Markov chain based demand prediction model for stations in bike sharing systems," *Math. Problems Eng.*, vol. 2018, Jan. 2018, Art. no. 8028714.

[21] D. Li and Y. Zhao, "A multi-categorical probabilistic approach for short-term bike sharing usage prediction," *IEEE Access*, vol. 7, pp. 81364–81369, 2019.

[22] C. Feng, J. Hillston, and D. Reijsbergen, "Moment-based probabilistic prediction of bike availability for bike-sharing systems," in *Proc. 13th Int. Conf. Quant. Eval. Syst. (QEST)*, Québec City, QC, Canada, vol. 9826, Aug. 2016, pp. 139–155.

[23] M. G. Karlaftis and E. I. Vlahogianni, "Statistical methods versus neural networks in transportation research: Differences, similarities and some insights," *Transp. Res. C, Emerg. Technol.*, vol. 19, no. 3, pp. 387–399, Jun. 2011.

[24] B. L. Smith and R. K. Oswald, "Meeting real-time traffic flow forecasting requirements with imprecise computations," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 18, no. 3, pp. 201–213, May 2003.

[25] J. Liu, L. Qiao, Q. Meng, W. Chen, and Y. Fu, "Station site optimization in bike sharing systems," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2016, pp. 883–888.

[26] H. Xu, J. Ying, H. Wu, and F. Lin, "Public bicycle traffic flow prediction based on a hybrid model," *Appl. Math. Inf. Sci.*, vol. 2, no. 7, p. 667, 2013.

[27] Y. Li, Y. Zheng, H. Zhang, and L. Chen, "Traffic prediction in a bike-sharing system," in *Proc. 23rd SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, no. 33, Nov. 2015, pp. 1–10.

[28] Y. Yin, C. Lee, and Y. Wong, "Demand prediction of bicycle sharing systems, Stanford University," Stanford Univ., Stanford, CA, USA, Tech. Rep., 2014. [Online]. Available: http://cs.229stanford.edu/proj2014/YuchunYin,ChiShuenLee,Yu-PoWong,DemandPredictionofBicycleSharingSystems.pdf

[29] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[31] E. Hsueh-Chan Lu and Z.-Q. Lin, "Rental prediction in bicycle-sharing system using recurrent neural network," *IEEE Access*, vol. 8, pp. 92262–92274, 2020.

[32] Y. Pan, R. C. Zheng, J. Zhang, and X. Yao, "Predicting bike sharing demand using recurrent neural networks," *Procedia Comput. Sci.*, vol. 147, pp. 562–566, Dec. 2019.

[33] C. Xu, J. Ji, and P. Liu, "The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets," *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 47–60, Oct. 2018.

[34] J. Ke, H. Zheng, H. Yang, and X. Chen, "Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach," *Transp. Res. C, Emerg. Technol.*, vol. 85, pp. 591–608, Dec. 2017.

[35] J. Zhang, Y. Zheng, and D. Qi, "Spatio-temporal residual networks for citywide crowd flows prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 1655–1661.

[36] B. Du, H. Peng, S. Wang, M. Z. A. Bhuiyan, and L. Wang, "Deep irregular convolutional residual LSTM for urban traffic passenger flows prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 972–985, Mar. 2019.

[37] Y. Ai, Z. Li, M. Gan, Y. Zhang, D. Yu, W. Chen, and Y. Ju, "A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system," *Neural Comput. Appl.*, vol. 31, no. 5, pp. 1665–1677, May 2019.

[38] D. Ljubenkov, F. Kon, and C. Ratti, "Optimizing bike sharing system flows using graph mining, convolutional and recurrent neural networks," in *Proc. IEEE Eur. Technol. Eng. Manage. Summit (E-TEMS)*, Dec. 2020, pp. 1–6.

[39] T. S. Kim, W. K. Lee, and S. Y. Sohn, "Graph convolutional network approach applied to predict hourly bike-sharing demands considering spatial, temporal, and global effects," *PLoS ONE*, vol. 14, no. 9, Sep. 2019, Art. no. e0220782.

