

Received September 19, 2018, accepted October 2, 2018, date of publication October 16, 2018, date of current version November 14, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2875239

Identification and Prediction of Urban Traffic Congestion via Cyber-Physical Link Optimization

DARONG HUANG¹, (Member, IEEE), ZHENGPING DENG¹, SHAOHUA WAN², (Member, IEEE), BO MI¹, AND YANG LIU¹

¹School of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China

²School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073, China

Corresponding authors: Darong Huang (drhuang@mails.cqjtu.edu.cn) and Shaohua Wan (shaohua.wan@ieee.org)

This work was supported in part by the National Science Foundation of China (NSFC) under Grant 61573076, Grant 61663008, and Grant 61703063, in part by the Scientific Research Foundation for the Returned Overseas Chinese Scholars under Grant 2015-49, in part by the Program for Excellent Talents of Chongqing Higher School under Grant 2014-18, in part by the Science and Technology Research Project of Chongqing Municipal Education Commission of China under Grant KJ1600518, Grant KJ1705121, Grant KJ1705139, and Grant KJZD-K201800701, in part by the Chongqing Natural Science Foundation of China under Grant CSTC2017jcyjA1665, in part by the Innovation Foundation of Chongqing Postgraduate Education under Grant CYS18228, and in part by the Program of Chongqing Innovation and Entrepreneurship of Returned Overseas Scholars of China under Grant cx2018110.

ABSTRACT In this paper, to accurately predict and evaluate the traffic states in local urban areas, an effective as well as efficient algorithm combined with leakage integral echo state network (LiESN) and Pearson correlation analysis is presented. First, aiming at the parameter optimization of LiESN, the differential evolution algorithm is used to calibrate the key parameters. Second, the road weight allocation is carried out by Pearson correlation analysis considering the specific topology characteristics of local road nets. On this basis, the corresponding flow-frameworks are designed. Finally, the congestion delay data of local road nets in Chongqing Nan'an district, acquired from AutoNavi, are used to verify the validity and rationality of our method. Experimental results indicate that the proposed identification and prediction mechanism relates the congestion indicators to actual traffic operating mechanism in a more reasonable way.

INDEX TERMS Traffic congestion, traffic state, leakage integral echo state network (LiESN), Pearson correlation analysis, differential evolution algorithm.

I. INTRODUCTION

The evolution of urban road traffic states is a nonlinear, dynamic and complex process. Since most road topologies are heterogeneous and the operations of traffic are time-related, the congestion would in turn lead to change in travel way, travel routes and travel plans. Therefore, the distribution pattern of urban road nets is also dynamic [1], [2]. To maximize the availability of road nets and minimize the commute time of travelers, it is of great social significance to study the methods for urban road traffic states evolution and to establish a reasonable prediction model for traffic states regulation.

In recent years, how to accurately and effectively characterize the traffic states of urban road nets has been widely researched by scholars both in China and many other countries. Some outstanding research findings are reviewed as below. Zhao *et al.* have analyzed the traffic flow signals based on mountain road structure in [3]. They have formulated the representation of non-stationary signals from the traffic flow signals under sub-health states. In fact, Zhao and her

coauthors have provided a good theoretical support for mountain road traffic control. Moreover, she and her coauthors have constructed and established a functional relationship between traffic congestion sources and congestion evaluation points based on data visualization. To meet the needs of the research, they have introduced the spatial and temporal cumulative indicators of traffic congestion in [4]. Their experimental results demonstrated that the method can comprehensively evaluate the operation states of urban road nets on spatial scale. Either way, the case study and many projects have illustrated that the traditional microscopic congestion evaluation indicators may not fully reflect the operational states of road nets. Thus, to overcome the shortage, the characteristics and influencing factors of urban road traffic have been in-depth analyzed and chosen based on massive traffic data in [5]. In this, a reasonable and comprehensive evaluation algorithm of traffic jam have proposed to make auxiliary decision support by optimizing the traffic resources for traffic management.

However, with the development of information technology, how to accurately distinguish the traffic congestion level is key problem to achieve the precise of traffic inducement and control. Then, the K-means clustering method was introduced to classify all of traffic congestion patterns in terms of the vehicle driving state in [6]. To better represent and analyze the evolution process of the traffic state in real road network, the cluster density and attractivity were used to justify the traffic congestion pattern. Meanwhile, the real simulation examples were used to verify the effectiveness of the evolution of traffic states condition. But as we all know, the traffic congestion level of real road network is affected by the different factors such as travel time, numbers of stops, total delays and so on. Therefore, how to accumulate the related data is very important to improve the traditional identification performance of traffic congestion. Based on this, Bachman and his research group have presented the measuring method of the regional traffic congestion by used the travel data acquired from the Global Positioning System (GPS) in [7]. Specially, they pointed out that the congestion level and pattern were influenced by the precision of the measurement equipment and the significance of the regional area in whole road network. Further, considering that the traffic congestion' evolution of urban road traffic network was influenced by the multiple factors such as weather, holidays, major events and so on, Wei *et al.* have in detail analyzed and discussed the prediction and evaluation indexes of urban traffic performance based on K-nearest neighbor algorithm in [8]. To guarantee the effectiveness and rationality of overall evolution mechanism of traffic state, the key parameters of the evaluating algorithm were optimized combined with congestion delay index. Of course, the accuracy of traffic congestion prediction has been greatly improved.

In fact, the above research results indicated that the traffic jam state of road network was described in space and time. In general, the traffic signal is usually used to induce and control the traffic jam in urban road network. So, how to properly redeploy the traffic information resource is very important to ensure the smooth traffic under the digital network environment. In addition, the case study showed that the synthetical evaluation of traffic safety performance is the basic sector to ease traffic congestion of whole road network. Based on this, Huang *et al.* have presented that the selection of evaluation index and rule are the core problem of mitigating traffic jam in [9]. Moreover, if the evaluation model and algorithm of traffic conditions have been only constructed and established according to historical traffic data, it is difficult to recognize and simulate the future trend of traffic states evolution. Thus, the evaluating techniques, which have constructed according to static historical traffic-flow data, have some limitations in real traffic systems. Especially, with the popularization of vehicle-mounted terminals and intelligent traffic equipment, the kinds of traffic data are acquired and collected based on floating car data increase rapidly, and then the demands on quantity and quality of the data have also become higher than before for the traffic management in [8] and [10].

To meet the needs of people's outgoing, many major cities provide a real-time post about the traffic congestion state on web or other information media. This can bring great convenience to our lives and play an important role in aid decision-making support of traffic management. However, because of the diversification of transmitted mechanism of traffic information amongst heterogeneous traffic equipment, the instantaneous and strong fluctuation of traffic flow are usually triggered by the traffic emergent events such as the temporary change of the road network' structure, the traffic accidents, the severe weather, the heterogeneity of the trip distribution and so on.

