

Inatorial forecasting method considering macro and micro characteristics of chaotic traffic flow

Yue Hou(侯越), Di Zhang(张迪), Da Li(李达), and Ping Yang(杨萍)
Citation: Chin. Phys. B, 2023, 32 (**10**): 100508. DOI: 10.1088/1674-1056/acd3df
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Modeling differential car-following behavior under normal and rainy conditions: A memory-based deep learning method with attention mechanism

Hai-Jian Bai(柏海舰), Chen-Chen Guo(过晨晨), Heng Ding(丁恒), Li-Yang Wei(卫立阳), Ting Sun(孙婷), and Xing-Yu Chen(陈星宇)
Chin. Phys. B, 2023, 32 (**6**): 060507. DOI: 10.1088/1674-1056/acaa2f

Bifurcation analysis of visual angle model with anticipated time and stabilizing driving behavior

Xueyi Guan(管学义), Rongjun Cheng(程荣军), and Hongxia Ge(葛红霞)
Chin. Phys. B, 2022, 31 (**7**): 070507. DOI: 10.1088/1674-1056/ac5606

Traffic flow prediction based on BILSTM model and data denoising scheme

Zhong-Yu Li(李中昱), Hong-Xia Ge(葛红霞), and Rong-Jun Cheng(程荣军)
Chin. Phys. B, 2022, 31 (**4**): 040502. DOI: 10.1088/1674-1056/ac3647

Lagrangian analysis of the formation and mass transport of compressible vortex rings generated by a shock tube

Haiyan Lin(林海燕), Yang Xiang(向阳, Hong Liu(刘洪), and Bin Zhang(张斌)
Chin. Phys. B, 2021, 30 (**3**): 030501. DOI: 10.1088/1674-1056/abc67f

An extended smart driver model considering electronic throttle angle changes with memory

Congzhi Wu(武聪智), Hongxia Ge(葛红霞), and Rongjun Cheng(程荣军)
Chin. Phys. B, 2022, 31 (**1**): 010504. DOI: 10.1088/1674-1056/ac0eec

---

Inatorial forecasting method considering macro and micro characteristics of chaotic traffic flow

Yue Hou(侯越)†, Di Zhang(张迪), Da Li(李达), and Ping Yang(杨萍)

School of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, *China*

(Received 11 January 2023; revised manuscript received 15 April 2023; accepted manuscript online 10 May 2023)

Traffic flow prediction is an effective strategy to assess traffic conditions and alleviate traffic congestion. Influenced by external non-stationary factors and road network structure, traffic flow sequences have macro spatiotemporal characteristics and micro chaotic characteristics. The key to improving the model prediction accuracy is to fully extract the macro and micro characteristics of traffic flow time sequences. However, traditional prediction model by only considers time features of traffic data, ignoring spatial characteristics and nonlinear characteristics of the data itself, resulting in poor model prediction performance. In view of this, this research proposes an intelligent combination prediction model taking into account the macro and micro features of chaotic traffic data. Firstly, to address the problem of time-consuming and inefficient multivariate phase space reconstruction by iterating nodes one by one, an improved multivariate phase space reconstruction method is proposed by filtering global representative nodes to effectively realize the high-dimensional mapping of chaotic traffic flow. Secondly, to address the problem that the traditional combinatorial model is difficult to adequately learn the macro and micro characteristics of chaotic traffic data, a combination of convolutional neural network (CNN) and convolutional long short-term memory (ConvLSTM) is utilized for capturing nonlinear features of traffic flow more comprehensively. Finally, to overcome the challenge that the combined model performance degrades due to subjective empirical determined network parameters, an improved lightweight particle swarm is proposed for improving prediction accuracy by optimizing model hyperparameters. In this paper, two highway datasets collected by the Caltrans Performance Measurement System (PeMS) are taken as the research objects, and the experimental results from multiple perspectives show that the comprehensive performance of the method proposed in this research is superior to those of the prevalent methods.

Keywords: traffic flow prediction, phase space reconstruction, particle swarm optimization algorithm, deep learning models

PACS: 05.60.–k, 05.45.–a, 07.05.Mh **DOI:** 10.1088/1674-1056/acd3df

1. Introduction

With the fast development of the urban road network, traffic congestion is becoming a hot issue of increasing concern. As a major research area in intelligent transportation systems (ITSs),[1–3] road network traffic data forecasting can provide some auxiliary decision bases for citizens to plan travel routes and reduce travel costs, which can alleviate traffic congestion to a certain extent. Affected by external uncertainties such as vehicle running state and adverse weather, traffic flow data exhibit nonlinear chaotic dynamics characteristics. At the same time, different spatial road network structures are differentially distributed, which makes the traffic flow at a macroscopic level possess complex spatial and temporal correlation performance. The key to modeling the chaotic traffic flow prediction is how to fully and effectively extract the macro and micro characteristics of traffic flow data.

Modeling chaotic traffic flow prediction is primarily comprised of macro and micro characteristic analyses and traffic flow prediction stages. Macro and micro analyses of features are the crucial basis for modeling chaotic traffic flow. Most of the existing studies analyze the macro spatiotemporal characteristics of data by embedding the time dimension and spatial dimension. Deng *et al.*[4] constructed the spatiotemporal matrix from the perspective of spatiotemporal correlation to realize the conversion of spatiotemporal characteristics of road networks to Euclidean space. In terms of microscopic chaotic characteristics, the main focus is placed on phase space reconstruction technique. The phase space reconstruction algorithms are classified as two major categories. One category is to solve independently embedding dimension m and delay time τ, such as the G–P method,[5] the mutual information method,[6] and the pseudo-nearest neighbor method.[7] The other category is to combine the embedding window theory to simultaneously compute m and τ, like the C–C method.[8] Peng *et al.*[9] used the C–C method to calculate the delay time and embedding dimension of traffic flow sequence to portray the univariate phase space information of the traffic system with higher efficiency. For the phase space reconstruction technique, some studies have shown that models incorporating multivariate information of the system have better prediction effects than those considering only univariate information.[10,11] Consequently, multivariate phase space reconstruction is widely applied to system dynamics feature extraction. Ma *et al.*[12] conducted phase space reconstruction for average occupancy, average speed, and time series within the traffic system, respectively, and demonstrated that

†Corresponding author. E-mail: houyue@mail.lzjtu.cn

© 2023 Chinese Physical Society and IOP Publishing Ltd

considering multivariate approach has a better prediction effect than considering only single variable. The above research discussed the spatiotemporal data organization form of traffic flow and multivariate phase space reconstruction algorithm in the macro and micro feature analysis stage from multiple perspectives, which provides strong theoretical support for fully extracting the spatiotemporal characteristics of data and chaotic features in modeling the chaotic traffic flow prediction. However, there are still some important points that have not been taken into account in these studies. The authors in Ref. [4] showed the construction of the spatiotemporal matrix only from the spatiotemporal perspective and does not consider the realistic road network node correlation. In Ref. [12], based on solving the multivariate reconstruction parameters separately, the maximum embedding dimension and the minimum delay time are selected for multivariate phase space reconstruction. However, owing to the large number of realistic road network nodes, the method is difficult to apply to the multivariate phase space reconstruction of regional road networks. In addition, when using phase space reconstruction technique to characterize the nonlinear evolution of traffic system, the univariate phase space reconstruction method cannot evolve the changing pattern of multi-node traffic data, and the organization of reconstructing multi-node phase space one by one greatly increases the complexity of the model. Therefore, it is necessary to study data organization form that can efficiently characterize both macro and micro aspects of urban road networks.

