2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Shared Bicycle Demand Prediction Based on Hierarchical Spatiotemporal Graph Convolution Network

Zikang Dai, Lifeng Yang, Liming Jiang (✉), and Huanyu Wang

School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China
liming@hnust.edu.cn

Abstract. Shared bicycle demand prediction plays a crucial role in urban public transportation planning, citizen mobility, and environmental protection. However, existing demand prediction models have certain limitations in modeling the coupling relationship between the inflow and outflow of shared bicycle stations. Furthermore, the performance of the model is highly influenced by sparse data, especially for fine-grained shared bicycle demand prediction tasks. To address these challenges, this paper proposes a Hierarchical Spatiotemporal Graph Convolutional Network (HST-GCN). Specifically, we apply a hierarchical spatial-temporal feature learning framework to capture both coarse-grained and fine-grained shared bicycle demand features, and utilize a feature transformation matrix to achieve cross-scale fusion of these demand features, alleviating the impact of data sparsity on fine-grained feature modeling. We also design a dynamic coupling graph convolution module to better model the dynamic spatial dependencies between the inflow and outflow of shared bicycle stations. On this basis, we integrate temporal convolution networks and temporal attention mechanisms to capture the spatial-temporal correlations of shared bicycle demand. Extensive experiments are conducted on the Citi Bike dataset from New York and the Divvy dataset from Chicago. The results show that the proposed model outperforms the baseline models in prediction accuracy.

Keywords: Transportation planning, Demand prediction, Temporal convolution network, Graph convolution network

1 Introduction

Shared bicycles are a low-carbon, eco-friendly, and convenient mode of transportation that solves the problem of the "last mile" in urban areas. They have become an essential component of modern urban transport systems [1]. However, with the rapid development of the industry, issues such as excessive bicycle deployment, disorganized parking, and supply-demand imbalance have become urgent problems that need to be addressed in bicycle-sharing systems. Accurately predicting the inflow and outflow demand of shared bicycle stations is the key to addressing the supply-demand imbalance and further promoting the sustainable development of the bicycle-sharing system.

In recent years, researchers have proposed various deep learning-based prediction methods for shared bicycle demand, including methods considering temporal, spatial, and spatial-temporal correlations. Pan et al. [2] used a two-layer Long Short-Term Memory (LSTM) network, while Yang et al. [3] proposed a combined prediction model based on Prophet and Bidirectional Long Short-Term Memory (BiLSTM) networks. Although effective in modeling temporal dependencies, these methods often neglect spatial correlations, limiting their applicability in systems with complex network structures.

To address spatial dependencies, Qiao et al. [4] designed a dynamic CNN-based model that incorporates weather data and represents station status as a 2D feature matrix. However, CNNs are limited in capturing long-range dependencies in non-Euclidean spaces. Graph Neural Networks (GNNs), particularly Graph Convolutional Networks (GCNs), have since been introduced to model these complex spatial structures. For instance, Liang et al. [5] proposed a GNN-based trip generation model, and Qin et al. [6] developed RESGCN to capture long-range spatial dependencies using directed weighted graphs.

However, most current spatiotemporal graph convolutional network prediction models have the following shortcomings: (1) Neglecting inflow-outflow coupling: The mutual influence between inflow and outflow at and across stations is often overlooked. For example, inflow at a station is affected by the outflow of its neighbors, and vice versa. Ignoring this coupling weakens prediction accuracy. (2) Data sparsity: Many stations exhibit zero demand during most time intervals, causing underfitting and reduced model performance.

To address the above issues, this paper proposes a Hierarchical Spatiotemporal Graph Convolutional Network (HST-GCN) model. Firstly, spectral clustering is applied to partition the stations in bicycle-sharing systems into regions, thus constructing demand data at both the station-level and region-level granularity. Furthermore, for demand data with different granularities, a dynamic coupling graph convolution module is employed to capture the coupling relationship and spatial dependencies between inflow and outflow for shared bicycle demand, while temporal convolution and attention mechanisms are incorporated to capture the temporal features. Finally, a hierarchical dynamic interaction module is constructed to facilitate cross-scale interactions between demand features, alleviating the issue of data sparsity in fine-grained demand prediction tasks.

2 RELATED WORK

2.1 Region-Based Shared Bicycle Demand Predictions

This approach divides bicycle stations into regions, then uses deep learning methods to predict the shared bicycle demand for each region. Zhang et al. [7] proposed a CNN-based deep prediction model by dividing Beijing city into several grids, achieving accurate predictions of shared bicycle demand in each grid. Although this method achieved notable results in spatial feature modeling, it overlooks the impact of temporal features on demand prediction. Yu et al. [8] proposed a hybrid model that combines

Seasonal Autoregressive Integrated Moving Average (SARIMA) and Long Short-Term Memory (LSTM), which primarily captured the temporal dynamics of shared bicycle demand while ignoring the spatial features. To capture both temporal and spatial features, Ai et al. [9] proposed a deep prediction model based on ConvLSTM, which divided the Chengdu geographical area into multiple grids of specified sizes to fully capture the temporal dynamics and multi-dimensional feature dependencies in shared bicycle demand data. Li et al. [10] proposed a region-level shared bicycle demand prediction model based on the Irregular Convolution Long Short-Term Memory (IrConv+LSTM) model. This model extracts features over regions that are both spatially adjacent and far apart, and enhances prediction accuracy by introducing the concept of semantic neighbors. In terms of handling complex spatiotemporal dependencies, Tang et al. [11] proposed the Multi-community Spatiotemporal Graph Convolution Network (MC-STGCN) framework, which uses heterogeneous graphs to capture geographical adjacency and functional similarity between regions, thus enabling more accurate region-level demand predictions. Compared to traditional methods, this model performs better in modeling complex spatiotemporal dependencies.

