Asia Pacific
Academy of Science Pte. Ltd.

Mathematics and Systems Science 2024, 2(2), 2957.
https://doi.org/10.54517/mss2957

Article

A novel tucker decomposition driven taxi travel demand forecasting algorithm

Benjia Chu*, **Hongyu Yan**

College of Computer Science and Technology, Qingdao University, Qingdao 266701, China
* **Corresponding author:** Benjia Chu, chubenjia@ubinet.cn

CITATION

Chu B, Yan H. A novel tucker decomposition driven taxi travel demand forecasting algorithm. Mathematics and Systems Science. 2024; 2(2): 2957. https://doi.org/10.54517/mss2957

ARTICLE INFO

Received: 23 September 2024
Accepted: 22 October 2024
Available online: 10 November 2024

Abstract: This research aims to enhance taxi travel demand forecasting for sustainable urban traffic management and planning. We extend the Seasonal Autoregressive Integrated Moving Average model into a high-dimensional tensor form by treating the urban transport network as a Euclidean space and introducing Tucker decomposition. This novel approach, both theoretically significant and practically applicable, represents time series data as tensors to better capture multimodal structures and correlations, improving predictive accuracy. Tucker decomposition reduces computational complexity and memory requirements, making it ideal for large-scale urban traffic network prediction. Experiments on six real-world datasets show that the model's MAE and RMSE are reduced by about 39.43% and 27.01% on average, respectively, compared to the baseline model. Notably, the model is computationally very efficient and takes only a relatively short time to train., suitable for real-time traffic management, congestion mitigation, and resource optimization. In summary, this work innovates time series analysis, providing an efficient and precise tool for urban traffic management and planning, contributing to sustainable urban transportation advancement.

Keywords: taxi travel demand forecasting; tucker decomposition; seasonal autoregressive integrated moving average

1. Introduction

Taxis have long been an important component of urban transport. Taxis are a flexible transport option that passengers can hail anytime, anywhere, without the need to plan ahead or wait for public transport [1–3]. Timely and accurate demand forecasting can reduce traffic congestion [4,5], ease traffic pressure, improve overall traffic efficiency [6], reduce environmental pollution [7], improve people's traveling efficiency [8], and save time and costs [9]. Taxi travel demand forecasting is of great practical importance to policy makers and urban planners. It can help optimize the dynamic allocation of transportation resources, predict changes in demand at different time periods and regions, and help dispatch taxis more efficiently, reduce idling rates, and ease traffic pressure during peak periods. Through accurate demand forecasting, traffic management authorities can adjust the number and location of taxis according to the actual situation, thus improving the overall efficiency of traffic operation. Cab travel demand forecasting can provide valuable data support for urban infrastructure development. Urban planners can use the forecast to understand which areas have consistently high demand, and then optimize road design, plan new transportation hubs or parking lots, and rationally lay out public transportation facilities in order to alleviate traffic congestion and enhance the travel experience of residents.

Seasonal Autoregressive Integrated Moving Average (SARIMA) is a time series analysis model for modeling and forecasting time series data. However, most of the current SARIMA models are unable to model and forecast multiple time series at the same time and can only be calculated one by one. Therefore, its computational cost is high. In addition, SARIMA cannot capture the correlation between multiple time series, which may adversely affect its modeling effect.

In taxi travel demand forecasting, if the transport network is modeled as a Euclidean space (grid format), then the time series data can be represented in the form of a tensor. Processing time series in tensor form preserves the correlation within the time series [10,11].

Tensor decomposition is a technique to decompose high-dimensional data [12,13], which is widely used in the fields of data dimensionality reduction, feature extraction, data compression, and pattern recognition [14,15]. Through tensor decomposition, a multi-dimensional tensor can be represented as a set of low-rank core tensors and modal matrices in the form of a product, in order to extract and represent the key information and structure in the tensor [16,17].

In this paper, this work deal with time series in the form of a tensor by modeling the traffic network as a Euclidean space and applying Tucker decomposition to the tensor to obtain the core tensor and the factor matrix. The temporal correlation between the core tensor can be fully preserved by the factor matrix. Then, this work extends SARIMA to the tensor form and apply it to the core tensor. Finally, the core tensor at the next moment is predicted by the SARIMA model, and the inverse Tucker decomposition is performed to obtain the predicted value at the next moment. The contributions of this paper are 2 points:

With the help of Tucker decomposition, this work extends the classical SARIMA model to tensor form, which enables the modified SARIMA model not only to deal with multiple time series simultaneously but also reduces the computational complexity as well as the memory requirement.

This work used six real-world datasets to evaluate the proposed prediction model and demonstrated that the proposed model not only has higher prediction accuracy but also requires less time cost.

2. Literature review

Traditional statistical methods are recognized and widely used by researchers due to their advantages such as strong interpretability and solid theoretical foundation. Cyril et al. [18] used a univariate time series Seasonal Autoregressive Integrated Moving Average (ARIMA) model to predict the demand for inter-district public transport travel from Trivandrum to five other districts of Kerala. Jiao et al. [19] proposed a modified Kalman filter model based on error correction coefficients, a modified Kalman filter model based on historical bias, a modified Kalman filter model based on Bayesian combination and nonparametric regression based on the traditional Kalman filter method and applied it to short-term rail traffic forecasting. Milenković et al. [20] chose the least square residuals based on the SARIMA model to predict train passenger flow. However, traditional statistical methods not only can only model a single time series but also have difficulty in capturing the correlation between multiple time series data sequences. Therefore, the traditional statistical model cannot fully reflect the spatial characteristics of historical traffic data.

As computer performance continues to improve, researchers are increasingly choosing machine-learning approaches for travel demand modeling and forecasting. These algorithms include support vector machines, random forests, Bayesian networks, neural networks, and deep learning models [21–23]. Among them, deep learning has received a lot of attention from researchers in the last decade. Deep learning models such as convolutional neural network, recurrent neural network and its variants, Transformer model for sequence prediction, attention mechanism, and graph convolutional neural network have been used to predict the state of the city [24]. Li et al. [25] based on Pearson's correlation coefficient and dynamic time regularization to calculate the correlation between different regions, using irregular convolution to capture the spatial correlation between the data, and Long Short-Term Memory to capture the temporal correlation between the data, thus accurately completing the demand prediction of shared bikes. Irregular convolution associates regions with similar temporal usage patterns, so the IrConv+LSTM approach is superior to the general Conv+LSTM model. Zhao et al. [26] viewed the built environment as a dynamic feature and proposed a graph deep learning-based approach coupled with the spatio-temporal influence of the built environment (GDLBE) to enhance short-term transit travel demand prediction. Xu et al. [27] integrated natural environmental factors and socio-economic factors for predicting urban travel demand. Tang et al. [28] used a combination of geographically-weighted regression (GWR) and generalized linear modeling (GLM) combination to identify the factors affecting travel demand. Wang et al. [29] extracted the spatio-temporal features of taxi travel demand through the ConvLSTM model to predict the taxi travel demand. Liang et al. [30] proposed a multi-task adaptive recurrent graphical attention network that can predict travel demand at the regional level.