Traffic flow prediction is mainly modeled by using statistical learning modeling techniques, traditional machine learning modeling techniques, and deep learning modeling techniques. The statistical learning techniques emerged earlier and mainly include the historical average (HA) method, autoregressive integrated moving average (ARIMA),[13,14] *etc*. Although these methods are simple to operate, they are ineffective in analyzing the nonlinear laws of traffic data. Compared with statistical methods, traditional machine learning methods have strong feature abstraction capabilities, mainly including support vector regression (SVR),[15,16] Bayesian networks,[17] and feedforward neural networks,[18,19] which can effectively extract the nonlinear laws of traffic data. With the generation of massive multi-source traffic data, the capability of machine learning modeling methods is inadequate in large-scale data feature extraction, while deep learning is extensively used owing to its complex feature abstraction capability. In deep learning modeling, recurrent neural network (RNN) is commonly utilized to capture temporal dependence characteristics in traffic data, but it is ineffective in extracting long-time dependence features. Long short-term memory (LSTM)[20] and gated recurrent unit (GRU)[21,22] can compensate effectively for RNN, but these methods have the drawback that data space features cannot be extracted effectively. To extract spatial features effectively, CNN[23] is widely applied. Zhuang and Cao[24] performed the extraction of space characteristics of traffic data by utilizing the CNN and enhanced the model prediction performance. Zhang *et al*.[25] utilized convolutional neural networks to extract spatiotemporal features from the spatiotemporal matrix of traffic flow and achieved better prediction results. Nevertheless, the single modeling approach is hard to effectively characterize the spatiotemporal correlation in traffic flow sequences. In recent years, therefore, combined methods of modeling the spatiotemporal traffic flow have been proposed. Zheng *et al.*[26] combined attention components and ConvLSTM to achieve short-term prediction of traffic flow by consistent extraction of spatiotemporal features of traffic flow. Li *et al*.[27] combined Bayesian theory and data based on phase space reconstruction of multi-source traffic data, and then utilized radial basis function neural networks to extract the internal chaotic features. Huang *et al*.[28] extracted the nonlinear features of chaotic time series by using CNN based on phase space reconstruction of chaotic time series and utilized GRU to learn temporaldependence in a tandem network manner. The above studies extracted temporal features, spatial features and chaotic features of time-series data by using single modeling and combined modeling, respectively, which provides a multi-angle modeling idea for macro and micro feature extraction of traffic data. However, there are still some points that have not been fully considered in these studies. In Ref. [24] CNN was utilized to extract the spatial features of traffic data, but the process ignored temporal features. In Ref. [25] the CNN was used to extract spatiotemporal features based on the construction of spatiotemporal matrix, which can capture the spatiotemporal correlation to a certain extent but is not sufficient for the extraction of time-series features. In Ref. [26], the spatiotemporal features of traffic data were extracted by taking the advantage of ConvLSTM consistent extraction of spatiotemporal features, but it is difficult to effectively ensure the stability of the model by considering only macro spatiotemporal features. In Ref. [27] the phase space reconstruction theory was utilized to investigate the chaotic characteristics of traffic flow, but did not consider the macroscopic spatiotemporal correlation features of traffic flow. In Ref. [28] a combined modeling approach was adopted to learn the chaotic features and temporal features on time series data, yet ignored macroscopic spatial features. For the macro spatiotemporal characteristics and micro chaotic characteristics of traffic flow data, most of the existing studies only consider the macro spatiotemporal characteristics and ignore the micro chaotic characteristics, or only model the micro chaotic characteristics and ignore the macro spatiotemporal characteristics, resulting in inadequate feature extraction. Therefore, in the aspect of multi-node macro and micro feature extraction for large-scale road networks, it is necessary to study the modeling strategies for

extracting spatiotemporal features and multi-node chaotic features in a comprehensive and efficient manner.

In addition, the traditional combinatorial modeling approach frequently relies on subjective experience to determine the model hyperparameters, however, with the complexity of the model parameters, this approach is prone to degrade the model performance. For such problems, the existing methods often utilize intelligent algorithms to search for model optimal parameters, and particle swarm optimization (PSO)[29] algorithm is more frequently used in model optimization because of its highly collaborative and efficient. Sheikhan and Mohammadi[30] used the PSO to search for the optimal neuron number for multilayer perceptron and reduced the effect of subjective aspects on modeling. Mohamad *et al.*[31] utilized the PSO for optimizing the hyperparameters of feedback neural network and achieved better results. The aforementioned literature used the PSO for optimizing the parameters of complex neural network models, which effectively improves the generalization performance of complex models to a certain extent. However, the combination of PSO and deep learning networks increases the complexity of model training, and therefore highly time-efficient optimization strategies for PSO improvement need studying.

In summary, most of the existing traffic flow prediction modeling methods focus on spatiotemporal characteristics, and there are few studies on the micro phase space evolution characteristics of traffic flow, especially those studies about the chaotic traffic flow of regional road networks have not been found. Therefore, we believe that comprehensive research considering the spatiotemporal characteristics and chaotic characteristics of the regional road network is necessary. At the same time, efficient data organization of macro and micro characteristics and high optimization efficiency of the model are essential to fully extract the macro and micro characteristics of traffic flow. Therefore, this paper constructs a combined traffic flow prediction model based on the improved PSO optimization from the perspective of comprehensive consideration of spatio-temporal characteristics and chaotic phase space evolution characteristics, aiming to improve the traffic flow prediction accuracy and model stability, while enhancing the applicability of the model in realistic scenarios. Several important means applied to this research are listed below.

(i) To tackle the problem that multivariate phase space reconstruction constructs node phase space one by one, leading the algorithm to take longer time, an improved multivariate phase space reconstruction idea of filtering global representative nodes is proposed to realize the construction of data organization forms based on microscopic chaos.

(ii) To address the problem that the combined modeling approach is difficult to effectively extract the more adequate data features at both macro and micro levels, this paper proposes a CNN-ConvLSTM combined modeling method, which reveals the complex features of chaotic traffic flow from different perspectives and can effectively extract the hidden correlation characteristics between features.

(iii) For improving the time efficiency of PSO in optimizing the combined model, this research proposes a lightweight particle swarm optimization (ALPSO) idea for the challenge that model performance degrades, owing to subjective empirically determined parameters.

The remainder of this research is organized below. Section 2 describes each submodule of the combinatorial model. Section 3 introduces experimental data sources, and conducts experimental simulations from multiple perspectives. Some conclusions are drawn from this research in Section 4.

2. Method of CNN-ConvLSTM-ALPSO (CCA) model construction

The components of the CNN-ConvLSTM-ALPSO chaotic traffic flow intelligent combination prediction model proposed in this research are shown in Fig. 1, whose main modules are described below.

(I) **Chaotic feature extraction module** Firstly, the improved multivariate phase space reconstruction (IMPSR) algorithm is used to reconstruct phase space of multiple nodes data of the road network, then the chaotic features from the reconstructed phase space tensor are extracted by utilizing CNN, and finally the chaotic features are input into the feature connection layer for stitching with macroscopic spatiotemporal features.

(II) **Spatiotemporal feature extraction module** Firstly, the time series with different periods are constructed as spatiotemporal matrix (STM), then the spatiotemporal features in the matrix are extracted by ConvLSTM, and finally, the model output is obtained by inputting the spliced spatiotemporal characteristics and chaotic characteristics into the fully connected network.

(III) **Improved particle swarm algorithm optimization module** The model hyperparameters are regarded as an individual particle, and the error value of the validation dataset is treated as an algorithmic fitness value. The optimization process of combined model hyperparameters is realized by repeated iterations.

In this paper, the detector nodes at different locations within the same road network are considered as system multivariate parameters. Assuming that the number of detector nodes in the real road network is N, the L-dimensional multivariate traffic data sequence is $x(1)$, $x(2)$, ..., $x(N)$, where $x(i) = \left(x_i^1, x_i^2, \ldots, x_i^L\right)$ is the traffic data sequence of node i and $Tf(t) = (x_1^t, x_2^t, \ldots, x_N^t)$ denotes the traffic flow data of the N nodes at moment t. In Fig. 1, $\text{Input}_{\text{mic}} = (x(1), x(2), \ldots, x(N)) \in R^{(L,N)}$ indicates the traffic flow data

of all nodes, $\text{Input}_{\text{week}} = Tf(\text{TP}_{\text{week}})$ is the traffic flow at the moment last week corresponding to the prediction time point, $\text{Input}_{\text{day}} = Tf(\text{TP}_{\text{day}})$ refers to the traffic flow at the moment last day corresponding to the prediction time point, and $\text{Input}_{\text{recent}} = (Tf(\text{TP}_{\text{recent_1}}), Tf(\text{TP}_{\text{recent_2}}), Tf(\text{TP}_{\text{recent_3}}))$ indicates the traffic flow 15 minutes before the first prediction time point. In order to describe the input data of the spatiotemporal feature extraction module more clearly, the time distribution of the predicted traffic flow data at the moment of TP_6 and the corresponding inputs are given in Fig. 2. Combining with the distribution of time points in the figure, it can be seen that when predicting the traffic flow data at the moment of TP_6, the inputs of the spatiotemporal feature extraction module are represented as $\text{Input}_{\text{week}} = Tf(\text{TP}_1)$, $\text{Input}_{\text{day}} = Tf(\text{TP}_2)$, and $\text{Input}_{\text{recent}} = (Tf(\text{TP}_3), Tf(\text{TP}_4), Tf(\text{TP}_5))$.

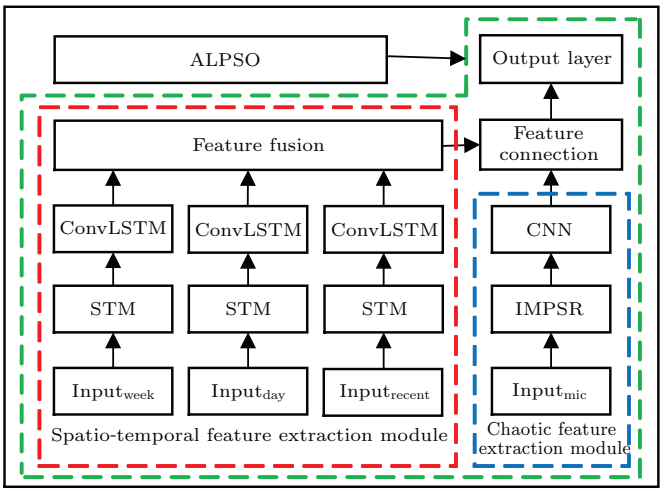


Fig. 1. Basic structure of CCA model.

