Intelligent Group Prediction Algorithm of GPS Trajectory Based on Vehicle Communication

Guobin Chen, Lukun Wang, *Member, IEEE*, Muhammad Alam, *Senior Member, IEEE*, and Mohamed Elhoseny, *Member, IEEE*

***Abstract*— With the rapid development of in-vehicle communication technology and the integration of big data intelligent technology, intelligent algorithms for vehicle communication used to predict traffic flow and location information have been widely used. Aiming at the problem that the gravitational algorithm is difficult to minimize the complex function and easily fall into the local optimum, this paper proposes an improved IGSA algorithm. First, a gridding algorithm is introduced to initialize the population, and under the premise of ensuring the randomness of the initial individuals, improving the ergodicity of the population is conducive to improving the quality of the solution; then, an adaptive location-based update strategy of decreasing inertia weights is proposed. this strategy inherits the advantages of linearly decreasing weights, and adaptively adjusts the weights according to the fitness value to further improve the optimization performance. The optimization simulation of 8 classic test functions shows that the IGSA algorithm is an effective algorithm for solving complex optimization problems. Finally, the IGSA algorithm is used to predict the geographic location problem in the vehicle GPS data. The IGSA algorithm is used to optimize the extreme learning method to optimize the hyperparameters and establish a vehicle GPS data prediction model. Simulation results verify the feasibility of the method.**

***Index Terms*— Universal gravitational optimization algorithm, position prediction, GPS, extreme learning methods.**

I. Introduction

WITH the application of on-board communication technology and the highly integrated development of intelligent big data technology, the pace of smart city construction continues to accelerate and the development of the automobile industry, the popularity of family cars is getting higher and higher, resulting in many traffic problems, such as traffic jams, vehicle accidents, etc. In order to alleviate the traffic problems in the city, in addition to the conventional construction of roads to increase traffic volume, traditional emergency measures such as setting up road traffic barriers and markings can also be used to improve the smoothness of road sections so as to improve the traffic operation environment of roads, which only alleviates the congestion problem in the city in a short period of time. However, with the explosive growth of traffic demand, the traffic network system is increasingly complicated. It is not only costly and expensive to solve this thorny problem only by building roads and traffic control, but also the effect of these measures on easing traffic congestion and improving road capacity is not immediate. In order to solve the traffic problem, the participants of traffic work comprehensively apply computer, communication, control and other technologies to the traffic operation system to improve the efficiency of transportation, thus generating the intelligent transportation system [1]. Its essence is to make use of today's cutting-edge technology to make an attempt to find a better traffic operation system for the existing problems that cannot be solved [2]–[4].

In the field of trajectory data mining, Researchers have presented some research results, Includes a position-based recommendation system [5] to mine hidden motion laws [6] in animal tracks [7] Trajectory Data Mining for Road Network Drawing [8] Application of Parallel Computing in Large Number of Trajectory Data Mining [9] and Route Planning [10] In addition, The researchers also proposed the research direction of mining users' daily behavior patterns by using users' personal motion tracks. Such as personal semantic location (home, Company, etc.) found that [11] destination prediction [12], [13] trajectory prediction [14], public transportation condition prediction [15], [16] privacy protection [17] and other data mining based on user motion trajectory have great application value and practical significance in the field of smart city environmental monitoring and information protection [18]. In these application scenarios, the most critical step is vehicle location prediction.

Reference [19] excavates and extracts the hottest and most attractive shopping malls in the city on the basis of in-depth mining of taxi track record sets. Similarly, [20] establishes an algorithm to quantify the attraction of shopping malls using taxi track data. Through a full analysis of the historical track of urban taxis, researchers found that taxi drivers'

Manuscript received January 29, 2020; revised April 27, 2020; accepted June 3, 2020. This work was supported in part by the Science and Technology Research Program of Chongqing Municipal Education Commission under Grant KJZD-K201902101, in part by the Open Fund of Chongqing Key Laboratory of Spatial Data Mining and Big Data Integration for Ecology and Environment, in part by the Humanities and Social Sciences Project of Rongzhi College of Chongqing Technology and Business University under Grant 20197004, and in part by the National Natural Science Foundation of Shandong Province under Grant ZR2018BF005. The Associate Editor for this article was S. Mumtaz. *(Corresponding author: Lukun Wang.)*

Guobin Chen is with the Chongqing Key Laboratory of Spatial Data Mining and Big Data Integration for Ecology and Environment, Rongzhi College of Chongqing Technology and Business University, Chongqing 401320, China (e-mail: d150201001@stu.cqupt.edu.cn).

