Data-Driven Optimisation of Supply-Demand Balance for Shared Bicycles

Xixiang Zhang

Department of Computer Science, University of Warwick Coventry, UK

Abstract—With the rapid proliferation of shared bicycles in urban transport, the imbalance between supply and demand distribution has become increasingly pronounced. The coexistence of vehicle shortages in high-demand areas and idle bicycles in low-demand zones is particularly evident during morning and evening rush hours. Existing research typically treats hotspot identification, demand forecasting, and dispatch optimisation as independent tasks, hindering the formation of a unified decision chain and constraining improvements in dispatch efficiency. To address this, this paper constructs an operational optimisation framework integrating spatial clustering, demand forecasting, and intelligent dispatch. First, irregularly distributed hotspot areas are identified through grid processing and DBSCAN density clustering. Second, a high-precision short-term demand forecast is achieved using an XGBoost model that integrates spatio-temporal features with historical data. Finally, based on the forecast results, an ant colony optimisation algorithm generates dispatch plans that balance both path efficiency and utilisation rates. Experiments using five days of real-world cycling data from Xiamen demonstrate that DBSCAN captures hotspot structures more accurately than K-Means, XGBoost achieves a 34% reduction in RMSE compared to the baseline model, and the ant colony algorithm reduces dispatch distances by 18% relative to a greedy strategy. These findings validate the framework's effectiveness in mitigating supply-demand imbalances and enhancing operational efficiency.

1. Introduction

As the development of smart cities advances, shared bicycles have gradually become a vital mode of short-distance urban travel due to their convenience, eco-friendliness and flexibility. However, in actual operation, the spatial distribution of shared bicycles exhibits marked imbalance: certain areas experience a shortage of bicycles during morning and evening rush hours, while low-demand zones see large numbers of vehicles left idle. This imbalance between supply and demand not only compromises the user experience but also increases operational and maintenance costs for bike-sharing dispatchers. Consequently, accurately identifying high-demand locations, forecasting short-term demand for shared bicycles, and devising efficient dispatch strategies have become critical challenges in managing bike-sharing systems[1][2].

Existing research primarily unfolds along two directions. The first concerns demand forecasting, encompassing traditional time series models and machine learning methods widely adopted in recent years. However, many approaches can only handle linear trends or exhibit limited performance in non-linear scenarios[3]. The second focuses on scheduling and rebalancing optimisation, frequently employing genetic algorithms, ant colony algorithms, or local greedy strategies to plan delivery routes. Yet, these methods lack deep integration with demand forecasting[4][5]. Overall, existing approaches tend to treat spatial hotspot identification, demand forecasting, and scheduling optimisation as separate processes. This absence of a unified end-to-end framework makes it challenging to achieve satisfactory results in complex urban environments.

To address the aforementioned issues, this paper proposes a shared bicycle operation framework integrating spatial clustering, machine learning prediction, and intelligent scheduling optimisation. Firstly, spatial clustering of shared bicycle travel data is attempted using K-Means and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithms to optimally identify hotspot areas of varying density and morphology[6]. Secondly, an eXtreme Gradient Boosting (XGBoost) model integrates temporal, spatial, and historical features to achieve high-precision forecasting of short-term demand across regions[7]. Finally, a multi-objective optimisation model is constructed based on these predictions, employing an ant colony algorithm for route planning to minimise travel distances and enhance vehicle availability[8].

The research contributions of this paper include:

- By integrating grid-based statistics with DBSCAN clustering, a refined spatial distribution structure for shared bicycles was constructed.
- Utilising XGBoost to fuse multi-source features enabled high-precision regional-level forecasting of non-linear demand fluctuations.
- Combining forecast outcomes with system constraints, an ant colony optimisation-based dispatch model was proposed to effectively enhance operational service efficiency.

Xixiang.Zhang@warwick.ac.uk

The remainder of this paper is structured as follows: Section II introduces the methodology; Section III presents experimental results; Section IV provides conclusions and future research directions.

2. Data and Methods

2.1 Data Source and Preprocessing

The research data originates from actual operational records of a shared bicycle platform in Xiamen City, Fujian Province, China, covering the period from 21st to 25th December 2020. The raw dataset comprises approximately 200,000 records, containing multiple fields including order ID, vehicle ID, start and end times, ride duration, GPS coordinates, and distance travelled.

To ensure data quality, preprocessing comprised four steps:

1) Journey chain construction: Matching complete ride sequences based on vehicle unlock and lock events, while excluding records with missing events or logical inconsistencies.

2) Anomaly filtering: Removing outliers such as zero ride durations, distances exceeding 20 km, or coordinates beyond geographical boundaries.

3) Time standardisation: Converted all timestamps to the hour level for subsequent time-series modelling.