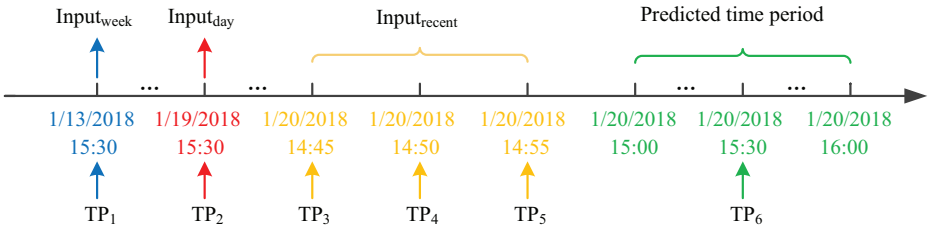


Fig. 2. Time point distribution of input data and predicted data.

2.1. Improved multivariate phase space reconstruction

In the real scenario, the traffic system has chaotic characteristics, such as some roads are congested while other roads related to them run normally, and the traffic flow varies unsteadily owing to the vehicles going and stopping on the congested road, *etc*., which are the manifestations of the chaotic characteristics of traffic flow. At the same time, the traffic system has the initial value sensitivity of a chaotic system, and a small behavior of a vehicle while driving may lead to traffic congestion, which in turn leads to abnormal changes in traffic flow on that road. For the chaotic dynamics characteristics exhibited among the data within the traffic flow, the internal nonlinear relationship cannot be revealed in a one-dimensional time series, and a phase space reconstruction is needed to unfold the structure of the singular attractor to analyze and discriminate the chaos. However, since the traditional phase space reconstruction method is primarily based on univariate systems, applying it to multivariate systems is likely to result in greater complexity, which puts forward higher requirements for the efficiency of algorithm. Therefore, this paper proposes an algorithm to construct the correlation matrix of the road network by combining the gray correlation coefficients among the detector nodes and proposes an improved multivariate phase space reconstruction algorithm based on this matrix to achieve the purpose of reconstructing the multi-node phase space of the road network more efficiently.

The IMPSR algorithm takes $\text{Input}_{\text{mic}}$ as the input data to reconstruct the phase space of the road network multi-detector node data. The detailed steps are as follows.

Step 1 The road network nodes data are normalized, and the gray correlation coefficient matrix is calculated based on Eqs. (1) and (2) to generate the gray correlation degree matrix, which is calculated as follows:

$$\boldsymbol{V} = \frac{\rho \cdot \max_j \max_k \left|x_i^k - x_j^k\right| + \min_j \min_k \left|x_i^k - x_j^k\right|}{\left|x_i^k - x_j^k\right| + \rho * \max_j \max_k \left|x_i^k - x_j^k\right|}, \tag{1}$$

$$\boldsymbol{A} = \frac{\sum_{k=1}^{L} \boldsymbol{V}_k}{L}, \tag{2}$$

where k represents the data series index, i and j are the node numbers, ρ denotes the resolution coefficient, $\boldsymbol{V}$ is the gray correlation coefficient matrix, and $\boldsymbol{A}$ refers to the gray correlation matrix.

Step 2 On the basis of generating the gray correlation matrix, the row count u and column count v of road network correlation matrix are calculated based on Eqs. (3)–(5) as shown below:

$$v \le 2 \cdot u, \tag{3}$$

$$u \le 2 \cdot v, \tag{4}$$

$$\min(u \cdot v - N) \ge 0. \tag{5}$$

To ensure that the topology of the road network nodes has a

strong correlation, firstly, any node of the road network is selected as the initial element of the matrix, and the most relevant node of the current node is found, with the row as the benchmark. Secondly, with the column as the benchmark, the relevant nodes are searched for on the basis of the row elements in the direction of the vertical axis, and finally, the road network correlation matrix with the strong correlation between rows and columns is obtained as follows:

$$\boldsymbol{C}=\begin{bmatrix} \varepsilon_{1,1} & \varepsilon_{1,2} & \cdots & \varepsilon_{1,v} \\ \varepsilon_{2,1} & \varepsilon_{2,2} & \cdots & \varepsilon_{2,v} \\ \cdots & \cdots & \cdots & \cdots \\ \varepsilon_{u1} & \varepsilon_{u,2} & \cdots & \varepsilon_{u,v} \end{bmatrix}, \tag{6}$$

where $\varepsilon_{u,v}$ denotes the serial number of road detector nodes after sorting using the row-column strong correlation rule.

Step 3 Based on the building of the road network correlation matrix, u global representative nodes are chosen on the counter-diagonal, and m and τ of the u nodes data are obtained by utilizing the C–C method respectively. By applying the BDS test statistic to the reasonable estimate of m and the search radius r, $S(T)$, $\Delta S(T)$, and $S_{\text{cor}}(T)$ are calculated from the following equations:

$$S(T)=\frac{1}{16}\sum_{m=2}^{5}\sum_{z=1}^{4}S(m,r_z,T), \tag{7}$$

$$\Delta S(T)=\frac{1}{4}\sum_{m=2}^{5}\Delta S(m,T), \tag{8}$$

$$S_{\text{sor}}(T)=\Delta S(T)+|S(T)|. \tag{9}$$

The 1st minimum value point for $\Delta S(T)$ or the 1st zero value point for $S(T)$ is taken as the optimal value of τ. The point of minimum value of $S_{\text{cor}}(T)$ is taken as the value of sliding window, and the m is obtained according to the embedding window theory. Finally, the global representative node reconstruction parameters are solved by the mean rounding method, which is calculated from the following equations:

$$m=F\left(\frac{1}{u}\sum_{l=1}^{u}m_l\right), \tag{10}$$

$$\tau=F\left(\frac{1}{u}\sum_{l=1}^{u}\tau_l\right), \tag{11}$$

where $F(\cdot)$ denotes the mean rounding method.

Step 4 The phase space reconstruction tensor of all nodes in the road network is obtained by utilizing m and τ solved in Step 3, and the reconstructed phase point of all nodes at time t is represented below:

$$\begin{aligned}\boldsymbol{Y}(t)=&\left[\left[x_1^t,x_1^{t+\tau},\ldots,x_1^{t+(m-1)\tau}\right],\ldots,\right.\\&\left.\left[x_N^t,x_N^{t+\tau},\ldots,x_N^{t+(m-1)\tau}\right]\right].\end{aligned} \tag{12}$$

The phase space tensor after the phase space reconstruction of all node data is shown as follows:

$$\begin{aligned}&[\boldsymbol{Y}(1),\boldsymbol{Y}(2),\ldots,\boldsymbol{Y}(t),\ldots,\boldsymbol{Y}(N-(m-1)\tau)]\\&\in R^{(N-(m-1)\tau,m,N)}.\end{aligned} \tag{13}$$

2.2. Deep learning methods

2.2.1. Structure of CNN model

Because of the relatively high complexity of the internal phase space structure for road network node data and the high dimensionality of the data, it is difficult to match the high-dimensional phase space data by using traditional neural networks suitable for processing one-dimensional time series. In addition, the data structure of phase space resembles that of an image matrix, therefore we can extract phase space characteristics by utilizing CNN, which is efficient in processing tensor data. The structure of the CNN used in this paper is depicted in Fig. 3.

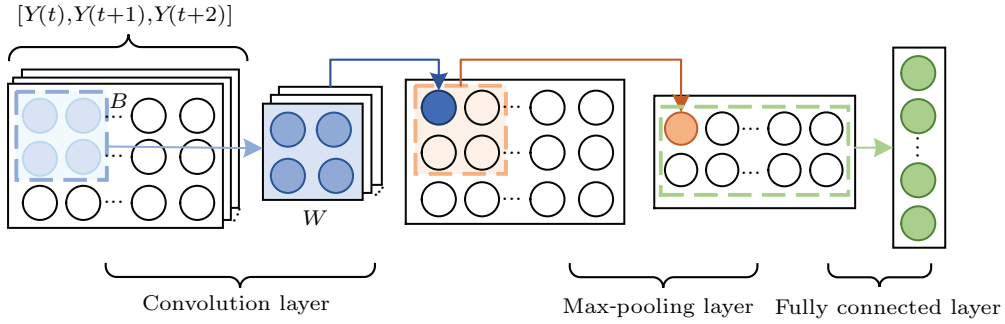


Fig. 3. Basic structure of CNN.

When extracting chaotic feature, redundant data may degrade the performance of the model and increase the complexity of the model. Therefore, we select only the three phase points closest to the predicted target value in the phase space tensor output from IMPSR as the input to the CNN, and the specific expression is $[\boldsymbol{Y}(t),\boldsymbol{Y}(t+1),\boldsymbol{Y}(t+2)]\in R^{(3,m,N)}$. On this basis, the expression $((t+2)+(m-1)\tau)<pre_1$ can better circumvent the problem of leakage of forecast data, where pre_1 denotes the time point corresponding to the first time slice in the forecast of future one-hour traffic flow. The CNN convolution kernel conducts convolutional operations on different regions in the phase space tensor one by one to ob-

tain the initial extracted spatial features, then the feature information is highly aggregated by multiple convolutional layer networks and finally, the extraction of phase space structure features is realized. The specific convolution calculation formula is given below:

$$\boldsymbol{Q} = \sigma(\boldsymbol{W} \odot \boldsymbol{B} + b), \tag{14}$$

where $\boldsymbol{Q}$ refers to the eigenvalue extracted by the convolution operation, b the bias value of the neurons, $\boldsymbol{W}$ the convolution kernel weight matrix, $\odot$ the Hadamard product operation, $\boldsymbol{B}$ the local phase space tensor involved in the convolution calculation, and σ the S-type function.

