Transfer Learning for Cross-Modal Demand Prediction of Bike-Share and Public Transit

Mingzhuang Hua [a, *], Francisco Camara Pereira [b], Yu Jiang [b, c], Xuewu Chen [d], and Junyi Chen [d]

[a] *College of General Aviation and Flight, Nanjing University of Aeronautics and Astronautics*

[b] *Department of Technology, Management and Economics, Technical University of Denmark*

[c] *Lancaster University Management School, Lancaster University*

[d] *School of Transportation, Southeast University*

* Corresponding author: Binhe East Road #29, 213300, Liyang, China
E-mail: huamingzhuang@nuaa.edu.cn

Mingzhuang Hua: Conceptualization, Writing-Original draft, Data curation. **Francisco Camara Pereira:** Methodology, Writing-Reviewing and Editing, Validation. **Yu Jiang:** Writing-Reviewing and Editing, Visualization, Project administration. **Xuewu Chen:** Supervision, Resources, Funding acquisition. **Junyi Chen:** Writing-Reviewing and Editing, Visualization.

Transfer Learning for Cross-Modal Demand Prediction of Bike-Share and Public Transit

Abstract: The urban transportation system is a combination of multiple transport modes, and the interdependencies across those modes exist. This means that the travel demand across different travel modes could be correlated as one mode may receive/create demand from/for another mode, not to mention natural correlations between different demand time series due to general demand flow patterns across the network. It is expected that cross-modal ripple effects will become more prevalent with Mobility as a Service. Therefore, by propagating demand data across modes, a better demand prediction could be obtained. To this end, this study explores various transfer learning strategies and machine learning models for cross-modal demand prediction. The trip data of bike-share, metro, and taxi are processed as the station-level passenger flows, and then the proposed prediction method is tested in the large-scale case studies of Nanjing and Chicago. The results suggest that prediction models with transfer learning perform better than unimodal prediction models. Fine-tuning without freezing strategy performs the best among all transfer learning strategies, and the split-brain strategy can handle the data missing problem. Furthermore, the 3-layer stacked Long Short-Term Memory model performs particularly well in cross-modal demand prediction. These results verify our deep transfer learning method's forecasting improvement over existing benchmarks and demonstrate the good transferability for cross-modal demand prediction in multiple cities.

Keywords: Transport demand, Cross-modal prediction, Transfer learning, Bike-share, Public transit

1. Introduction

Compared with only a few decades back, today's transportation system is much more large-scale, heterogeneous, and dynamic. It is large-scale because of having a large quantity of spatial objects (Lee et al., 2002) within each transport network, e.g., bike-share has thousands of stations. It is heterogeneous because a myriad of new services exists, such as car sharing, ride sharing, shared micro-mobility of bike-share and scooter-share, and Mobility as a Service (MaaS) (Wong et al., 2020) . Autonomous shuttles already operate in several places, and even existing traditional modes, such as metro, bus, or taxi, have their modernized versions, often with smartphone apps, increased electrification, and autonomy. The options for travellers are certainly more varied today than before. It is also dynamic because these new technologies allow for within-day (sometimes real-time) repositioning/control. For example, taxi and ride sharing companies often redistribute their fleets during the day; shared micro-mobility and car sharing companies often rebalance (Chen et al., 2018) at least during the night; and pricing of services can vary by time and zone.

Many cities around the world have operated these new mobility services that can provide users with convenient options, cost-saving benefits, and safe services (Hua et al., 2021) . As a typical service of micro-mobility, bike-share has been proven to mitigate traffic congestion (Fan and Zheng, 2020), improve health benefits (Babagoli et al., 2019), and protect the environment (A. Li et al., 2021). Furthermore, these novel mobility services are promising to solve the first or last-mile problems of public transit such as metro and bus (Yang et al., 2019). In such a dynamic system, the risk of supply-demand misalignment is intuitively greater than in a static system. Demand prediction is a prerequisite and basis for supply-demand rebalancing (Chen et al., 2021). Thus, having an accurate demand prediction is vital for efficient responsiveness to demand.

Most existing studies focus on separately predicting real-time demand, e.g., shared micro-mobility and public transit. Thus, even though users can interchange between various transport modes, operators have difficulty in providing collaborative operation of new mobility services and public transit. The key challenge of the collaboration is the lack of demand information exchange of multiple transport modes (Cleophas et al., 2019) . For example, bike-share companies cannot know and then predict metro passenger flow, and metro companies also cannot know and then predict bike-share passenger flow. In the absence of demand information from other transport modes, it is virtually impossible to provide collaborative service and build the MaaS platform. Therefore, it is essential but very hard to address the large-scale network forecasting problem with thousands of spatial objects. This study is dedicated to filling this research gap, particularly considering cross-modal predictions between bike-share, metro, and taxi.

The concept of cross-modal predictions demands clarification: it refers to predictions of a certain transport mode (target) that use information from another transport mode (source). This includes situations where there is data missing for the target transport mode but rich data for the source transport mode (e.g., using metro data to predict bike-share demand soon after it is introduced in a city). It also includes situations of jointly predicting two or more transport modes given the aggregation of their datasets in a combined model. Therefore, cross-modal prediction is ultimately about data fusion across traditional and new transport modes, taking advantage of inter-modal correlations to enhance predictability and data quality.

Of interest is also the concept of Transfer Learning, which is the improvement of learning in a new task through the transfer of knowledge from a related task that has already been learned (Olivas et al., 2009) . Transfer learning has the potential to resolve the above two problems because it can pass the forecasting knowledge into new mobility services such as MaaS from other traditional transport modes, which may have been in operation for many years. There is a research gap in cross-modal demand prediction among multiple transport modes. In the existing studies of transport demand prediction, transfer learning has gained initial applications, enabling knowledge transfer across time and space. However, these studies (Wang et al., 2019; J. Li et al., 2021) focus only on a single transport mode, missing the opportunity for knowledge transfer between different transport modes.

In order to build a transfer learning model among different transport modes, we need first to determine the input-output frames. There are two input-output frames: single-station-input single-station-output (SISO), and multiple-station-input multiple-station-output (MIMO). The SISO frame is popular in demand prediction, and existing MIMO papers are limited. However, the SISO framework has two insurmountable drawbacks. The first drawback is that SISO consumes too much computing time in large-scale cases. The actual operation should consider hundreds or thousands of stations (or other spatial objects) in the transport modes and predicting the station demand one by one would cost too much time. The second drawback is that SISO is not flexible enough to accommodate cases of different sizes. Therefore, in this study, we establish transfer learning methods that can adapt to different input amounts and adopt the MIMO frame, which is more suitable for the cross-modal prediction problem.

There is a trend of collaborative operation for multimodal mobility services in urban transportation, so cross-modal demand prediction is very important in this trend. Besides, demand prediction studies face difficulty in forecasting new services and often suffer from the problem of data missing (Nie et al., 2022). Therefore, this paper proposes a cross-modal forecasting framework that incorporates deep learning and transfer learning based on trip data of shared micro-mobility (bike-share) and public transit (metro and taxi). The main contributions of this paper are threefold: a) The cross-modal method is applied to get better prediction results of multimodal transport demands, overcoming the key challenge of collaboration between multiple transport modes. b) A transfer learning framework is established aiming at the data missing problem by passing the forecasting knowledge from traditional modes to new mobility services. c) The proposed MIMO models address the difficulty of large-scale network prediction, as demonstrated through two case studies conducted in Nanjing and Chicago. This study, to the best of our knowledge, is the first paper to output large-scale cross-modal transport prediction results with the transfer learning approach.

The rest of this study is organized as follows. Section 2 reviews the related works of machine learning prediction and transfer learning methods. Section 3 elaborates on machine learning models and transfer learning strategies in the cross-modal prediction framework. Section 4 describes the data sources and presents the results of the developed method. Finally, Section 5 discusses and concludes this study.

