Proceedings of CONF-APMM 2025 Symposium: Controlling Robotic Manipulator Using PWM Signals with Microcontrollers
DOI: 10.54254/2753-8818/2025.AD26273

Ablation Study of Spatiotemporal Features and Bidirectional LSTM in Bicycle Demand Forecasting

Tianhe Yang

School of Economics and Management, Beijing University of Chemical Technology, Beijing, China
yangend123@outlook.com

Abstract. The penetration rate of shared bikes in urban transportation continues to rise, yet scheduling challenges caused by temporal and spatial supply-demand mismatch hinder industry development. Accurate demand forecasting is key to optimizing operational strategies and improving service efficiency, with significant research value. This paper focuses on bike-sharing demand forecasting, proposing an enhanced spatio-temporal attention model to boost prediction accuracy. It uses Mobike riding order data from Shanghai, extracting temporal features like hours and weekdays, and constructs a matrix with spatial features (average neighborhood demand after latitude-longitude network division). Synthetic Minority Over-Sampling Technique (SMOTE) is adopted to oversample peak data for data imbalance. Processed data is input into a bidirectional Long Short-Term Memory (LSTM) to capture temporal correlations in the sequence, combined with a spatio-temporal attention mechanism to focus on key features. Model performance is evaluated via ablation experiments (removing spatial features, bidirectional LSTM, or attention mechanism) using indicators such as RMSE, MAE, MAPE for non-zero demand, zero-value prediction accuracy, and peak hit rate. Results show the model effectively captures spatio-temporal features, providing strong support for bike-sharing scheduling management.

Keywords: Bike - sharing demand forecasting, Spatio - temporal features, Bidirectional LSTM network

1. Introduction

Since their inception, shared bicycles have rapidly gained popularity in cities worldwide, becoming an integral part of urban transportation systems. They not only provide residents with a convenient and eco-friendly mode of short-distance travel, effectively addressing urban traffic congestion and the "last-mile" dilemma, but also promote the concept of green travel. As of 2025, the shared bicycle market has reached a scale of billions of yuan, with over 200 million users and daily rides exceeding 10 million. Accurate demand forecasting is crucial for operators to optimize deployment and scheduling, refine user experience and boost efficiency.

In the field of demand forecasting, traditional methods such as time series analysis and regression analysis, although widely used, have limitations in handling data with highly nonlinear and dynamic characteristics [1]. The theoretical basis of these traditional methods can be traced back to the classical statistical learning framework [2], but they have insufficient adaptability when dealing with

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2025.AD26273

complex spatiotemporal data. In recent years, deep learning technology has brought new breakthroughs to this field, and its theoretical system has been systematically elaborated in the research by Goodfellow et al. [3]. The Long Short-Term Memory (LSTM) network, a special type of recurrent neural network proposed by Hochreiter and Schmidhuber [4], solves the gradient vanishing problem of traditional RNNs through a gating mechanism, and has shown significant advantages in time series forecasting [5]. Rao et al. were the first to apply LSTM to bicycle rental forecasting, verifying its ability to capture temporal features [6].

To improve forecasting accuracy, researchers have attempted model fusion. The introduction of the attention mechanism is an important breakthrough, which originated from the pioneering work of Vaswani et al. [7]. The AM-LSTM model built based on this mechanism dynamically focuses on key time steps and features, and combined with the temporal processing capability of LSTM, it significantly improves forecasting accuracy compared to single models and baseline methods [8]. In terms of feature extraction, the idea of rich feature hierarchies proposed by Girshick et al. has provided inspiration for cross-modal fusion. For example, the CNN-BiLSTM model constructed by Yang et al., which Unites the feature identification of convolutional neural networks and the sequential data processing skill of bidirectional long short-term memory models, has also achieved good results [9,10]. In addition, the residual network idea provides support for the training of deep models [11], while the spatiotemporal frameworks proposed by Ding et al. and Wang et al. respectively emphasize the collaborative modeling of spatial dependencies and temporal dynamics [12,13], providing a multi-dimensional perspective for shared bicycle forecasting. This paper adopts a spatiotemporal attention bidirectional LSTM model, which captures the front-back dependencies of time series through bidirectional LSTM and combines the attention mechanism to automatically learn feature importance, so as to enhance model expression and forecasting accuracy.