In Fig. 3, the main function of the max-pooling layer is to reduce the redundant chaotic features by retaining the maximum feature value of the sub-region, while the fully connected layer is mainly used to highly abstract the features after the max-pooling layer has been processed.

2.2.2. Description of ConvLSTM network

The temporal features and spatial features of traffic data are the critical factors in portraying the evolution of data, and whether the spatiotemporal characteristics of data can be adequately extracted is the key to determining the performance of the prediction model. Traditional combinatorial forecasting methods mostly form network models by serially connecting multiple components. Although these methods can extract part of the spatiotemporal information, they destroy the spatiotemporal correlation within the data, which is not highly time-efficient. ConvLSTM is proposed to better solve this problem, and it can efficiently extract spatial correlations while learning the temporal dependency of the data by embedding convolutional operations into the gating unit. The ConvLSTM internal cell structure in this research is shown in Fig. 4.

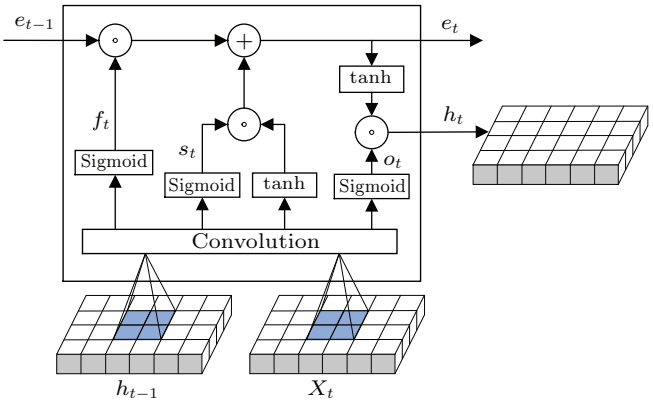


Fig. 4. Basic structure of ConvLSTM.

In Fig. 4, $\boldsymbol{X}_t$ denotes STM, whose structure is given below:

$$\boldsymbol{X}_t = \begin{bmatrix} x^t_{\varepsilon_{1,1}} & x^t_{\varepsilon_{1,2}} & \cdots & x^t_{\varepsilon_{1,v}} \\ x^t_{\varepsilon_{2,1}} & x^t_{\varepsilon_{2,2}} & \cdots & x^t_{\varepsilon_{2,v}} \\ \cdots & \cdots & \cdots & \cdots \\ x^t_{\varepsilon_{u,1}} & x^t_{\varepsilon_{u,2}} & \cdots & x^t_{\varepsilon_{u,v}} \end{bmatrix}, \tag{15}$$

where $x^t_{\varepsilon_{u,v}}$ denotes the traffic flow value for detector $\varepsilon_{u,v}$. If there is an empty value in the last row of STM, the mean value of the current column is chosen to fill it.

The ConvLSTM combines the gating units for fine-grained extraction of spatiotemporal characteristics of traffic data to ensure that the internal correlation of the original data is not split as much as possible, and the specific calculation process is shown as follows:

$$\boldsymbol{o}_t = \sigma\left(\boldsymbol{W}_{ox}\Theta\boldsymbol{X}_t + \boldsymbol{W}_{oh}\Theta\boldsymbol{h}_{t-1} + b_o\right), \tag{16}$$

$$\boldsymbol{f}_t = \sigma\left(\boldsymbol{W}_{fx}\Theta\boldsymbol{X}_t + \boldsymbol{W}_{fh}\Theta\boldsymbol{h}_{t-1} + b_f\right), \tag{17}$$

$$\boldsymbol{s}_t = \sigma\left(\boldsymbol{W}_{sx}\Theta\boldsymbol{X}_t + \boldsymbol{W}_{sh}\Theta\boldsymbol{h}_{t-1} + b_s\right), \tag{18}$$

$$\boldsymbol{e}_t = \boldsymbol{f}_t \circ \boldsymbol{e}_{t-1} + \boldsymbol{s}_t \circ \tanh\left(\boldsymbol{W}_{ex}\Theta\boldsymbol{X}_t + \boldsymbol{W}_{eh}\Theta\boldsymbol{h}_{t-1} + b_e\right), \tag{19}$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \circ \tanh\left(\boldsymbol{e}_t\right), \tag{20}$$

where $\circ$ denotes the element-by-element multiplication. $\boldsymbol{o}_t$, $\boldsymbol{f}_t$, and $\boldsymbol{s}_t$ are the output gate, input gate, and forget gate. $\boldsymbol{e}_t$ and $\boldsymbol{h}_t$ denote the cell information and hidden state information at time t, respectively. $\boldsymbol{W}_{sx}$, $\boldsymbol{W}_{sh}$, $\boldsymbol{W}_{fx}$, $\boldsymbol{W}_{fx}$, $\boldsymbol{W}_{ox}$, $\boldsymbol{W}_{oh}$, $\boldsymbol{W}_{ex}$, and $\boldsymbol{W}_{eh}$ are the weight matrix. b_s, b_f, b_o, and b_e refer to the bias of the neurons. Θ represents the convolution operation and $\boldsymbol{X}_t$ denotes the spatiotemporal matrix.

2.3. Improved lightweight particle swarm optimization

The PSO utilizes heuristic search ideas to obtain the best solution in the optimization scenario, which is widely used to optimize the parameters of traffic flow prediction models.[32–34] When performing parametric optimization, an individual particle is taken as a solution of the problem optimized, and position and velocity are two properties of the particle, which are updated below:

$$\begin{aligned} \boldsymbol{\xi}_q(n+1) &= \omega\boldsymbol{\xi}_q(n) + \eta_1\mu_1(\boldsymbol{p}_q(n) - \boldsymbol{\gamma}_q(n)) \\ &\quad + \eta_2\mu_2(\boldsymbol{g}(n) - \boldsymbol{\gamma}_q(n)), \end{aligned} \tag{21}$$

$$\boldsymbol{\gamma}_q(n+1) = \boldsymbol{\gamma}_q(n) + \boldsymbol{\xi}_q(n+1), \tag{22}$$

where ω is the inertia weight, η_1 and η_2 denote the learning factors, μ_1 and μ_2 are constants between 0 and 1, $\boldsymbol{p}_q(n)$, $\boldsymbol{\xi}_q(n)$, and $\boldsymbol{\gamma}_q(n)$ represent the individual historical optimal solution, velocity vector, and position vector of particle q at the n-th iteration, respectively, and $\boldsymbol{g}(n)$ refers to the optimal solution of the n-th generation particle swarm.

The traditional particle swarm algorithm has fixed values of ω, η_1, and η_2. It cannot dynamically change the particle search step, resulting in an imprecise solution. In view of this, this research proposes the ALPSO with adaptive update parameters, and its improved form is given as follows:

$$\mathrm{dist}(n) = \frac{\sqrt{\sum_{q=1}^{E}(\boldsymbol{\gamma}_q - \boldsymbol{g}(n))^2}}{E}, \tag{23}$$

$$\alpha = \begin{cases} \beta^* \log \dfrac{\text{dist}(n)}{\text{dist}(n-1)}, & n > 1, \\ 0, & n = 1, \end{cases} \tag{24}$$

$$\omega = \begin{cases} 0.4, & \omega \le 0, \\ \omega + \omega \cdot \alpha, & 0 < \omega < 1, \\ 0.85, & \omega \ge 1, \end{cases} \tag{25}$$

$$\eta_1 = 2 \times \omega, \tag{26}$$

$$\eta_2 = 2 - \eta_1, \tag{27}$$

where $\text{dist}(n)$ denotes the mean Euclidean distance among all particles of the n-th generation and the global optimal solution, E refers to the count of all particles, and β is the random number in a range from 0 to 1.

The above improvement methods can further enhance the efficiency of PSO, so this paper optimizes the combined model parameters based on ALPSO, and the detailed flow is depicted in Fig. 5.