2. Literature review

This section reviews existing studies on transport prediction and focuses on two categories, namely, transfer learning application and machine learning prediction, as they can be applied to solve the challenging problems in transport forecasting: data missing problem and the inability problem to cope with new mobility services.

2.1. Transfer learning in related studies

Transfer learning applies the knowledge of one domain to another related domain, which can provide better performance. It has been widely applied in image processing (Guo et al., 2023), classification (Liu et al., 2022), and prediction (Fawaz et al., 2018; Zhang et al., 2019).

But transfer learning does not get enough attention in transport demand predicting. The existing transfer learning research for transport prediction primarily focuses on road traffic forecasting (Wang et al., 2018; Li et al., 2022) and mostly is limited to single-mode prediction.

Transfer learning is a promising method in cross-modal transport prediction and solving data missing or lacking problems. Many studies have demonstrated the reliability and effectiveness of transfer learning in addressing insufficient data (Coston et al., 2019; Ma et al., 2020). Besides, transfer learning can be used to transfer forecasting knowledge not only between different areas on the same transport mode but also between different transportation modes in the same area. If the forecasting knowledge is transferred from a small-scale mode to a large-scale mode, transfer learning could also save the running time of large-scale demand prediction. Therefore, transfer learning could be a tool for predicting newly-operated micro-mobility services, especially bike-share and e-scooter sharing, using available public transit data. This is explored in this study.

2.2. Machine learning for transport demand forecasting

In general, there are two types of machine learning models for transport demand forecasting: single-mode demand model and cross-modal demand model. A single-mode demand model is for a selected transport mode, and a cross-modal demand model (also considered as a part of the cross-modal demand modal) is for multiple transport modes.

There are many papers discussing single-mode demand models, especially for bike-share and public transit. In bike-share demand prediction, novel machine learning methods (Li and Zheng, 2020) and deep learning methods, especially deep neural networks (Cho et al., 2021; Chai et al., 2018) have been widely used. In public transit prediction, there are also many other studies applying decision trees and neural networks (J. Zhang et al., 2020; C. Li et al., 2021). However, the single-mode demand model is based on sufficient data from one transport mode and therefore cannot integrate the data from other transport modes and perform data imputation both for better predictions and for compensating for insufficient/missing data.

The collaborative service of urban transportation is based on demand prediction in multiple transport modes (Pan et al., 2019). For example, Li et al. (2021) first considered users' preferred time windows in developing a user-friendly demand-responsive transit system. Yang et al. (2023) conducted short-term passenger flow prediction for multi-traffic modes. Among the relevant studies, bike-share is a popular type of shared transport, and its correlation with public transit is significant (Chu et al., 2020). Bike-share data can provide broader spatial information for public transit prediction, and, in turn, public transit data can provide different demand features for bike share prediction. Hence, it is necessary to develop a cross-modal demand prediction methodology, which integrates multiple transport modes into one general prediction framework.

Despite the necessity in practical operation, cross-modal demand prediction, or joint demand prediction in multiple transport modes, has not received sufficient attention in existing studies. Only a few papers (Toman et al., 2020; Wang et al., 2022; Liang et al., 2021; Li et al., 2023) discuss joint demand prediction of multimodal urban transport. Yet these joint prediction studies only discuss a small-scale problem or do not have a suitable spatial object, which is not practical in large-scale transport operations. Besides, these existing studies all use New York City as the case study, which ignores the urban transport demand in developing countries. Urban transport, especially MaaS, should consider the large-scale network with hundreds of stations or even thousands of stations. Therefore, a large-scale method for joint demand prediction is required, which is also the focus of our paper.

2.3. Summary

In short, existing studies of demand prediction for multiple transport modes have research gaps in three aspects: data missing problems, large-scale network applications, and cross-modal demand forecasting. This study aims to fill these three research gaps by proposing a transfer learning approach with a MIMO framework. Transport demand forecasting faces the challenge of large-scale networks (Lin et al., 2018), and cross-modal demand forecasting makes this problem even more complicated. Both shared micro-mobility and public transit have hundreds or even thousands of stations. The demands for these stations need to be predicted and output at the same time. As a prerequisite for cross-modal demand forecasting, the input-output framework of MIMO (Hua et al., 2020) needs to be given full attention. But most studies focus on SISO prediction, and the research related to MIMO forecasting is relatively insufficient. Therefore, the novel MIMO and cross-modal models need to be established by combining transfer learning and deep learning.

3. Methodology

In this study, for the cross-modal demand prediction, machine learning and transfer learning are combined for the forecasting knowledge transfer between bike-share and public transit. A framework of machine-learning-first transfer-learning-second is adopted for this cross-modal prediction problem. The process of transferring bike-share knowledge to public transit prediction is as follows. Firstly, we use machine learning to predict the dynamic demand of public transit;

Secondly, we apply different transfer learning strategies to transfer the forecasting knowledge of public transit; Thirdly, we build the machine learning models with transferred knowledge to predict the large-scale dynamic demand of bike-share. Using a similar process, the knowledge of bike-share is also transferred to the machine learning prediction of public transit. In what follows, Section 3.1 describes the transfer learning framework; Section 3.2 presents the proposed deep learning model; and, finally, Section 3.3 introduces prediction model benchmarks and performance index.

3.1. Transfer learning framework for joint prediction

For the cross-modal demand prediction of bike-share and public transit, a general framework has been established among the various cities and scenarios. The architecture of transfer learning for cross-modal demand prediction is shown in Fig. 1. The detailed stages of cross-modal prediction are as follows. Firstly, the forecast models without transfer of bike-share and public transit are established separately. Secondly, transferring public transit knowledge to bike-share prediction and transferring bike-share knowledge to public transit prediction is to build forecast models with transfer. Lastly, compare the results of the forecast models without and with transfer.

There are four transfer strategies: feature extraction (does not change model), fine-tuning without freezing transferred layers (FT), fine-tuning with freezing transferred layers (FTF), and split-brain (SB). Because the input/output of the origin and target domains are different, the feature extraction strategy cannot be applied in the cross-modal demand prediction. Therefore, three strategies FT, FTF, and SB are applied in this paper.

Firstly, the details of the FTF strategy are as follows. Because the input and output amounts of bike-share and public transit are different, the input and output layers of the prediction model should be tuned and cannot be frozen. The hidden layers are the transferred layers in transfer learning, which can be tuned or frozen. If the weights of the hidden layers from the related task are frozen in the target task, it is the transfer learning strategy of fine-tuning with freezing transferred layers. Secondly, FT is different from the FTF strategy. If the weights of the hidden layers from the related task are tuned in the target task, it is the transfer learning strategy of fine-tuning without freezing transferred layers. Thirdly, the SB strategy is a novel transfer learning type with two sub-tasks. The split-brain strategy splits the prediction task into two disjoint sub-tasks and predicts the output of one subset with the model of another subset. In the split-brain prediction study of Zhang et al. (2017) one sub-task predicts depth from images, while the other predicts images from depth. The structures of these three transfer learning strategies are shown in Fig. 2.

3.2. Stacked LSTM model for demand forecasting

In this study, the stacked LSTM method is selected as the deep learning prediction model. LSTM is an elegant type among many RNN models, which has been widely used in transport prediction (Liu et al., 2019; Ma et al., 2019; Hao et al., 2019; Petersen et al., 2019). The RNN method has the feature of network delay recursion, which could grasp the patterns of dynamic systems. LSTM improves the basic RNN model with internal mechanisms called gates and the memory cell. The LSTM gates consist of the forget gate, input gate, and output gate. This LSTM model can solve the problem of vanishing or exploding gradients and better deal with short-term memory conditions.