On the other hand, urban road net can be regarded as Cyber-Physical Systems (CPS), which consists of several interconnected roads and the corresponding detect data of each road can be obtained by sensors. With the development of Intelligent Transportation System (ITS), one of the biggest challenges in urban transport system is how to predict and evaluate traffic states with multi-source data from local road network and road information network. Especially, in the multi-resources collaborative adjustment of traffic jam, the dynamic change trend of traffic congestion mode is high uncertainty in [11] and [12]. So, the analysis of multi-source and heterogeneous spatial data becomes more and more important in ITS. Moreover, it is difficult to abstract and quantify traffic data from the non-linear, dynamic and complex inter-connected cyber-physical system. In actual scenarios, the local traffic states are always influenced by the road layout structure in different extent. As we all know, different information structure of road network affects the activity of traffic capacity. Thus, finding a reasonable and effective method to analyze and predict the traffic trend is an urgent need to provide the rational basis for the traffic management. At the same time, it should be to ensure the normal operation of road network.

A. CONTRIBUTIONS OF THIS PAPER

Based on this, to better deal with multisource data, the Leakage Integral Echo State Network (i.e. LiESN) (see [13], [14]) is introduced to train the entire data set because of the testing ability to simplify the complexity of traditional training rule and overcome the local optimization of whole traffic state's change. Notice that LiESN not only has excellent feature learning abilities, but also improves the generalization performance and real-time prediction accuracy in [15] and [16]. Thus, the LiESN may accomplish the exchange and synchronization of heterogeneous data. In addition, considering that the multisource data will impact on the whole training performance of traffic evolution, the differential evolution algorithm is used to optimize the key parameters of LiESN for improving the ability of parallel disposal and unsupervised self-adaptive adjustment in real traffic systems. On this basis, the prediction algorithm and rule of urban traffic congestion states are described combined with the road-based topology relationship and correlation analysis. At the same time,

the corresponding flow-framework are designed and analyzed.

B. ORGANIZATION OF THIS PAPER

The layout of the rest of the paper is arranged as follows. In Section II, the Echo State Network (ESN) and LiESN were introduced briefly. Meanwhile, the differential evolution algorithm was used to optimize the key parameters of LiESN. Section III described the prediction algorithm and rule of urban traffic congestion states of local road network based on LiESN. In addition, the corresponding flow-framework was designed according to improved congestion index. In Section IV, some real traffic data from Chongqing-Based AutoNavi was used to verify and simulate the effectiveness and reasonability of the presented algorithm and rule. Finally, Section V concluded this paper with inferences and directions for future work.

II. INTRODUCTION OF LEAKAGE INTEGRAL ECHO STATE NETWORK

A. BASIC ECHO STATE NETWORK

As described in [17] and [18], the Echo State Network (ESN) is a new type of recursive neural network proposed by Jaeger and Haas. The advantage of ESN is that the mechanism of the “internal reservoir” is used to construct a direct-predicting rule, which relates the prediction origin and forecast horizon. The basic topological structure of ESN is shown in Fig. 1.

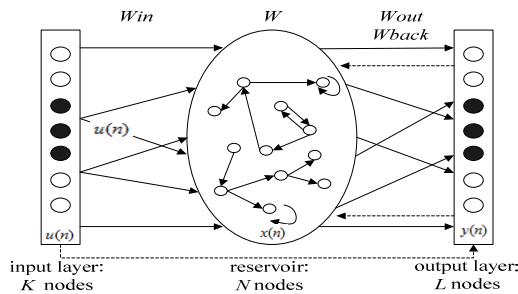


FIGURE 1. The structure of basic ESN.

In Fig. 1, the left cells constitute input layer of ESN, which is composed of K input nodes; the middle represents a reservoir network with N internal nodes and sparse connection weights; the right side stands for the output layer which is composed of L outlets. Notice that the solid lines indicate obligatory connections of the network and the dashed lines are optional (i.e. whether the links are activated or not depends on the different application situations).

Throughout this study, suppose that the input sequence at time t is described as $u(t) = [u_1(t), u_2(t), \dots, u_K(t)]^T$, the neuron state of reservoir is $x(t) = [x_1(t), x_2(t), \dots, x_N(t)]^T$, and the output is indicated as $y(t) = [y_1(t), y_2(t), \dots, y_L(t)]^T$.

In general, the state of the reservoir can be updated according to the following formula.

$$x(t + 1) = f(W_{in}u(t + 1) + \rho Wx(t) + W_{back}y(t)) \quad (1)$$

And, the output of ESN turns to be

$$y(t + 1) = g(W_{out}(u(t + 1), x(t + 1), y(t))) \quad (2)$$

Where, W_{in} , W and W_{back} are weight matrixes of input nodes, internal nodes of reservoir and output feedback link, respectively. $f(\cdot)$ represents the activation function of reservoir nodes, which is generally chosen as Sigmoid or tanh. Similarly, $g(\cdot)$ denotes the activation function of output layer, which is usually chosen as tanh or linear function. Notice that W_{in} , W and W_{back} will be not unchanged once they were generated stochastically during the training process of ESN, and then W_{out} converges along with training. Without loss of generality, the output weight value is usually calculated by ridge regression algorithm and model in [19]. The regression model is described as following.

$$W_{out} = (x^T x + \lambda I)^{-1} x^T y \quad (3)$$

Where, x denotes a matrix, which is composed of state vectors of reservoir, and λ is the regularization coefficient.

Obviously, at least in theory, the internal reservoir may produce diverse and nonlinear state space when the external input sequences activate the ESN. Thus, the output objective may be quickly achieved and generated excellent performance via a simple readout network. Moreover, the nonlinear features in the original space may be translated via the reservoir rules to linear features in the high-dimension reservoir state space. Then, to improve the reservoir learning algorithm, the nonlinear processing mechanism of reservoirs should be utilized for ESN.

B. LEAKAGE INTEGRAL ECHO STATE NETWORK

To revise and improve the reservoir learning algorithm and rule of ESN, the Leakage Integral Echo State Network (LiESN) was presented in [20]–[22]. Its most significant improvement is that the reservoir is composed of time-continuous leaky integral neurons group. In general, the continuous differential equation of dynamic neuron $x(t)$ can be defined as following.

$$\frac{dx(t)}{dt} = \beta(-\alpha x(t) + f(W_{in}u(t) + \rho Wx(t) + W_{back}y(t))) \quad (4)$$

Where, $\beta > 0$, $\alpha > 0$ are leakage rate parameters of the neurons of reservoir.

In real application, if the step is set up as 1, the Eq. (4) may be discretized and rewritten as follows.

$$x(t + 1) = (1 - \beta\alpha)x(t) + \beta f(W_{in}u(t + 1) + \rho Wx(t) + W_{back}y(t)) \quad (5)$$

Where $u(t)$ represents the discrete input of LiESN.