Lukun Wang is with the College of Intelligent Equipment, Shandong University of Science and Technology, Tai'an 271019, China, and also with the School of Cyber Science and Engineering, Southeast University, Nanjing 211189, China (e-mail: wanglukun@sdust.edu.cn).

Muhammad Alam is with the Department of Computer and Software Engineering, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China (e-mail: m.alam@xjtlu.edu.cn).

Mohamed Elhoseny is with the Faculty of Computers and Information, Mansoura University, Mansoura, Egypt (e-mail: Mohamed_elhoseny@mans.edu.eg).

Digital Object Identifier 10.1109/TITS.2020.3001188

income is closely related to the way they choose routes [21]. Reference [22] focuses on the relationship between the change rule of taxi track data with time and urban road traffic conditions, and designs a spatio-temporal three-dimensional track visualization model. Reference [23] In order to obtain a better path planning algorithm, based on in-depth analysis and mining of taxi track data, an empirical model for driver route planning is established, and on this basis, urban roads are classified to achieve better path planning. Reference [24] proposes a recommendation algorithm based on user trajectory clustering analysis on the basis of combining passenger interest with taxi GPS trajectory data, and verifies its accuracy through real taxi data sets. With the support of GPS historical track data set of urban taxis [25], the change of taxi track with time is analyzed, the functional definition of urban areas is carried out, and on this basis, a conventional taxi service recommendation model is provided. However, [26] predicts the urban road conditions in a short period of time on the basis of real-time taxi GPS records, and the experimental results show good results. Reference [27] focuses on the accurate passing time of the vehicle through a fixed area. In order to help the transportation department to find out the abnormal situation of road traffic in time and quickly and to carry out targeted guidance, [28] has carried out research on the identification of traffic flow abnormalities and road traffic congestion. Most of the methods proposed in the above paper focus on the research of GPS data used to predict commercial recommendations and route optimization, but there are differences in the effect of recommendations. Location information prediction is only aimed at some hot spots, and there is no research on non-hot spots. Therefore, the proposed GPS trajectory big data algorithm can accurately predict the position information.

Therefore, in order to accurately predict the vehicle trajectory, combined with the characteristics of GPS data trajectory big data, a research on the prediction algorithm of swarm intelligence algorithm in GPS trajectory data is proposed. The structure of this paper is as follows: 1) Describe the data source of vehicle data communication; 2) Introduce the basic GSA algorithm and the improved GSA algorithm; 3) Compare the performance of IGSA algorithm with other algorithms in function testing; 4) Use GPS large data set to predict the trajectory under IGSA algorithm and compare the experiments.

From the data processing center in Figure 1, data of different dimensions are processed, including data of different dimensions such as time, speed, distance and position. First of all, the original acquired data should be pre-analyzed: processing judgment, missing value filling, data transformation and data protocol. Then, the processed data are trained and predicted by the proposed model.

Practical Application of Vehicle Information Big Data Processing Platform, as mentioned earlier, the on-board information system has two attributes. It is not only used as a data acquisition unit, but also can accept services provided by the on-board information big data processing platform. Road navigation is a common function of on-board information systems. With the help of on-board information big data processing platform, better road navigation can be realized. For example, considering the impact of future traffic congestion on the best

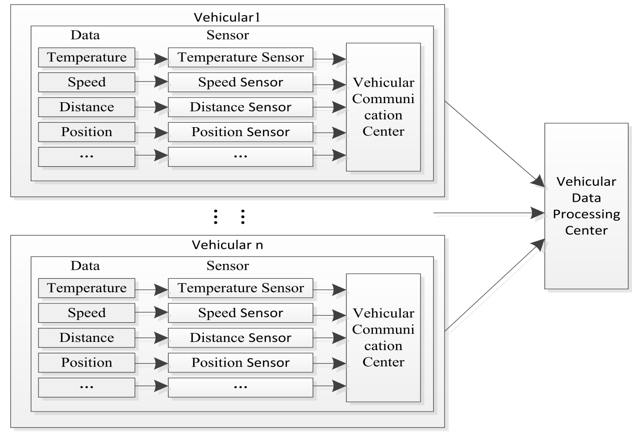


Fig. 1. Vehicle sensor data processing flowchart.