2. Data sources and feature dimensions addition

2.1. Data sources

The data source is the shared bicycle riding data in Shanghai in August 2016, with a total of more than 100,000 pieces of data. The original variable information is shown in Table 1.

Table 1. Description of original data

Variable	Code	Example
Start Time of Riding	Start Time	2016/8/31 21:03
Starting Longitude	Start Longitude	121.533
Starting Latitude	Start Latitude	31.295
End Time of Riding	End Time	2016/8/31 21:27
Ending Longitude	End Longitude	121.53
Ending Latitude	End Latitude	31.289
Bicycle Riding Track	Riding Track	121.4,31.27#121.4,31.20

2.2. Expansion of feature dimensions

Since the original data only contains information such as time and latitude/longitude, the feature dimensions are relatively limited. Therefore, it is essential to mine local weather information based

DOI: 10.54254/2753-8818/2025.AD26273

on time data and analyze the impact of weather features on shared bicycle riding patterns, which is crucial for improving the accuracy of shared bicycle riding demand forecasting.

2.2.1. Time and rainy/sunny weather features

As can be seen from Figure 1, the number of rides surges in the morning and evening rush periods (7 to 9 AM and 5 to 7 PM), and this phenomenon is more pronounced on sunny days than on rainy days; the number of rides is very low in the late night (0-6 AM), with little difference between sunny and rainy days; the number of rides is very low in the late night (0-6 AM), with little difference between sunny and rainy days. It can be concluded that sunny days can better promote cycling during the morning and evening rush hours, while rainy days will inhibit travel demand during peak hours.

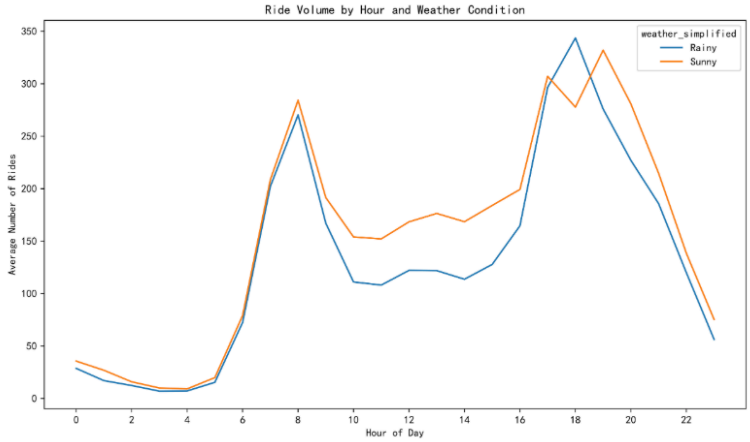


Figure 1. The relationship between weather and riding volume at each moment (picture credit: original)

2.2.2. Features of relative humidity and temperature

As can be seen from Figures 2 and 3, when the humidity is in the range of 60%-90%, the number of rides is the most concentrated (with the densest data points), indicating that this is a "comfortable humidity range" where people are more willing to ride bicycles. When the humidity is too high (>90%) or too low (<60%), the overall number of rides decreases. Especially when the humidity exceeds 95%, the trend line drops significantly, suggesting that excessively high humidity (such as stuffiness or rainy conditions) will reduce people's willingness to ride.

When the temperature is between 29-32°C, the overall number of rides is relatively high (with data points concentrated around 4,000 rides). After the temperature exceeds 32°C, the trend line declines. Particularly in the range of 33-34°C, the number of rides decreases significantly (some data points drop below 2,000 rides), indicating that high temperatures will inhibit the demand for cycling. The pink band (fluctuation range) first widens and then narrows, which shows that the number of rides fluctuates greatly when the temperature is moderate (possibly affected by other factors), while people's opinions on "whether to ride a bicycle" are more consistent in high temperatures (no one wants to ride).