4) Spatial grid-based discretisation: Divided the study area into fixed-scale grids to reduce GPS noise and provide structured input for spatial clustering.

Following preprocessing, approximately 186,000 valid ride records were obtained.

2.2 Spatial Clustering Method

Bicycle-sharing docking stations exhibit characteristics such as uneven distribution density and irregular spatial patterns within urban environments. To more accurately identify high-demand areas, this paper sequentially employs two clustering methods: K-Means and DBSCAN. The final clustering strategy is selected based on the results generated by both algorithms.

5) K-Means: K-Means clustering was first applied to the grid-based spatial data, with the number of clusters K selected using the elbow method and contour coefficient. However, owing to K-Means' reliance on Euclidean distance and the convexity assumption, its performance proved unsatisfactory when handling real-world non-spherical, cross-block bicycle parking points laden with substantial noise. Cluster boundaries exhibited instability, rendering it difficult to obtain interpretable hotspot area delineations.

6) DBSCAN: Given the limitations of K-Means, this paper employs DBSCAN as the final clustering model. DBSCAN identifies high-density regions through density-reachability relationships and automatically eliminates isolated noise points, rendering it suitable for irregular spatial structures. Parameters $\varepsilon = 0.005$ and $MinPts = 10$ were selected based on the k-distance curve, ultimately yielding a hotspot partition that reflects the actual distribution of urban functional zones. This outcome provides a sound spatial foundation for subsequent regional-level demand forecasting.

2.3 Demand Forecasting Model

To achieve accurate forecasting of future demand in hotspot areas, this paper employs the XGBoost model for short-term demand modelling. Based on the gradient boosting framework, this model possesses strong non-linear fitting capabilities and excellent generalisation performance, making it well-suited for shared bicycle regional demand forecasting scenarios[9].

7) Feature Construction: Input features comprise temporal characteristics (hour, day of the week, public holidays), spatial attributes (area codes and historical average demand in neighbouring zones), and historical data (ride demand within a sliding window of the preceding 1–2 hours).

8) Model Training: Using MSE as the loss function, hyperparameters were selected via grid search: learning rate 0.05, maximum depth 6, number of trees 200, and subsampling ratio 0.8. Model performance was evaluated using five-fold cross-validation. Experimental results demonstrate that XGBoost outperforms both linear regression and random forests in terms of RMSE and MAE metrics, proving more effective at capturing the non-linear fluctuations in shared bicycle demand.

2.4 Dispatch Optimization Algorithm

Based on demand forecasting results, this paper constructs a vehicle rebalancing scheduling model and employs the Ant Colony Optimisation (ACO) algorithm for solution. The scheduling objectives encompass minimising vehicle travel distance and enhancing the availability rate of bicycles within the service area.

The ACO algorithm achieves global optimisation by simulating the pheromone-based optimisation mechanism of ants searching for paths. Its primary workflow comprises: initialising the pheromone matrix and heuristic factors, constructing a regional distance matrix, generating scheduling tasks based on forecasted demand, and subsequently updating pheromone weights through multiple iterative rounds to progressively converge upon high-quality scheduling solutions. Experiments demonstrate that a stable feasible scheduling solution can be obtained after 50 iterations[10].

3. Results and Discussion

By integrating five days of actual shared bicycle trajectory and order data from Xiamen, this study empirically evaluates the proposed ‘spatial clustering-demand forecasting-dispatch optimisation’ framework across three dimensions: spatial distribution characteristics, demand forecasting performance, and dispatch optimisation effectiveness. This validation confirms the methodology's efficacy.

3.1 Spatial Distribution Characteristics Analysis

To identify hotspots for shared bicycles and potential dispatch centres, this study first constructed a spatial grid density map of shared bicycle distribution based on Xiamen's actual area. Figure 1 illustrates the initial bicycle demand per grid cell after dividing Xiamen into 25x25 equally proportioned grids, revealing a typical distribution pattern characterised by central density and peripheral sparsity.

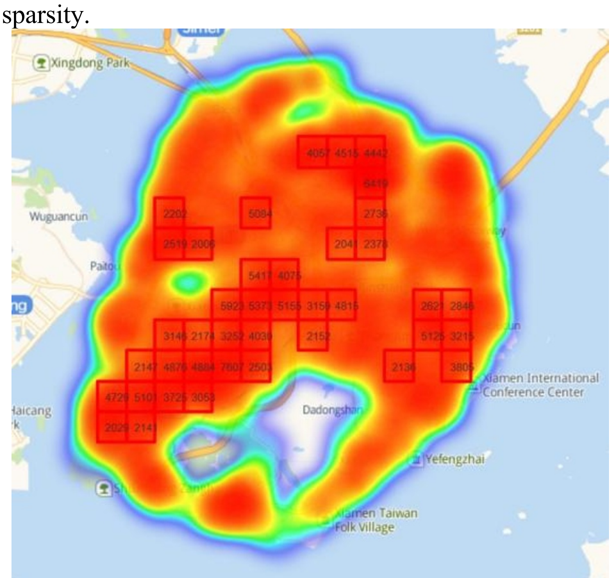

Fig. 1. Grid-based distribution map of shared bicycles

This pronounced spatial imbalance is also the core reason behind the difficulties in both borrowing and returning bicycles.