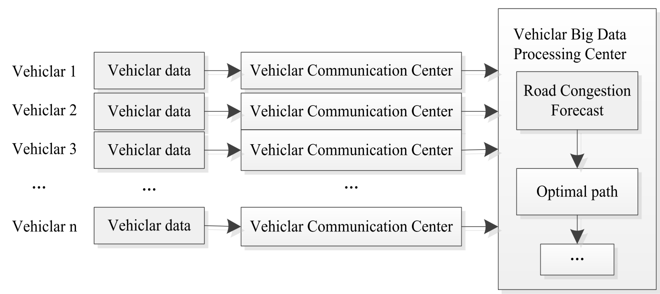


Fig. 2. Optimal processing process based on vehicle information path.

driving route, this is completely impossible for traditional on-board information systems, as shown in Figure 2. In this chapter, the application of vehicle-mounted information big data processing platform in road navigation is discussed in detail.

Road navigation is to recommend the driving route to the destination according to the current position of the user. The best road navigation refers to the best driving route. Generally, only the length of the path is considered in road navigation, and the real-time congestion of the road is not considered. In fact, road congestion will have a great impact on the route.

Based on these deficiencies, the weights that affect the driving route on each road are represented. The vehicle-mounted information big data processing platform has strong distributed computing capability, which enables us to better use distributed parallel programming model to predict road congestion. IGSA-ELM model is established to predict the location information of roads in the whole network, and road navigation is used under the vehicle-mounted information big data processing platform, which has a brand-new thinking mode.

II. Basic Principle of Improved GSA Algorithm

A. GSA Basic Algorithm

In GSA algorithm, the mass of a particle is calculated by the fitness value of the particle according to a certain relationship. The fitness value of a particle is closely related to its position. The better the position, the greater the fitness value of the particle and the greater it's mass. In GSA algorithm, the mass $M_i(t)$ of particle i at time t.

$$\begin{cases} m_i(t) = \dfrac{fit_i(t) - worst(t)}{best(t) - worst(t)} \\ M_i(t) = \dfrac{m_i(t)}{\sum_{j=1}^{N} m_j(t)} \end{cases} \tag{1}$$

where fit_i is the fitness value of the particle aby the particle at time t.

The formula for calculating the acting force of particle i on particle j in the d-dimensional space at time is as follows:

$$F_{ij}^{d}(t) = G(t)\frac{M_i(t) \times M_j(t)}{R_{ij}(t) + \varepsilon}\left(x_j^d(t) - x_i^d(t)\right) \tag{2}$$

where M_i and M_j are the masses of particles i and j at time t; A constant where ε is not 0;$R_{ij}(t)$ is the Euclidean distance in the population; $G(t)$ is that gravitational constant. It is defined that the resultant force of particle i in D-dimensional space is:

$$F_i^d(t) = \sum_{j=1, j\neq i}^{N} rank_j F_{ij}^d(t) \tag{3}$$

where $rank_j \in [0, 1]$.

At time t, the acceleration of particle i in the d-th dimensional space is $a_i^d(t)$:

$$a_i^d(t) = \frac{F_i^d(t)}{M_i(t)} \tag{4}$$

The velocity and latest position of particles in the population should be following.

$$\begin{cases} v_i^k(t+1) = rank_i \times v_i^k(t) + a_i^k(t) \\ x_i^k(t+1) = x_i^k(t) + v_i^k(t+1) \end{cases} \tag{5}$$

In this which, $rank_j \in [0, 1]$.

Acceleration a in formula (5) updates the velocity and position of the population in K-dimensional space. The new velocity is based on the previous velocity multiplied by $rank_i$ plus acceleration a, and the new position is the previous position plus velocity. It can be seen that acceleration in formula 5 is very important to be able to update the velocity and position.

B. Initialization Population Based on Grid Algorithm

The GSA algorithm must first initialize the population. Initializing all the space of the GSA algorithm is beneficial to improving the speed and quality. If GSA algorithm cannot cover all the space in a limited time, the algorithm will easily converge prematurely. In this paper, grid algorithm is used to initialize the population, and its expression is shown in formula (6):

$$x_i^k = min(k) + \frac{max(k) - min(k)}{n-1} \times i \tag{6}$$

where x_i^k is the Kth component of the moving object i, $max(k)$ and $min(k)$ are the maximum and minimum values.