DOI: 10.54254/2753-8818/2025.AD26273

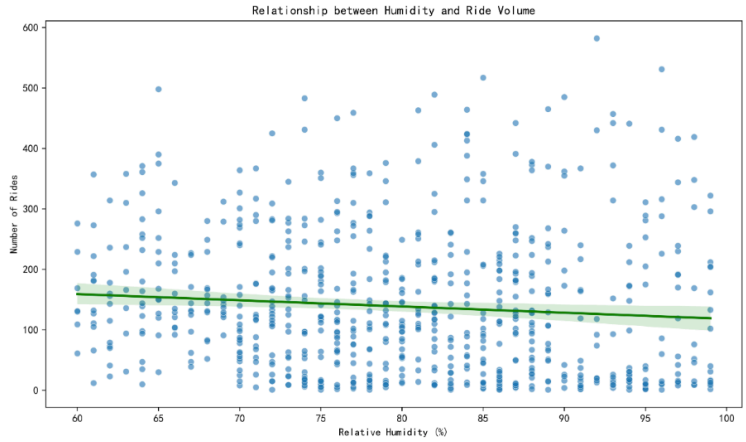


Figure 2. The relationship between Humidity and Ride Volume (picture credit: original)

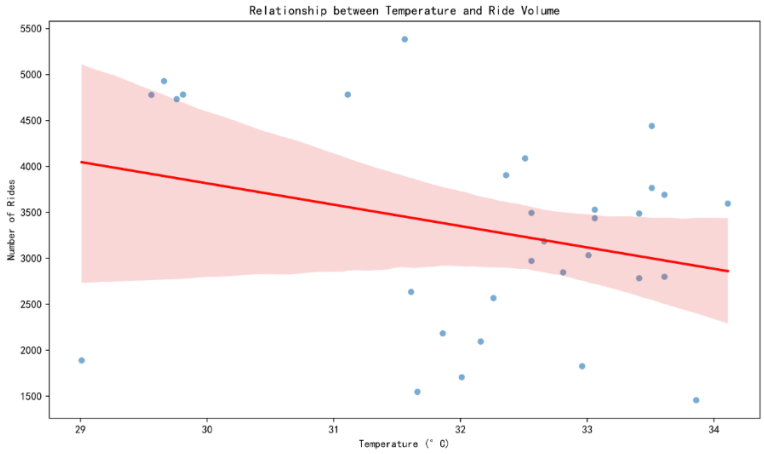


Figure 3. The relationship between Temperature and Ride Volume (picture credit: original)

3. Data preprocessing and feature engineering

3.1. Data filtering

For the convenience of model computation, this paper selects a latitude range of 31.18°N to 31.22°N and a longitude range of 121.43°E to 121.48°E, which corresponds to the central area of Shanghai, China. This region covers the southeast part of Huangpu District, including core business districts such as the Bund and Yuyuan Garden. After data cleaning and deduplication, the number of data entries is reduced to 7,798. A grid-time demand matrix of 660×29 is constructed, with zero-value demands accounting for 70.89%.

3.2. Feature extraction

3.2.1. Temporal features

Basic temporal features extracted include hour, weekday, and whether it is a weekend. To capture the periodicity of time, sine/cosine encoding is applied to hours and weekdays. This encoding maps linear time to a 2D plane, maintaining spatial continuity between adjacent time points (e.g., 23:00

DOI: 10.54254/2753-8818/2025.AD26273

and 00:00 are adjacent after encoding), which effectively enhances the model's ability to recognize periodic patterns.

3.2.2. Spatial features

The Haversine formula is used to calculate the spherical distance between grids. For each grid, the 3 nearest neighboring grids are selected, and statistics of their demand at the current time are computed. This feature captures spatial dependencies—for instance, neighboring grids with high demand may indicate commercial centers or transportation hubs.