2.2 Station-Based Shared Bicycle Demand Prediction

This approach treats each station as an individual entity and predicts the demand for bicycles at each station. Traditional statistical models for shared demand prediction typically use regression models such as Auto-regressive Integrated Moving Average (ARIMA) [12] and Linear Regression (LR) [13] to reveal the relationship between bicycle demand and historical data. However, as demand patterns become more complex, the prediction accuracy of traditional statistical models becomes limited. Subsequent research explored advanced machine learning models to address the limitations. Ashqar et al. [14] employed Random Forest and Least Squares Boosting to model the demand at each station. Feng et al. [15] proposed a hierarchical demand prediction model based on Gradient Boosting Regression Trees. However, these methods still face challenges in efficiently processing large-scale, unstructured data. In recent years, deep learning models have shown significant advantages in capturing complex nonlinear relationships in demand prediction tasks. Li et al. [16] combined Convolutional Long Short-Term Memory (Conv-LSTM) networks with feature engineering techniques to effectively capture spatiotemporal dependencies at the station level. Yin et al. [17] conducted a one-year data analysis of the Divvy bicycle-sharing system in Chicago, quantifying the predictability of station-level shared bicycle demand and emphasizing the importance of time-varying patterns for accurate predictions. Lin et al. [18] proposed a model based on Graph Convolutional Networks (GCN), combined with a recursive network, to enhance the modeling ability of spatiotemporal dependencies. Liang et al. [19] proposed a cross-mode knowledge adaptation method based on multi-relational graph neural networks, effectively improving demand prediction accuracy by combining historical data with domain adversarial learning.

Based on the above work, a Hierarchical Spatiotemporal Graph Convolutional Network model (HST-GCN) is proposed. First, the dynamic coupling module captures the correlation effects between inflow and outflow. Then, spatiotemporal features are extracted from coarse and fine-grained data, and the hierarchical dynamic interaction module integrates features of different granularities to solve the data sparsity problem.

3 PRELIMINARIES

3.1 Definitions

This paper focuses on predicting future inflow and outflow demand based on historical shared bicycle inflow and outflow data. To properly define this prediction problem, we use the following definitions:

Definition 1 (Station Graph). Unlike traditional transportation systems with fixed infrastructure, bicycle-sharing systems are inherently dynamic, with new stations frequently added over time. To ensure the stability of the network structure during model training, we adopt a long observation window and focus on nodes with relatively consistent historical presence, thereby reducing the impact of recently added or short-lived stations. We define the station network as an unweighted undirected graph $G^s = (V^s, E^s, A^s)$, where V^s represents the set of nodes, E^s represents the set of edges, and A^s is the adjacency matrix of graph G^s. Let $N = |V^s|$ be the number of stations. Shared bicycle stations correspond to the nodes in V^s, with each node recording the inflow and outflow information of the corresponding station.$A^S \in \mathbb{R}^{N \times N}$ reflects the spatial proximity effects between connected stations.

Definition 2 (Region Graph). A region is a larger spatial unit formed by aggregating stations through spectral clustering, with each region consisting of multiple stations that are highly correlated. We define the region network as an unweighted undirected graph $G^R = (V^R, E^R, A^R)$, where V^R represents the set of regions,E^R represents the set of edges, and A^R is the adjacency matrix of graph G^R. Let $N^R = |V^R|$ be the number of regions. Shared bicycle regions correspond to the nodes in V^R, with each node recording the inflow and outflow information of the corresponding region. $A^R \in \mathbb{R}^{N^R \times N^R}$ reflects the spatial proximity effects between connected regions. The region demand data at time$t$ is denoted as x_t^R.

Definition 3 (Out/In-demand). The inflow X_{in}^t and outflow X_{out}^t at time t represent the number of users arriving and leaving at each station, respectively. The historical demand over T time slices is denoted as $X = [X_{\text{in}}, X_{\text{out}}] = (X_{t-T+1}, X_{t-T+2}, \dots, X_t) \in \mathbb{R}^{N \times 2F \times T}$, where F is the feature dimension. This encapsulates the fine-grained shared demand feature. Region-wise features can be denoted as$X^R = [X_{\text{in}}^R, X_{\text{out}}^R] = (X_{t-T+1}^R, X_{t-T+2}^R, \dots, X_t^R) \in \mathbb{R}^{N^R \times 2F \times T}$, which characterizes the demand features of urban areas at a coarse-grained level.

Problem Definition. Based on historical inflow and outflow demand data X for T time slices and the bicycle-sharing system's station structure graph G, let the model's learning mapping function be g(·). The function g(·) is used to learn the inflow and outflow demand features from the historical data for T time slices to predict the inflow and outflow demand $\hat{X} = (X_{t+1}, X_{t+2}, \dots, X_{t+M})$ for the next M time slices. Therefore, the shared bicycle demand prediction problem can be formulated as follows:

$$\left(\hat{X}_{t+1}, \hat{X}_{t+2}, \dots, \hat{X}_{t+M}\right) = g\left(G; (X_{t-T+1}, X_{t-T+2}, \dots, X_t)\right) \tag{1}$$

3.2 Spatial Graph Convolution

Graph convolution is a convolution operation based on graph-structured data, used to capture the spatial dependencies between nodes and their neighborhoods. In Graph Neural Networks (GNNs), graph convolution learns node representations by interacting node features with the adjacency matrix, effectively modeling the spatial correlation characteristics in complex graph structures. This paper adopts a spatial-based graph convolution method, with the model based on the spatial method proposed by Kipf et al. [20], defined as:

$$H^{(l+1)} = \sigma\left(\widehat{D}^{\frac{1}{2}}\hat{A}\widehat{D}^{\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{2}$$

In the formula, $H^{(l)} \in \mathbb{R}^{N\times F}$ represents the hidden state matrix at layer. The adjacency matrix $\hat{A}$ is the self-loop enhanced adjacency matrix, which adds the identity matrix I to the original adjacency matrix A to enhance the self-loop information; $\widehat{D}$ is the degree matrix derived from $\hat{A}$, with diagonal elements $\widehat{D}_{ii} = \sum_j \hat{A}_{ij}$. $W^{(l)}$ is the learnable weight parameter matrix at layer l. The nonlinear activation function σ (such as ReLU or Sigmoid) is applied to the model's output to enhance its nonlinear expressive power.