Remark 1: If let $\alpha = 1$, the Eq. (5) may be simplified as following.

$$x(t + 1) = (1 - \beta)x(t) + \beta f(W_{in}u(t + 1) + \rho Wx(t) + W_{back}y(t)) \quad (6)$$

Moreover, if $\alpha = 1$ and $\beta = 1$ are simultaneously meet, the Eq. (5) may be regarded as the basic ESN.

Notice that the leakage parameter β is used to control and ensure the integrality of neuron's information at the previous time. In fact, the leakage integral neurons may memorize the historical state trend. Thus, smaller β can make the reservoir state $x(t)$ change slowly, which enhances the ability of short-term memory of LiESN. Base on this feature, the advantage of improved version is that the activated function can be spreading at an exponential rate as time goes on. Obviously, the shortage of LiESN may be improved to a certain extent in the process of time series prediction. But, the selection of parameters is an important problem for the whole performance of LiESN in real engineering practice. In addition, the Eq. (5) and (6) reveal that W_{out} is also affected by other factors such as spectrum radius ρ , node number N of reservoir and sparse connectivity SD of reservoir. Thus, finding a reasonable algorithm to optimize the parameters of LiESN is urgent needs to solve in [23] and [24]. Next, we will analyze and discuss the question in next section.

C. LIESN PARAMETER OPTIMIZATION BASED ON DIFFERENTIAL EVOLUTION ALGORITHM

Due to that the reservoir of LiESN is consisted of group neurons, the group optimization algorithm should be used to optimize the parameters. In that case, the algorithm and rules should be easy to implement. Of course, the algorithm should be of a simpler structure and a stronger robustness. To solve the problem, the differential evolution algorithm was introduced to determinate the key parameters of LiESN combined with congestion delay index of road network.

In order to analyze conveniently, the j -th dimension of the i -th individual in the g -th generation is defined as $x_{ij}(g)$; the population size of group neurons is set up as NP ; the maximum evolutionary generation is depicted as T . F indicates the scale factor, which is used to control the influence extent of difference vector. The crossover probability is defined as cr . Especially, the search ranges of different dimension are described as x_{ij}^L and x_{ij}^U , respectively.

According to the feature of LiESN, the model and algorithm of differential evolution group may be depicted and constructed as following.

$$x_i(g) = (x_{i1}(g), x_{i2}(g), \dots, x_{in}(g))$$

$$i = 1, 2, \dots, NP; \quad g = 1, 2, \dots, T \quad (7)$$

Based on the idea of evolution, the differential evolution algorithm and rule may be designed as follows.

Step 1 (Initialization): initializing the basic parameters of differential evolution algorithm and search range of variables of LiESN. Setting the individual coding format in Table 1.

TABLE 1. Individual encoding format of differential evolution algorithm.

$x_i(g)$	β	N	ρ	SD
----------	---------	-----	--------	------

Step 2: Generating the initial population of LiESN. The initial model of individual evolution may be modeled in

n -dimensional space as following.

$$x_{ij}(0) = x_{ij}^L + rand(0, 1)(x_{ij}^U - x_{ij}^L)$$

$$i = 1, 2, \dots, NP; \quad j = 1, 2, \dots, n \quad (8)$$

Step 3: Calculating the fitness values of the individuals by using the fitness function $f(\cdot)$. In this step, the Mean Square Error (MSE) of LiESN during training process is chosen as the fitness function. The smaller the fitness value of individual, the better the individual.

Step 4: If the maximum times of evolutions generation reaches the upper bound T , the best individual is outputted and the evolution is immediately terminated. Otherwise, proceed to the next step.

Step 5: Mutation operation of reservoir. Selecting randomly 3 individuals $x_{p1}(k), x_{p2}(k), x_{p3}(k)$ at the k -th iteration, and $p1 \neq p2 \neq p3$. So, the mutated individual can be described as follows.

$$h_{ij}(k) = x_{p1j}(k) + F(x_{p2j}(k) - x_{p3j}(k))$$

$$i = 1, 2, \dots, NP; \quad j = 1, 2, \dots, n \quad (9)$$

Step 6: Crossover operation of reservoir. Increasing the diversity of the population of LiESN.

$$v_{ij}(g+1) = \begin{cases} h_{ij}(g), & rand(0, 1) \leq cr \mid j = rand(1, n) \\ x_{ij}(g), & rand(0, 1) > cr \mid j \neq rand(1, n) \end{cases} \quad (10)$$

Step 7: Selection operation. Selecting MSE as fitness function $f(\cdot)$ to compare $v_i(g+1)$ with $x_i(g)$.

$$x_i(g+1) = \begin{cases} v_i(g+1), & f(v_i(g+1)) < f(x_i(g)) \\ x_i(g), & f(v_i(g+1)) \geq f(x_i(g)) \end{cases} \quad (11)$$

Step 8: Increasing evolutionary generation and go to **Step 3**.

In the actual training, there is an obvious fact F and cr are very important to guarantee the performance of the algorithm. To solve the problem, the settings $F \in [0.5, 1]$ and $cr \in [0.8, 1]$ are recommended in [25]. Moreover, some scholars have proved that the setting of $F \geq 0.6$ and $cr \geq 0.6$ can leads to better performance in many application environments in [26] and [27]. Thus, to achieve the desired result of the different evolution algorithm, some sensible values of the key parameters are gotten through a lot of program simulations. These values are listed in Table 2.

Obviously, whether the search scope is reasonable or not is very important to optimize the parameters of LiESN by using different evolution algorithm. In theory, the size of reservoir should be large enough to maintain memory capacity, but have sparse connection SD [28]. At the same time, the leakage rate parameter β is used to control the reservoir's state information at the previous time. The amount settings is usually set as $\beta \in (0, 1]$. What is more, to ensure the echo state property (ESP) of LiESN, the spectrum radius is set as $0 \leq \rho < 1$ [15], [22]. Without loss of generality, in our application operation, the search scope of parameters of the LiESN is shown in Table 3.

TABLE 2. Parameter settings of differential evolution algorithm.

Parameter	Parameter Description	Value
T	the maximum evolutionary generation	400
cr	crossover probability	0.7
F	scaling factor	0.8
NP	population number	40

TABLE 3. Parameter search scope of leakage integral echo states network.

Parameter	Parameter Description	Search Scope
β	leakage rate	(0, 1]
N	reservoir size	[100, 1000]
ρ	spectrum radius	[0, 1)
SD	sparse connectivity	(0, 1]

III. IDENTIFICATION AND PREDICTION ALGORITHM OF URBAN TRAFFIC JAM WITH ROAD CONGESTION DELAY INDEX

A. PREDICTION MODEL OF SINGLE ROAD PATH WITH ROAD CONGESTION DELAY INDEX

To ensure the reasonability of congestion states prediction, the congestion delay index was used to describe the running state and changing rule of the road traffic network from the perspective of travel time cost. In general, the congestion delay index is defined as following.

$$CDI = \frac{T_{congestion}}{T_{free}} \tag{12}$$

Where $T_{congestion}$ denotes the travel time in congested states and T_{free} represents the travel time in the smooth flow of traffic.