Neural network depth is generally attributed to the success of many challenging predictions. In particular, stacked Long Short-Term Memory (Stacked LSTM) is defined as an LSTM model comprised of multiple LSTM layers. Stacked LSTM can increase the depth of LSTM neural networks and has a better performance of prediction tasks. Multiple LSTM layers make the model deeper and could capture the high-dimensional non-linear patterns of transport demand. Meanwhile, various machine learning methods have also been used to compare with the stacked LSTM model and the corresponding transfer learning strategies, whose details are introduced in Section 3.3.

The architecture of our stacked LSTM model is shown in Fig. 3. In a stacked LSTM model, there are three types of layers: input layer, hidden layer, and output layer. In the input layer, the spatiotemporal flow data (the actual demand amounts in the former intervals of all stations) are used as the input matrix. As for the hidden layer, several LSTM layers are stacked and fused into the prediction model, and each LSTM layer has many units. In the output layer (fully connected, FC), the predicted demand amounts in the future interval of all stations are the output results.

3.3. Baselines and metrics

Five benchmarks are used to compare the forecasting performance of the proposed method, and the brief introduction of these baselines is as follows:

(1) One step

It takes the observed value at the former interval as the predicted value at the next interval.

(2) Historical average (HA)

The average observed value during the past weeks at the same period in time and the same station is calculated as the predicted value.

(3) Vector autoregression (VAR)

It uses a vector to generalize the autoregressive model and therefore is suitable for MIMO prediction. Its advantage is capturing the interdependencies of multiple time series.

(4) Random forest (RF)

It consists of multiple decision trees and outputs the average prediction of these trees. A decision tree is a tree-like model that learns the decision rules from root to leaf. The root is the entire sample, the branch is the feature conjunction, and the leaf is the target value.

(5) Graph Convolutional Network (GCN)

It is a type of CNN that can work directly on graphs, which means that filter parameters are typically shared over all locations in the graph. GCN combines GNN and CNN, which is good at learning graph representations and has achieved superior performance in many tasks.

Mean Absolute Error (MAE), root mean square error (RMSE), and R-square (R^2) are adopted as the evaluation metrics of the model performance. The calculation of MAE, RMSE, and R^2 are given in Equations (1), (2), and (3).

$$\mathrm{MAE} = \frac{1}{N \times T} \sum_{i=1}^{N} \sum_{t=1}^{T} \left| y_{i,t} - \hat{y}_{i,t} \right| \#(1)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N\times T}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t}-\hat{y}_{i,t}\right)^2} \#(2)$$

$$\mathrm{R}^2 = 1-\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t}-\hat{y}_{i,t}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t}-\bar{y}\right)^2} \#(3)$$

where N represents the number of stations, T represents the amount of prediction intervals, $y_{i,t}$ represents the observed flow in the i-th station in the t-th interval, $\bar{y}$ represents the mean value of $y_{i,t}$, and $\hat{y}_{i,t}$ represents the predicted flow in the i-th station in the t-th interval.

4. Results

4.1. Dataset description

The trip data used in this study are from Nanjing City and Chicago City. As shown in Table 1, these two cities both have large-scale multimodal transport services with more than 1,000 spatial objects. Specifically, a) in the case of Nanjing, China, there are 1.475 stations in bike-share mode and 160 stations in metro mode; b) in the case of Chicago, US, there are 50 stations in bike-share mode and 511 zones in taxi mode. The two case studies ensure that the methods and findings in this paper are not exclusive to one specific scenario, but rather more widely appropriate. For the Nanjing case, trip data in March 2019 of metro and bike-share are provided by the Nanjing Metro Group and Nanjing Public Bicycle Company separately. For the Chicago case, trip data in March 2019 of taxi and bike-share are obtained from the open data website. Among them, in March 2019, metro trips (37.9 million) were about fourteen times bike-share trips (2.8 million) in Nanjing, and taxi trips (1.5 million) were about ten times bike-share trips (165.6 thousand) in Chicago. The fields of all four trip datasets contain departure time and position, end time, and position.

Considering the difference in spatial attributes of various transport modes, taxi zone centroids, bike-share stations, and metro stations are selected as the spatial objects in this study. Fig. 4 shows the spatial distributions and temporal characteristics of Nanjing and Chicago transport. As for Nanjing, some bike-share stations are near to metro, but others are far from the metro stations. In Chicago, taxi and bike-share services have a similar spatial distribution. For taxi analysis, Chicago city is divided into five hundred small zones.

Hourly passenger flows of the typical stations at Nanjing and Chicago are also displayed in Fig. 4 (c) and (d). In Nanjing, the morning peak is 9 am, and the evening peak is 3 to 7 pm. The evening peak of bike mode is a little earlier than the evening peak of metro mode. The reason for this small difference may be that the major users of Nanjing Public Bicycle are elderly people who would go grocery shopping from 3 to 4 pm. But metro users mainly travel for commuting, so the metro peak is 7 pm. In Chicago, the morning peak is 9 to 10 am, and the evening peak is 5 to 6 pm. The bike peak is synchronized with the taxi peak.

The correlation matrix of the station-level demands for metro and bike-share in Nanjing is shown in Fig. 4 (e), and the correlation matrix of the station-level demands for taxi and bike-share in Chicago is shown in Fig. 4 (f). In these two subfigures, “b” for bike share, “m” for metro, and “t” stands for taxi. The correlation results of Nanjing are different from those in Chicago. In Nanjing, not only are there mostly positive correlations within the same transport mode, but the passenger flows of metro stations and bike-share stations are also basically positively correlated. It reflects the cooperative characteristic of metro and bike-share and the collaborative potential of MaaS. However, in Chicago, the passenger flows of a taxi centroid and other taxi centroids are positively correlated. The passenger flows of a bike-share station and other bike-share stations are also mostly positively correlated. But the passenger flows of some taxi centroids and some bike-share stations are negatively correlated. It reflects the competitive characteristics of taxi and bike-share.

4.2. Experiment settings

The models of this study are coded with TensorFlow in Python, and their target is to predict the inflow values at all stations in the next horizon. Before estimating model results, the experiment setting needs to be determined, which consists of the spatial object, time interval, training, and test datasets. Firstly, the spatial object in each case would be reduced. The reason for spatial object reduction is that some stations have very low passenger flows. For example, many hourly flow values of these stations are zero. If the average hourly flow is less than 0.1, this station would be removed from spatial objects. After spatial object reduction, the case studies still are large-scale networks. The Nanjing case has 1361 bike-share stations & 159 metro stations, and the Chicago case has 374 bike-share stations & 88 taxi centroids. Secondly, four-time intervals are chosen by trip data conversion. Trip data processing converts trip data into passenger flow data and is conducted in SQL Server Management Studio with three processing steps. a) Bike-share stations, metro stations, and taxi centroids are selected as spatial objects. b) Four forecast horizons or time intervals are set for the dynamic prediction, including 15 min, 30 min, 45 min, and 60 min. c) Passenger flows at each station in each horizon are calculated and output. Thirdly, this paper uses the first 70% of the data as the training dataset, the following

10% of the data as the validation dataset, and the last 20% as the test dataset in all machine learning prediction models.

Hyperparameters in the proposed models include input horizon amount, hidden layer amount, layer unit amount, training epoch, and batch size. To achieve a balance between underfitting and overfitting, hyperparameter tuning is based on the validation dataset. After hyperparameter tuning, four horizons are selected as the suitable input horizons, as shown in Fig. 5 (a). It means that the model input is the inflow values at all stations in the past four horizons. In the stacked LSTM model, the objective loss function is MAE, and the optimization algorithm uses Adam. As shown in Fig. 5 (b), the model performs the best when the value of batch size is set to 150. Generally, the model performance keeps improving as the training epoch increases. The model results become stable when the training epoch reaches 200, which is therefore set as the value of the training epoch. Besides, the hidden layer amount and unit amount of each layer are two important and related hyperparameters that build the basic architecture of the neural network. Grid search is used to optimize these two hyperparameters in this deep learning model: 3 stacked LSTM layers and 100 units in each hidden layer provide the best performance.