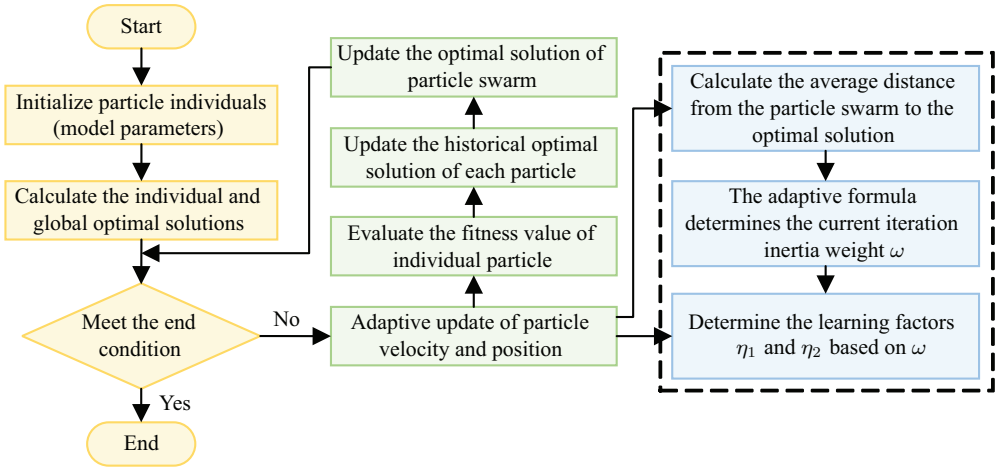


Fig. 5. Flow diagram of ALPSO optimization model.

3. Experiments

3.1. Dataset source

Two real highway data sets, PeMSD4 and PeMSD8, are used in this research as research objects, and the traffic parameters include traffic speed, traffic time occupancy, and traffic flow. When processing data, 5 min is taken as a time slice, and sum up the traffic flow every 5 min. Thus, 24 h of data can be split into 288 time slices of data. The PeMSD4 represents the traffic data dataset that is collected in the San Francisco Bay Area and contains 307 detectors spanning January to February 2018, with as many as 16992 time slices of data in total. The PeMSD8 represents the traffic dataset that is collected in the San Bernardino area and contains 170 detectors spanning from July to August 2016, with as many as 17856 time slices of data in total. The first 50 days of data from the dataset is a training dataset and the rest is a test dataset. The 300 time slices of data were selected from the training dataset as the validation dataset for assessing the model performance.

3.2. Experiment setups

In this research, we conduct model validation experiments in combination with the Pycharm experimental platform. In the micro chaos feature extraction module, this research sets the total number of network layers of CNN to 4 and assumes the convolution number and detector number to be equal. The convolution kernel has a dimension of 2 along both horizontal direction and vertical direction. The hyperparameters in the spatiotemporal features extraction module are optimized by using ALPSO, and the total number of particles and loop iteration updates are valued at 6 and 20 respectively in the optimization process. For model training, Adam is chosen as the optimizer and the model learning rate is taken as 0.0001. On this basis, the mean absolute error (MAE) of the training set is used as the feedback result of the model training to update the network weight parameters. To circumvent the problems of overtraining and overfitting of the model, we use the early stop method to control the model training process. Meanwhile, to make the balance between model prediction accuracy and training efficiency, the model training is terminated when the validation set error rises four times continuously, and the model structure with the lowest validation set error is saved. In this paper, traffic flow is predicted for 12 time slices in the next hour by using the time series of the previous week, the previous day, the recent 15 minutes, and the reconstructed phase space tensor data of the road network detector nodes.

To comprehensively validate the accuracy and generalization capability of CCA, we choose statistical models, deep learning modeling methods, and combined models, which are widely used in the traffic data modeling field, as our baseline models. The baseline models are introduced below.

(i) HA: predict the next time period by using the mean value of the time series historical data.

(ii) SVR: achieve prediction of regression problems by utilizing the idea of continuous optimization of model bias.

(iii) RNN: a neural network model widely used to capture

time series data dependence.

(iv) LSTM: an improved recurrent neural network used to extract long-time dependency features of data.

(v) GRU: an improved recurrent neural network with fewer network parameters compared with LSTM.

(vi) CNN-LSTM: CNN is used to extract chaotic characteristics of traffic data, meanwhile combining LSTM to capture temporal characteristics of chaotic sequence data.

(vii) Diffusion convolutional recurrent neural network (DCRNN): capture spatiotemporal features of data in a random wandering manner on the road network topology map.

(viii) Attention based spatial-temporal graph convolutional network (ASTGCN): it uses multiple component networks to more comprehensively capture the space features and time features in traffic data.

3.3. Evaluation of model prediction performance

To validate the comprehensive effect of the CCA, take MAE, R-square, and root-mean-square error (RMSE) as performance indexes. The specific calculation formulas are given below:

$$\mathrm{RMSE}(y,\hat{y}) = \sqrt{\frac{1}{Z}\sum_{I=1}^{Z}(y_I-\hat{y}_I)^2}, \tag{28}$$

$$\mathrm{MAE}(y,\hat{y}) = \frac{1}{Z}\sum_{I=1}^{Z}|y_I-\hat{y}_I|, \tag{29}$$

$$R\text{-square} = 1-\frac{\sum_{I=1}^{Z}(\hat{y}_I-y_I)^2}{\sum_{I=1}^{Z}(\bar{y}-y_I)^2}, \tag{30}$$

where Z represents the input sample's total number, y and $\bar{y}$ denote the true value, and its mean value, respectively, and $\hat{y}$ refers to the prediction result.

3.4. Experimental verification

3.4.1. Multivariate phase space reconstruction of chaotic traffic flow

Taking the PeMSD8 dataset as the research object, the road network correlation matrix of the PeMSD8 dataset is generated based on the road network correlation matrix construction algorithm proposed in Subsection 2.1.1, as shown in Fig. 6.

To verify that the nodes in the road network correlation matrix are strongly associated, 10 nodes and 4 nodes are respectively selected in the vertical axis region and local matrix region in Fig. 6 for node association analysis, and the results are shown in Fig. 7.

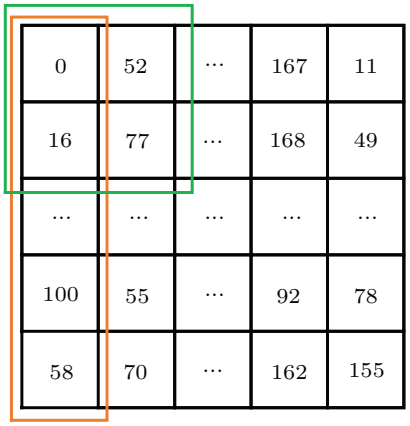


Fig. 6. PeMSD8 road network correlation matrix.

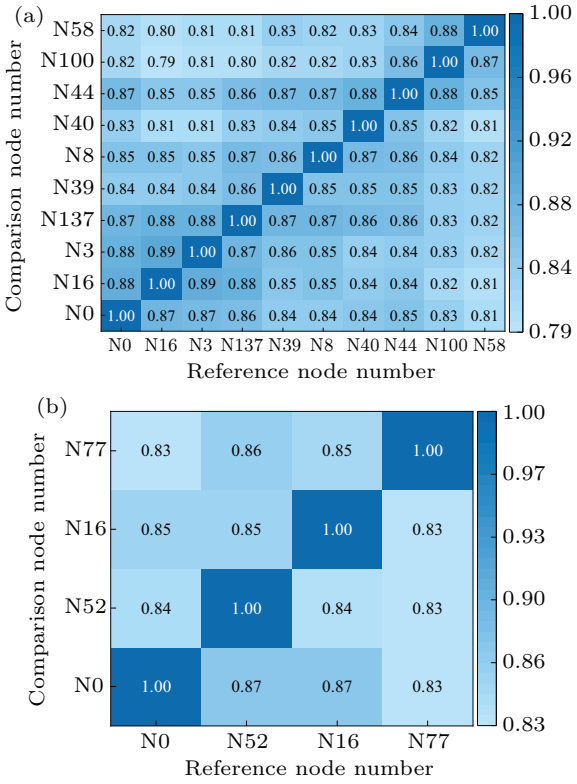


Fig. 7. (a) Vertical axis road network correlation and (b) matrix region nodes correlation of PeMSD8.

The horizontal coordinate in Fig. 7 indicates the reference node number and the vertical coordinate refers to the comparison node number. Combining the analysis of Fig. 6 with that of Fig. 7(a), it is evident that when node 0 is the reference sequence and the rest of the nodes are the comparison sequence, the correlation coefficient between node 0 and node 16 is as high as 0.87, while the correlation coefficients between node 0 and other nodes are all greater than or equal to 0.81, indicating that the space structure of nodes in the road network correlation matrix has a strong correlation. Combining the node data in Fig. 6 and the correlation coefficient analysis in Fig. 7(b), it is obvious that the correlation coefficient between node 0 and the neighboring nodes 16 and 52 in the horizontal direction and vertical direction are all 0.87, and the correlation coefficient between node 16 in the diagonal position and node 52 at

this time is 0.85, which further indicates that the nodes in different regions of the correlation matrix are strongly correlated. Since the node association mechanism is utilized to construct the road network association matrix, the nodes on the diagonal of this matrix also have a close association with the nodes in the upper triangle region and lower triangle region.