In fact, the higher the ratio of the delay to the travel time, the more severe the traffic congestion. And then the congestion index is an important reference index in the stage of trip route selection demand and trip planning when users use the navigation. For examples, the length of every road is supposed to be constant, the travel time in congestion states and free flow states may be defined as below.

$$\begin{aligned} T_{congestion} &= S/v_{congestion} \\ T_{free} &= S/v_{free} \end{aligned} \tag{13}$$

Where, $v_{congestion}$ denotes the mean speed in the congestion states, v_{free} stands for the mean speed in the free-flow states and S indicates the length of single road.

Thus, the CDI can also be rewritten as following.

$$CDI = v_{free}/v_{congestion} \tag{14}$$

In fact, different roads have different traffic states and congestion index because of the complexity of transposition system. Therefore, it is necessary to learn and optimize the

network parameters of each road when LiESN model is concerned. To indicate and depict the concrete connotation of congestion identification by using LiESN, the basic flow chart of the LiESN prediction model may be described and designed as Fig. 2.

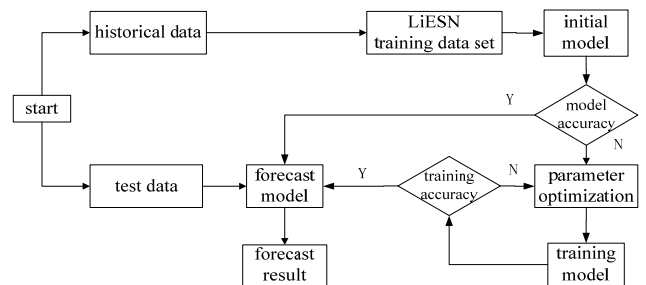


FIGURE 2. The prediction process of LiESN for single road Congestion delay index.

B. CONGESTION PREDICTION MODEL FOR URBAN LOCAL ROAD NETS

To the entire road network, the travel efficiency of trippers and urban residents are influenced by the traffic state of whole urban network. But at the same time, the urban traffic flow system is a human-joined, open and complex system. It is high nonlinear and uncertain. So, it is necessary to design a congestion prediction rule for the whole road network.

To effectively predict and regulate the traffic state of local area network and whole area network, the integrated identification and prediction algorithm should be improved and constructed based on the jam prediction rule of single path of road network. Obviously, how to compute the weight value of every path in road net is what should be solved. Meanwhile, the correlation coefficient between the traffic congestion levels of road path should be known. Next, we will discuss and analyze the basic index for the whole road system.

In actual road network, there are direct or indirect links between different road paths in road nets. Moreover, the road traffic flow can be imported and exported at intersections. So, the traffic state of local road network will be affected by the significance of every road-path in whole road nets. To overcome and solve the problem, the correlation coefficient was computed by following formula [29].

$$P_{ij} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (15)$$

Where, P_{ij} denotes the Pearson correlation coefficient of the j -th road and the i -th road, n denotes the number of data, x_i and y_i are the operation states data of the road section i and road section j , \bar{x} and \bar{y} are their averages, respectively.

Further, the road similarity is defined as follows.

$$Sup(R_i) = \sum_{j=1, j \neq i}^n P_{ij} \quad (i, j = 1, 2, \dots, n) \quad (16)$$

Where, $R_i(i = 1, 2, \dots, n)$ denotes the major traffic trunk road of the local road nets.

To identify the significance of single road path in whole road network, the weight coefficient is computed by the following formula.

$$\omega_i = \frac{Sup(R_i)}{\sum_{j=1}^n Sup(R_j)} \quad (i = 1, 2, \dots, n) \quad (17)$$

Obviously, the higher the value of $Sup(\cdot)$, the larger the weighted coefficient value. The results mean that this road-path is more significant in whole road network. Thus, we have designed the significant rule of the road-path as below.

Rule 1 (Significant Rule): If $\omega_i \geq \omega_j$, thus the i -th road-path is more significant in whole road network. Otherwise, the j -th road-path is more significant.

Base on the idea, the congestion delay index of whole road network had been defined and selected by some researchers in [30] and [31] as follows.

$$CDI_{nets} = \frac{\sum_{i=1}^n CDI_i}{n} \quad (18)$$

Where, CDI_i denotes the congestion states of single road-path, n denotes the number of road-path of local road nets.

Unfortunately, the urban transportation network system is usually a time-varying and non-linear system. In real traffic management, the congestion delay index need to be revised and improved according to the topological structure of local road nets. Therefore, the weighted coefficient between the traffic congestion delay levels of the different road-paths should be recomputed and revised according to the real traffic data. Obviously, the significance of single road path may be

calculated by Eq. (17). So, to better fit the real traffic state, the local congestion delay index may be rewritten as follows.

$$CDI_{nets} = \sum_{i=1}^n \omega_i CDI_i \quad (19)$$

In practice application, the correlation coefficient between different road paths may be computed according to historical traffic data. At the same time, to guarantee the effectiveness and accuracy of identifying and predicting congestion mode, the traffic congestion level may be divided to 5 sections including very smooth, smooth, light congestion, moderate congestion and severe congestion.

Based above analysis, the identification and prediction rule may be designed in the following pseudo code.

Algorithm 1 The Identification and Prediction Algorithm of Traffic States

```

in:  $X = \{x_1, x_2, \dots, x_n\}$ ,  $x_i$  denotes the CDI data set of  $i$ -th road;
out:  $CDI = \{CDI_1, CDI_2, \dots, CDI_n\}$ ,  $CDI_i$  denotes the predicted result of  $i$ -th road;  $CDI_{nets}$  denotes the CDI result of local road network.
1: for  $i \leftarrow 1$  to  $n$  do
2:    $t \leftarrow 0$ ;
3:   while  $t \leq T$  do
4:     Initial parameter of LiESN and EDA according to Table II and III;
4:     Get the training model according to Section II;
6:      $t \leftarrow t + 1$ ;
7:   end while
8:   Calculate the predicted output  $y_i$  based on Eq. (2);
9: end for
10: Calculate  $P_{ij}$  and  $Sup(R_i)$  based on Eq. (15) and Eq. (16);
11: Calculate similarity weight:  $\omega_i = \frac{Sup(R_i)}{\sum_{j=1}^n Sup(R_j)}$ 
    ( $i = 1, 2, \dots, n$ );
12: Calculate  $CDI_{nets}$  of local road nets:
     $CDI_{nets} = \sum_{i=1}^n \omega_i CDI_i$ ;
13: return  $CDI_{nets}$ .
    
```

The flow framework of identification and prediction is designed as Figure.3.