4.3. Prediction results

Table 2 summarizes the prediction results of different transfer learning strategies for Nanjing bike-share and metro. Table 3 summarizes the prediction results of different transfer learning strategies for Chicago bike-share and taxi. It can be found that fine-tuning without freezing transferred layers (FT) strategy has the best performance among all transfer learning strategies. Fine-tuning without freezing transferred layers (FTF) strategy performs slightly worse than FT strategy. This is because changes to the parameters of the transferred layers can further improve the prediction model for different tasks. Besides, the results of the split-brain (SB) strategy are not stable, which basically shows good performance in transferring public transit knowledge to bike-share prediction and sometimes has bad performance in transferring bike-share knowledge to public transit prediction. In summary, transfer learning greatly improves the prediction performance of large-scale transport networks.

The improvement effect of our proposed transfer learning method varies for different scenarios in cross-modal demand prediction. For all strategies, the prediction error MAE values of public transit are much bigger than bike-share. The possible reason could be that public transit has much more trips than bike-share in both Nanjing and Chicago. Meanwhile, the transfer learning method under the 30-minute forecast horizon does not perform as well as other forecast horizons. The possible reason is that the marginal distribution or conditional distribution of bike-share and public transit is more different in the 30-minute time interval.

Transfer learning has been proven to be effective in cross-modal demand prediction of bike-share and public transit. Bike-share and public transit have relations of causality and similarity. With regard to causality, the users of bike-share and public transit can transfer between these two modes of urban transport. With regard to similarity, the spatiotemporal characteristics of the two modes are similar, such as the same spatial agglomeration in the city center and the same temporal agglomeration in the morning and evening peaks. In short, the conditional distribution or marginal distribution of bike-share and public transit are similar, so the cross-modal prediction by transfer learning is effective. Therefore, these findings in this study suggest that both public transit and bike-share modes can get prediction benefits from the transferred knowledge of each other.

In order to deal with the data missing problem, the transfer learning method for cross-modal demand prediction has been proposed and tested in this study. There are two transfer learning solutions for data missing discussed in this paper. The first solution is using longer-term data of one transport mode to build a more reliable model and then predicting the demand of another mode. In this study, we also used three-month and six-month passenger flow data to build the corresponding models. The results of these two models did not improve significantly compared to the model based on one-month passenger flow data. Therefore, the solution of longer-period data is not applicable. The second solution is to directly use the SB strategy, with one mode of transport passenger flow as input and another mode of passenger flow as output. SB strategy is found to be useful in missing-data prediction. For Nanjing and Chicago bike-share, the prediction results of the SB strategy are relatively good. It shows that if the data of large-scale spatial objects (bike-share) are missing, the transport mode with small-scale spatial objects (public transit: taxi or metro) can be used as a substitute for predicting the transport mode with large-scale spatial objects. By using one transport mode with sufficient data for prediction modeling, the data missing problem of another transport mode can be solved.

The proposed LSTM-FT model in this study is compared with other benchmarks for the two-city cases. Table 4 shows the prediction results of different t different forecasting models for Nanjing bike-share and metro. Table 5 shows the prediction results of different forecasting models for Chicago bike-share and taxi. As shown in Tables 4 and 5, the proposed LSTM-FT model in our study shows pretty good performance in most short-term demand forecasting scenarios in Nanjing and Chicago. Besides, the LSTM-FT model also performs the best among all demand prediction models, in the passenger flow forecasting for the bike-share cases. The

LSTM-FT model is suitable for large-scale dynamic demand prediction with more than one thousand stations.

At the same time, the RF model is suitable for passenger flow prediction of public transit (taxi or metro). Meanwhile, one-step and HA perform worse than the other benchmark models, which indicates the traditional models are not good at predicting the dynamic flows and the nonlinear spatiotemporal demands. The VAR model performs very poorly in the dynamic demand prediction of bike-share but shows pretty good performance for the dynamic demand prediction of public transit. This finding reflects that the autoregressive time series model can effectively describe the passenger flow patterns of public transit, but it is not suitable for bike-share services with drastic changes in demand. The performance of the GCN model is not outstanding, which is not consistent with the existing research. The possible reason is that the GCN model needs to input more spatial and temporal data, such as land use and weather conditions, to build a better spatiotemporal GCN model.

5. Discussion and Conclusions

This study focuses on the cross-modal demand prediction for multiple transport modes, which plays a key part in promoting service cooperation, increasing transfer experience, and improving operators' dynamic operation efficiency. Transport demand prediction has a long-standing problem of data missing or lacking and cannot effectively adapt the demand information from other transport modes. To deal with these challenges, a combined framework of deep learning and transfer learning is proposed. Specifically, the stacked long short-term memory model with fine-tuning strategy is established for cross-modal demand prediction. For estimating the model performance, real-world case studies are conducted on bike-share and public transit services in

Nanjing and Chicago. Generally, this work provides insights into how to combine deep learning and transfer learning for multimodal demand prediction.

The key findings of this study are as follows: a) The cross-modal framework of deep transfer learning is effective for large-scale demand prediction; b) Valuable transferred knowledge could be gained through demand information of bike-share and public transit for predicting the passenger demand of each other; c) Among all transfer learning strategies, fine-tuning without freezing transferred layers strategy performs the best; d) The split-brain strategy is effective in handling the missing-data problem; e) The stacked LSTM model can be combined with a suitable transfer learning strategy for solving the cross-modal demand prediction problem. Besides, the spatiotemporal distribution and correlation of bike-share and public transit are discovered by the visualization analysis. It is found that bike-share and metro are mainly in the cooperation state, while bike-share and taxi have a certain degree of competitive relationship.

This study of cross-modal demand prediction can be applied in several aspects. Firstly, the transfer learning approach of this study enables demand forecasting when the data of a transport mode is missing. For example, the MaaS management platform may face data transmission failures of bike-share, and only the passenger flow data of public transport is available. In this case, the model of this study can be used to predict the future passenger flow of bike-share to guide the city-wide MaaS management. Moreover, cross-modal demand forecasting can guide the dynamic operation of multiple transportation modes. For example, the increase in metro passenger flow demand can result in dispatching more shared bikes into the vicinity of metro stations to meet users' travel and transfer needs. Besides, cross-modal demand forecasting can be applied to collaborative transportation services. The cross-modal demand forecast results can be used to infer the in-vehicle crowded information of public transit and the supply-demand balance of bike-share. Furthermore, the information can recommend users to make more reasonable travel choices to better use multimodal transportation services.

The proposed method of cross-modal demand prediction can be further improved or extended in the following directions. Firstly, many new mobility services of MaaS do not have fixed stations, which makes the large-scale problem for cross-modal demand prediction even more challenging. For example, ride- and e-scooter sharing have no physical stations and move freely around the city. These new mobility services lack a default spatial object such as the bike-share station in our study, making their demand prediction more difficult. Therefore, a new concept of the clustering-based virtual station can be applied as the spatial object for MaaS demand prediction. Secondly, our cross-modal forecasting models need to be compatible with other mobility service datasets, such as car sharing, station-less bike-share, and shared automated vehicles. These new mobility services have similarities and differences with bike-share in user travel patterns. Transferring the cross-modal model of this study to these new mobility services requires more research efforts. Lastly, the cross-modal forecasting model should consider incorporating multi-source data, such as land use and weather conditions. It could be helpful to assess the impact of inputting more information and build a spatiotemporal demand forecasting model.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Babagoli, M.A., Kaufman, T.K., Noyes, P., Sheffield, P.E., 2019. Exploring the health and spatial equity implications of the New York City Bike share system. J Transp Health 13, 200–209. https://doi.org/10.1016/j.jth.2019.04.003