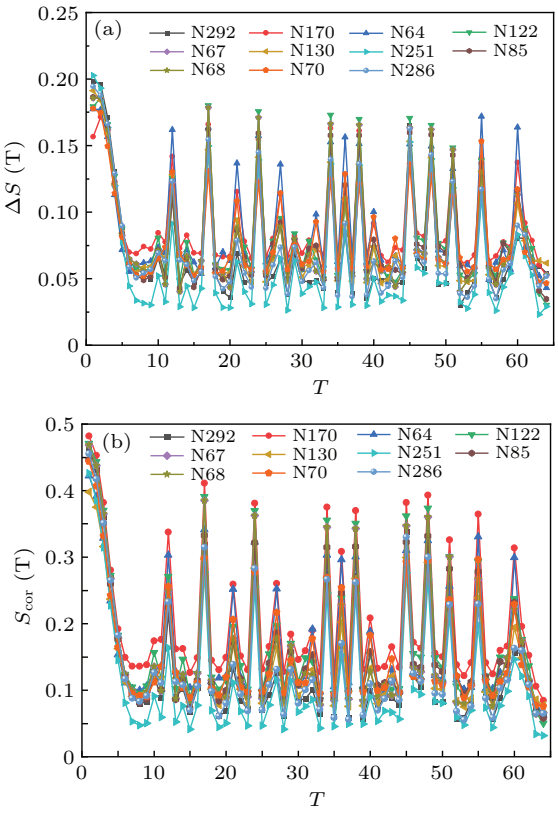


Fig. 8. Statistical data distribution of representative nodes of PeMSD4 road network: (a) $\Delta S(T)$ and (b) $S_{\rm cor}(T)$.

To optimize the reconstruction efficiency of the multivariate phase space, the IMPSR proposed in Subsection 2.2.1 is utilized to construct the phase space tensor of the road network nodes. Figures 8 and 9 exhibit the data distribution of the statistic $\Delta S(T)$ and $S_{\rm cor}(T)$ for the global representative nodes.

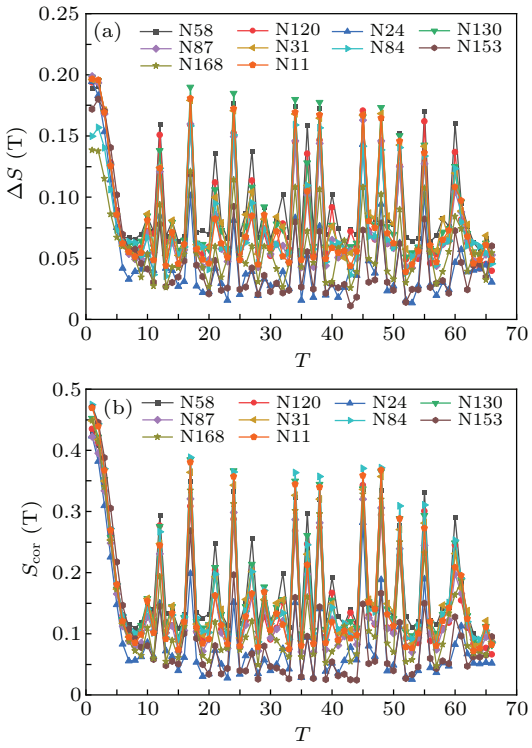


Fig. 9. Statistical data distribution of representative nodes of PeMSD8 road network: (a) $\Delta S(T)$ and (b) $S_{\rm cor}(T)$.

From the peaks and valleys of the data in Figs. 8 and 9, we can see that the distribution of the statistical curves is more similar to that of the representative nodes, indicating that the representative nodes have similar data change laws. On this basis, reconfiguration parameters m and τ of representative nodes are solved by applying C–C algorithm. Tables 1 and 2 give the detailed calculation results. Meanwhile, the mean rounding method is utilized for averaging reconstruction variables of representative nodes, and then the m and τ of all nodes are obtained. Table 3 lists the specific values.

Table 1. Phase space reconstruction parameters of PeMSD4 representative nodes.

	Global representative nodes									
	58	120	24	130	87	31	84	153	168	11
τ	8	8	7	8	8	8	9	11	9	8
m	9	9	8	9	6	9	7	5	7	6

Table 2. Phase space reconstruction parameters of PeMSD8 representative nodes.

	Global representative nodes										
	292	170	64	122	67	130	251	85	68	70	286
τ	9	7	7	8	9	7	9	8	9	8	7
m	7	10	10	9	8	8	8	9	8	9	8

From the data in Table 1, the maximum value of the reconstruction parameter m is 11 and the minimum value is 7 in the PeMSD4 dataset. Analysis of Table 2 indicates that the difference between maximum value and minimum value of m and τ for the representative nodes of PeMSD8 are 2 and 3, respectively, and their differences are small. Therefore, the

mean values of the reconstruction parameters of the representative nodes of the road network in Table 3 can be utilized to represent the values of m and τ for all nodes in this road network region, and this approach greatly improves the efficiency of multivariate reconstruction.

Table 3. Phase space reconstruction parameters of road network nodes.

	PeMSD4	PeMSD8
τ	8	8
m	9	8

To determine the chaotic features of PeMSD4 and PeMSD8 datasets, this paper conducts a phase space reconstruction of all node data based on the reconstruction parameters obtained. Then the largest Lyapunov exponent (LLE) is obtained by using the wolf[35] method to discriminate the chaotic characteristics of the traffic flow. When the largest Lyapunov exponent is a positive value, the phase point evolution law no longer follows the periodic law of continuous convergence of adjacent phase points, and the phase point trajectory in high-dimensional space obviously follows the non-periodic evolution law, and the dynamical system is considered chaotic at this time. Figures 10(a) and 10(b) represent the LLE for node data of PeMSD4 dataset and PeMSD8 dataset, respectively. The horizontal and vertical coordinates in the figures represent the node serial number and the largest Lyapunov exponent, respectively.

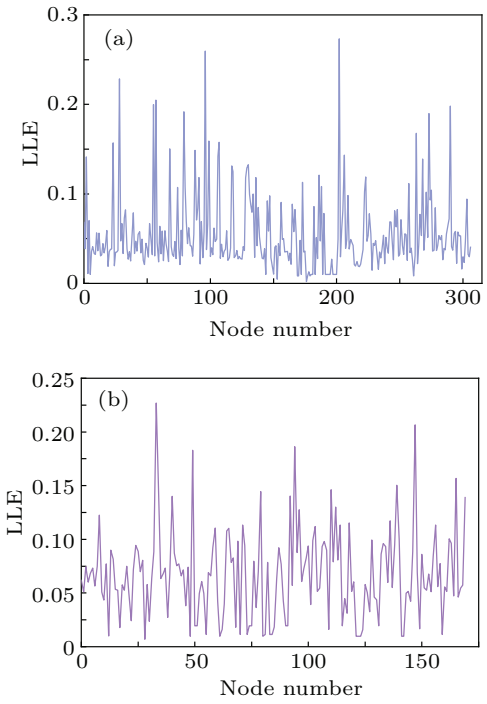


Fig. 10. Largest Lyapunov exponents of node data of (a) PeMSD4 and (b) PeMSD8.

From the data in Fig. 10, we can observe that the LLEs of all nodes in the PeMSD4 and PeMSD8 datasets are greater than 0, indicating that the data of each node in the road network all possess chaotic characteristics.

3.4.2. Model optimization experiments based on ALPSO

To verify the efficiency of the ALPSO algorithm, a one-dimensional vector optimal solution optimization problem is constructed. The $\boldsymbol{G}_1 = (a_1, a_2, \ldots, a_J)$ and $\boldsymbol{G}_2 = (a_1, a_2, \ldots, a_J)$ are assumed to represent the solution vector and the target vector, respectively, where J represents the total number of vector elements. The solution vector refers to the individual particle. Therefore, the optimization problem is expressed as $\boldsymbol{G}_1$ continuously approximating $\boldsymbol{G}_2$ through multiple iterative updates, and the algorithm stops iterating when $\boldsymbol{G}_1$ equals $\boldsymbol{G}_2$, which means that the particle finds the optimal solution. The experiment sets J to 10 and the particle count to 500. Meanwhile, the values of all elements of the target vector $\boldsymbol{G}_2$ are initialized to 2.2. The values of individual particles are randomly initialized in a range from -10 to 10. The optimization performances of different algorithms are given in Table 4.

Combination with the analysis in Table 4 evidently shows that the iteration number to searching for optimal solutions is 880 for traditional PSO and 681 for traditional IPSO, and their corresponding algorithm running times are 14 s and 10 s, respectively. The ALPSO finds the optimal solution in only 81 iterations, which corresponds to an algorithm running time of 2 s, less than all the compared algorithms. Therefore, the ALPSO algorithm can better balance the relationship between the local search step and the global search step for particles and has higher optimization efficiency.

Table 4. Optimal results from different PSO algorithms.

Algorithm	Total number of iterations	Running time (s)
Traditional PSO	880	14
IPSO[36]	681	10
ALPSO	81	2

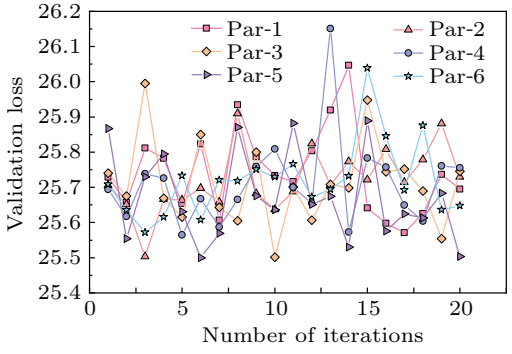


Fig. 11. Model optimization based on ALPSO.