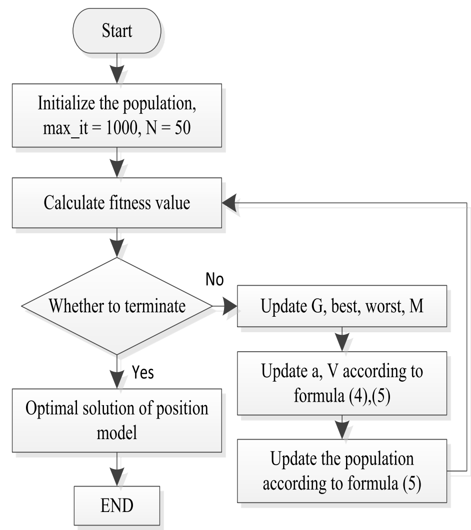


Fig. 3. The program flowchart of IGSA.

C. Adaptive Weights Based on Fitness Values

To solve the premature phenomenon of GSA algorithm, an adaptive inertia weight update mechanism based on fitness.

$$\begin{cases} t_i = (t_{start} - t_{end})\dfrac{maxiter - i}{maxiter} + t_{end} \\ \Phi_{fit} = \dfrac{2}{1 + e^{-\frac{fit_i}{\max iter}}} \\ w_i = \Phi_{fit} * t_i \end{cases} \tag{7}$$

where t_{start} and t_{end} are the initial and final values of the algorithm, and $0 < t_{end} < t_{start} < 1$, and $maxiter$ are the maximum iteration times and current algebra of GSA respectively, t_i is the inertia weight of linear decreasing, Φ_{fit} is the fitness value factor that changes with the change of fitness value *fit*, and w_i is the inertia weight of adaptive decreasing.

The weighted object position update formula is updated to Equation (8):

$$x_i^k(t+1) = w_i * x_i^k(t) + v_i^k(t+1) \tag{8}$$

D. IGSA Optimization Steps

Figure 3 is an optimization step for proposing an improved GSA algorithm:

The initialization of the population is carried out by using formula (6), formula (1) fitness value, and fitness factor and inertia weight are calculated by formula (7). Formula (4) and formula (5) calculate acceleration and velocity, and formula (8) calculates weighted position. The above formula process is calculated, and the gravity value, the best fitness value, the worst fitness value and the mass of particles are continuously updated. Therefore, the improvement of GSA algorithm is realized and the global optimization is realized.

III. Numerical Experiment and Analysis of IGSA Algorithm

It is verified that IGSA algorithm has advantages in convergence and efficiency, and compares with teaching optimization

TABLE I
BENCHMARK FUNCTION

NO	Function	Value Range	Theory Optimal Position	Theory Optimal Value
F1	Sphere	$[-100,100]^n$	$[0]^n$	0
F2	Schwefel2	$[-10,10]^n$	$[0]^n$	0
F3	Schwefel5	$[-30,30]^n$	$[1]^n$	0
F4	Step	$[-100,100]^n$	$[-0.5]^n$	0
F5	Rastrigin	$[-5.12,5.12]^n$	$[0]^n$	0
F6	Ackley	$[-32,32]^n$	$[0]^n$	0
F7	Griewank	$[-600,600]^n$	$[0]^n$	0
F8	Penalty1	$[-50,50]^n$	$[0]^n$	0

algorithm (TLBO), gravity search algorithm (GSA) and particle swarm optimization (PSO) algorithm in convergence. The final experimental results are analyzed.

A. Test Parameter

This paper uses 8 test functions to test the performance of IGSA algorithm. The 8 functions are shown in Table I. F1-F4 is unimodal and F5-F8 is multimodal.

The learning factors of PSO and IGSA algorithms are set in the experiment. TLBO and GSA related settings were normal. In order to be unified in comparison, the population size is 50 and the maximum iterations number is 1000. Several algorithms were run 30 times respectively, and their mean value and mean square deviation were counted. The experiment was carried out on Ubuntu 16 system of DLL server, E5-2670V2∗2 128G Memory and Matlab 2016a was run.

B. Result and Analysis of Simulation Experiment

Table 2 tests the above four algorithms on the 8 functions proposed in this paper in 30, 50 and 100 dimensions. The test results are as follows:

1) Among the 8 test functions, IGSA, TLBO, GSA and PSO obtained 12, 9, 9 and 1 optimal results respectively;

2) For unimodal function, for function F4 with dimension 30, the four algorithms have achieved optimal results. For the residual function of unimodal function, IGSA can have good optimization results. Among them, TLBO and GSA algorithms have the same optimization performance, which is better than the standard PSO algorithm.