3.2.3. Time-series features

The demand sequence of the 12 hours prior to the current moment is intercepted as the input to the LSTM. A sequence length of 12 hours balances the ability to capture short-term fluctuations (such as hourly peaks) and long-term trends (such as weekday patterns).

3.2.4. Weather features

Weather features extracted include weather type (binary encoded as sunny/rainy), temperature (29-32℃ as a comfortable range, with a high-temperature threshold set at >32℃), and relative humidity (60%-90% as a comfortable range, with a high-humidity threshold set at >95%). Additionally, "rainy day humidity" and "high temperature with high humidity" features are constructed to reflect the impact of sunny/rainy conditions on rush-hour rides and the inhibitory effect of extreme temperature/humidity on riding willingness.

4. Ablation experiment design and model architecture

4.1. Benchmark model architecture

The benchmark model adopts a "Bidirectional LSTM + Attention Mechanism" architecture, integrating spatiotemporal and weather features:

4.1.1. Input layer

Time-series features: Historical 12-hour demand (12×1); Spatial features: Statistical metrics of neighboring demand (1×4, including mean, maximum value, etc.);Time encoding features: Sine/cosine encoding of hours/weekdays (12×2);Weather features: 7-dimensional vector, including binary encoding of sunny/rainy conditions, standardized temperature, normalized humidity, high-temperature/high-humidity thresholds, and one-hot encoding of weather types.

4.1.2. Core components

Bidirectional LSTM (32 units), spatial fully connected layer (16 units), attention mechanism, and the output layer uses the Huber loss function.

Unified training parameters: Adam optimizer (lr=0.001), early stopping strategy (patience=10), batch size 32; Data preprocessing: Training and test sets are split at an 8:2 ratio, and numerical features are standardized.

DOI: 10.54254/2753-8818/2025.AD26273

4.2. Ablation experiment setup

This experiment adopts the Ablation Study method. By removing the key components of the model one by one and comparing the changes in model performance, it quantitatively verifies the contribution of each module to the demand forecasting task. Specifically, as shown in Table 2, the experiment sets up 5 groups of controls to disassemble the gains of five factors, namely spatial correlation, time periodicity, bidirectional time series, attention focusing, and environmental impact, on the model, so as to provide a basis for simplifying or enhancing the model structure.

Table 2. Ablation experiment setup

Experimental Group	Removed Component
Without Spatial Features	Neighborhood demand statistics
Without Time Encoding	Sine/cosine time features
Without Bidirectional LSTM	Bidirectional structure
Without Attention Mechanism	Attention weight calculation
Without Weather Features	All weather-related features

5. Experiments and results

5.1. Ablation experiment setup

Root Mean Square Error (RMSE):

The average degree of deviation between predicted values and actual values is measured by this metric.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{1}$$

In this context, n denotes the total count of samples, with y_i representing the actual value corresponding to the i-th sample, while $\hat{y}_i$ signifies its predicted counterpart.

Mean Absolute Error (MAE):

The average absolute divergence between predicted values and their actual counterparts is indicated by it.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \tag{2}$$

In this context, n denotes the total count of samples, with y_i representing the actual value corresponding to the i-th sample, while $\hat{y}_i$ stands for its predicted value.

Mean Absolute Percentage Error for Non-zero Demand (MAPE):

This indicator quantifies the error proportion of predicted values in relation to actual values. Within this research, MAPE calculations are restricted to non-zero demand scenarios alone.

DOI: 10.54254/2753-8818/2025.AD26273

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| *100\% \tag{3}$$

In this context, n denotes the number of samples with non-zero demand, with y_i representing the actual value corresponding to the i -th sample with non-zero demand, while $\hat{y}_i$ signifies its predicted counterpart.