The prediction performance of deep learning models is excellent, whether the traffic road network is modeled as a form of Euclidean space or non-Euclidean space. However, there is an unavoidable problem with deep learning models; the computational complexity and time cost it requires for modeling and prediction is high, despite the current computational performance of computers making new breakthroughs every year.

3. Preliminaries

3.1. Notations

Lowercase bold letters denote vectors, e.g. $\boldsymbol{x} \in \square^{N}$ .Capitalised bold letters denote matrices, e.g. $\boldsymbol{X} \in \square^{M \times N}$ .Bold Roman letters denote higher-order (N ≥ 3) tensors, e.g., $\mathbf{X} \in \square^{I_1 \times I_2 \times \cdots \times I_N}$. The (m,n)-th entity of the tensor $\boldsymbol{\mathcal{X}}$ is denoted as $\boldsymbol{x}_{m,n}$.The Frobenius paradigm for the tensor $\boldsymbol{\mathcal{X}}$ is denoted as $\|\boldsymbol{\mathcal{X}}\|_{\boldsymbol{F}} = \sqrt{\sum_{\boldsymbol{m,n}} \boldsymbol{x}_{\boldsymbol{m,n}}^{\boldsymbol{2}}}$. The n-mode product of a tensor $\boldsymbol{\mathcal{X}} \in \square^{I_1 \times I_2 \times \cdots \times I_N}$ with a matrix $\boldsymbol{U} \in \square^{I_n \times R_n}$ is defined as $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} \times_{\boldsymbol{n}} \boldsymbol{U}^{\boldsymbol{T}} \in \square^{I_1 \times \cdots \times I_{n-1} \times R_n \times I_{n+1} \times \cdots \times I_N}$. The mode-n expansion of the tensor $\boldsymbol{\mathcal{X}}$ is denoted as $\boldsymbol{X}(n) = Unfold(\boldsymbol{\mathcal{X}}) \in \square^{I_n \times \prod_{i \neq n} I_n}$, The inverse of the tensor expansion is fold, which is denoted as $\boldsymbol{\mathcal{X}} = \mathit{Fold}(\boldsymbol{X}^{(\mathrm{n})})$.

3.2. Tucker decomposition

The Tucker decomposition allows a higher-order tensor $\boldsymbol{\mathcal{X}_t} \in \square^{I_1 \times I_2 \times \cdots \times I_N}$ to be expressed as a product of a projection matrix and a factor matrix, as shown in Equation (1).

$$\begin{gathered}\boldsymbol{\mathcal{X}_t} \approx \boldsymbol{\mathcal{G}_t} \times_1 \boldsymbol{U}^{(1)} \times_2 \boldsymbol{U}^{(2)} \cdots \times_N \boldsymbol{U}^{(N)} \\ \boldsymbol{s.t.\, U^{(n)^T} U^{(n)} = I, n = 1, \ldots, N}\end{gathered} \tag{1}$$

where $\boldsymbol{\mathcal{G}_t} \in \square^{I_1 \times I_2 \times \cdots \times I_N}$ is a low-rank core tensor, $\boldsymbol{U^{(n)}} \in \square^{I_n \times R_n}$, $R_n < I_n$, $\boldsymbol{I}$ is identity matrix. The Tucker rank of a tensor $\boldsymbol{\mathcal{X}_t}$ is represented by a vector containing N non-negative integers: $[R_1, \ldots, R_n, \ldots, R_N]$, Each element of the Tucker rank vector corresponds to the rank of each factor matrix in the Tucker decomposition. **Figure 1** shows the Tucker decomposition diagram of the third-order tensor

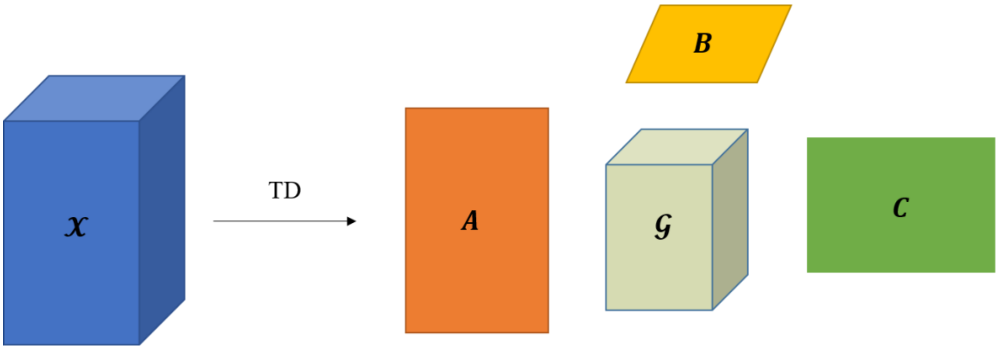


Figure 1. Tucker decomposition.

Note: A, B, C are factor matrices, $\mathcal{G}$ is the core tensor.

3.3. SARIMA

SARIMA is an extension of the ARIMA model to effectively handle seasonal time series data. It is represented as SARIMA (p, d, q) (P, D, Q, S). For a given time, series matrix $\boldsymbol{X} \in \square^{M \times N}$, the SARIMA model can be represented as Equation (2).

$$\Delta^D \Delta^d \boldsymbol{x_t} = \sum_{i=1}^{p} \alpha_i \, \Delta^D \Delta^d \boldsymbol{x_{t-i}} - \sum_{i=1}^{q} \beta_i \, \boldsymbol{\xi_{t-i}} + \sum_{i=1}^{P} \gamma_i \, \Delta^D \Delta^d \boldsymbol{x_{t-Si}} - \sum_{i=1}^{Q} \vartheta_i \, \boldsymbol{\xi_{t-Si}} + \boldsymbol{\xi_t} \tag{2}$$

where ξ_t is the error term at time t, S is Seasonal cycle, $\{\alpha_i\}_{i=1}^{p}$, $\{\beta_i\}_{i=1}^{q}$, $\{\gamma_i\}_{i=1}^{P}$, $\{\vartheta_i\}_{i=1}^{Q}$ are the coefficients of AR, MA, SAR and SMA respectively. *P*, *q*, *P*, *Q*, *d* and D are the orders of AR, MA, SAR, SMA, difference and seasonal difference, respectively.

4. Methodology

In this section, this work extends the SARIMA model to the tensor form and incorporate the Tucker decomposition technique. SARIMA is an extension of the ARIMA model to effectively handle seasonal time series data. It is represented as

SARIMA (p, d, q) (P, D, Q, S). The tensor form of SARIMA can be expressed in the form of Equation (3):

$$\Delta^D\Delta^d\boldsymbol{\mathcal{G}_t} = \sum_{i=1}^{p}\alpha_i\,\Delta^D\Delta^d\boldsymbol{\mathcal{G}_{t-i}} - \sum_{i=1}^{q}\beta_i\,\boldsymbol{\varepsilon_{t-i}} + \sum_{i=1}^{P}\gamma_i\,\Delta^D\Delta^d\boldsymbol{\mathcal{G}_{t-Si}} - \sum_{i=1}^{Q}\vartheta_i\,\boldsymbol{\varepsilon_{t-Si}} + \boldsymbol{\varepsilon_t} \tag{3}$$

where ε is the random error and $\mathcal{G}_t$ is the core tensor of $\boldsymbol{\mathcal{X}_t}$.