3.3 Temporal One-Dimensional Convolution

Temporal one-dimensional convolution is a convolution operation used to capture local patterns in time-series data by sliding the convolution kernel along the temporal dimension to extract the dynamic changes in node or region features over time. Let the input feature be $\mathrm{H} \in \mathbb{R}^{N\times 2F\times T}$. The calculation process of the temporal one-dimensional convolution can be represented as:

$$H_{\text{out}} = \mathrm{ReLU}(W_{\text{conv}} * H + b) \tag{3}$$

In the formula above, W_{conv} represents the learnable convolution kernel, which extracts the local dynamic features along the temporal dimension. The length of the convolution kernel is k; $*$ denotes the one-dimensional convolution operation, which essentially slides the convolution kernel along the time axis and performs a dot product with the input feature; b is the bias term, which adjusts the feature distribution; $\mathrm{ReLU}(\cdot)$ is a nonlinear activation function, which enhances the model's ability to express complex temporal dependencies. After the convolution operation, the time dimension length of the output feature decreases from T to $T - k + 1$, reflecting the constraint of the convolution kernel length on the local receptive field of the feature, while ensuring the model's ability to model short-term dependencies.

4 METHODOLOGY

This paper proposes a Hierarchical Spatiotemporal Graph Convolutional Network (HST-GCN) model for shared bicycle demand prediction, with the overall framework shown in Fig. 1. The model primarily consists of five modules: spectral clustering-based shared bicycle region graph generation, regional spatiotemporal graph convolution, station spatiotemporal graph convolution, hierarchical dynamic fusion module, and prediction module. The hierarchical graph generation module is used to generate the region-level graph structure and data representation. The station (region) spatiotemporal graph convolution network is used to learn the spatiotemporal features of demand data, with the DCA-GCN integrated with a dynamic coupling module to extract spatial dependencies and capture the mutual influence between inflow and outflow; temporal attention and temporal one-dimensional convolution capture temporal correlations. The hierarchical dynamic interaction module is used to integrate and fuse coarse and fine-grained demand features, addressing the data sparsity issue. Finally, the prediction module generates the final prediction output. Both regional spatiotemporal graph convolution and station spatiotemporal graph convolution consist of multiple spatiotemporal blocks, and they have the same structure.

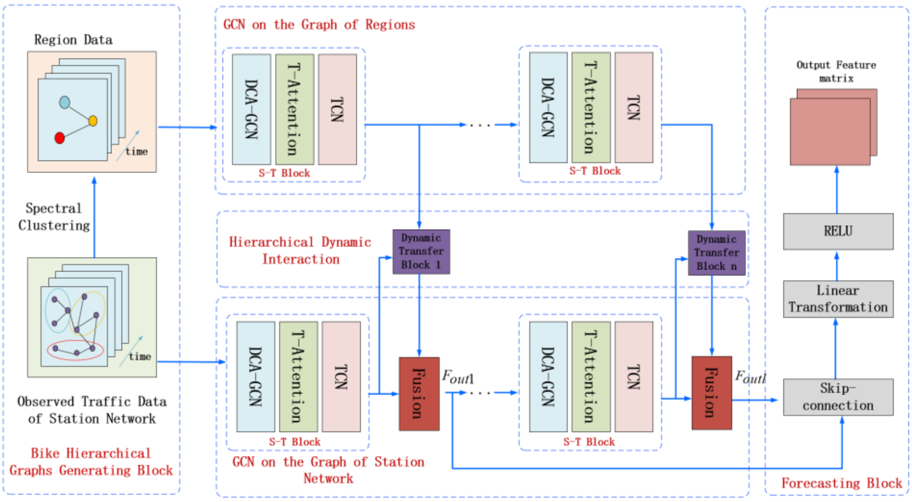

Fig. 1. The framework of HST-GCN

In the subsequent subsections of this section, we will first take the spatiotemporal block in regional spatiotemporal graph convolution as an example and provide a detailed introduction to its internal components, including the Dynamic Coupling Graph Convolution (DCA-GCN) module and temporal attention. Then, we will introduce the dynamic interaction module for station and region features.

4.1 Dynamic Coupling Graph Convolution Module

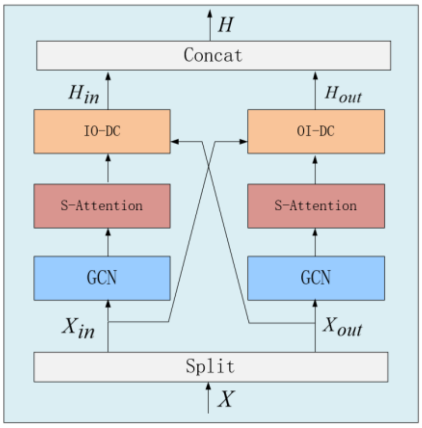

Fig. 2. GCNGRU unit structure

The Dynamic Coupling Graph Convolutional Network (DCA-GCN) learns the spatial dependencies in shared bicycle data and simulates the coupling influence between inflow and outflow. The structure of the DCA-GCN is shown in Fig. 2. The model mainly consists of a Graph Convolutional Network (GCN), spatial attention, and a dynamic coupling module. This subsection will elaborate on these modules in detail