Zero Prediction Accuracy:

The ratio of samples where both predicted and actual values are zero to those with actual values of zero is what zero prediction accuracy is defined as.

$$Zero\ Accuracy = \frac{\sum_{i-1}^{n} I(y_i<\epsilon \wedge \hat{y}_i<\epsilon)}{\sum_{i-1}^{n} I(y_i<\epsilon)} \tag{4}$$

In this context, n denotes the total count of samples, with y_i representing the actual value corresponding to the i -th sample, while $\hat{y}_i$ signifies its predicted counterpart. ϵ acts as a small threshold (set at 0.1), whereas I functions as an indicator function that outputs 1 when the condition is satisfied and 0 when it is not.

Spike Hit Rate:

The ratio of spike samples with predicted values exceeding 0.7 times their actual values to the total count of spike samples is what spike hit rate is defined as.

$$Spike\ Hit\ Rate = \frac{\sum_{i-1}^{n} I(y_i>\tau \wedge \hat{y}_i>0.7y_i)}{\sum_{i-1}^{n} I(y_i>\tau)} \tag{5}$$

In this context, n denotes the total count of samples, with y_i representing the actual value corresponding to the i -th sample, while $\hat{y}_i$ signifies its predicted counterpart, τ acts as the threshold employed to define spikes (set at 1.5), whereas I functions as an indicator function that yields 1 when the condition is satisfied and 0 in all other cases.

DOI: 10.54254/2753-8818/2025.AD26273

5.2. Comparison of ablation experiment results

Table 3. Ablation experiment setup

Model	RMSE	MAE	Non-zero MAPE (%)	Zero Prediction Accuracy (%)	Spike Hit Rate (%)
Benchmark Model	0.8165	0.4969	22.26	24.99	87.47
Model Without Spatial Features	0.8201	0.5012	22.76	21.39	87.49
Model Without Time Encoding	0.8560	0.5407	24.43	6.39	85.49
Model Without Bidirectional LSTM	0.8900	0.5772	24.67	21.79	86.03
Model Without Attention Mechanism	0.8353	0.5193	22.16	23.89	87.54
Model Without Weather Features	0.8452	0.5123	23.15	19.27	86.12

As can be seen from Table 3, The bidirectional LSTM's necessity is evident in its substantial role in boosting the model's predictive performance. After removing this structure, the RMSE increases by 7.35% and the MAE rises by 8.03%, which fully indicates that the bidirectional structure is crucial in capturing long-term temporal dependencies. For example, it can effectively identify the demand differences between weekdays and weekends.

The role of time encoding cannot be ignored either. After its removal, the zero prediction accuracy drops sharply by 18.6%, showing that time features are the core for identifying zero-value demands. For instance, the regularity of zero demand at night relies on the effective extraction of time encoding.

The impact of spatial features presents a certain degree of complexity. After removal, the zero prediction accuracy increases by 3.6%, but the non-zero MAPE rises by 0.5%. This indicates that spatial neighborhood features play a positive role in predicting zero values for low-activity grids, though their contribution to the accuracy of non-zero demands is relatively limited.

The value of the attention mechanism is two-sided. After its removal, the spike hit rate increases by 0.7%, but the RMSE rises by 1.8%, implying that while the attention mechanism focuses on spike demands, it may introduce certain noise to the prediction of stable demands.

Weather features also have a significant impact on model performance. After their removal, the overall error increases, indicating that weather factors have strong explanatory power for demand fluctuations. The 1.35% drop in the spike hit rate further suggests that features such as comfort indices can assist in identifying demand peaks during favorable weather conditions.

6. Presentation of prediction results

The comparison of shared bicycle demand forecasts is shown in Figure 4. The revised forecast (orange) is highly synchronized with the overall fluctuations of the actual demand (blue), with a good fit of peak and valley trends. The predictions for spikes (such as the 250-270 time period) and troughs (such as the 50 and 100 time points) are highly consistent, which can effectively capture the periodic patterns of demand and extreme values. Although there are slight deviations in some local

DOI: 10.54254/2753-8818/2025.AD26273

areas (such as the 150-200 interval), it can accurately assist in operational scheduling and provide reliable support for resource allocation and supply-demand matching.