IV. SIMULATION EXAMPLE

A. EVALUATION INDEXES

To evaluate the effectiveness and rationality of presented algorithm, we had selected 5 indexes because the accuracy of the identification and prediction would directly influence of the real-time traffic optimal control. These evaluating indexes are shown as following.

(1) Mean Absolute Error

$$MAE = \frac{1}{n} \sum_{t=1}^n |X_t - \hat{Y}_t| \quad (20)$$

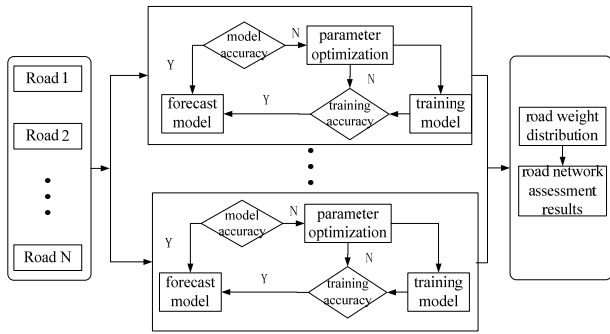


FIGURE 3. Identification and prediction framework of road nets based on LiESN.

(2) Mean Square Error

$$MSE = \frac{1}{n} \sqrt{\sum_{t=1}^n (X_t - \hat{Y}_t)^2} \quad (21)$$

(3) Mean Absolute Percentage Error

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - \hat{Y}_t}{X_t} \right| \quad (22)$$

(4) Mean Square Percentage Error

$$MSPE = \frac{1}{n} \sqrt{\sum_{t=1}^n \left(\frac{X_t - \hat{Y}_t}{X_t} \right)^2} \quad (23)$$

Where, X_t and \hat{Y}_t are the real value and predicted value of the congestion index at time t , respectively, and n stands for the number of samples of the congestion index.

B. PREDICTION OF ROAD CONGESTION DELAY INDEX

In our experiment, the real-time traffic data of Nan'an district of Chongqing Municipality in China have been selected to verify the actual prediction effectiveness of presented algorithm and rule. The data set have been acquired from AutoNavi and every data package contains the historical running data of each road for one month.

Firstly, to identify the performance of single road path by LiESN, the congestion delay index of Guangdian Road was selected to implement the corresponding algorithm. The topological structure of Guangdian road is shown as Figure 4.



FIGURE 4. The structure of Guangdian road (Map source: Google Maps).

The schematic diagram of the section observation is shown in Fig. 5. During the experiment, the real-time velocity of the floating vehicle within observation section can be transmitted back to the server. Then, the congestion index may be gotten according to the average detection speed and free flow velocity.

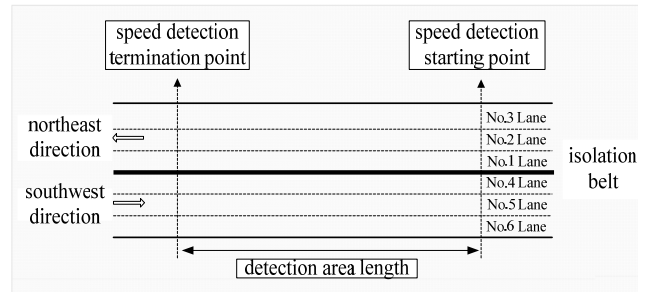


FIGURE 5. Road observation diagram.

After the original data has been processed by simulating real situations, the monthly trend of the congestion delay index from Feb. 13 to Mar. 13, 2016 is shown in Fig. 6.

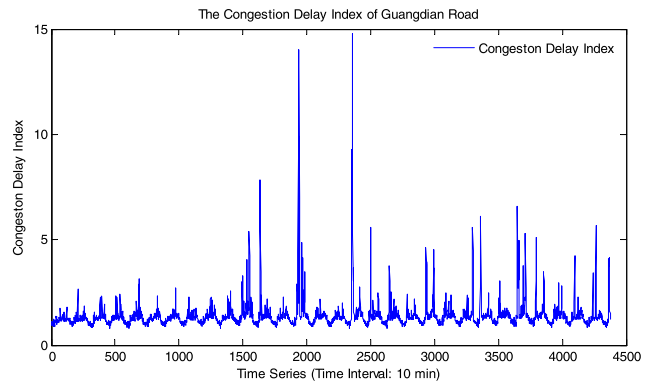


FIGURE 6. The monthly congestion delay index trend of Guangdian road in Nan'an district.

As we all know, with the rapid development of ITS and the increase of requirement of traffic information, all kinds of uncertainty factors increase. The road-based transport system has become more and more complicated. Especially, some fortuitous events will result in that the data fluctuation characteristics on working days and weekends are different. Through simulation and analysis, the trend of congestion delay index of Guangdian road in Nan'an district on working days and weekends from Feb. 13 to Mar. 13, 2016 are shown in Fig. 7 and Fig. 8, respectively.

Obviously, Fig. 7 and Fig. 8 show that the trend of congestion delay index is different on working days and weekends. The conclusion is that some data points on working days higher than weekends. This is because some uncertain events (such as traffic accident, rush hour commute lingered and so on) lead to traffic congestion.

Therefore, the main challenge and crucial problem are to ensure that LiESN have good learning ability and low training

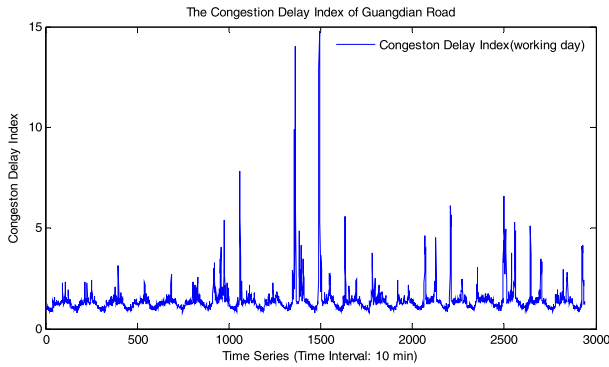


FIGURE 7. The daily congestion delay index trend of Guangdian road in Nan'an district (working days).

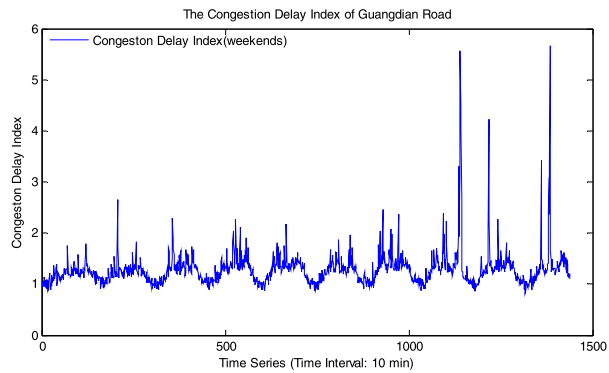


FIGURE 8. The daily congestion delay index trend of Guangdian road in Nan'an district (weekends).

error in different application environment. In other words, the evaluation premise of traffic states in local urban areas is that the congestion delay index of single road path may be accurately predicted and evaluated.