This work defines TD-SARIMA as the following optimization problem:

$$\begin{aligned} min_{\{\mathcal{G}_t,\boldsymbol{U}^{(k)},\varepsilon_{t-i},\alpha_i,\beta_i,\gamma_i,\vartheta_i\}} \sum\nolimits_{t=SP+SQ+1}^{T}\Bigg(\frac{1}{2}\Bigg\|&\Delta^D\Delta^d\boldsymbol{\mathcal{G}_t} - \sum\nolimits_{i=1}^{p}\alpha_i\,\Delta^D\Delta^d\boldsymbol{\mathcal{G}_{t-i}} \\ &+\sum\nolimits_{i=1}^{q}\beta_i\,\boldsymbol{\varepsilon_{t-i}} - \sum\nolimits_{i=1}^{P}\gamma_i\,\Delta^D\Delta^d\boldsymbol{\mathcal{G}_{t-Si}} + \sum\nolimits_{i=1}^{Q}\vartheta_i\,\boldsymbol{\varepsilon_{t-Si}}\Bigg\|_F^2 \\ &+\frac{1}{2}\left\|\Delta^D\Delta^d\boldsymbol{\mathcal{G}_t} - \Delta^D\Delta^d\boldsymbol{\mathcal{X}}_{t_{\times 1}}\boldsymbol{U}^{(1)^T}\cdots_{\times N}\boldsymbol{U}^{(N)^T}\right\|_F^2\Bigg) \end{aligned}$$

$$s.t.\,U^{(k)^T}U^{(k)} = I, k = 1,2,\cdots,N \tag{4}$$

Then, for ease of derivation, this work subjects the tensor in the optimization problem (4) to a mode-n expansion operation, and this work obtain the following form:

$$\begin{aligned} min_{\{G_t^{(k)},\boldsymbol{U}^{(k)},E_{t-i}^{(k)},\alpha_i,\beta_i,\gamma_i,\vartheta_i\}} \sum\nolimits_{t=SP+SQ+1}^{T}\sum\nolimits_{k=1}^{N}\Bigg(\frac{1}{2}\Bigg\|&\Delta^D\Delta^d\boldsymbol{G_t}^{(\boldsymbol{k})} \\ &-\sum\nolimits_{i=1}^{p}\alpha_i\,\Delta^D\Delta^d\boldsymbol{G_{t-i}}^{(k)} + \sum\nolimits_{i=1}^{q}\beta_i\,\boldsymbol{E_{t-i}}^{(k)} \\ &-\sum\nolimits_{i=1}^{P}\gamma_i\,\Delta^D\Delta^d\boldsymbol{G_{t-Si}}^{(k)} + \sum\nolimits_{i=1}^{Q}\vartheta_i\,\boldsymbol{E_{t-Si}}^{(k)}\Bigg\|_F^2 \\ &+\frac{1}{2}\left\|\Delta^D\Delta^d\boldsymbol{G_t}^{(k)} - \boldsymbol{U}^{(k)^{\mathrm{T}}}\boldsymbol{\mathcal{X}}_t\boldsymbol{U}^{(-k)^{\mathrm{T}}}\right\|_F^2\Bigg) \end{aligned}$$

$$s.t.\,\boldsymbol{U}^{(k)^T}\boldsymbol{U}^{(k)} = I, k = 1,2,\cdots,N \tag{5}$$

where $\boldsymbol{U}^{(-k)^{\mathrm{T}}} = \boldsymbol{U}^{(k)^{\mathrm{T}}}\otimes\cdots\otimes\boldsymbol{U}^{(k+1)^{\mathrm{T}}}\otimes\boldsymbol{U}^{(k-1)^{\mathrm{T}}}\otimes\cdots\boldsymbol{U}^{(1)^{\mathrm{T}}} \in \square^{\prod_{j\neq k}R_j\times\prod_{j\neq k}I_j}$. Next, this work derives the update method for each parameter.

Update $\boldsymbol{G_t}^{(\boldsymbol{k})}$

The part of Equation (5) with respect to $\boldsymbol{G_t}^{(\boldsymbol{k})}$ is:

$$\begin{aligned} &min_{\{G_t^{(k)}\}}\sum\nolimits_{t=SP+SQ+1}^{T}\sum\nolimits_{k=1}^{N}\frac{1}{2}\Bigg\|\Delta^D\Delta^d\boldsymbol{G_t}^{(\boldsymbol{k})} - \sum\nolimits_{i=1}^{p}\alpha_i\,\Delta^D\Delta^d\boldsymbol{G_{t-i}}^{(\boldsymbol{k})} + \sum\nolimits_{i=1}^{q}\beta_i\,\boldsymbol{E_{t-i}}^{(\boldsymbol{k})} \\ &-\sum\nolimits_{i=1}^{P}\gamma_i\,\Delta^D\Delta^d\boldsymbol{G_{t-Si}}^{(\boldsymbol{k})} + \sum\nolimits_{i=1}^{Q}\vartheta_i\,\boldsymbol{E_{t-Si}}^{(\boldsymbol{k})}\Bigg\|_F^2 + \frac{1}{2}\left\|\Delta^D\Delta^d\boldsymbol{G_t}^{(\boldsymbol{k})} - \boldsymbol{U}^{(k)^{\mathrm{T}}}\boldsymbol{\mathcal{X}}_t\boldsymbol{U}^{(-k)^{\mathrm{T}}}\right\|_F^2\Bigg) \end{aligned} \tag{6}$$

By taking the partial derivatives of Equation (6) and making them zero, this work can update $\boldsymbol{G_t}^{(\boldsymbol{k})}$ by Equation (7):

$$\Delta^D \Delta^d \boldsymbol{G_t}^{(\boldsymbol{k})} = \frac{1}{2}(\boldsymbol{U}^{(k)^{\mathrm{T}}} \boldsymbol{\mathcal{X}}_t \boldsymbol{U}^{(-k)^{\mathrm{T}}} + \sum_{i=1}^{p} \alpha_i \, \Delta^D \Delta^d \boldsymbol{G_{t-i}}^{(\boldsymbol{k})} {}_{-} \sum_{i=1}^{q} \beta_i \, \boldsymbol{E_{t-i}}^{(\boldsymbol{k})} + \sum_{i=1}^{P} \gamma_i \, \Delta^D \Delta^d \boldsymbol{G_{t-Si}}^{(\boldsymbol{k})} {}_{-} \sum_{i=1}^{Q} \vartheta_i \, \boldsymbol{E_{t-Si}}^{(\boldsymbol{k})}) \tag{7}$$

Update $\boldsymbol{U}^{(\boldsymbol{k})}$,

The part of Equation (5) with respect to $\boldsymbol{U}^{(\boldsymbol{k})}$ is:

$$min_{\{U^{(k)}\}} \sum_{t=SP+SQ+1}^{T} \sum_{k=1}^{N} \frac{1}{2} \left\| \Delta^D \Delta^d G_t{}^{(k)} - U^{(k)^{\mathrm{T}}} \mathcal{X}_t U^{(-k)^{\mathrm{T}}} \right\|_F^2$$

$$s.t. \, U^{(k)^T} U^{(k)} = I, k = 1,2, \cdots, N \tag{8}$$

The optimization problem of Equation (8) is equivalent to the classical Orthogonal Procrustes Problem, which has a globally optimal solution:

$$U^{(k)} = U^{*(k)} V^{*(k)^{\mathrm{T}}} \tag{9}$$

where $U^{*(k)}$ and $V^{*(k)}$ are the left singular value matrix and right singular value matrix of $\sum_{t=SP+SQ+1}^{T} \boldsymbol{\mathcal{X}}_t{}^{(k)} \boldsymbol{U}^{(-k)^{\mathrm{T}}} \Delta^D \Delta^d \boldsymbol{G_t}^{(\boldsymbol{k})^T}$, respectively.