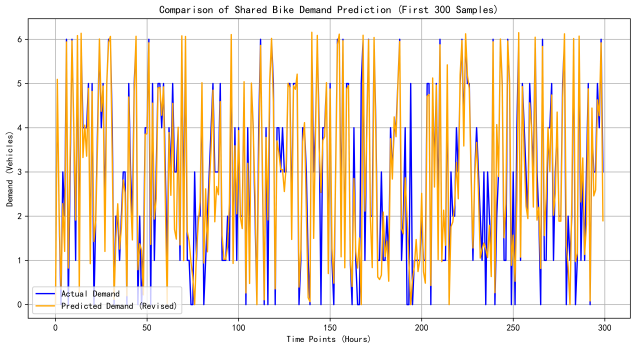


Figure 4. Display of shared bicycle prediction results (picture credit: original)

7. Conclusions

Aiming at the problem of shared - bike demand prediction, this paper proposes an enhanced attention model that integrates spatiotemporal features and bidirectional LSTM. The research is based on the 2016 Mobike riding data in Shanghai. Through feature engineering, it expands features such as time (sine/cosine encoding of hours/weeks), space (neighborhood demand statistics), time series (historical 12 - hour demand), and weather (sunny/rainy, temperature and humidity thresholds). The SMOTE method is used to deal with the problem of unbalanced spike data, and ablation experiments are designed to verify the role of each core component.

The experimental results show that the bidirectional LSTM is crucial for capturing long - term temporal dependencies, time encoding is the core feature for identifying zero - value demands, weather features have significant explanatory power for demand fluctuations, and the spatiotemporal attention mechanism shows certain advantages in focusing on spike demands. Comprehensively, the proposed model can effectively integrate multi - dimensional features and outperforms the comparison models with key components removed in indicators such as RMSE, MAE, and spike hit rate. It can provide reliable demand prediction support for the refined scheduling of shared bikes.

Future research can be expanded in three aspects. First, introduce more fine - grained external features (such as traffic control, large - scale event information) to further improve the prediction accuracy in special scenarios. Second, try to expand the model to multi - city comparative research to analyze the differences in the impact of traffic features of different cities on demand patterns. Third, combine real - time data stream processing technology to realize dynamic prediction updates and enhance the timeliness of the model in actual operations. In addition, exploring the integration method of advanced architectures such as Transformer with spatiotemporal features may further break through the limitations of existing models in capturing long - term dependencies.

References

[1] Li, P. (2024). Research on Shared - bike Flow Calculation Method Based on Spatiotemporal Features. Southwest University of Finance and Economics.
[2] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning (2nd ed.). Springer.
[3] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
[4] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.

DOI: 10.54254/2753-8818/2025.AD26273

[5] Shi, Y. S. (2024). Application of RNN and LSTM Online Learning Framework in Time Series Prediction. Yunnan Normal University.
[6] Rao, Y. (2019). Applying LSTM for Bicycle Rental Prediction. Journal of Transportation Research.
[7] Vaswani, A., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.
[8] Xu, M., Liu, H. F., Chu, K. (2020). Spatiotemporal Demand Prediction of Shared Bikes Based on AM - LSTM Model.
[9] Girshick, R., et al. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 580-587.
[10] Yang, F., Che, X. H., Wang, Y., et al. (2023). Multi - feature CNN - BiLSTM Prediction Method for Urban Shared - bike Demand. Bulletin of Surveying and Mapping.
[11] He, K., et al. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770-778.
[12] Ding, C., et al. (2018). A spatio-temporal framework for bike-sharing demand prediction with deep learning. Transportation Research Part C: Emerging Technologies, 93, 205-218.
[13] Wang, H., et al. (2020). ST-ResNet: Spatio-temporal residual networks for citywide crowd flows prediction. 21(9), 3837-3848.