To verify presented algorithm and rule, the parameter settings of LiESN based on differential evolution algorithm for working days and weekends are shown in Table 4. Note that all the experiments are carried out on Intel(R) Core(TM) i5-3210M CPU@2.5GHz processor and Matlab R2012b.

TABLE 4. The parameter settings of LiESN for working days and weekends.

Item	N	β	ρ	SD
Working days	120	0.003	0.005	0.080
weekends	115	0.02	0.011	0.100

According to the prediction process of LiESN, the prediction result and evaluation index on Mar. 10~Mar. 11 2016 (working days) and Mar. 12~Mar.13, 2016 (weekends) are shown in Fig. 9, Fig. 10 and Table 5, respectively.

According the evaluation index from Table 5, the LiESN combined with differential evolution algorithm can achieve a good prediction result for both working days and weekends. At the same time, the result is also satisfactory from the

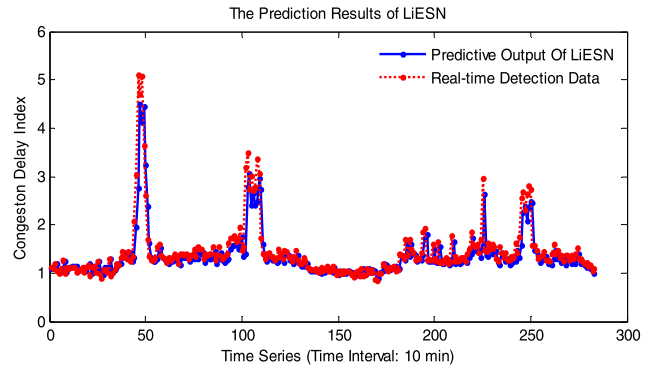


FIGURE 9. The prediction result of congestion delay index for Guangdian road (Mar. 10~11, working days).

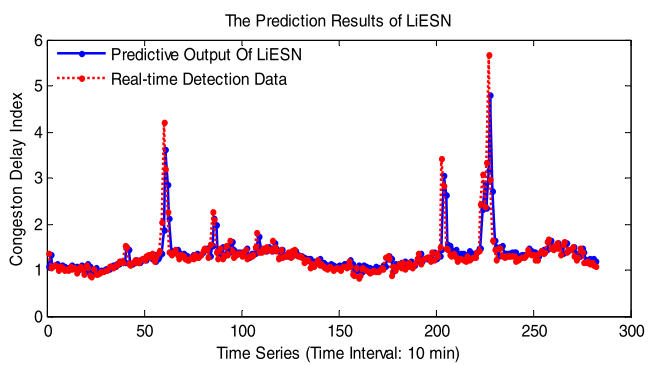


FIGURE 10. The prediction result of congestion delay index for Guangdian road (Mar. 12~13, weekends).

TABLE 5. The evaluation index of prediction result for single road path.

Date	Week	MAE	MSE	MAPE	MSPE
3.10~3.11	Thur., Fri.	0.1718	0.0184	0.1068	0.0091
3.12~3.13	Sat., Sun.	0.1437	0.0195	0.0872	0.0081

fitting-figures that shown in Fig. 9 and Fig. 10. Of course, the method can be applied to traffic management. In the next section, the states of local traffic network may be predicted and evaluated based on the correlation analysis.

C. PREDICTION OF LOCAL ROAD NET CONGESTION DELAY INDEX

To identify the congestion performance of the whole or local area, we have selected the local road network of Nan'an district to implement our algorithm. In fact, the traffic state and performance of the major road path should be computed in local road network. In order to facilitate the analysis, the congestion performance of local road network need to be integrated by using Eq. (19). Notice that the topological structure of local road nets is shown below.

In Nan'an District of Chongqing, the key avenues of the local region include Guangdian road, Huayuan road,

TABLE 6. The basic statistical information of local roads.

Road ID	Road Name	Road Level	Number of Lanes	Road Length (unit: m)	Free Speed (unit: km/h)
Road1	Guangdian road	Trunk Road	3	6905	44.686
Road2	Huayuan road	Trunk Road	3	3545	43.780
Road3	Nancheng tunnel	Trunk Road	2	5880	62.600
Road4	Nanping west road	Trunk Road	4	9725	42.502



FIGURE 11. Local road structure of Nan'an district (Map source: Google Maps).

Nancheng tunnel and Nanping west road. In Fig. 11, the major avenues are denoted with red broken lines. According to the data package provided by AutoNavi, the basic statistical information of the roads is shown in Table 6.

The urban road nets can be considered as cyber-physical link, which consists of several interconnected roads. However, it is also different from simple links. During the process of traffic flow aggregation in terms of human, vehicle, and environment interaction, the current traffic congestion states are often affected by previous moments in time and space scales. Thus, the correlation analysis based on historical data is necessary.

Moreover, according to Eq. (16), the correlation coefficient of major avenues of local road nets is shown in Table 7. At the same time, the similarity and its weight coefficient may be computed by using Eq. (17), i.e. as shown in Table 8.

TABLE 7. The correlation analysis of local road nets.

Road ID	Road1	Road2	Road3	Road4
Road1	1.0000	0.2562	0.4541	0.2190
Road2	0.2562	1.0000	0.3725	0.3785
Road3	0.4541	0.3725	1.0000	0.0258
Road4	0.2190	0.3785	0.0258	1.0000

TABLE 8. The similarity and weight of local road nets.

Item	Road1	Road2	Road3	Road4
similarity	0.9293	1.0072	0.8523	0.6233
weight	0.2724	0.2952	0.2498	0.1827

In addition, considering the diversity of traffic flow trend and the actual environment in each street, the parameters of LiESN have been optimized by the historical operation states.

The corresponding parameters of different avenues are shown in Table 9.

TABLE 9. The parameters of liesn for different roads.

Item	Road1	Road2	Road3	Road4
N	120	110	130	125
β	0.0030	0.0030	0.0180	0.0190
ρ	0.0050	0.0020	0.0030	0.0040
SD	0.0800	0.1200	0.1100	0.1400

To analyze and fit the actual trend of traffic congestion delay index, the prediction result for local road network from Mar. 10 to Mar. 11, 2016 is shown as Fig 12.

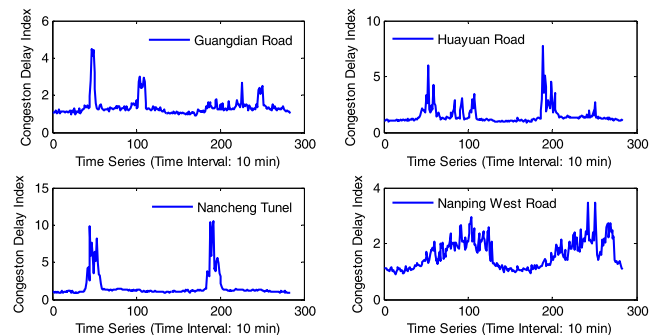


FIGURE 12. The prediction results of congestion delay index in local road network (Mar.10~Mar.11, 2016).