Update$E_{t-i}{}^{(k)}$,

The part of Equation (5) with respect to $E^{(k)}$ is:

$$min_{\{E^{(k)}\}} \sum_{t=SP+SQ+1}^{T} \sum_{k=1}^{N} \frac{1}{2} \left\| \Delta^D \Delta^d \boldsymbol{G_t}^{(\boldsymbol{k})} - \sum_{i=1}^{p} \alpha_i \, \Delta^D \Delta^d \boldsymbol{G_{t-i}}^{(\boldsymbol{k})} + \sum_{i=1}^{q} \beta_i \, \boldsymbol{E_{t-i}}^{(\boldsymbol{k})} - \sum_{i=1}^{P} \gamma_i \, \Delta^D \Delta^d \boldsymbol{G_{t-Si}}^{(\boldsymbol{k})} + \sum_{i=1}^{Q} \vartheta_i \, \boldsymbol{E_{t-Si}}^{(\boldsymbol{k})} \right\|_F^2 \tag{10}$$

Calculating the partial derivatives of Equation (10) and making them zero allows us to update the parameters in the same way as Equation (11).

$$\boldsymbol{E}_{\boldsymbol{t-i}}^{(\boldsymbol{k})} = -\frac{1}{\beta_i} \Big(\Delta^D \Delta^d \boldsymbol{G}_{\boldsymbol{t}}^{(\boldsymbol{k})} - \sum_{i=1}^{p} \alpha_i \, \Delta^D \Delta^d \boldsymbol{G_{t-i}}^{(\boldsymbol{k})} + \sum_{i=1}^{q} \beta_i \, \boldsymbol{E_{t-i}}^{(\boldsymbol{k})} - \sum_{i=1}^{P} \gamma_i \, \Delta^D \Delta^d \boldsymbol{G_{t-Si}}^{(\boldsymbol{k})} + \sum_{i=1}^{Q} \vartheta_i \, \boldsymbol{E_{t-Si}}^{(\boldsymbol{k})} \Big) \tag{11}$$

Update $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\vartheta}$,

For the coefficients $\alpha, \beta, \gamma, \vartheta$of the SARIMA model in the objective function (5), we estimate them based on the Yule-Walker equation improved by the least squares method, which first calculates the self-covariance of the time series, then uses the series of self-covariance to construct the Yule-Walker equation and estimates the parameters of the model by matrix solving. Specifically, we use the self-covariance with lag p as $\zeta(p)$,and the seasonal covariance with lag P and seasonal term s as $\zeta(\mathrm{P} \cdot \mathrm{s})$, then the non-seasonal self-covariance matrix can be expressed as:

$$\mathrm{R}_{no-seasonal} = \begin{bmatrix} \zeta(0) & \zeta(1) & \cdots & \zeta(\mathrm{p}-1) \\ \zeta(1) & \zeta(0) & \cdots & \zeta(\mathrm{p}-2) \\ \vdots & \vdots & \ddots & \vdots \\ \zeta(\mathrm{p}-1) & \zeta(\mathrm{p}-2) & \cdots & \zeta(0) \end{bmatrix}$$

Denote the seasonal self-covariance matrix as:

$$\mathrm{R}_{seasonal} = \begin{bmatrix} \zeta(0) & \zeta(\mathrm{s}) & \cdots & \zeta(\mathrm{P}\cdot\mathrm{s}) \\ \zeta(\mathrm{s}) & \zeta(0) & \cdots & \zeta((\mathrm{P}-1)\cdot\mathrm{s}) \\ \vdots & \vdots & \ddots & \vdots \\ \zeta((\mathrm{P}-1)\cdot\mathrm{s}) & \zeta((\mathrm{P}-2)\cdot\mathrm{s}) & \cdots & \zeta(0) \end{bmatrix}$$

We combine the non-seasonal and seasonal self-covariance matrices to obtain the total self-covariance matrix:

$$\mathrm{R} = \begin{bmatrix} \mathrm{R}_{no-seasonal} & 0 \\ 0 & \mathrm{R}_{seasonal} \end{bmatrix}$$

Denote the vector of non-seasonal self-covariances as:

$$\mathrm{r}_{no-seasonal} = \begin{bmatrix} \zeta(1) \\ \zeta(2) \\ \vdots \\ \zeta(\mathrm{p}) \end{bmatrix}$$

The seasonal self-covariance vector is denoted as:

$$\mathrm{r}_{seasonal} = \begin{bmatrix} \zeta(\mathrm{s}) \\ \zeta(2\mathrm{s}) \\ \vdots \\ \zeta(\mathrm{P}\cdot\mathrm{s}) \end{bmatrix}$$

We merge the non-seasonal and seasonal self-covariance vectors to obtain the total self-covariance vector expressed as:

$$\mathrm{r} = \begin{bmatrix} \mathrm{r}_{no-seasonal} \\ \mathrm{r}_{seasonal} \end{bmatrix}$$

Finally, we solve the improved Yule-Walker equation by the least squares method, into the equation shown in Equation (12):

$$\Phi = \mathrm{R}^{-1} r \tag{12}$$

where Φ is a vector containing non-seasonal and seasonal autoregressive parameters.

5. Experiments

5.1. Datasets and experimental setup

We use six real-world datasets to evaluate the performance of TD-SARIMA, and the details of the datasets are shown in **Table 1**. The TaxiBJ dataset is provided by Zhang et al. [31], which includes Beijing taxi traffic data for four time periods. The whole city of Beijing is divided into a 32 × 32 grid. The NYC-Taxi dataset is provided by Yao et al. [32] for the taxi traffic data of New York City from 01/01/2015 to 02/10/2015. The entire New York City is divided into 10 × 20 grids. The Manhattan-Taxi dataset is provided by Liu et al. [33] for 09/01/2015–12/31/2015 taxi flow data for the Manhattan neighborhood of New York City. We use the taxi outflow data as the travel demand number for the area.

Table 1. Description of the dataset.