Spatial Attention. In the spatial dimension, different locations of shared bicycle nodes exert mutual influence, which has high dynamicity. The spatial attention mechanism can adaptively adjust the weights between nodes based on the locations, thus more accurately reflecting the spatial relationships between nodes. Therefore, this model adopts the spatial attention mechanism to learn the dynamic influence relationships in the spatial dimension. The formula for the spatial attention mechanism is expressed as follows:

$$S = V_s \cdot \sigma((XW_1)W_2(W_3X)^\top + b_s) \tag{4}$$

$$S'_{(i,j)} = \frac{\exp(s_{(i,j)})}{\sum_{j=1}^{N} \exp(s_{(i,j)})} \tag{5}$$

$$Satt(X) = S'X \tag{6}$$

where $X \in \mathbb{R}^{N^R \times F \times T}$ represents the input to the spatial attention module, F is the feature dimension of the nodes, and T is the length of the time series. In the formula, V_s, $b_s \in \mathbb{R}^{N^R \times N^R}$, $W_1 \in \mathbb{R}^{T}$, $W_2 \in \mathbb{R}^{F \times T}$ and $W_3 \in \mathbb{R}^{F}$ are learnable parameters, and σ is an activation function. First, the spatial attention weight matrix S is obtained according to equation (5). Then, the Softmax function is applied to normalize the weight matrix S, ensuring that the sum of attention weights for each node is 1, yielding the final spatial attention matrix $S' \in \mathbb{R}^{N^R \times N^R}$, where $S'_{(i,j)}$ indicates the strength of the correlation between node i and node j in the weight matrix. Finally, the spatial attention matrix is multiplied with the input features to dynamically adjust the spatial dependencies between nodes, obtaining the final result $Satt(X) \in \mathbb{R}^{N^R \times F \times T}$.

Dynamic Coupling Module. In bicycle-sharing systems, there is a correlation between the inflow and outflow of vehicles at the nodes. We establish two criteria to describe such correlations. The first inflow criterion: the inbound demand at a node at time $t+1$ is influenced by the outbound demand of its neighboring nodes at time t. The second outflow criterion: the outbound demand at a node at time $t+1$ is influenced by its inbound demand at time t. Based on the inflow criterion, the In-Out Dynamic Coupling module (IO-DC) is constructed to capture the coupling correlation between outflow and inflow demand. Based on the outflow criterion, the Out-In Dynamic Coupling module (OI-DC) is constructed to simulate the influence of inflow on outflow.

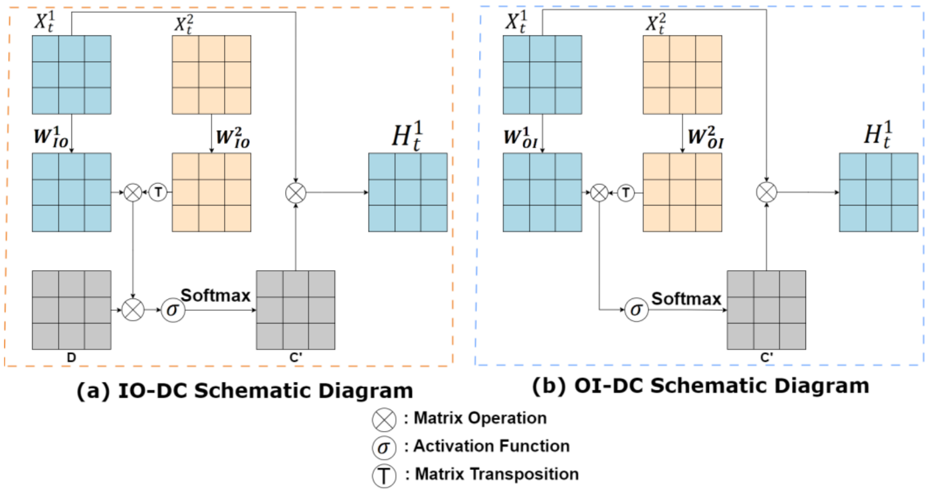

Fig. 3. Two variants of dynamic coupling, namely IO-DC and OI-DC

The structure of the IO-DC module is shown in Fig. 3 (a). This module has two input data sources: the feature $X_1 = (X_t^1, X_{t+1}^1, \dots, X_{t+T}^1) \in \mathbb{R}^{N^R \times F \times T}$ after spatial attention and the raw feature $X_2 = (X_t^2, X_{t+1}^2, \dots, X_{t+T}^2) \in \mathbb{R}^{N^R \times F \times T}$. X_1 represents inflow information, and X_2 represents outflow information. Based on the inflow criterion, the coupling operation is applied to the features of the T time slices. For example: $X_t^1 \in \mathbb{R}^{N^R \times F}$ represents the inflow feature at time $t+1$, and $X_t^2 \in \mathbb{R}^{N^R \times F}$ represents the outflow feature at time t. The operation process is shown in Fig. 3 (a), and its calculation formula is defined as follows:

$$C_{IO} = \sigma((X_t^1 W_{IO}^1)(X_t^2 W_{IO}^2)^\top D) \tag{7}$$

$$C'_{(i,j)} = \frac{\exp(C_{(i,j)})}{\sum_{j=1}^{N} \exp(C_{(i,j)})} \tag{8}$$

$$H_t^1 = C'X_t^1 \tag{9}$$

where $W_{IO}^1 \in \mathbb{R}^{F\times 1}$, $W_{IO}^2 \in \mathbb{R}^{F\times 1}$ are learnable parameters, $D \in \mathbb{R}^{N^R\times N^R}$ is the normalized matrix of the historical inflow and outflow data between nodes, and $D_{(i,j)}$ represents the probability of vehicles flowing from node i to node j. According to Eq. (8), the correlation matrix C is obtained, and it is normalized to obtain a normalized matrix $C' \in \mathbb{R}^{N^R\times N^R}$. The larger the value of $C'_{(i,j)}$, the stronger the correlation between nodes i and j. The correlation coefficient matrix C' is used for matrix operations with X_t^1, obtaining the inflow feature $H_t^1 \in \mathbb{R}^{N^R\times F}$ after the coupling impact at the current time step. The features of the T time slices are passed to the coupling module, producing T outputs, which are then concatenated to obtain the final inflow feature $H_{in} \in \mathbb{R}^{N^R\times F\times T}$. The formula is as follows:

$$H_{\text{in}} = \text{Concat}(H_t^1, H_{t+1}^1, \dots, H_{t+T}^1) \tag{10}$$

As shown in Fig. 3 (b), the OI-DC structure has two inputs: X_1 and X_2. X_1 provides outflow information for time $t+1, t+2, \dots, t+T+1$, and X_2 provides inflow demand information for time $t, t+1, \dots, t+T$. The formula for calculating the coupling correlation coefficient is similar to Eq. (7):

$$C_{OI} = \sigma((X_t^1 W_{OI}^1)(X_t^2 W_{OI}^2)^\top) \tag{11}$$

as in Eq. (8), (9), and (10), the weight coefficient matrix is normalized, and the outflow demand data is used in matrix operations with the weight coefficient matrix to obtain the outflow demand feature. Multiple time slice data are then concatenated to obtain the final outflow feature $H_{out} \in \mathbb{R}^{N^R\times F\times T}$. Finally, H_{in} and H_{out} are concatenated to obtain the final feature.

4.2 Temporal Attention

In bicycle-sharing systems, demand data across different time slices exhibit temporal correlations, and this correlation has fluctuating dynamics. The temporal attention mechanism can adaptively assign different weights to the data, thereby more accurately reflecting temporal dependencies. Therefore, this model uses the temporal attention mechanism to learn the dynamic influence relationships in the temporal dimension. The temporal attention calculation formula is as follows:

$$E = V_t \cdot \sigma(((X)^\top U_1) U_2 (U_3 X)^\top + b_t) \tag{12}$$

$$E'_{(i,j)} = \frac{\exp(E_{(i,j)})}{\sum_{j=1}^{N} \exp(E_{(i,j)})} \tag{13}$$

$$Tatt(X) = E'X \tag{14}$$

where $V_t, b_t \in \mathbb{R}^{T\times T}$, $U_1 \in \mathbb{R}^{N^R}$, $U_2 \in \mathbb{R}^{2F\times N^R}$ and $U_3 \in \mathbb{R}^{2F}$ are learnable parameters, and X represents the input to the temporal attention mechanism. Matrix operations are performed on the input data to obtain the temporal attention weight matrix E, and the softmax function is applied to normalize the weight matrix E, resulting in the normalized weight matrix, where $E'_{(i,j)}$ indicates the strength of the correlation between node i and node j.

4.3 Hierarchical Dynamic Interaction Module

The above structure extracts the spatiotemporal features of stations $F^S \in \mathbb{R}^{N\times 2F\times T}$ and the spatiotemporal features of regions $F^R \in \mathbb{R}^{N^R\times 2F\times T}$, but the station sparsity problem has not yet been addressed. Interacting station features with region features can enhance the expression of station features and solve the data sparsity problem. Therefore, this paper constructs a dynamic interaction module, which involves two steps: feature mapping and feature fusion. Feature mapping refers to transforming the coarse-grained region data into fine-grained station data with consistent dimensions. Feature fusion refers to performing a weighted aggregation of the features from both granularity levels.

The feature mapping step is as follows: First, a transfer matrix $Tran \in \mathbb{R}^{N\times N^R}$ is defined to represent the correspondence between stations and regions. The formula is expressed as:

$$Tran_{i,j} = \begin{cases} 1, & \text{if the node } i \text{ belongs to the region } j \\ 0, & \text{else} \end{cases} \tag{15}$$

Region features are the sum of all station features within the region. For example, if a region contains five stations, the demand for the region should be $x = x_1 + x_2 + \cdots + x_5$. The contribution of each station generated in the region's demand is different and changes over time. To achieve dynamic feature mapping, we propose constructing a dynamic transfer matrix $Tran^d \in \mathbb{R}^{N\times N^R}$ to learn the contribution matrix of stations to region features dynamically. First, a contribution coefficient matrix $Tran^d$ is constructed using station and region features and normalized. Then, the transfer matrix Tran is used to adjust the correspondence between the positions of stations and regions in the matrix $Tran^d$, and finally, the region feature $F^R \in \mathbb{R}^{N^R\times 2F\times T}$ is mapped to $F^R_{tran} \in \mathbb{R}^{N\times 2F\times T}$. The formula is expressed as:

$$Tran^d = \sigma((F^S U_1)U_2(F^R U_3)^\top + b_d) \tag{16}$$

$$Tran^d = Tran^d - \text{mean}(Tran^d, \text{axis} = 0) \tag{17}$$

$$Tran^d = \sigma(Tran^d) * Tran \tag{18}$$

$$F^R_{tran} = (Tran^d)(F^R) \tag{19}$$

where $U_1 \in \mathbb{R}^T$,$U_2 \in \mathbb{R}^{2F\times 2F}$,$U_3 \in \mathbb{R}^T$ and $b_d \in \mathbb{R}^{N\times N^R}$ are learnable parameters.

The fusion of station features and region features $F_{\text{out}} \in \mathbb{R}^{N\times 2F\times T}$ can be defined as:

$$F_{\text{out}} = W_{f_1} \odot F^S + W_{f_2} \odot F^R_{\text{tran}} \tag{20}$$

where W_{f_1} and W_{f_2} are learnable parameters, and $\odot$ denotes the Hadamard product.

5 Experiments

5.1 Datasets

The experiments in this paper use two real-world shared bicycle demand datasets: the Citi Bike dataset from New York City and the Divvy dataset from Chicago. The preliminary information of the two datasets includes station IDs and times for bike rentals and returns, as well as the latitude and longitude of each station. The experiment in this paper selects shared bicycle data from September 1, 2014, to August 31, 2015, covering 171 shared bicycle stations in Manhattan, New York City. The Divvy shared bicycle dataset contains user bike rental records from January 1, 2017, to December 31, 2017, in Chicago, with 133 shared bicycle stations selected as experimental stations for this study. In the dataset configuration, 80% of the data is used as the training dataset, while the remaining 20% is used as the testing dataset.