Chai, D., Wang, L., Yang, Q., 2018. Bike Flow Prediction with Multi-Graph Convolutional Networks, in: Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL 18). New York, NY, USA, pp. 397–400. https://doi.org/10.1145/3274895.3274896

Chen, J., Yang, Z., Cheng, P., Shu, Y., 2021. Rebalancing Bike-Sharing System With Deep Sequential Learning. IEEE Intelligent Transportation Systems Magazine 13, 92–98. https://doi.org/10.1109/MITS.2019.2926252

Chen, Q., Liu, M., Liu, X., 2018. Bike Fleet Allocation Models for Repositioning in Bike-Sharing Systems. IEEE Intelligent Transportation Systems Magazine 10, 19–29. https://doi.org/10.1109/MITS.2017.2776129

Cho, J.-H., Ham, S.W., Kim, D.-K., 2021. Enhancing the Accuracy of Peak Hourly Demand in Bike-Sharing Systems using a Graph Convolutional Network with Public Transit Usage Data. Transp Res Rec 2675, 554–565. https://doi.org/10.1177/03611981211012003

Chu, J., Duan, Y., Yang, X., Wang, L., 2020. The Last Mile Matters: Impact of Dockless Bike Sharing on Subway Housing Price Premium. Manage Sci 67, 297–316. https://doi.org/10.1287/mnsc.2019.3550

Cleophas, C., Cottrill, C., Ehmke, J.F., Tierney, K., 2019. Collaborative urban transportation: Recent advances in theory and practice. Eur J Oper Res 273, 801–816. https://doi.org/10.1016/j.ejor.2018.04.037

Coston, A., Ramamurthy, K.N., Wei, D., Varshney, K.R., Speakman, S., Mustahsan, Z., Chakraborty, S., 2019. Fair Transfer Learning with Missing Protected Attributes, in: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES 19). New York, NY, USA, pp. 91–98. https://doi.org/10.1145/3306618.3314236

Ding, Z., Shao, M., Fu, Y., 2015. Missing Modality Transfer Learning via Latent Low-Rank Constraint. IEEE Transactions on Image Processing 24, 4322–4334. https://doi.org/10.1109/TIP.2015.2462023

Fan, Y., Zheng, S., 2020. Dockless bike sharing alleviates road congestion by complementing subway travel: Evidence from Beijing. Cities 107, 102895. https://doi.org/10.1016/j.cities.2020.102895

Fawaz, H.I., Forestier, G., Weber, J., Idoumghar, L., Muller, P., 2018. Transfer learning for time series classification, in: Proceedings of the 2018 IEEE International Conference on Big Data (Big Data). Seattle, WA, USA, pp. 1367–1376. https://doi.org/10.1109/BigData.2018.8621990

Guo, W., Fang, X., Jiang, L., Han, N., Teng, S., 2023. Low-rank constraint-based multiple projections learning for cross-domain classification. Knowl Based Syst 276, 110734. https://doi.org/10.1016/J.KNOSYS.2023.110734

Hao, S., Lee, D.-H., Zhao, D., 2019. Sequence to sequence learning with attention mechanism for short-term passenger flow prediction in large-scale metro system. Transp Res Part C Emerg Technol 107, 287–300. https://doi.org/10.1016/j.trc.2019.08.005

Hua, M., Chen, J., Chen, X., Gan, Z., Wang, P., Zhao, D., 2020. Forecasting usage and bike distribution of dockless bike-sharing using journey data. IET Intelligent Transport Systems. https://doi.org/10.1049/iet-its.2020.0305

Hua, M., Chen, X., Cheng, L., Chen, J., 2021. Should bike-sharing continue operating during the COVID-19 pandemic? Empirical findings from Nanjing, China. J Transp Health 23, 101264. https://doi.org/10.1016/j.jth.2021.101264

Kang, Z., Yang, B., Yang, S., Fang, X., Zhao, C., 2020. Online transfer learning with multiple source domains for multi-class classification. Knowl Based Syst 190, 105149. https://doi.org/10.1016/J.KNOSYS.2019.105149

Lee, T., Sheng, L., Bozkaya, T., Balkir, N.H., Özsoyoglu, Z.M., Özsoyoglu, G., 2002. Querying Multimedia Presentations Based on Content, in: Jeffay, K., Zhang, H.B.T.-R. in M.C. and N. (Eds.), The Morgan Kaufmann Series in Multimedia Information and Systems. Morgan Kaufmann, San Francisco, pp. 413–437. https://doi.org/https://doi.org/10.1016/B978-155860651-7/50122-4

Li, A., Gao, K., Zhao, P., Qu, X., Axhausen, K.W., 2021. High-resolution assessment of environmental benefits of dockless bike-sharing systems based on transaction data. J Clean Prod 296, 126423. https://doi.org/10.1016/j.jclepro.2021.126423

Li, C., Bai, L., Liu, W., Yao, L., Waller, S.T., 2021. A multi-task memory network with knowledge adaptation for multimodal demand forecasting. Transp Res Part C Emerg Technol 131, 103352. https://doi.org/https://doi.org/10.1016/j.trc.2021.103352

Li, J., Guo, F., Sivakumar, A., Dong, Y., Krishnan, R., 2021. Transferability improvement in short-term traffic prediction using stacked LSTM network. Transp Res Part C Emerg Technol 124, 102977. https://doi.org/10.1016/j.trc.2021.102977

Li, J., Xie, N., Zhang, K., Guo, F., Hu, S., Chen, X. (Michael), 2022. Network-scale traffic prediction via knowledge transfer and regional MFD analysis. Transp Res Part C Emerg Technol 141, 103719. https://doi.org/https://doi.org/10.1016/j.trc.2022.103719

Li, X., Liu, W., Qiao, J., Li, Y., Hu, J., 2023. An Enhanced Semi-Flexible Transit Service with Introducing Meeting Points. Netw Spat Econ 23, 487–527. https://doi.org/10.1007/S11067-022-09583-8/FIGURES/20

Li, X., Wang, T., Xu, W., Hu, J., 2021. A Novel Model for Designing a Demand- Responsive Connector (DRC) Transit System with Consideration of Users' Preferred Time Windows. IEEE Transactions on Intelligent Transportation Systems 22, 2442–2451. https://doi.org/10.1109/TITS.2020.3031060

Li, Y., Zheng, Y., 2020. Citywide Bike Usage Prediction in a Bike-Sharing System. IEEE Trans Knowl Data Eng 32, 1079–1091. https://doi.org/10.1109/TKDE.2019.2898831

Liang, Y., Huang, G., Zhao, Z., 2021. Joint Demand Prediction for Multimodal Systems: A Multi-task Multi-relational Spatiotemporal Graph Neural Network Approach [WWW Document]. https://doi.org/10.48550/arxiv.2112.08078

Liu, B., Liu, C., Xiao, Y., Liu, L., Li, W., Chen, X., 2022. AdaBoost-based transfer learning method for positive and unlabelled learning problem. Knowl Based Syst 241, 108162. https://doi.org/10.1016/J.KNOSYS.2022.108162

Liu, Y., Liu, Z., Jia, R., 2019. DeepPF: A deep learning based architecture for metro passenger flow prediction. Transp Res Part C Emerg Technol 101, 18–34. https://doi.org/10.1016/j.trc.2019.01.027

Ma, J., Cheng, J.C.P., Jiang, F., Chen, W., Wang, M., Zhai, C., 2020. A bi-directional missing data imputation scheme based on LSTM and transfer learning for building energy data. Energy Build 216, 109941. https://doi.org/10.1016/j.enbuild.2020.109941

Ma, X., Zhang, J., Du, B., Ding, C., Sun, L., 2019. Parallel Architecture of Convolutional Bi-Directional LSTM Neural Networks for Network-Wide Metro Ridership Prediction. IEEE Transactions on Intelligent Transportation Systems 20, 2278–2288. https://doi.org/10.1109/TITS.2018.2867042