Dataset	TaxiBJ13	TaxiBJ14	TaxiBJ15	TaxiBJ16	NYC-Taxi	Manhattan-Taxi
Location	Beijing				New York	Manhattan, New York
Time Span	7/1/2013–10/30/2013	3/1/2014–6/30/2014	3/1/2015–6/30/2015	11/1/2015–4/10/2016	01/01/2015–02/10/2015	09/01/2015–12/31/2015
Sampling interval	30 minutes					
Gird map size	(32, 32)				(10, 20)	(15, 5)
Min-Max demand	[0, 1230]	[0, 1285]	[0, 1267]	[0, 1151]	[0, 1283]	[0, 1688]

Before the experiments, we performed Z-score normalization on these six datasets, by which the Z-score normalization can eliminate the data scale differences and reduce the influence of outliers on the model, thus enhancing the robustness of the model. In the experiments, the ratio of training, validation, and test sets used for the model is 7:1:2. The number of training rounds is 200, and the size of the Tucker decomposition core tensor rank is 5.

5.2. Baselines

This work compares TD-SARIMA to the following five baselines:

- HA: Historical Average Method, based on data collected over a period of time in the past, calculates the average of these data as a forecast value, often used for baseline comparisons.
- SARIMA: An extended ARIMA model that captures seasonal patterns in time series. By introducing a seasonal difference term, SARIMA can model time series data with cyclical variations.
- VAR: A multivariate time series model for the interdependence of multiple time series, which is predicted by regressing each time series on its own lagged value as well as the lagged values of other time series.
- SVR: a regression method based on Support Vector Machines for both linear and non-linear regression problems. SVR minimizes the prediction error by constructing a hyperplane and applies to high-dimensional feature spaces.
- GRU: an improved recurrent neural network unit that solves the gradient vanishing problem of traditional RNN by introducing a gate mechanism, and is suitable for dealing with sequential data, e.g., traffic flow prediction in time series.
- CNN+GRU: A model that combines CNN and GRU, which uses CNN to extract local spatio-temporal features first, and then performs temporal dependency modeling by GRU for complex spatio-temporal data prediction.
- ConvGRU: A model that incorporates convolutional operations into GRU, combining the advantages of CNN and GRU for processing sequence data

with spatio-temporal dependencies.

- IrConvGRU: a GRU variant for processing irregular spatio-temporal data. By irregular convolutional operations, IrConvGRU is able to capture complex dependencies in irregular spatio-temporal data more efficiently.
- DeepST: a deep learning model that consists of two parts: a spatio-temporal component and a global component. The spatio-temporal component uses a convolutional neural network framework to simultaneously model spatial near and far dependencies, and temporal correlations. The global component is used to capture global factors.

5.3. Evaluation criteria

To evaluate the accuracy of the prediction model, this work used two common evaluation criteria to measure the error, including mean absolute error (MAE) and root mean square error (RMSE). They are defined as shown in Equations (13) and (14).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{13}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{14}$$

where n is the number of validation data, $\hat{y}_i$ denotes the predicted value and y_i denotes the true value of travel demand.

5.4. Results

Table 2. Comparison of MAE of different models with different data sets.

Baseline	**Dataset**					
	TaxiBJ13	**TaxiBJ14**	**TaxiBJ15**	**TaxiBJ16**	**NYC-Taxi**	**Manhattan-Taxi**
HA	32.94	34.66	54.09	35.40	20.59	32.43
SARIMA	20.64	21.77	21.40	18.07	16.79	24.66
VAR	18.81	21.34	18.10	17.82	16.24	24.47
SVR	18.66	18.88	18.76	16.84	14.67	24.64
GRU	18.31	18.28	18.74	16.33	14.76	24.76
CNN+GRU	17.68	18.61	15.66	16.18	12.27	23.56
ConvGRU	16.73	15.21	13.89	14.37	9.13	16.88
IrConvLSTM	12.59	13.38	13.07	11.89	9.62	15.16
DeepST	11.96	12.71	12.42	11.30	9.14	14.40
TD-SARIMA	**11.91**	**11.38**	**11.53**	**10.86**	**8.61**	**13.96**

In this subsection, the experimental results of TD-SARIMA and the benchmark model for taxi travel demand forecasting under different datasets are analyzed and discussed as a way to validate the effectiveness of TD-SARIMA in the task of taxi travel demand forecasting. The results of the comparative experiments of TD-SARIMA and the benchmark model are shown in **Tables 2** and **3**.

Table 3. Comparison of MAE of different models with different data sets.

Baseline	Dataset					
	TaxiBJ13	**TaxiBJ14**	**TaxiBJ15**	**TaxiBJ16**	**NYC-Taxi**	**Manhattan-Taxi**
HA	42.03	40.55	62.34	42.25	34.47	42.03
SARIMA	29.61	31.04	31.25	28.05	29.39	44.71
VAR	28.10	30.46	27.58	27.98	28.47	42.00
SVR	28.37	28.62	28.54	27.49	26.94	36.44
GRU	27.42	28.11	26.74	26.56	25.60	34.71
CNN+GRU	25.32	26.99	24.59	25.32	24.03	32.65
ConvGRU	23.40	25.10	25.54	24.87	23.76	31.43
IrConvLSTM	22.88	24.33	23.69	21.01	23.55	27.30
DeepST	21.74	23.11	22.51	19.96	23.32	25.94
TD-SARIMA	**21.09**	**22.42**	**21.83**	**16.39**	**22.63**	**23.80**

The HA model takes the average of historical data as the forecast value by assuming that future taxi travel demand is similar to the average of a certain period in the past. Taxi travel demand usually exhibits complex nonlinear patterns, such as changes in demand during peak and trough periods. The HA model assumes that future demand is similar to the historical average and fails to take these nonlinear relationships into account, so the HA model has the worst prediction effect in the comparison experiments. SARIMA is an extended ARIMA model, and by introducing a seasonal difference term, SARIMA is able to predict the demand for the modeling of time series data with cyclical variations, and it is able to capture the cyclical features in time series. Compared with the HA model, the MAE and RMSE of SARIMA are reduced by about 41.30 and 26.40%, respectively. Compared to SARIMA, the VAR model can handle multiple correlated time series simultaneously. Therefore, in the task of taxi demand forecasting, VAR can incorporate the demand between different regions into the modeling scope, that is, VAR is able to capture the spatial correlation between the demand for taxi trips in different regions. Compared to the SARIMA model, the MAE and RMSE of the VAR model are reduced by 5.31% and 4.88%, respectively.