[40] A. Yoshida, Y. Yatsushiro, N. Hata, T. Higurashi, and K. Fujisawa, "Practical end-to-end repositioning algorithm for managing bike-sharing system," in *Proc. IEEE Int. Conf. Big Data*, Dec. 2019, pp. 1251–1258.

[41] R. Guo, Z. Jiang, J. Huang, J. Tao, C. Wang, and J. Li, "BikeNet: Accurate bike demand prediction using graph neural networks for station rebalancing," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innovation*, 2019, pp. 686–693.

[42] G. Xiao, R. Wang, C. Zhang, and A. Ni, "Demand prediction for a public bike sharing program based on spatio-temporal graph convolutional networks," *Multimedia Tools Appl.*, vol. 13, pp. 1–9, Mar. 2020.

[43] H. Xu, F. Duan, and P. Pu, "Dynamic bicycle scheduling problem based on short-term demand prediction," *Int. J. Speech Technol.*, vol. 49, no. 5, pp. 1968–1981, May 2019.

[44] K. Mellou and P. Jaillet, "Dynamic resource redistribution and demand estimation: An application to bike sharing systems," *SSRN Electron. J.*, pp. 1–58, Feb. 2019.

[45] C. Federico, P. Chiara, Z. Andrea, and Z. Michele, "A dynamic approach to rebalancing bike-sharing systems," *Sensors-Basel*, vol. 2, no. 18, pp. 512–534, 2018.

[46] A. A. Kadri, I. Kacem, and K. Labadi, "Lower and upper bounds for scheduling multiple balancing vehicles in bicycle-sharing systems," *Soft Comput.*, vol. 23, no. 14, pp. 5945–5966, Jul. 2019.

[47] J. Brinkmann, M. W. Ulmer, and D. C. Mattfeld, "The multi-vehicle stochastic-dynamic inventory routing problem for bike sharing systems," *Bus. Res.*, vol. 13, no. 1, pp. 69–92, Apr. 2020.

[48] J. Brinkmann, M. W. Ulmer, and D. C. Mattfeld, "Dynamic lookahead policies for stochastic-dynamic inventory routing in bike sharing systems," *Comput. Oper. Res.*, vol. 106, pp. 260–279, Jun. 2019.

[49] L. Zu, *A Study of the Static Bicycle Reposition Problem With a Single Vehicle*. Clemson, SC, USA: Clemson University, 2016.

[50] D. F. de Lima Silva and A. T. de Almeida Filho, "Sorting with TOPSIS through boundary and characteristic profiles," *Comput. Ind. Eng.*, vol. 141, Mar. 2020, Art. no. 106328.

[51] V. Salehi, H. Zarei, G. A. Shirali, and K. Hajizadeh, "An entropy-based TOPSIS approach for analyzing and assessing crisis management systems in petrochemical industries," *J. Loss Prevention Process Industries*, vol. 67, Sep. 2020, Art. no. 104241.

[52] Y. Ji, X. Ma, M. He, Y. Jin, and Y. Yuan, "Comparison of usage regularity and its determinants between docked and dockless bike-sharing systems: A case study in nanjing, China," *J. Cleaner Prod.*, vol. 255, May 2020, Art. no. 120110.

[53] M. Hua, X. Chen, S. Zheng, L. Cheng, and J. Chen, "Estimating the parking demand of free-floating bike sharing: A journey-data-based study of nanjing, China," *J. Cleaner Prod.*, vol. 244, Jan. 2020, Art. no. 118764.

[54] Y. Shen, X. Zhang, and J. Zhao, "Understanding the usage of dockless bike sharing in Singapore," *Int. J. Sustain. Transp.*, vol. 9, no. 12, pp. 1–15, 2018.

[55] A. Pal, Z. Yu, and C. Kwon, *Analyzing Mobility Patterns and Imbalance of Free-Floating Bike Sharing Systems*. Tampa, FL, USA: University of South Florida, 2017.

[56] D. Zhao, G. P. Ong, W. Wang, and X. J. Hu, "Effect of built environment on shared bicycle reallocation: A case study on nanjing, China," *Transp. Res. A, Policy Pract.*, vol. 128, pp. 73–88, Oct. 2019.