Olivas, E.S., Guerrero, J.D.M., Martinez-Sober, M., Magdalena-Benedito, J.R., López, A.J.S., 2009. Handbook of research on machine learning applications and trends: Algorithms, methods, and techniques. IGI Global. https://doi.org/10.4018/978-1-60566-766-9

Pan, S., Trentesaux, D., Ballot, E., Huang, G.Q., 2019. Horizontal collaborative transport: survey of solutions and practical implementation issues. Int J Prod Res 57, 5340–5361. https://doi.org/10.1080/00207543.2019.1574040

Petersen, N.C., Rodrigues, F., Pereira, F.C., 2019. Multi-output Deep Learning for Bus Arrival Time Predictions. Transportation Research Procedia 41, 138–145. https://doi.org/10.1016/j.trpro.2019.09.025

Toman, P., Zhang, J., Ravishanker, N., Konduri, K.C., 2020. Dynamic predictive models for ridesourcing services in New York City using daily compositional data. Transp Res Part C Emerg Technol 121, 102833. https://doi.org/10.1016/j.trc.2020.102833

Wang, B., Yan, Z., Lu, J., Zhang, G., Li, T., 2018. Road traffic flow prediction using deep transfer learning, in: Proceedings of the 13th International FLINS Conference (FLINS 2018). Belfast, Northern Ireland, UK, pp. 331–338. https://doi.org/10.1142/9789813273238_0044

Wang, Q., Guo, B., Ouyang, Y., Cheng, L., Wang, L., Yu, Z., Liu, H., 2022. Learning Shared Mobility-Aware Knowledge for Multiple Urban Travel Demands. IEEE Internet Things J 9, 7025–7035. https://doi.org/10.1109/JIOT.2021.3115174

Wei, X., Liu, S., Xiang, Y., Duan, Z., Zhao, C., Lu, Y., 2020. Incremental learning based multi-domain adaptation for object detection. Knowl Based Syst 210, 106420. https://doi.org/10.1016/J.KNOSYS.2020.106420

Wong, Y.Z., Hensher, D.A., Mulley, C., 2020. Mobility as a service (MaaS): Charting a future context. Transp Res Part A Policy Pract 131, 5–19. https://doi.org/10.1016/j.tra.2019.09.030

Yang, Y., Heppenstall, A., Turner, A., Comber, A., 2019. A spatiotemporal and graph-based analysis of dockless bike sharing patterns to understand urban flows over the last mile. Comput Environ Urban Syst 77, 101361. https://doi.org/10.1016/j.compenvurbsys.2019.101361

Yang, Yongjie, Zhang, J., Yang, L., Yang, Yang, Li, X., Gao, Z., 2023. Short-term passenger flow prediction for multi-traffic modes: A Transformer and residual network based multi-task learning method. Inf Sci (N Y) 642, 119144. https://doi.org/10.1016/J.INS.2023.119144

Zhang, C., Zhang, H., Qiao, J., Yuan, D., Zhang, M., 2019. Deep Transfer Learning for Intelligent Cellular Traffic Prediction Based on Cross-Domain Big Data. IEEE Journal on Selected Areas in Communications 37, 1389–1401. https://doi.org/10.1109/JSAC.2019.2904363

Zhang, R., Isola, P., Efros, A.A., 2017. Split-Brain Autoencoders: Unsupervised Learning by Cross-Channel Prediction, in: Proceedings of the 2017 IEEE Conference on Computer Vision and

Pattern Recognition (CVPR). Honolulu, HI, USA, pp. 1058–1067.

https://doi.org/10.1109/CVPR.2017.76

Table 1. Trip data of Chicago and Nanjing cases.

Transport mode in the city	Nanjing		Chicago	
	Metro	Bike-share	Taxi	Bike-share
Travel amount	37,918,284	2,786,634	1,515,941	165,611
Vehicle amount	-	50,963	4,691	4,331
Station amount	160	1,475	511	590
Travel time (min)	29.7	18.4	13.9	16.8

Table 2. Transfer learning results in Nanjing.

a) Nanjing Bike-share

Period	LSTM-Base	FT		FTF		SB	
	MAE	MAE	Improving rate	MAE	Improving rate	MAE	Improving rate
15min	0.502	0.465	**7.4%**	0.473	5.7%	0.473	5.8%
30min	0.774	0.724	**6.5%**	0.729	5.8%	0.751	3.0%
45min	1.023	0.935	8.6%	0.932	**8.9%**	1.001	2.2%
60min	1.275	1.141	**10.5%**	1.149	9.9%	1.284	-0.7%

b) Nanjing Metro

Period	LSTM-Base	FT		FTF		SB	
	MAE	MAE	Improving rate	MAE	Improving rate	MAE	Improving rate
15min	13.391	12.231	**8.7%**	12.628	5.7%	17.625	-31.6%
30min	27.998	23.190	**17.2%**	23.885	14.7%	31.006	-10.7%
45min	51.102	38.636	**24.4%**	40.186	21.4%	43.509	14.9%
60min	74.500	57.725	**22.5%**	59.047	20.7%	57.780	22.4%

Table 3. Transfer learning results in Chicago.

a) Chicago Bike-share

Period	LSTM-Base	FT		FTF		SB	
	MAE	MAE	Improving rate	MAE	Improving rate	MAE	Improving rate
15min	0.206	0.161	**21.8%**	0.165	20.0%	0.168	18.4%
30min	0.322	0.300	**6.9%**	0.307	4.8%	0.318	1.2%
45min	0.474	0.431	**9.1%**	0.437	7.9%	0.459	3.2%
60min	0.613	0.555	**9.4%**	0.562	8.3%	0.590	3.8%

b) Chicago Taxi

Period	LSTM-Base	FT		FTF		SB	
	MAE	MAE	Improving rate	MAE	Improving rate	MAE	Improving rate
15min	2.497	1.623	**35.0%**	1.724	31.0%	3.424	-37.1%
30min	2.596	2.379	**8.3%**	2.460	5.2%	4.896	-88.6%
45min	4.120	3.311	**19.6%**	3.454	16.2%	6.788	-64.8%
60min	5.835	4.606	**21.1%**	4.622	20.8%	8.386	-43.7%

Table 4. Prediction Results of different models in Nanjing.

a) Nanjing Bike-share

Model	15min			30min			45min			60min		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
One Step	0.631	1.324	0.291	1.029	2.137	0.440	1.423	3.003	0.459	1.862	4.071	0.410
HA	0.546	1.086	0.523	0.829	1.582	0.693	1.055	1.988	0.763	1.255	2.358	0.802
VAR	1.067	1.807	-0.322	4.415	11.869	-16.271	1.921	3.220	0.378	1.933	3.248	0.625
RF	0.508	**0.949**	0.635	0.757	**1.389**	**0.763**	0.967	1.775	0.811	1.179	2.191	0.829
GCN	0.527	**0.949**	**0.636**	0.808	1.468	0.736	1.004	1.822	0.801	1.236	2.237	0.822
LSTM-FT	**0.465**	0.966	0.622	**0.724**	1.401	0.759	**0.935**	**1.769**	**0.812**	**1.141**	**2.134**	**0.838**

b) Chicago Taxi

Model	15min			30min			45min			60min		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
One Step	15.516	40.269	0.939	40.690	116.483	0.871	76.811	217.980	0.794	120.272	335.085	0.720
HA	12.436	36.429	0.950	21.484	68.806	0.955	30.213	100.331	0.956	**38.804**	131.110	0.957
VAR	10.856	**23.590**	**0.979**	20.560	**44.921**	**0.981**	35.945	**77.467**	**0.974**	54.780	**116.836**	**0.966**
RF	**10.836**	29.907	0.966	**18.809**	56.760	0.969	**27.904**	88.076	0.966	42.049	140.617	0.951
GCN	14.095	30.905	0.964	26.563	61.356	0.964	39.543	93.338	0.962	62.260	145.476	0.947
LSTM-FT	12.231	33.415	0.958	23.190	68.451	0.955	38.636	108.197	0.949	57.725	154.908	0.940