Based on the verification that the ALPSO algorithm has better performance, this research achieves the model optimization process by taking the number of model input batches, ConvLSTM convolutional kernel number, and convolutional

kernel dimension as an individual particle. Figure 11 shows the validator error for 6 particles during 20 iterations. The analysis of the overall particle search process indicates that the particle Par-5 finds the best value of the model hyperparameters at the 6th iteration. At this time, the count for model input batches is 52, the number of convolution kernels for ConvLSTM is 51, and the convolution kernel along the horizontal axis and vertical axis have three dimensions.

3.4.3. Combined model experiments

In this paper, the optimization results of ALPSO are used to set some hyperparameters of the model. To compare average effect of CCA with that of the baseline model, two real data sets, PeMSD4 and PeMSD8, are used as the research objects. The error indexes of different models in predicting the next one-hour traffic flow are given in Table 5.

The analysis of the data in Table 5 indicates that the traditional traffic flow prediction models HA and SVR only extract simple linear features from traffic data, so the comprehensive performances of the models are relatively poor. The LSTM and GRU can well extract the time dependence in the data, and outperform RNN in average effect on both datasets. The analysis from the perspective of modeling micro chaotic features and macro temporal features indicate that the CNN-LSTM performance indexes are superior to the indexes of modeling approach with considering only chaotic sequence temporal features. Analysis from the perspective of spatiotemporal feature modeling shows that DCRNN and ASTGCN both can extract the spatiotemporal features and outperform CNN-LSTM in RMSE, but are slightly inferior to CNN-LSTM in MAE, indicating that spatiotemporal features and chaotic characteristics can enhance model prediction performance from different feature perspectives. Compared with the baseline model, the CCA model can fully extract the microscopic chaotic features while considering the spatiotemporal features. The RMSE, MAE, and *R*-square of CCA on PeMSD4 are 32.22, 19.43, and 0.932, respectively. The RMSE, MAE, and *R*-square of CCA on PeMSD8 are 22.94, 13.95, and 0.943, respectively, which outperforms the counterparts of all baseline models. In particular, in the PeMSD8 dataset, the RMSE and MAE of CCA are reduced by at least 9.5% and 12.4%, respectively, compared with those of the optimal baseline model. In the PeMSD4 dataset, the MAE of CCA is reduced by 4.9% compared with that of the ASTGCN model with better comprehensive performance in the baseline, and the *R*-square is improved by 1.3% compared with the counterpart of CNN-LSTM that only considers chaotic and temporal characteristics, further demonstrating that the CCA model achieves a fine-grained extraction process of macro and micro features of traffic data.

Figures 12 and 13 show the index curves of CCA and the baseline model in predicting the traffic flow of 12 time slices in the next hour. A comprehensive analysis of the index curves in the figure indicates that in the baseline model, the traditional statistical model HA has poor prediction performance and the SVR has improved prediction performance on the PeMSD8 dataset. The RNN, GRU, and LSTM only consider the time characteristics of data, and the performances of the three models are similar. The CNN-LSTM considers the chaotic characteristics and time characteristics, and the comprehensive performances of its index curves are superior to those of the model that only considers the temporal characteristics of chaotic sequences. The DCRNN and ASTGCN both consider the spatiotemporal characteristics which, in the first time slice, both have small RMSE values. With increasing time steps, the RMSE curve and MAE curve of the CCA model always increase at a relatively flat rate, which shows that the CCA model has better stability in predicting the future traffic flow at multiple time steps. Meanwhile, the RMSE, MAE, and *R*-square of the CCA model predicting the 12th time slice on the PeMSD8 dataset are 24.71, 14.99, and 0.928, respectively; and the RMSE, MAE, and *R*-square of the 12th time slice on the PeMSD4 dataset are 33.44, 20.17, and 0.927, respectively, which are superior to the counterparts of all other baseline models. The aboveanalyses demonstrate that the CCA model can extract the macro and micro features of traffic flow more completely and has better prediction performance in predicting the future traffic flow in multiple time slices.

Table 5. Comparison of performance between CCA and baseline.

Model	PeMSD4			PeMSD8		
	RMSE	MAE	*R*-square	RMSE	MAE	*R*-square
HA	49.09	31.48	0.807	42.63	28.10	0.743
SVR	41.47	28.43	0.766	28.98	20.12	0.843
RNN	36.60	21.90	0.913	29.00	17.93	0.913
GRU	35.69	21.10	0.917	25.92	15.93	0.929
LSTM	35.62	20.86	0.917	27.42	16.59	0.921
CNN-LSTM	35.05	20.69	0.920	26.56	16.41	0.926
DCRNN	34.28	22.27	0.916	25.63	16.73	0.928
ASTGCN	32.80	20.44	0.919	25.35	16.49	0.928
CCA	32.22	19.43	0.932	22.94	13.95	0.943

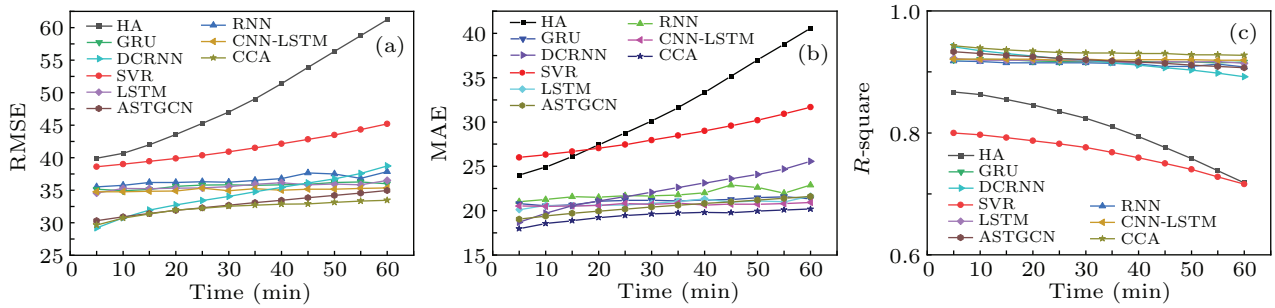


Fig. 12. Comparison of performance among different models on PeMSD4: (a) RMSE, (b) MAE, and (c) *R*-square.

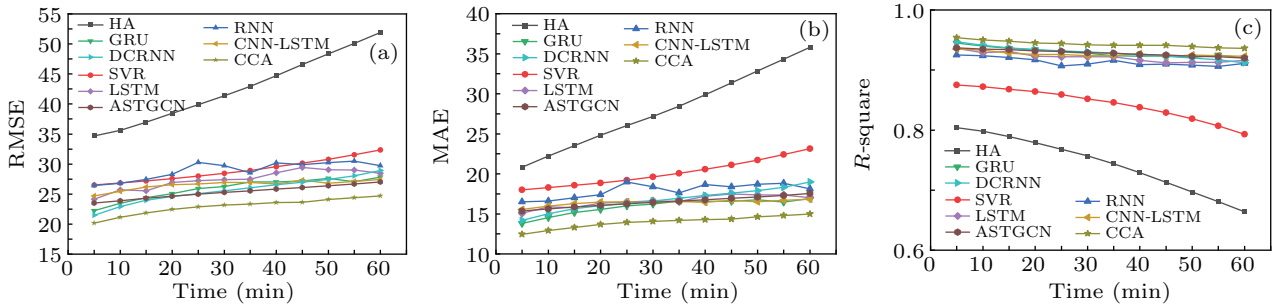


Fig. 13. Comparison of performance among different models on PeMSD8: (a) RMSE, (b) MAE, and (c) *R*-square.

3.4.4. Ablation experiments

To fully validate the feasibility and innovation of the modeling ideas in this research, this subsubsection takes the PeMSD8 dataset as the research object, and analyzes the ablation experiment of the model. Taking CCA as the base model, the model with removing the macro spatiotemporal feature extraction module is named CCA-ST, and the model with removing the micro chaotic feature extraction module is called CCA-M. Table 6 describes the index values for the ablation experiments.

Analysis of the data in Table 6 indicates that the CCA-ST removes the macro spatiotemporal feature extraction module of traffic flow and its RMSE, MAE, and *R*-square on the PeMSD8 dataset have higher values of 34.93, 22.83, and 0.881, respectively. The CCA-M removes the traffic flow microscopic chaotic feature extraction module, and its RMSE, MAE, and *R*-square on the PeMSD8 dataset are 24.27, 14.56, and 0.937, respectively. The RMSE and MAE of CCA decrease by 34.3% and 38.8%, respectively, and the *R*-square is improved by 7% compared with those of the CCA-ST, which further illustrates that the modeling idea of this paper can better improve the model overall performance by comprehensively considering the micro features and macro spatiotemporal features of chaotic traffic flow.