Subsequently, this study attempted spatial clustering of trajectory points using both K-Means and DBSCAN. The experiments revealed that K-Means tends to generate uniform, regular clusters. When applied to the irregular spatial distribution and highly uneven density of shared bicycles, K-Means often incorrectly merges scattered noise points into clusters, producing boundaries that do not correspond to actual urban functional zones, as illustrated by the clustered parking markers in Fig. 2. In contrast, DBSCAN identifies high-density parking areas through density-based criteria and simultaneously filters out isolated noise points, thereby more accurately capturing the spontaneously formed parking hotspots commonly observed in real-world bike-sharing systems, as shown in Fig. 3.

Ultimately, DBSCAN yielded 256 cluster centres covering nearly all significant activity areas, providing a spatial foundation for subsequent prediction and dispatch operations.

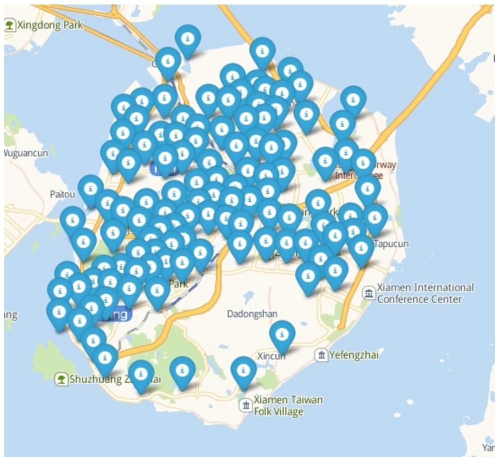

Fig. 2. Distribution of rental hotspots using K-Means

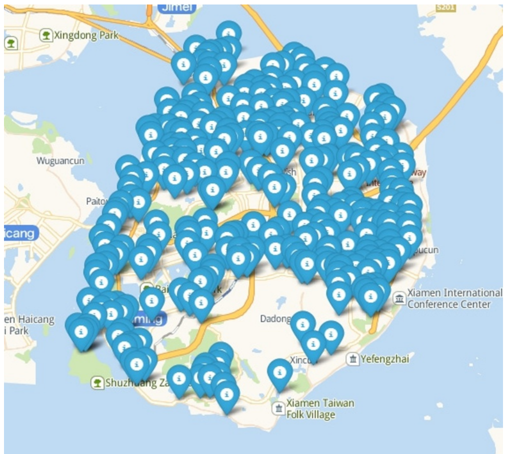

Fig. 3. Distribution of rental hotspots using DBSCAN

3.2 Short-Term Demand Forecasting Performance for Shared Bicycles

Following the acquisition of dispatch centres, this study constructed XGBoost predictive models for each centre to forecast future bicycle borrowing demand and return supply. The experimental results in Table 1 demonstrate that XGBoost outperforms traditional methods in capturing non-linear patterns. Compared to baseline models (linear regression, random forest), XGBoost achieved significant reductions in both RMSE and MAE, with its prediction curves maintaining synchronisation with actual demand trends.

Table 1. Performance comparison

Metric	Model		
	Linear Regression	***Random Forest***	***XGBoost***
RMSE	12.84	10.96	**8.52**
MAE	9.31	7.24	**5.80**

3.3 Dispatch Optimization Results

Upon obtaining demand forecast results, this paper constructs a shared bicycle dispatch route optimisation model based on the ACO algorithm. The dispatch process incorporates positive and negative demand values from the forecast as constraints: positive demand indicates bicycles available for dispatch, while negative demand signifies bicycles required for dispatch.

As illustrated in Fig. 4, the constructed optimisation model generates rational and efficient route plans within urban environments characterised by complex spatial structures and significant demand fluctuations. The dispatch paths exhibit compact local movement patterns, forming short-distance dispatch chains between adjacent stations that effectively reduce vehicle idling rates. The model-generated routes prioritise serving the highest-demand stations while balancing overall travel costs, significantly improving the equilibrium between incoming and outgoing vehicles.