Table 5. Prediction Results of different models in Chicago.

a) Chicago Bike-share

Model	15min			30min			45min			60min		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
One Step	0.236	0.648	-0.207	0.409	0.967	0.039	0.561	1.255	0.141	0.722	1.611	0.123
HA	0.197	0.540	0.160	0.337	0.804	0.335	0.454	1.019	0.433	0.558	1.228	0.491
VAR	0.290	0.550	0.128	0.536	0.930	0.111	0.824	1.402	-0.072	1.236	2.086	-0.469
RF	0.211	**0.484**	**0.325**	0.350	**0.720**	**0.467**	0.467	**0.927**	**0.532**	0.571	**1.140**	**0.561**
GCN	0.224	0.489	0.311	0.374	0.742	0.434	0.499	0.962	0.496	0.598	1.176	0.533
LSTM-FT	**0.161**	0.525	0.206	**0.300**	0.788	0.362	**0.431**	1.016	0.438	**0.555**	1.217	0.500

b) Chicago Taxi

Model	15min			30min			45min			60min		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
One Step	**1.259**	3.666	0.977	2.208	6.814	0.965	3.347	11.151	0.947	4.871	16.874	0.922
HA	2.296	7.112	0.915	3.169	10.101	0.923	4.039	13.156	0.926	4.872	16.008	0.930
VAR	1.311	**3.393**	**0.981**	**2.195**	**5.883**	**0.974**	3.246	9.047	0.965	4.587	13.365	0.951
RF	1.629	4.314	0.969	2.360	6.738	0.966	**3.181**	9.981	0.957	**4.174**	14.161	0.945
GCN	1.808	3.531	0.979	2.756	6.377	0.969	3.474	**8.237**	**0.971**	4.646	**11.753**	**0.962**
LSTM-FT	1.623	4.098	0.972	2.379	6.681	0.966	3.311	9.697	0.960	4.606	14.086	0.946

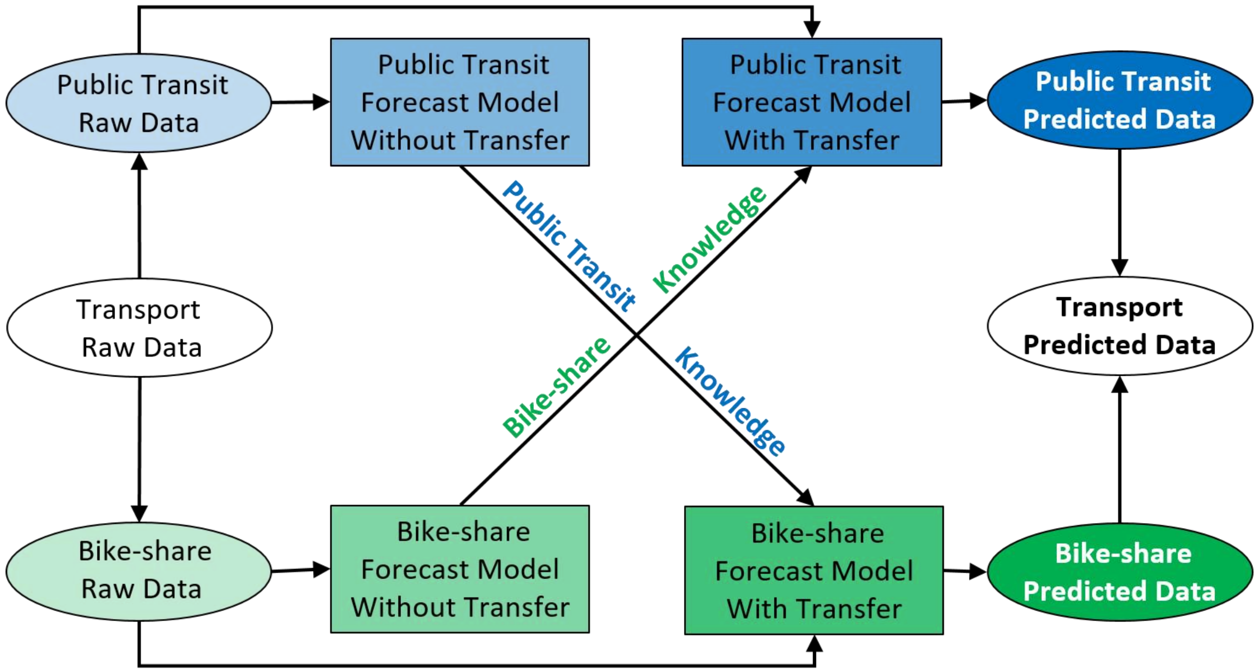


Fig. 1. Transfer learning for cross-modal demand prediction.

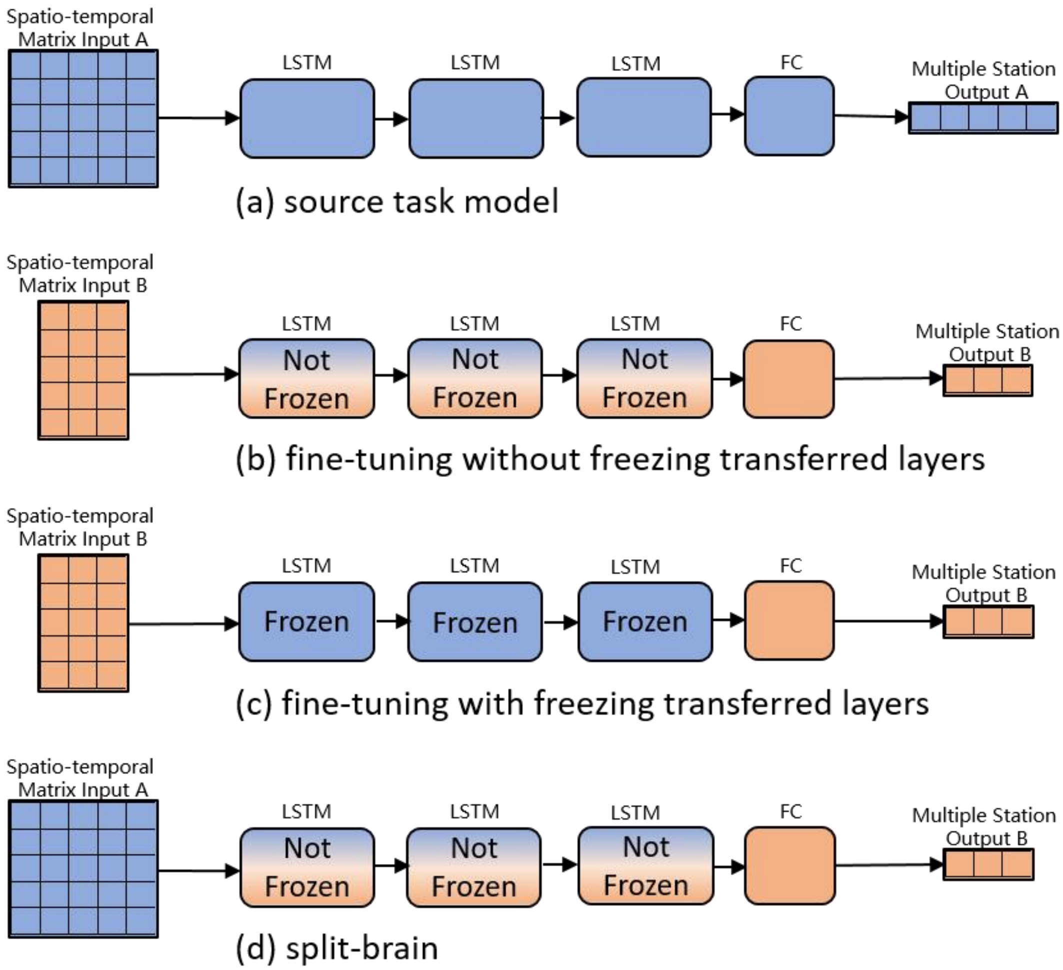


Fig. 2. The structure of different transfer learning strategies.