5.2 Baseline Models

We compare the proposed model with the following 9 baseline models:

- **HA**: Predicts future inflow and outflow demand based on the average values of historical demand data.
- **LASSO**: A regression analysis method that performs feature selection and regularization simultaneously.
- **XGBOOST** [21]: An efficient gradient boosting tree algorithm used for regression and classification problems, known for its excellent predictive performance and interpretability.
- **RNN** [22]: A neural network used for learning continuous time-series data, capable of capturing temporal dependencies in data of any time sequence length.
- **GRU** [23]: A recurrent neural network model used for sequence data, featuring a simpler structure and stronger memory capacity. It resolves the gradient vanishing and exploding issues present in RNNs.
- **GAT**: A graph neural network model that uses attention mechanisms between nodes to effectively learn and represent the relationships between nodes and edges in graph data. It is widely applied in shared bicycle demand prediction and traffic flow prediction.
- **MH-GCN** [24]: A novel hierarchical spatiotemporal graph convolutional network for solving traffic flow prediction problems.
- **DSTH-GCN** [25]: A dynamic local graph and corresponding adaptive local graph convolution network that captures spatial dependencies in traffic flow data.
- **TS-STN** [26]: A new deep learning architecture for OD flow prediction in urban rail systems.

5.3 Overall Model Prediction Performance Analysis

We compare HST-GCN with other baseline models on the Citi Bike and Divvy datasets. Table 1 demonstrates the experimental results of each model for outflow and inflow demand. As Table 1 shows, HST-GCN achieves the best prediction results for both outflow and inflow demand under two evaluation metrics. Furthermore, the experimental results show that deep learning-based prediction models outperform traditional statistical learning and machine learning methods. MH-GCN and DSTH-GCN yield better results than RNN, GRU, and GAT models, confirming that hierarchical feature learning methods effectively improve prediction accuracy. This phenomenon occurs because hierarchical feature learning can address the data sparsity issue in shared bicycle demand data. TS-STN outperforms RNN, GRU, and GAT in both evaluation metrics, demonstrating that prediction performance can be significantly improved by fully considering the relationships between outflow and inflow demand.

Table 1. Performance comparison of different models on the SZ-Taxi and Los-Loop dataset

Datasets	Methods	Inflow		Outflow	
		MAE	RMSE	MAE	RMSE
Citi Bike	HA	3.84	6.47	3.62	6.03
	LASSO	3.42	6.08	3.27	5.61
	XGBOOST	3.29	5.58	3.09	5.15
	RNN	3.12	5.16	2.98	4.97
	GRU	2.97	4.83	2.64	4.51
	GAT	2.89	4.72	2.61	4.48
	MH-GCN	2.61	4.27	2.51	4.15
	DSTH-GCN	2.47	4.12	2.44	4.01
	TS-STN	2.51	4.19	2.45	4.08
	HST-GCN	**2.44**	**3.96**	**2.40**	**3.81**
Divvy	HA	1.17	2.14	1.22	2.29
	LASSO	1.07	2.01	1.13	2.07
	XGBOOST	1.03	1.92	1.06	1.94
	RNN	1.01	1.86	1.02	1.88
	GRU	0.94	1.69	0.96	1.71
	GAT	0.92	1.61	0.93	1.65
	MH-GCN	0.87	1.54	0.91	1.57
	DSTH-GCN	0.83	1.46	0.84	1.49
	TS-STN	0.84	1.49	0.85	1.53
	HST-GCN	**0.81**	**1.39**	**0.83**	**1.44**

Compared to MH-GCN, DSTH-GCN, and TS-STN, the HST-GCN model proposed in this paper thoroughly examines the impact of hierarchical features and simultaneously takes into account the coupling relationships between outflow and inflow demand at each level. The experimental results show that the HST-GCN model outperforms MH-GCN, DSTH-GCN, and TS-STN in predicting the outflow and inflow demand. This verifies the effectiveness of the experimental framework proposed in this paper. The HST-GCN model models and learns features from different levels, enabling it to more comprehensively capture the spatial and temporal characteristics of shared bicycle demand. Additionally, considering the coupling relationships between outflow and inflow demand helps the model make more accurate predictions. The experimental results demonstrate that the HST-GCN model performs better in predicting outflow and inflow demand than other models. This proves the superiority and effectiveness of the HST-GCN model in shared bicycle demand prediction tasks.

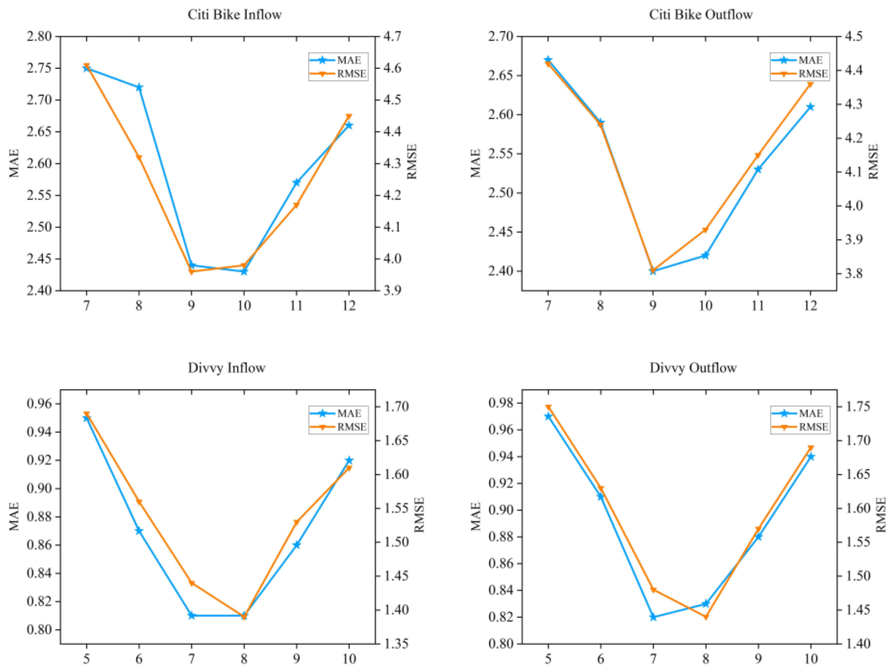

Fig. 4. Comparison of MAE and RMSE for different region numbers on the Citi Bike and Divvy datasets