3) For multimodal functions, IGSA, TLBO and GSA have found optimal solutions for functions F5 and F7 with dimensions of 30, 50 and 100. For functions F6 and F8, TLBO and GSA perform the same performance, while IGSA performs slightly better than TLBO and GSA. PSO algorithm is not as good as the above three algorithms in 4 multimodal test functions.

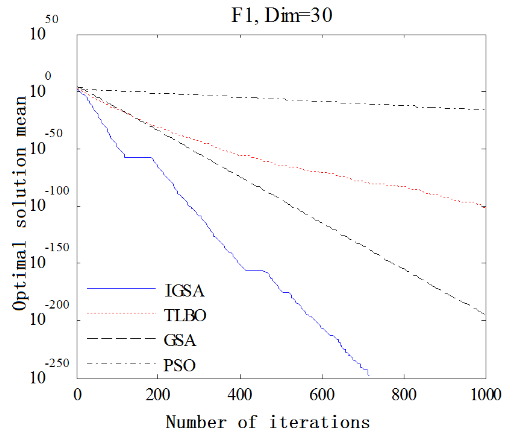


Fig. 4. Four algorithms comparison in F1 function.

4) No matter the single-peak test function or the multi-peak test function, the average value and standard deviation show that IGSA is better than GSA and PSO algorithms and IGSA algorithm is also the best in searching performance. In most test function optimization problems, the optimal value or close to the theoretical optimal value is found.

5) Whether the optimization problem is 30-dimensional, 50-dimensional or 100-dimensional, the accuracy of the optimal solution searched by IGSA is relatively stable, and the search accuracy of the algorithm does not decrease with the increase of the dimension of the optimization problem. On the whole, TLBO and GSA algorithms are less affected by the change of optimization problem dimension, of which IGSA algorithm is the most stable.

It can be seen from Table II that the test results of F4, F5 and F7 functions in IGSA, TLBO and GSA are basically 0, and the theoretical optimal values of F4, F4 and F7 functions are 0. From this result, it can only show the optimal performance of IGSA, TLBO and GSA algorithms in these three functions, and cannot show that the test performance has not been achieved. If tested by other functions, there may be differences, but IGSA tests also perform the best performance compared with other algorithms.

Figures 4-11 are 30-dimensional optimization curves of IGSA, TLBO, GSA and PSO in 8 benchmark test functions in Table II. The 50 and 100-dimensional optimization curves obtained in the experiment are similar to those in Figures 4-11.

When IGSA algorithm optimizes F1 and F8 functions, there are many inflection points, indicating that IGSA's ability to jump out of local optimization is enhanced. As can be seen intuitively from Figures 4-11, IGSA algorithm is more effective than the other three algorithms for most of the test functions in Table II with dimension 30. Among them, IGSA algorithm obtains the optimal values in four test functions (F1, F4, F5 and F7). Except F5, the convergence speed of IGSA is faster than that of the other three algorithms on the other 7 test functions. As can be seen from Figures 4-11, IGSA quickly searches for satisfactory solutions for most optimization problems.

In short, IGSA algorithm has greatly improved its optimization capability compared with standard PSO. For most