Table 6. Comparison of performance among variant models on PeMSD8

Model	PeMSD8		
	RMSE	MAE	*R*-square
CCA-ST	34.93	22.83	0.881
CCA-M	24.27	14.56	0.937
CCA	22.94	13.95	0.943

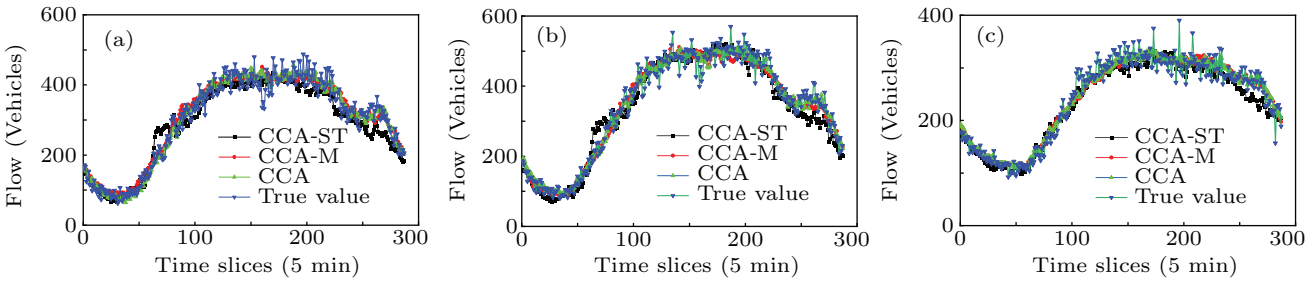


Fig. 14. Prediction results of different models in (a) detector 1, (b) detector 4, and (c) detector 131.

To further highlight the prediction performance of the proposed model in this research, we randomly select the predicted values of 288 time slices of detectors numbered 1, 4, and 131 in the test set and compare them with the corresponding true values, and the results are shown in Fig. 14. The combination of the data distribution of the true and predicted values in the figure shows that the true values maintain an ascending trend when the time slice is in a range from 50 to 120 and a descending trend when the time slice is in a range from 230 to 288, respectively. At this time, the predicted values of

CCA-ST on the detector numbered 1 and the detector numbered 4 deviate more from the true value, and the deviation of CCA-M decreases compared with that of CCA-ST, while the predicted value of CCA always fits the true value more closely, and it has the smallest overall deviation. The true value fluctuates more when the time slice is taken in a range from 150 to 220. At this time, CCA-ST deviation is still larger, especially in the detector numbered 4 it is more obvious, CCA-M floats above or below the middle boundary of the true value, while the predicted value of CCA as a whole remains at the middle boundary position of the fluctuation line, with better stability. The above analysis shows that the CCA model, which integrates spatiotemporal and chaotic features, has better prediction performance than the CCA-M and CCA-ST models, which consider features only from a single perspective.

4. Conclusions

By combining the nonlinear dynamic features and macro spatiotemporal features of chaotic traffic flow, this research proposes an intelligent combinatorial forecasting model that integrates the macro and micro characteristics of traffic flow. Through the analysis of CCA performance indexes from different perspectives, several reliable conclusions are drawn below.

(I) The ALPSO can better adaptively adjust the balanced relationship between the global search step and the local search step for particles. In a one-dimensional vector optimization problem, ALPSO only requires 81 iterations to find the optimal solution, outperforming all comparative algorithms.

(II) In the process of constructing the road network association matrix, it can be found that although the traffic network structure is unknown, the IMPSR can still filter the global representative nodes by constructing the road network correlation matrix, and then reconstruct the phase space tensor that can effectively portray the chaotic characteristics of traffic data.

(III) In comparative experiments on traffic data forecast, the CCA has the lowest error indexes on both PeMSD4 and PeMSD8 datasets. Taking the MAE as the evaluation metric, the CCA model reduces 4.5% and 12.4% respectively on the two datasets compared with the model with the best effect in the baseline and thus has superior prediction performance.

(IV) The analysis of the ablation experiment demonstrates that the CCA model outperforms the model that considers both macro spatiotemporal and micro chaotic characteristics alone. In particular, the RMSE and MAE indexes are reduced by 34.3% and 38.8%, and the R-square is enhanced by 7% compared with the counterparts of the CCA-ST model which only considers the chaotic characteristics, thus demonstrating the feasibility of this paper considering macro spatiotemporal characteristics for chaotic traffic data and micro chaotic characteristics comprehensively.

(V) Future work will further investigate the contributions of other traffic parameters like traffic speed and vehicle spacing in traffic data prediction, and introduce weather, holidays, and other external factors to enhance model robustness.

Acknowledgements

Project supported by the National Natural Science Foundation of China (Grant No. 62063014) and the Natural Science Foundation of Gansu Province, China (Grant No. 22JR5RA365).

References

[1] Zhang J P, Wang F Y, Wang K F, Lin W H, Xu X and Chen C 2011 *IEEE Trans. Intell. Transp. Syst.* **12** 1624
[2] Hamilton A, Waterson B, Cherrett T, Robinson A and Snell I 2013 *Transp. Plann. Technol.* **36** 24
[3] Poole A and Kotsialos A 2016 *Appl. Soft Comput.* **38** 134
[4] Deng S J, Jia S Y and Chen J. 2019 *Appl. Soft Comput.* **78** 712
[5] Grassberger P and Procaccia I 1983 *Physica D* **9** 189
[6] Fraser A M and Swinney H L 1986 *Phys. Rev. A* **33** 1134
[7] Kennel M B, Brown R and Abarbanel H D I 1992 *Phys. Rev. A* **45** 3403
[8] Kim H S, Eykholt R and Salas J D 1999 *Physica D* **127** 48
[9] Peng Y N and Xiang W L 2020 *Physica A* **549** 123913
[10] Zhang L Y, Wang G L, Tan G R and Wu Y 2022 *Acta Phys. Sin.* **71** 080502 (in Chinese)
[11] Zhang C T, Ma Q L, Peng H and Jiang Y Y 2011 *Acta Phys. Sin.* **60** 020508 (in Chinese)
[12] Ma Q L, Huang G H and Ullah S 2020 *IEEE Access* **8** 222774
[13] Williams B M and Hoel L A 2003 *J. Transp. Eng.* **129** 664
[14] Hamed M M, Al-Masaeid H R and Bani Said Z M 1995 *J. Transp. Eng.* **121** 249
[15] Castro-Neto M, Jeong Y S, Jeong M K and Han L D 2009 *Expert Syst. Appl.* **36** 6164
[16] Jeong Y S, Byon Y J, Castro-Neto M M and Easa S M 2013 *IEEE Trans. Intell. Transp. Syst.* **14** 1700
[17] Sun S L, Zhang C S and Yu G Q 2006 *IEEE Trans. Intell. Transp. Syst.* **7** 124
[18] Li C J, Xu Y, Yu X H, Ryan C and Huang T W 2017 *IEEE Trans. Ind. Inform.* **13** 2620
[19] Smith B L and Demetsky M J 1997 *J. Transp. Eng.* **123** 261
[20] Hochreiter S and Schmidhuber J 1997 *Neural Comput.* **9** 1735
[21] Zhang D and Kabuka M R 2018 *IET Intell. Transp. Syst.* **12** 578
[22] Dai G W, Ma C X and Xu X C 2019 *IEEE Access* **7** 143025
[23] Yang D, Li S J, Peng Z, Wang P, Wang J H and Yang H M 2019 *IEICE Trans. Inf. Syst.* **E102.D** 1526
[24] Zhuang W Q and Cao Y B 2022 *Appl. Sci.* **12** 8714
[25] Zhang W B, Yu Y H, Qi Y, Shu F and Wang Y H 2019 *Transportmetrica A: Transp. Sci.* **15** 1688
[26] Zheng H F, Lin F, Feng X X and Chen Y J 2020 *IEEE Trans. Intell. Transp. Syst.* **22** 6910
[27] Li Y F, Jiang X, Zhu H, He X Z, Peeta S, Zheng T X and Li Y G 2016 *Nonlinear Dyn.* **85** 179
[28] Huang W J, Li Y T and Huang Y 2020 *IEEE Access* **8** 159552
[29] Marini F and Walczak B 2015 *Chemometr. Intell. Lab. Syst.* **149** 153
[30] Sheikhan M and Mohammadi N 2013 *Neural Comput. Appl.* **23** 1185
[31] Mohamad E T, Armaghani D J, Momeni E, Yazdavar A H and Ebrahimi M 2018 *Neural Comput. Appl.* **30** 1635
[32] Hu W B, Yan L P, Liu K Z and Wang H 2016 *Neural Process. Lett.* **43** 155
[33] Yan H, Qi Y, Ye Q L and Yu D J 2022 *IEEE Trans. Intell. Transp. Syst.* **23** 14542
[34] Cai W H, Yang J J, Yu Y D, Song Y Y, Zhou T and Qin J 2020 *IEEE Access* **8** 6505
[35] Wolf A, Swift J B, Swinney H L and Vastano J A 1985 *Physica D* **16** 285
[36] Zhou Y, Wang N and Xiang W 2016 *IEEE Access* **5** 2241