5.4 Region Quantity Analysis Experiment

The number of regions m is a critical parameter in the experiment, as its variation significantly impacts the prediction results. To assess this effect, we set 6 different values of m for both datasets. For the Citi Bike dataset, the number of regions m is set from 7 to 12; for the Divvy dataset, m ranges from 5 to 10. Fig. 4 shows the experimental results for different m values in both datasets. As the figure shows, as the number of regions increases, MAE and RMSE gradually decrease. However, when the number of regions increases to a certain level, MAE and RMSE rise. The paper suggests that this happens because too few or too many regions can lead to a decline in the model's prediction performance. Only by setting m within a reasonable range can the prediction accuracy be optimized.

The figure shows that the predictions for outflow and inflow demand are relatively good for the Citi Bike dataset when m is 9 or 10. Considering the MAE and RMSE of both outflow and inflow demands from the two experimental groups, the region number m for the Citi Bike dataset is set to 9. For the Divvy dataset, the prediction results for m=7 and m=8 are better. After considering the overall results, the region number m for the Divvy dataset is set to 8.

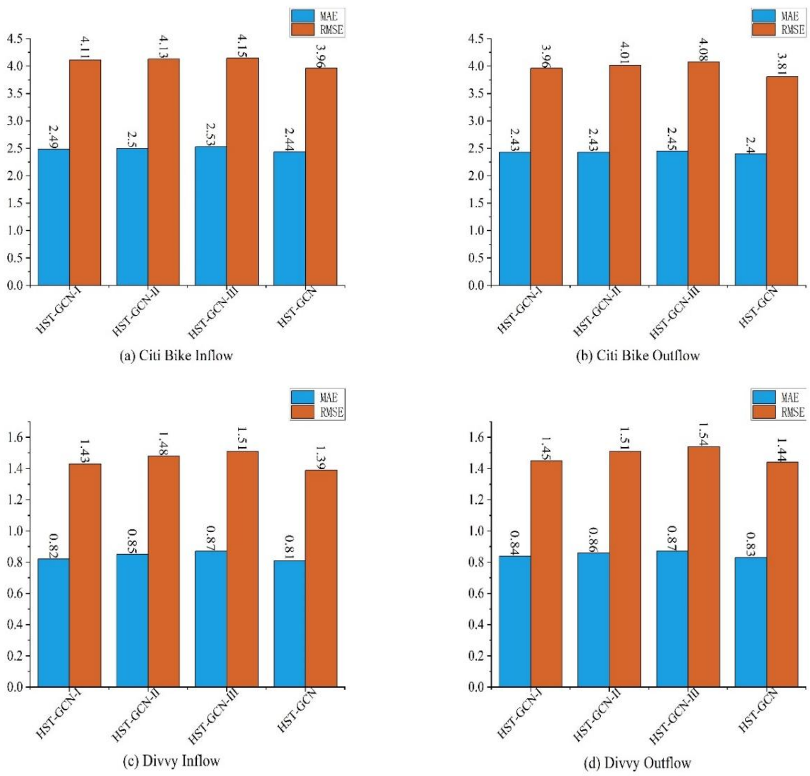

Fig. 5. The results of ablation experiments on the Citi Bike and Divvy dataset

5.5 Ablation Experiment

To further examine the effectiveness of the model proposed in this paper, this section will dismantle each HST-GCN module and perform ablation experiments on the variant models using the Citi Bike and Divvy datasets. The relevant variants of the model are as follows:

– **HST-GCN-I**: This variant is designed to evaluate the effect of the regional graph convolution on prediction accuracy. It eliminates the regional graph convolution from HST-GCN, while preserving the integrity of all other components.

– **HST-GCN-II**: This variant investigates the role of the dynamic transfer block in prediction performance. The dynamic transfer block in the hierarchical dynamic interaction layer is substituted with a static transfer matrix, leaving the remaining architecture unaltered.

– **HST-GCN-III**: This model removes the dynamic coupling module in HST-GCN while keeping the other structures unchanged.

Fig. 5 shows the results of the ablation experiments. As seen from the figure, when the regional graph convolution is removed, both MAE and RMSE of HST-GCN-I decrease to varying degrees, indicating that the regional graph convolution helps improve the prediction results for shared bicycle demand. When the Tran matrix replaces the dynamic transfer block, the results of HST-GCN-II show a declining trend, which suggests that the dynamic transfer block is effective. The dynamic coupling module is also an important component of the model. The experimental results show that when the dynamic coupling module is removed, the performance of HST-GCN-III worsens, proving the effectiveness of the dynamic coupling module. Based on the above comprehensive analysis, removing any of the modules in this chapter weakens the experimental results, indicating that each module plays an indispensable role in the overall framework. When all modules are present, we observe that all evaluation metrics achieve optimal levels on both datasets. The results of the ablation experiment confirm the effectiveness of the modules proposed in this chapter.

6 Conclusion and Future Work

This paper proposes a Hierarchical Spatiotemporal Graph Convolutional Network (HST-GCN) model for shared bicycle demand prediction. The model takes station and region-level demand data as input and jointly employs hierarchical spatiotemporal feature learning and dynamic coupling learning to better address fine-grained shared bicycle inflow and outflow demand prediction tasks. Relevant experiments were conducted on the Citi Bike dataset from New York and the Divvy dataset from Chicago. The experimental results show that the prediction results of HST-GCN outperform the baseline models, validating the effectiveness of the HST-GCN model. Future work includes: (1) Considering more factors that influence shared bicycle travel, such as unexpected events, to predict bicycle demand under extreme conditions; (2) Since user behavior is random, incorporating user-related information, such as social data, to analyze individual user cycling habits and further improve the accuracy of prediction results.