Due to that the actual traffic congestion mode of different roads are cross-linked and interplayed via cyber-physical link optimization, the congestion delay index of every road may be computed and simulated by Eq. (18) and Eq. (19). The experiment diagram is shown in Figure 13.

Obviously, as can be seen from Fig. 13, the performance of our algorithm is much better in the actual simulation. To visualize the overall congestion delay performance of identification and prediction, the congestion delay sequence diagram of major avenues in Nan'an district is also shown in Fig. 14.

Where, the horizontal axis represents the time axis and the vertical axis represents the synthesis predicted results of congestion delay index. In addition, different colors and shades stands for different degree of congestion.

In Fig. 13 and Fig. 14, the presented algorithm can effectively predict and evaluate the operation states of urban local road network. Therefore, the method proposed in this paper

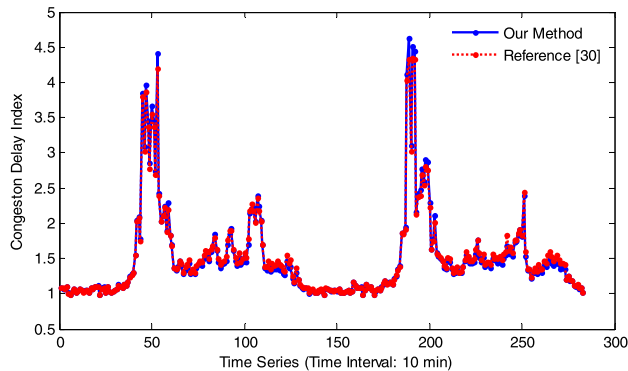


FIGURE 13. The synthesis evaluation result of local traffic congestion delay index (Mar.10~Mar.11, 2016).

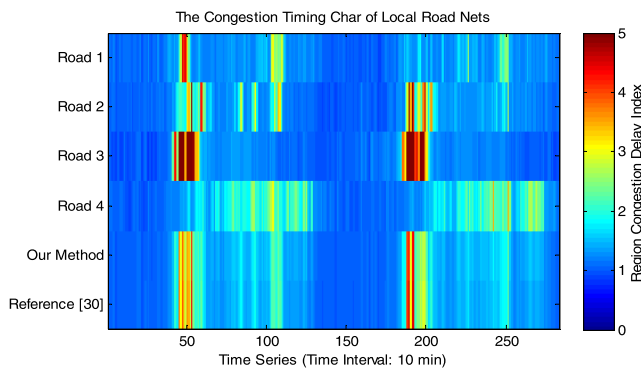


FIGURE 14. The congestion timing chart of Nan'an district (Mar.10~Mar.11, 2016).

can provide theoretical guidance for the regulation of traffic congestion in urban areas.

V. CONCLUSION

In this paper, the identifying and predicting algorithm of urban traffic congestion was depicted based LiESN via cyber-physical link optimization. Meanwhile, The simulation examples shown that the validity and rationality of presented method is very high. This indicates that the proposed identification and prediction mechanism can carry out the dynamic traffic management in real engineering.

However, owing to the fact that the future traffic congestion trend was identified and predicted according to the foregone traffic state and the current traffic state. The original data package should be processed by some new methods. Then, the state transition probability matrix may be introduced and applied in next study.

In fact, the trippers always hope that identification algorithm can forecast and recognize all congestion modes as precise as possible. However, due to the high uncertainty and non-linearity of urban traffic system, how to improve the performance and accuracy of the prediction and evaluation in the light of uncertain factors is also a hot research area. Although this study provides a solution for local road network, there is still a need to continually improve the accuracy of prediction result.

Last but not least, urban roads are not independent during the process of traffic state aggregation in terms of human, vehicle, and environment interaction, the current traffic congestion states are often affected by previous moments in time and space scales. Thus, how to fuse the incomplete and complete traffic information to analyze the congestion mode is also an important problem in urban transport system. All this will be our continuous research orientation.

ACKNOWLEDGMENT

The authors would like to thank the editors and reviewer for their insightful comments and suggestions on improving the quality of this publication. Also, they are grateful to AutoNavi Maps for its sharing the data used in the case study.

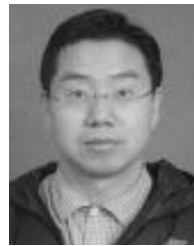
REFERENCES

- [1] S. Kitthamkesorn, A. Chen, and X. Xu, "Elastic demand with weibit stochastic user equilibrium flows and application in a motorised and non-motorised network," *Transp. A, Transp. Sci.*, vol. 11, no. 2, pp. 158–185, Feb. 2015.
- [2] D. R. Huang et al., "Traffic congestion status identification method for road network with multi-source uncertain information," *Acta Automatica Sinica*, vol. 44, no. 3, pp. 533–544, Mar. 2018.
- [3] L. Zhao, D. R. Huang, and J. Song, "Fractal characteristics of mountain cities' traffic flow with sub-health state," *Control Eng. China*, vol. 19, no. 4, pp. 583–586, Apr. 2012.
- [4] Z. C. He, Y. Q. Zhou, and Z. Yu, "Regional traffic state evaluation method based on data visualization," *J. Traffic Transp. Eng.*, vol. 16, no. 1, pp. 133–140, Feb. 2016.
- [5] Y. Y. Wang and Y. Huang, "Index system of traffic congestion evaluation in beijing based on big data," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 16, no. 4, pp. 231–240, May 2016.
- [6] L. L. Zhang, X. W. Li, and Z. L. Li, "An evaluation method of road operation condition based on two-stage K-means clustering," *J. Transp. Inf. Safety*, vol. 35, no. 5, pp. 99–105, Sep. 2017.
- [7] W. Bachman et al., "Household-level global positioning system travel data to measure regional traffic congestion," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2308, no. 1, pp. 10–16, Dec. 2012.
- [8] Q. B. Wei et al., "Prediction of urban traffic performance index considering multiple factors," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 17, no. 1, pp. 74–81, Feb. 2017.
- [9] D.-R. Huang, J. Song, S. Q. Li, and H. Y. Xiang, "Control technology review of traffic congestion in urban road network under networked dynamic scheduling and control," *J. Traffic Transp. Eng.*, vol. 13, no. 5, pp. 105–114, May 2013.
- [10] D. B. Alaigba, M. Soumah, and M. O. Banjo, "Heterogeneity index for the assessment of relationship between land use pattern and road traffic congestion in Apapa-Oworoshoki express way, Lagos metropolis," *Int. Arch. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 189–196, May 2017.
- [11] D. R. Huang, Z. P. Deng, and B. Mi, "A new synergistic forecasting method for short-term traffic flow with event-triggered strong fluctuation," *J. Control Sci., Eng.*, vol. 2018, no. 5, pp. 1–8, Mar. 2018.
- [12] D. Wang, W.-W. Che, H. Yu, and J.-Y. Li, "Adaptive pinning synchronization of complex networks with negative weights and its application in traffic road network," *Int. J. Control, Autom. Syst.*, vol. 16, no. 2, pp. 782–790, Mar. 2018.
- [13] L. C. Jiao, S. Y. Yang, F. Liu, S. G. Wang, and Z. X. Feng, "Seventy years beyond neural networks: Retrospect and prospect," *Chin. J. Comput.*, vol. 39, no. 8, pp. 1697–1716, Aug. 2016.
- [14] A. Graves et al., "Hybrid computing using a neural network with dynamic external memory," *Nature*, vol. 538, no. 7626, pp. 471–476, 2016.
- [15] H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," *Science*, vol. 304, no. 5667, pp. 78–80, Apr. 2004.
- [16] X. Luo, J. Li, and Z. Q. Sun, "Review on echo networks," *J. Univ. Sci. Technol. Beijing*, vol. 34, no. 2, pp. 217–222, Feb. 2012.