Fig. 4. Example of ACO-based scheduling routes

In certain high-demand areas, the algorithm employs a cross-regional vehicle acquisition strategy to meet large-scale boarding demands. This involves first concentrating sufficient vehicles for dispatch, then returning to fulfil critical station requirements. Although cross-regional routes increase local travel distances, experimental results demonstrate that the ant colony algorithm-based dispatch strategy reduces dispatch travel distances by 18% compared to unoptimised paths based on a greedy approach. Based on the industry's cost analysis, assuming standard truck operating costs of RMB 3-4 per kilometre, this approach could reduce average daily dispatch expenditure in the study area by approximately 15%-20%.

3.4 Discussion

The clustering-forecasting-scheduling optimisation strategy constructed herein demonstrates sound applicability in operational analysis of shared bicycles. Firstly, during the spatial clustering phase, comparisons between K-Means and DBSCAN reveal that DBSCAN proves more adept at handling irregular shapes and noise points within shared bicycle distributions, enabling more accurate identification of genuine hotspot areas.

During demand forecasting, the XGBoost model effectively utilised temporal, spatial, and historical demand features to achieve stable predictions across all dispatch centres, capturing non-linear variations more effectively than linear models.

Dispatch optimisation results demonstrate that the ant colony optimisation algorithm, informed by forecasted demand, generates shorter routes covering primary high-demand zones while effectively reducing overall dispatch distances, significantly enhancing user satisfaction. Nevertheless, local loops persist within dispatch paths and some demand remains unmet, reflecting the limitations of heuristic algorithms when addressing extreme supply-demand imbalances.

4. Conclusion and Future Work

This study addresses spatial supply-demand imbalances in shared bicycle systems through an integrated strategy combining cluster analysis, demand forecasting, and scheduling optimisation. Employing a grid-based approach, we compared K-means and DBSCAN algorithms before selecting DBSCAN to handle irregular urban hotspot distributions. Subsequently, an XGBoost model incorporating spatio-temporal historical features was constructed to achieve precise short-term regional demand forecasting. Finally, an ant colony optimisation algorithm generates optimised dispatch routes to minimise costs while maximising vehicle availability.

Experiments using real urban datasets validate the framework's effectiveness: DBSCAN accurately captures hotspot patterns, the XGBoost model outperforms benchmark models, and the ant colony algorithm significantly reduces dispatch distances while enhancing availability rates. These findings demonstrate the practical value of this integrated approach and highlight the application potential of data-driven methods in shared bicycle operations management.

However, this study still holds room for improvement. The current prediction model employs a fixed time window, lacking real-time responsiveness; external factors such as weather and public holidays have not been fully incorporated into the model; and the scheduling optimisation component does not account for complex scenarios including multi-vehicle coordination, practical operational constraints, and dynamic demand fluctuations. Future research may explore several avenues for enhancement: incorporating advanced methodologies such as graph neural networks or reinforcement learning to improve predictive and decision-making capabilities; and scaling the system into a practical platform supporting multiple vehicles and dispatch points.

References

1. Y. Du, F. Deng, and F. Liao, "A model framework for discovering the spatio-temporal usage patterns of public free-floating bike-sharing systems," *Transportation Research Part C: Emerging Technologies,* vol. 103, pp. 39–55, 2019.
2. F. Chiariotti, C. Pielli, A. Zanella, and M. Zorzi, "A dynamic approach to rebalancing bike-sharing systems," *Sensors,* vol. 18, no. 2, p. 512, 2018.
3. E. Collini, P. Nesi, and G. Pantaleo, "Deep learning for short-term prediction of available bikes on bike-

sharing stations," *IEEE Access,* vol. 9, pp. 124337–124347, 2021.

4. X. Luo, L. Li, L. Zhao, J. Lin, and M. Lu, "Dynamic intra-cell repositioning in free-floating bike-sharing systems using approximate dynamic programming," *Transportation Science,* vol. 56, no. 4, pp. 799–826, 2022.
5. V. Mahmoodian, Y. Zhang, and H. Charkhgard, "Hybrid rebalancing with dynamic hubbing for free-floating bike sharing systems," *International Journal of Transportation Science and Technology,* vol. 11, no. 3, pp. 636–652, 2022.
6. M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining (KDD),* 1996, pp. 226–231.
7. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining,* 2016, pp. 785–794.
8. J. Liang, S. D. Jena, D. Liu, and A. Lodi, "A reinforcement learning approach for dynamic rebalancing in bike-sharing systems," *arXiv preprint arXiv:2402.03589,* 2024.
9. Z. Jin, K. Pan, Z. J. M. Shen, and W. Xu, "Integrated vehicle allocation and relocation for shared micromobility under competition and demand uncertainty," *IISE Trans., vol. 57, no. 8,* pp. 994–1013, 2025.
10. J. Ma, S. Zheng, S. Lin, and Y. Cheng, "A city shared bike dispatch approach based on temporal graph convolutional network and genetic algorithm," *Biomimetics,* vol. 9, no. 6, p. 368, 2024.