Acknowledgments. This work is supported by the Doctoral Scientific Research Foundation of Hunan University of Science and Technology (E51890).

References

1. Singhvi, D., Singhvi, S., Frazier, P.I., Henderson, S.G., O’Mahony, E., Shmoys, D.B., Woodard, D.B.: Predicting bike usage for new york city’s bike sharing system. In: Workshops at the twenty-ninth AAAI conference on artificial intelligence (2015)
2. Pan, Y., Zheng, R.C., Zhang, J., Yao, X.: Predicting bike sharing demand using recurrent neural networks. Procedia computer science **147**, 562–566 (2019)

3. Yang, G., Li, H.: Prediction method of shared bicycle traffic based on prophet-bilstm combined model. In: Proceedings of the 2022 10th International Conference on Information Technology: IoT and Smart City. pp. 251–256 (2022)
4. Qiao, S., Han, N., Huang, J., Yue, K., Mao, R., Shu, H., He, Q., Wu, X.: A dynamic convolutional neural network based shared-bike demand forecasting model. ACM Transactions on Intelligent Systems and Technology (TIST) **12**(6), 1–24 (2021)
5. Liang, Y., Ding, F., Huang, G., Zhao, Z.: Deep trip generation with graph neural networks for bike sharing system expansion. Transportation Research Part C: Emerging Technologies **154**, 104241 (2023)
6. Qin, T., Liu, T., Wu, H., Tong, W., Zhao, S.: Resgcn: Residual graph convolutional network based free dock prediction in bike sharing system. In: 2020 21st IEEE International Conference on Mobile Data Management (MDM). pp. 210–217. IEEE (2020)
7. Zhang, J., Zheng, Y., Qi, D., Li, R., Yi, X.: Dnn-based prediction model for spatio-temporal data. In: Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems. pp. 1–4 (2016)
8. Yu, L., Feng, T., Li, T., Cheng, L.: Demand prediction and optimal allocation of shared bikes around urban rail transit stations. Urban Rail Transit **9**(1), 57–71 (2023)
9. Ai, Y., Li, Z., Gan, M., Zhang, Y., Yu, D., Chen, W., Ju, Y.: A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system. Neural Computing and Applications **31**, 1665–1677 (2019)
10. Li, X., Xu, Y., Zhang, X., Shi, W., Yue, Y., Li, Q.: Improving short-term bike sharing demand forecast through an irregular convolutional neural network. Trans- portation research part C: emerging technologies **147**, 103984 (2023)
11. Tang, J., Liang, J., Liu, F., Hao, J., Wang, Y.: Multi-community passenger demand prediction at region level based on spatio-temporal graph convolutional network. Transportation Research Part C: Emerging Technologies **124**, 102951 (2021)
12. Yoon, J.W., Pinelli, F., Calabrese, F.: Cityride: a predictive bike sharing journey advisor. In: 2012 IEEE 13th international conference on mobile data management. pp. 306 – 311. IEEE (2012)
13. Singhvi, D., Singhvi, S., Frazier, P.I., Henderson, S.G., O’Mahony, E., Shmoys, D.B., Woodard, D.B.: Predicting bike usage for new york city’s bike sharing system. In: Workshops at the twenty-ninth AAAI conference on artificial intelligence (2015)
14. Ashqar, H.I., Elhenawy, M., Almannaa, M.H., Ghanem, A., Rakha, H.A., House, L.: Modeling bike availability in a bike-sharing system using machine learning. In: 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS). pp. 374–378. IEEE (2017)
15. Feng, S., Chen, H., Du, C., Li, J., Jing, N.: A hierarchical demand prediction method with station clustering for bike sharing system. In: 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC). pp. 829–836. IEEE (2018)
16. Li, X., Xu, Y., Chen, Q., Wang, L., Zhang, X., Shi, W.: Short-term forecast of bicycle usage in bike sharing systems: a spatial-temporal memory network. IEEE Transactions on Intelligent Transportation Systems **23**(8), 10923–10934 (2021)
17. Yin, Z., Hardaway, K., Feng, Y., Kou, Z., Cai, H.: Understanding the demand predictability of bike share systems: A station-level analysis. Frontiers of Engineering Management **10**(4), 551–565 (2023)
18. Lin, L., He, Z., Peeta, S.: Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. Transportation Research Part C: Emerging Technologies **97**, 258–276 (2018)

19. Liang, Y., Huang, G., Zhao, Z.: Cross-mode knowledge adaptation for bike sharing demand prediction using domain-adversarial graph neural networks. IEEE Trans- actions on Intelligent Transportation Systems (2023)
20. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
21. Hulot, P., Aloise, D., Jena, S.D.: Towards station-level demand prediction for effective rebalancing in bike-sharing systems. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 378– 386 (2018)
22. Pan, Y., Zheng, R.C., Zhang, J., Yao, X.: Predicting bike sharing demand using recurrent neural networks. Procedia computer science **147**, 562–566 (2019)
23. Wang, B., Kim, I.: Short-term prediction for bike-sharing service using machine learning. Transportation research procedia **34**, 171–178 (2018)
24. Li, Z., Ren, Q., Chen, L., Sui, X., Li, J.: Multi-hierarchical spatial-temporal graph convolutional networks for traffic flow forecasting. In: 2022 26th International Conference on Pattern Recognition (ICPR). pp. 4913–4919. IEEE (2022)
25. Li, H., Jin, D., Li, X., Qiao, S.: A dynamic heterogeneous graph convolution network for traffic flow prediction. The Computer Journal **67**(1), 31–44 (2024)
26. Jiang, W., Ma, Z., Koutsopoulos, H.N.: Deep learning for short-term origin–destination passenger flow prediction under partial observability in urban railway systems. Neural Computing and Applications pp. 1–18 (2022)