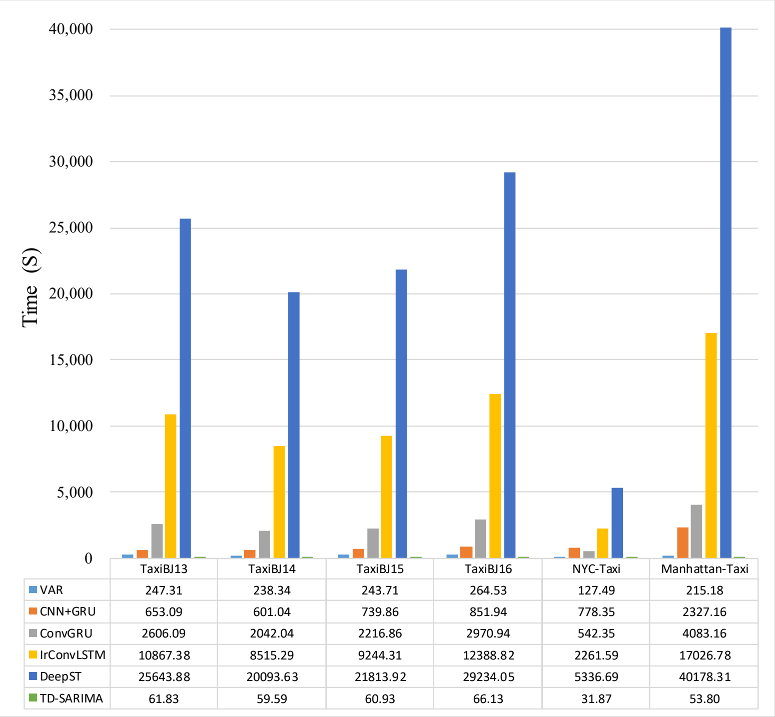
SVR is a regression method based on Support Vector Machines, which is mainly used to solve linear and nonlinear regression problems. SVR minimizes the prediction error by finding an optimal hyperplane. GRU is a variant of Recurrent Neural Networks, which is able to efficiently solve the problem of gradient vanishing of the traditional RNNs by introducing a gate mechanism, so that it can capture data temporal dependence in a long time series to capture the temporal dependence of the data. Both SVR and GRU have better nonlinear modeling capabilities, but GRU is more suitable for dealing with complex time series data, especially when there are long time dependencies or nonlinear dynamic changes in the data. The MAE and RMSE of GRU are reduced by 3.45% and 4.12% respectively compared to SVR.

CNN+GRU is a hybrid model that combines the feature extraction capability of CNN with the temporal modeling capability of GRU. ConvGRU is a model that integrates convolutional operations directly into GRU. Unlike the separated structure of CNN+GRU, ConvGRU introduces convolutional operations directly into the structure of GRU, enabling each GRU unit to capture temporal dependencies as well as handle spatial features. The state update mechanism of ConvGRU retains the time-series modeling capability of GRU while capturing spatial dependencies through convolutional operations. Compared with CNN+GRU, ConvGRU models both spatial and temporal dependencies, and thus the MAE and RMSE of ConvGRU are reduced by 17.07% and 3.02%, respectively.

IrConvLSTM introduces irregular convolution operation, which enables the model to capture complex dependencies more efficiently in irregular spatial dimensions. DeepST divides taxi travel demand into three types of features in temporal order according to the time gap of taxi travel demand, and captures the spatio-temporal features of taxi travel demand simultaneously through a multi-channel separated CNN. As a result, the MAE and RMSE of DeepST are reduced by 4.99% and 4.33%, respectively, compared to IrConvLSTM.

The proposed TD-SARIMA model compensates for the shortcomings of SARIMA, which can only deal with a single time series, by tensorizing the demand for taxi trips and then combining the Tucker decomposition with the SARIMA model. Meanwhile, by spatio-temporal modeling of the core tensor after Tucker decomposition, TD-SARIMA is able to better capture the multimodal structure and spatio-temporal correlation within the data, which greatly improves the predictive ability of the model. The experimental results show that compared with the benchmark model, TD-SARIMA has the highest prediction accuracy, and its MAE and RMSE are reduced by about 39.43% and 27.01% on average compared with the benchmark model.

As shown in **Figure 2**, we conducted a comparative experiment between the training time of TD-SARIMA and the benchmark model on each of the six datasets. It is very obvious from the figure that the TD-SARIMA model has the fastest training speed, even when compared with the VAR, which is the fastest training model in the benchmark model, the training speed of TD-SARIMA is four times that of the VAR, which is much faster than that of the deep learning-based prediction model. This is because deep learning models usually contain a large number of parameters and complex network structures, which need to be optimized by backpropagation algorithms. The higher the number of parameters, the greater the computational and storage requirements, resulting in a longer training process.



	TaxiBJ13	TaxiBJ14	TaxiBJ15	TaxiBJ16	NYC-Taxi	Manhattan-Taxi
VAR	247.31	238.34	243.71	264.53	127.49	215.18
CNN+GRU	653.09	601.04	739.86	851.94	778.35	2327.16
ConvGRU	2606.09	2042.04	2216.86	2970.94	542.35	4083.16
IrConvLSTM	10867.38	8515.29	9244.31	12388.82	2261.59	17026.78
DeepST	25643.88	20093.63	21813.92	29234.05	5336.69	40178.31
TD-SARIMA	61.83	59.59	60.93	66.13	31.87	53.80

Figure 2. Comparison of training time overhead for different models with different datasets.

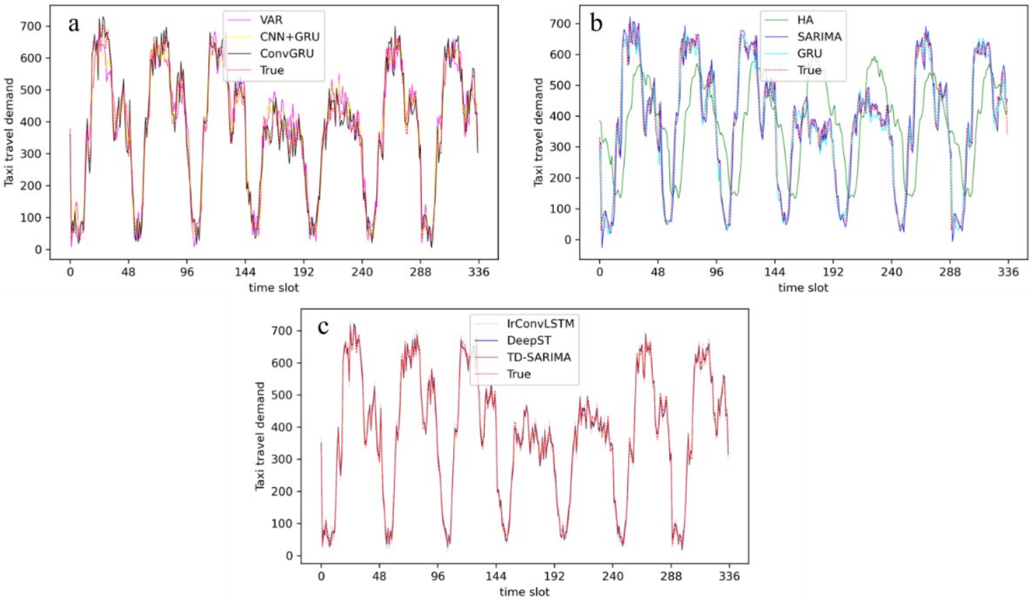


Figure 3. Comparison of prediction results of different models under TaxiBJ16 dataset.