TABLE II
SIMULATION RESULTS OF FOUR ALGORITHMS FOR BENCHMARK FUNCTIONS

Fun	Dim	IGSA		TLBO		GSA		PSO	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std
F1	30	0	0	6.37×10^{-102}	1.34×10^{-101}	2.48×10^{-195}	0	4.56×10^{-17}	1.55×10^{-17}
	50	0	0	2.87×10^{-107}	4.65×10^{-107}	1.33×10^{-194}	0	1.34×10^{-16}	3.84×10^{-17}
	100	0	0	7.38×10^{-101}	1.37×10^{-100}	2.34×10^{-193}	0	2.15×10^{2}	6.32×10^{1}
F2	30	3.18×10^{-166}	0	2.35×10^{-54}	3.36×10^{-54}	1.07×10^{-96}	9.30×10^{-97}	3.58×10^{-8}	7.66×10^{-9}
	50	4.26×10^{-161}	0	5.32×10^{-53}	1.42×10^{-52}	2.55×10^{-96}	4.37×10^{-96}	4.49×10^{-2}	6.89×10^{-2}
	100	1.39×10^{-144}	0	1.87×10^{-54}	2.68×10^{-54}	3.64×10^{-97}	3.62×10^{-97}	2.23	2.13
F3	30	3.23×10^{-5}	4.66×10^{-5}	3.26×10^{-4}	4.56×10^{-4}	2.89×10^{1}	1.44×10^{-2}	2.59×10^{1}	9.33×10^{-1}
	50	4.67×10^{-5}	3.48×10^{-5}	2.83×10^{-3}	3.31×10^{-3}	4.89×10^{1}	4.97×10^{-2}	1.25×10^{2}	1.10×10^{2}
	100	3.23	5.76	1.96×10^{1}	4.24×10^{1}	9.89×10^{1}	2.16×10^{-2}	4.88×10^{3}	3.09×10^{3}
F4	30	0	0	0	0	0	0	0	0
	50	0	0	0	0	0	0	11	1.16×10^{1}
	100	0	0	0	0	0	0	1.21×10^{3}	3.51×10^{2}
F5	30	0	0	0	0	0	0	2.61×10^{1}	5.25
	50	0	0	0	0	0	0	3.72×10^{1}	9.35
	100	0	0	0	0	0	0	1.16×10^{2}	1.93×10^{1}
F6	30	8.67×10^{-16}	0	4.32×10^{-15}	1.43×10^{-15}	4.44×10^{-15}	0	5.35×10^{-9}	1.21×10^{-9}
	50	8.67×10^{-16}	0	4.67×10^{-15}	0	3.73×10^{-15}	1.59×10^{-15}	6.19×10^{-9}	9.11×10^{-10}
	100	8.67×10^{-16}	0	5.24×10^{-15}	0	4.44×10^{-15}	0	1.96	4.53×10^{-1}
F7	30	0	0	0	0	0	0	6.51	2.53
	50	0	0	0	0	0	0	2.36×10^{1}	2.58
	100	0	0	0	0	0	0	7.82×10^{1}	8.94
F8	30	1.23×10^{-8}	2.54×10^{-8}	1.45×10^{-7}	2.15×10^{-7}	6.59×10^{-1}	2.36×10^{-1}	4.77×10^{-2}	6.11×10^{-2}
	50	1.45×10^{-8}	2.09×10^{-9}	4.58×10^{-7}	4.57×10^{-7}	7.66×10^{-1}	2.12×10^{-1}	1.63	6.46×10^{-1}
	100	1.47×10^{-8}	2.01×10^{-8}	8.76×10^{-7}	1.33×10^{-7}	9.34×10^{-1}	1.68×10^{-1}	3.01	9.82×10^{-1}

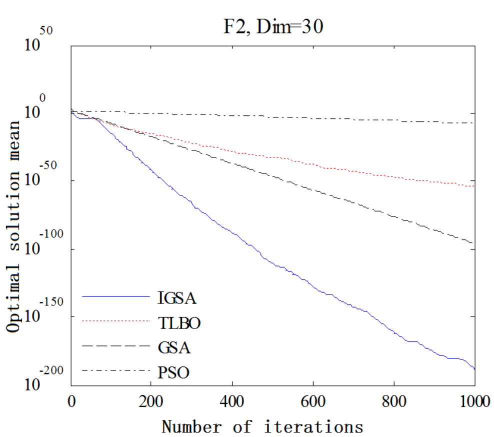


Fig. 5. Four algorithms comparison in F2 function.

optimization problems, the optimization results of IGSA algorithm are better than TLBO, GSA and standard PSO algorithms.

IV. EXTREME LEARNING MACHINE OPTIMIZATION MODEL

A. ELM Fundamentals

ELM model is described as follows.

$$\sum_{i=1}^{M} \boldsymbol{\beta}_i g(\boldsymbol{\omega}_i \cdot \boldsymbol{x}_i + b_i) = \boldsymbol{o}_j, \quad j = 1, 2, \ldots, N \tag{9}$$

where $\boldsymbol{\omega}_i$ is the input weight vector; $\boldsymbol{\beta}_i$ is an output weight vector; $\boldsymbol{o}_j$ is the network output value. Cost Function E Settings:

$$E(\boldsymbol{S}, \boldsymbol{\beta}) = \sum_{j=1}^{N} ||\boldsymbol{o}_j - \boldsymbol{t}_j|| \tag{10}$$