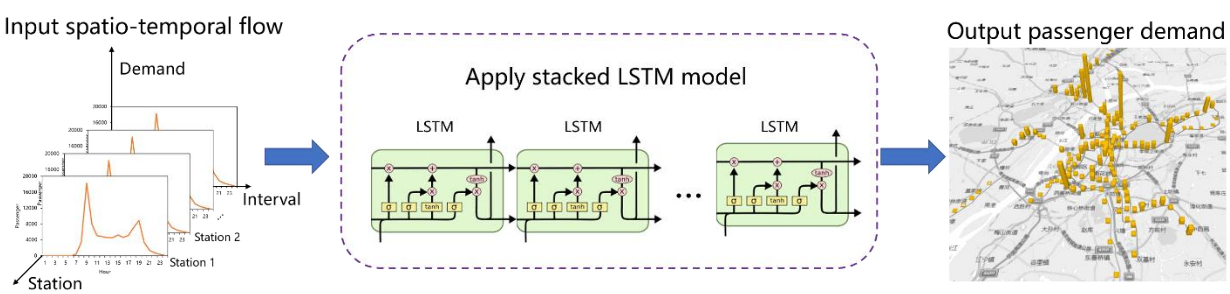
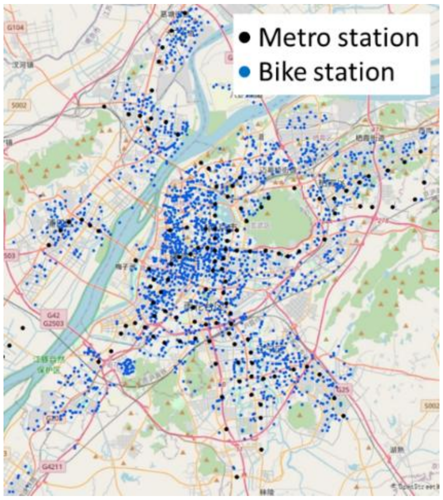
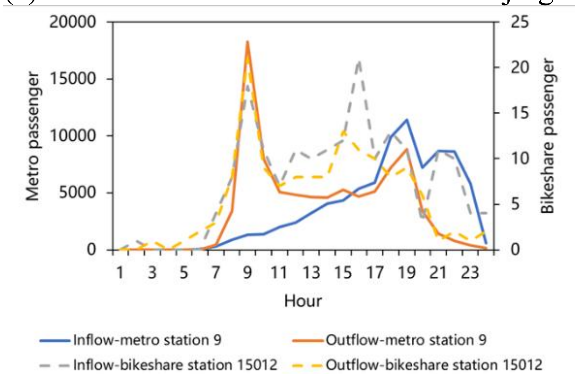


Fig. 3. The prediction approach of stacked LSTM model.

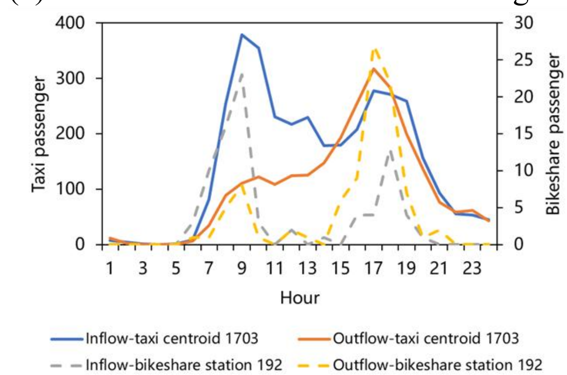


(a) Metro & bike-share services in Nanjing

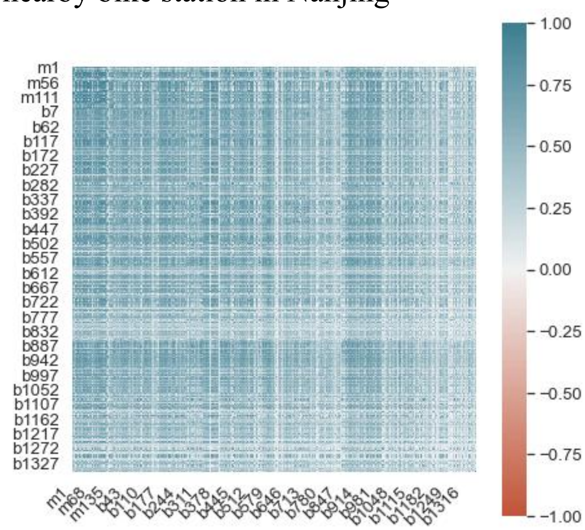
(b) Taxi & bike-share services in Chicago



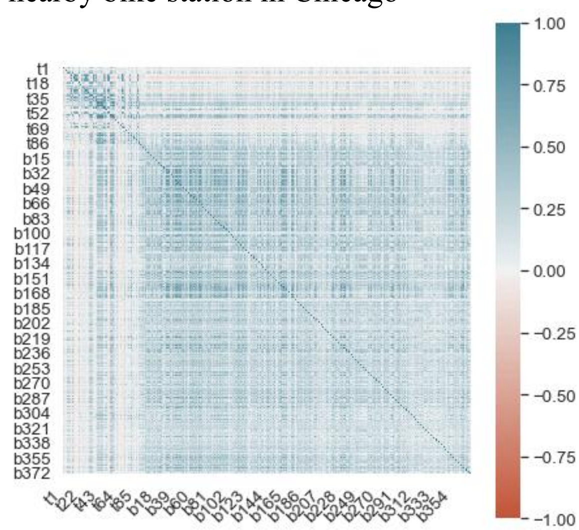
(c) Passenger flow of a metro station and its nearby bike station in Nanjing



(d) Passenger flow of a taxi centroid and its nearby bike station in Chicago

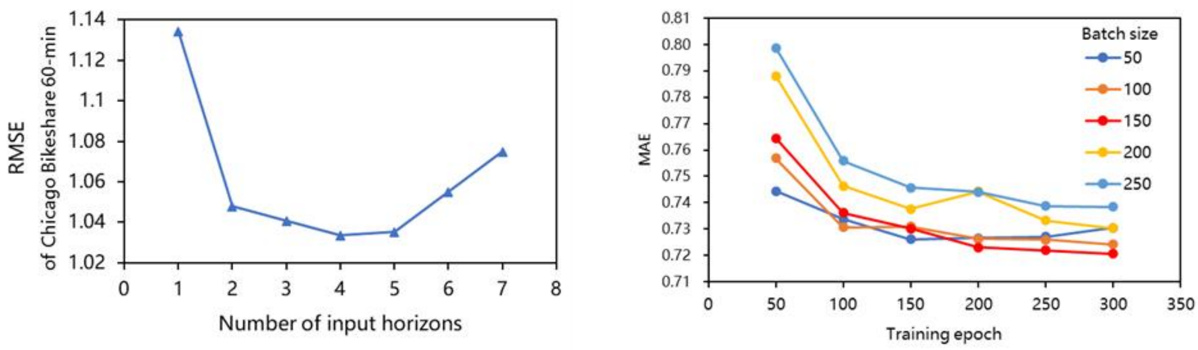


(e) Correlation matrix of station-level demands for metro and bike-share in Nanjing



(f) Correlation matrix of station-level demands for taxi and bike-share in Chicago

Fig. 4. Public transit and bike-share in Chicago and Nanjing.



(a) Input horizon amount (b) Batch size and training epoch

Fig. 5. Hyperparameter selection.