Figures 3 and **4** show the plots of the actual prediction results of the TD-SARIMA and benchmark models on the TaxiBJ16 and NYC-Taxi datasets. The HA model, which is based only on the average of the historical data, is not able to capture these real-time variations, which leads to a lack of flexibility and accuracy in the prediction results. Especially on the TaxiBJ16 dataset, there is an obvious gap between the prediction results of the HA model and the real values, which is due to the fact that there are some missing data in the TaxiBJ16 dataset, and the taxi travel demand data is not recorded in the isochronous time slot, which affects the HA model and leads to inaccurate prediction results. On the other hand, in the Manhattan-Taxi dataset, the taxi travel demand data is recorded with equal time slots, and the HA model is not affected by the missing data. The main advantage of the SARIMA model over the HA model is its ability to capture seasonal patterns in the time series. Taxi travel demand has significant cyclical variations, and thus SARIMA provides better forecasts in the short term. Compared to the HA and SARIMA models, GRU performs well in the face of complex nonlinearities and dynamic changes. This is because its structure can be flexibly adapted to changes in the input data. Therefore, GRU's forecasting ability is superior to that of HA and SARIMA models.

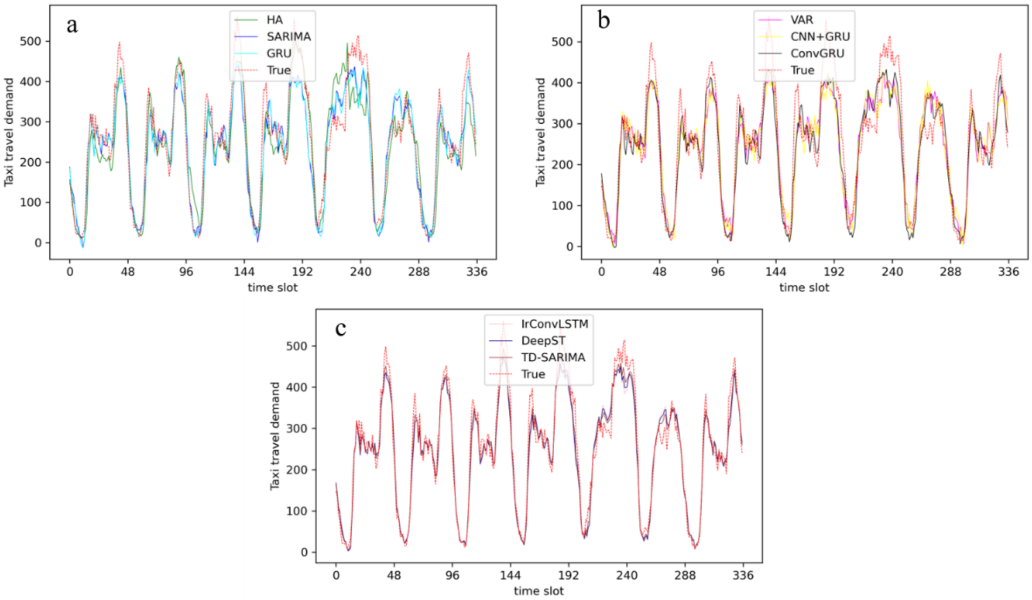


Figure 4. Comparison of prediction results of different models under Manhattan-Taxi dataset.

The VAR model is capable of modeling multiple time series at the same time, but it is not a strength of the VAR model in capturing the spatial dependence of the data, and thus the VAR model is unable to take into account the interactions of the demand for taxi trips between different regions. The CNN+GRU is a hybrid model combining the feature extraction capability of a CNN with the GRU's time-series modeling capability. Local spatio-temporal features in the input data are extracted by CNN. CNN can handle data with spatial structure well and capture spatial dependencies in the data. GRU are responsible for capturing long and short-term temporal dependencies for time-series prediction. ConvGRU directly introduces convolutional operations into the structure of GRUs, which makes each GRU unit capable of capturing temporal dependencies as well as handling spatial features. ConvGRU's state update mechanism retains the time series modeling capability of the GRU while capturing spatial dependencies through convolutional operations. ConvGRU performs better than CNN+GRU in the taxi travel demand forecasting task because it is able to capture the dependencies in the spatio-temporal data more closely.

IrConvLSTM is able to adapt to data with irregular spatio-temporal structure, such as missing data, non-uniform sampling, etc. through irregular convolutional operations. This makes it more robust in dealing with real-world taxi demand data. IrConvLSTM predicts better than ConvGRU because IrConvLSTM is more suitable for complex and dynamically changing scenarios in the real world. DeepST selects data at different times based on temporal attributes to concatenate them and then uses multiple convolutional layers to perform feature extraction in the process. The spatio-temporal dependence of the data is incorporated into the modeling scope by DeepST. Meanwhile, the global component part of DeepST enables it to capture the global features of taxi travel demand data. Therefore, DeepST has excellent prediction performance.TD-SARIMA model, which combines Tucker decomposition and SARIMA model, through the spatio-temporal modeling of the core tensor after Tucker decomposition, TD-SARIMA is able to better capture the multimodal structure and spatio-temporal correlation within the data, which greatly improves the prediction ability of the model.

6. Conclusions and discussions

In our experiments, this work selects five real-world datasets for evaluation, including data from different cities and different time periods. These datasets are chosen to verify the generalizability and robustness of the TD-SARIMA model. The results show that the TD-SARIMA model achieves excellent prediction accuracy on all datasets, which is an outstanding performance compared to traditional benchmark models. Not only that, the TD-SARIMA model also demonstrates robustness under various training data sizes, which provides more possibilities for its application in different scenarios and data conditions. Future research directions could include further improving the model performance, expanding the application areas, and exploring in depth the parameter tuning and practical deployment of the TD-SARIMA model. In addition, we will consider real-world constraints such as unexpected traffic patterns, road construction, or external factors such as weather conditions. This will help to better meet the needs of urban traffic management and promote sustainable urban development and intelligent transport systems. For larger urban environments, increasing the size of the tucker decomposition rank is necessary, which allows the core tensor to be more informative. In missing data scenarios, we can first use tensor complementation techniques to try to recover the data and then use predictive models to make predictions.

Author contributions: Conceptualization, BC and HY; methodology, BC; software, BC; validation, BC and HY; formal analysis, BC; investigation, BC; resources, BC; data curation, BC; writing—original draft preparation, BC; writing—review and editing, BC; visualization, HY; supervision, HY; project administration, HY; funding acquisition, HY. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. Cheng Z, Jian S, Rashidi T H, et al. Integrating household travel survey and social media data to improve the quality of od matrix: A comparative case study. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(6): 2628-2636.doi:

10.1109/TITS.2019.2958673

2. Xu Z, Li J, Lv Z, et al. A classification method for urban functional regions based on the transfer rate of empty cars. IET Intelligent Transport Systems, 2022, 16(2): 133-147. doi: 10.1049/itr2.12134
3. Lv Z, Li, J, Dong C, Li H, et al. Deep learning in the COVID-19 epidemic: A deep model for urban traffic revitalization index. Data & Knowledge Engineering, 2021,135, 101912. doi: 10.1016/j.datak.2021.101912
4. Li J, Lv Z, Ma Z, Wang X, et al. Optimization of spatial-temporal graph: A taxi demand forecasting model based on spatial-temporal tree. Information Fusion, 2024,104, 102178. doi: 10.1016/j.inffus.2023.102178
5. Lv Z, Ma Z, Xia F, et al. A transportation Revitalization index prediction model based on Spatial-Temporal attention mechanism. Advanced Engineering Informatics,2024, 61, 102519. doi: 10.1016/j.aei.2024.102519
6. Li H, Lv Z, Li J, et al. Traffic flow forecasting in the covid-19: A deep spatial-temporal model based on discrete wavelet transformation. ACM Transactions on Knowledge Discovery from Data, 2023, 17(5): 1-28.doi: 10.1145/3564753
7. Xu Z, Lv Z, Chu B, et al. Progress and prospects of future urban health status prediction. Engineering Applications of Artificial Intelligence, 2024,129, 107573. doi: 10.1016/j.engappai.2023.107573
8. Xu Z, Lv Z, Li J, et al. A novel approach for predicting water demand with complex patterns based on ensemble learning. Water Resources Management, 2022, 36(11): 4293-4312. doi: 10.1007/s11269-022-03255-5
9. Lv Z, Wang X, Cheng Z, et al. A new approach to COVID-19 data mining: A deep spatial–temporal prediction model based on tree structure for traffic revitalization index. Data & Knowledge Engineering, 2023, 146: 102193. doi: 10.1016/j.datak.2023.102193
10. Sheng Z, Lv Z, Li J, et al. Taxi travel time prediction based on fusion of traffic condition features. Computers and Electrical Engineering, 2023, 105: 108530. doi: 10.1016/j.compeleceng.2022.108530
11. Li Y, Li J, Lv Z, et al. GASTO: A fast adaptive graph learning framework for edge computing empowered task offloading. IEEE Transactions on Network and Service Management, 2023, 20(2): 932-944. doi: 10.1109/TNSM.2023.3250395
12. Lv Z, Li J, Xu Z, et al. Parallel computing of spatio-temporal model based on deep reinforcement learning,International Conference on Wireless Algorithms, Systems, and Applications. Cham: Springer International Publishing, 2021: 391-403.doi: 10.1007/978-3-030-85928-2_31
13. Xu Z, Lv Z, Chu B, et al. Fast autoregressive tensor decomposition for online real-time traffic flow prediction. Knowledge-Based Systems, 2023, 282: 111125. doi: 10.1016/j.knosys.2023.111125
14. Lv Z, Cheng Z, Li J, et al. TreeCN: time series prediction with the tree convolutional network for traffic prediction. IEEE Transactions on Intelligent Transportation Systems, 2023. doi: 10.1109/TITS.2023.3325817
15. Ye R, Lv Z, Xu Z, et al. MT-CNN: A Lightweight Spatial-Temporal Convolutional Neural Network for Deep Learning of Complex Trajectory Distributions based on Area Partitioning,2024 International Joint Conference on Neural Networks (IJCNN). IEEE, 2024: 1-8.doi: 10.1109/IJCNN60899.2024.10651261
16. Xu Z, Lv Z, Chu B, et al. A Fast Spatial-temporal Information Compression algorithm for online real-time forecasting of traffic flow with complex nonlinear patterns. Chaos, Solitons & Fractals, 2024, 182: 114852. doi: 10.1016/j.chaos.2024.114852
17. Cyril A, Mulangi R H, George V. Modelling and forecasting bus passenger demand using time series method,2018 7th international conference on reliability, infocom technologies and optimization (trends and future directions)(ICRITO). IEEE, 2018: 460-466.doi: 10.1109/ICRITO.2018.8748443
18. Jiao P, Li R, Sun T, et al. Three revised Kalman filtering models for short-term rail transit passenger flow prediction. Mathematical Problems in Engineering, 2016, 2016(1): 9717582.doi: 10.1155/2016/9717582
19. Milenković M, Švadlenka L, Melichar V, et al. SARIMA modelling approach for railway passenger flow forecasting. Transport, 2018, 33(5): 1113-1120.doi: 10.3846/16484142.2016.1139623
20. Sun Y, Leng B, Guan W. A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system. Neurocomputing, 2015, 166: 109-121.doi: 10.1016/j.neucom.2015.03.085
21. Wang X, An K, Tang L, et al. Short term prediction of freeway exiting volume based on SVM and KNN. International Journal of Transportation Science and Technology, 2015, 4(3): 337-352. doi: 10.1260/2046-0430.4.3.337
22. Hou Y, Edara P, Chang Y. Road network state estimation using random forest ensemble learning, 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2017: 1-6.doi: 10.1109/ITSC.2017.8317743
23. Joubert J W, De Waal A. Activity-based travel demand generation using Bayesian networks. Transportation Research Part C: Emerging Technologies, 2020, 120: 102804.doi: 10.1016/j.trc.2020.102804

24. Guo G, Zhang T. A residual spatio-temporal architecture for travel demand forecasting. Transportation Research Part C: Emerging Technologies, 2020, 115: 102639.doi: 10.1016/j.trc.2020.102639
25. Li X, Xu Y, Zhang X, et al. Improving short-term bike sharing demand forecast through an irregular convolutional neural network. Transportation research part C: emerging technologies, 2023, 147: 103984. doi: 10.1016/j.trc.2022.103984
26. Zhao T, Huang Z, Tu W, et al. Coupling graph deep learning and spatial-temporal influence of built environment for short-term bus travel demand prediction. Computers, Environment and Urban Systems, 2022, 94: 101776.doi: 10.1016/j.compenvurbsys.2022.101776
27. Xu Z, Lv Z, Li J, et al. A Novel Perspective on Travel Demand Prediction Considering Natural Environmental and Socioeconomic Factors. IEEE Intelligent Transportation Systems Magazine (April 2022), 2-25. 2022. doi: 10.1109/MITS.2022.3162901
28. Tang J, Gao F, Liu F, et al. Understanding spatio-temporal characteristics of urban travel demand based on the combination of GWR and GLM. Sustainability, 2019, 11(19): 5525.doi: 10.3390/su11195525
29. Wang D, Yang Y, Ning S. DeepSTCL: A deep spatio-temporal ConvLSTM for travel demand prediction,2018 international joint conference on neural networks (IJCNN). IEEE, 2018: 1-8.doi: 10.1109/IJCNN.2018.8489530
30. Liang J, Tang J, Gao F, et al. On region-level travel demand forecasting using multi-task adaptive graph attention network. Information Sciences, 2023, 622: 161-177. doi: 10.1016/j.ins.2022.11.138
31. Zhang J, Zheng Y, Qi D. Deep spatio-temporal residual networks for citywide crowd flows prediction, Proceedings of the AAAI conference on artificial intelligence. 2017, 31(1). doi: 10.1609/aaai.v31i1.10735
32. Yao H, Tang X, Wei H, et al. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction, Proceedings of the AAAI conference on artificial intelligence. 2019, 33(01): 5668-5675.doi: 10.1609/aaai.v33i01.33015668
33. Liu L, Qiu Z, Li G, et al. Contextualized spatial–temporal network for taxi origin-destination demand prediction. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(10): 3875-3887.doi: 10.1109/TITS.2019.2915525