- [17] J. Qiao et al., "Growing echo-state network with multiple subreservoirs," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 2, pp. 391–404, Feb. 2016.
- [18] S. X. Lun and H. F. Hu, "Parameter optimization of leak integrator echo state network with internal-point penalty function method," *Acta Automatica Sinica*, vol. 43, no. 7, pp. 1160–1168, Jul. 2017.
- [19] P. Exterkate, P. J. F. Groenen, C. Heij, and D. Dijk, "Nonlinear forecasting with many predictors using kernel ridge regression," *Int. J. Forecasting*, vol. 32, no. 3, pp. 736–753, Jul./Sep. 2016.
- [20] H. Jaeger, M. Lukoševičius, D. Popovici, and U. Siewert, "Optimization and applications of echo state networks with leaky-integrator neurons," *Neural Netw.*, vol. 20, no. 3, pp. 335–352, Apr. 2007.
- [21] S. X. Lun, X. S. Yao, H. Y. Qi, and H. F. Hu, "A novel model of leaky integrator echo state network for time-series prediction," *Neurocomputing*, vol. 159, no. 1, pp. 58–66, Jul. 2015.
- [22] J. Li and W. Q. Yue, "Dynamic soft sensor modeling and its application using leaky-integrator ESN," *CIESC J.*, vol. 65, no. 10, pp. 4004–4014, Oct. 2014.
- [23] V. Santucci, M. Bairoletti, and A. Milani, "A differential evolution algorithm for the permutation flowshop scheduling problem with total flow time criterion," *IEEE Trans. Evol. Comput.*, vol. 20, no. 5, pp. 682–694, Jan. 2016.
- [24] Q. Fan, W. Wang, and X. Yan, "Differential evolution algorithm with strategy adaptation and knowledge-based control parameters," *Artif. Intell. Rev.*, vol. 2017, pp. 1–35, Jun. 2017.
- [25] J. Liu and J. Lampinen, "A fuzzy adaptive differential evolution algorithm," *Soft Comput.*, vol. 9, no. 6, pp. 448–462, Jan. 2005.
- [26] M.-Y. Cheng and D.-H. Tran, "An efficient hybrid differential evolution based serial method for multimode resource-constrained project scheduling," *KSCE J. Civil Eng.*, vol. 20, no. 1, pp. 90–100, Apr. 2015.
- [27] M. Leon and N. Xiong, "Adapting differential evolution algorithms for continuous optimization via greedy adjustment of control parameters," *J. Artif. Intell. Soft Comput. Res.*, vol. 6, no. 2, pp. 103–118, Mar. 2016.
- [28] P. Tanisaro and G. Heidemann, "Time series classification using time warping invariant echo state networks," in *Proc. IEEE Int. Conf. Mach. Learn. Appl.*, Dec. 2017, pp. 831–836.
- [29] W. Wiedermann and M. Haggmann, "Asymmetric properties of the Pearson correlation coefficient: Correlation as the negative association between linear regression residuals," *Commun. Statist.-Theory Methods*, vol. 45, no. 21, pp. 6263–6283, 2015.
- [30] X. L. Zhang, L. Yu, and M. H. Liu, "Traffic demand-based traffic congestion measurement models for road networks," *Mod. Transp. Technol.*, vol. 5, no. 6, pp. 71–75, Dec. 2008.
- [31] S. Wan, Y. Zhang, and J. Chen, "On the construction of data aggregation tree with maximizing lifetime in large-scale wireless sensor networks," *IEEE Sensors J.*, vol. 16, no. 20, pp. 7433–7440, Oct. 2016.



ZHENPING DENG was born in 1992. He received the B.S. degree in information engineering from the Chongqing University of Arts and Sciences, Chongqing, China, in 2016. He is currently pursuing the master's degree with the College of Information Science and Engineering, Chongqing Jiaotong University, Chongqing, China. His research interests include informationization of transportation and complex networks.



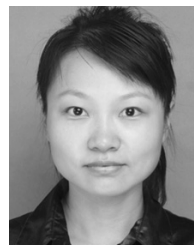
SHAOHUA WAN received the joint Ph.D. degree from the School of Computer, Wuhan University, and the Department of Electrical Engineering and Computer Science, Northwestern University, USA, in 2010. Since 2015, he has been holding a post-doctoral position at the State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology. From 2016 to 2017, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, Technical University of Munich, Germany. He is currently an Associate Professor and the Master Advisor with the School of Information and Safety Engineering, Zhongnan University of Economics and Law. His main research interests include massive data computing for Internet of Things and edge computing.



BO MI was born in 1982. He received the Ph.D. degree in computer system architecture from Chongqing University, China, in 2009. Since 2011, he has been an Associate Professor with the College of Information Science and Engineering, Chongqing Jiaotong University, China. His current research interests include intelligent transportation, vehicular ad hoc networks, and cryptography.



DARONG HUANG was born in 1978. He received the B.S. degree in applied mathematics from the Hubei National Institute, Hubei, China, in 2000, the M.S. degree in applied mathematics from Liaoning University, Liaoning, China, in 2003, and the Ph.D. degree in control theory and control engineering from Chongqing University, Chongqing, China, in 2006. Since 2011, he has been a Professor with the College of Information Science and Engineering, Chongqing Jiaotong University, Chongqing. His research interests include fault diagnosis and fault-tolerant control of dynamic systems, analysis and design of complex systems, big data analytic of transport systems, and reliability engineering.



YANG LIU was born in 1984. She received the Ph.D. degree in computer software and theory from the University Electronic Science and Technology of China, Chengdu, China, in 2013. Her research interests include formal verification, information safety, data processing, and fault-tolerant control.

• • •