[57] (2020). *Weather Underground*. [Online]. Available: https://www.wunderground.com

[58] S. Sohrabi, R. Paleti, L. Balan, and M. Cetin, "Real-time prediction of public bike sharing system demand using generalized extreme value count model," *Transp. Res. A, Policy Pract.*, vol. 133, pp. 325–336, Mar. 2020.

[59] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," 2013, *arXiv:1312.6203*. [Online]. Available: https://arxiv.org/abs/1312.6203

[60] J. Chen, X. Xu, Y. Wu, and H. Zheng, "GC-LSTM: Graph convolution embedded LSTM for dynamic link prediction," 2018, *arXiv:1812.04206*. [Online]. Available: http://arxiv.org/abs/1812.04206

[61] L. Lin, Z. He, and S. Peeta, "Predicting station-level hourly demand in a large-scale bikesharing network: A graph convolutional neural network approach.," *Transp. Res. C, Emerg. Technol.*, vol. 97, pp. 258–276, Dec. 2018.

[62] E. D. O'Mahony, *Smarter Tools For (CITI) Bike Sharing*. New York, NY, USA: Cornell Univ., 2015.

[63] H. Jia, H. Miao, G. Tian, M. Zhou, Y. Feng, Z. Li, and J. Li, "Multiobjective bike repositioning in bike-sharing systems via a modified artificial bee colony algorithm," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 2, pp. 909–920, Apr. 2020.

[64] I. A. Forma, T. Raviv, and M. Tzur, "A 3-step math heuristic for the static repositioning problem in bike-sharing systems," *Transp. Res. B, Methodol.*, vol. 71, pp. 230–247, Jan. 2015.

[65] R. W. Sinnott, "Virtues of the haversine," *Sky Telescope*, vol. 68, no. 2, p. 159, 1984.

[66] C. L. Hwang and K. Yoon, *Multiple Attribute Decision Making: Methods and Applications* (Lecture Notes in Economics and Mathematical Systems), vol. 186. 1981, p. 1.

[67] Q. Qin, *The Research on Repositioning in a Bike Sharing System*. Beijing, China: Beijing Jiaotong Univ., 2013.

[68] D. G. Marte and H. Kristine, "A column generation heuristic for the dynamic rebalancing problem in bike sharing systems," Norwegian Univ. Sci. Technol., Trondheim, Norway, Tech. Rep., 2018.

[69] W. Yu, "Study on intelligent scheduling optimization algorithm of urban public bicycle system," Shanghai Jiao Tong Univ., Shanghai, China, Tech. Rep. 1120320987, 2015.

[70] P. Feng, "Research on scheduling management optimization method of public bicycle system," Southeast Univ., Dhaka, Bangladesh, Tech. Rep. 139234, 2018.

[71] C. Zhang, J. J. Q. Yu, and Y. Liu, "Spatial-temporal graph attention networks: A deep learning approach for traffic forecasting," *IEEE Access*, vol. 7, pp. 166246–166256, 2019.

[72] Y. Ji, X. Jin, X. Ma, and S. Zhang, "How does dockless bike-sharing system behave by incentivizing users to participate in rebalancing?" *IEEE Access*, vol. 8, pp. 58889–58897, 2020.


**MINGJIA HE** received the bachelor's degree in transportation engineering from Southeast University, in 2020. Her research interests include shared mobility optimization, travel behavior analysis, and transportation economics.

**XINWEI MA** received the M.S. degree in transportation planning and management from the Hebei University of Technology, in 2013, and the Ph.D. degree in transportation engineering from the School of Transportation, Southeast University, in 2020. He is currently a Lecturer at the Hebei University of Technology. His research interests include public transport shared mobility, transport data analytics, and intelligent transport systems.

**YUCHUAN JIN** received the bachelor's degree in traffic engineering from Southeast University, in 2019. He is currently pursuing the master's degree in transport science with the KTH Royal Institute of Technology. His research interests include transportation, big data analysis and modeling, machine learning, data mining, and intelligent transportation systems.