In the formula, $\boldsymbol{S}$ includes the network input weight and the hidden layer node threshold. The training goal of ELM is to seek the best S, β.

$$\min E = \min ||\boldsymbol{H}(\boldsymbol{\omega}, b, \boldsymbol{x})\boldsymbol{\beta} - \boldsymbol{T}|| \tag{11}$$

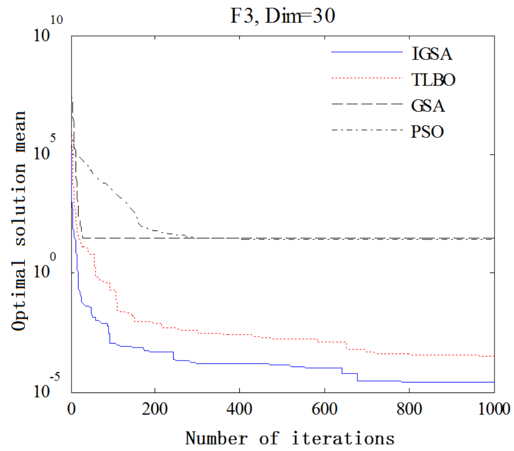


Fig. 6. Four algorithms comparison in F3 function.

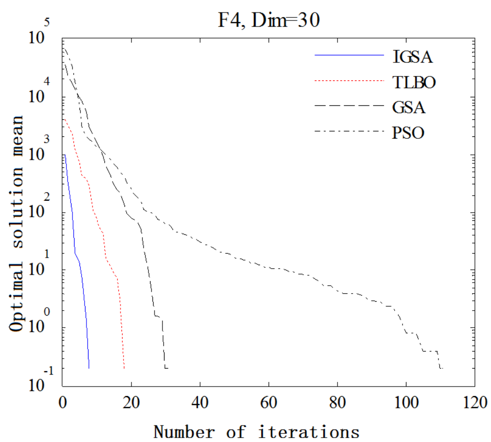


Fig. 7. Four algorithms comparison in F4 function.

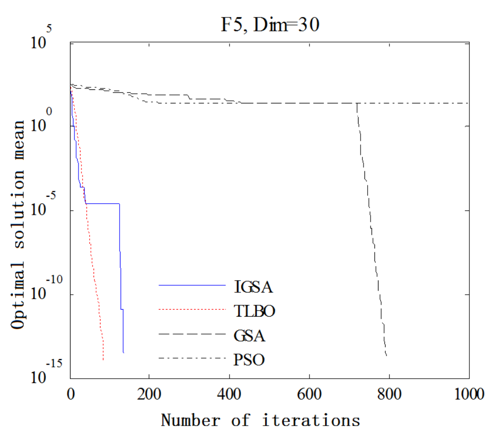


Fig. 8. Four algorithms comparison in F5 function.

where H is the hidden layer output matrix, $\boldsymbol{\beta}$ is the output weight matrix, and $\boldsymbol{T}$ is the target value matrix. $\boldsymbol{H}, \boldsymbol{\beta}, \boldsymbol{T}$ defined as follows:

$$H(\boldsymbol{\omega}, b, \boldsymbol{x}) = \begin{bmatrix} g(\boldsymbol{\omega}_1 \boldsymbol{x}_1 + b_1) & \cdots & g(\boldsymbol{\omega}_M \boldsymbol{x}_1 + b_M) \\ \vdots & & \vdots \\ g(\boldsymbol{\omega}_1 \boldsymbol{x}_N + b_1) & \cdots & g(\boldsymbol{\omega}_M \boldsymbol{x}_N + b_M) \end{bmatrix}_{N \times M} \tag{12}$$

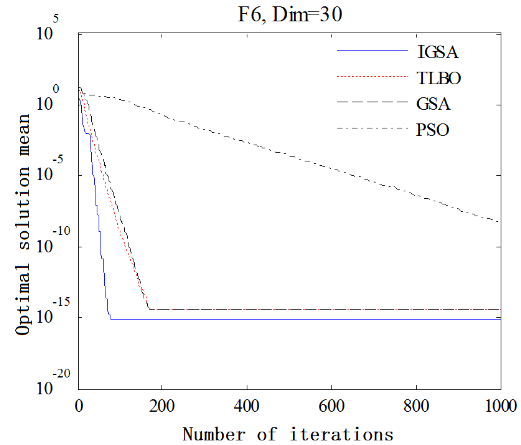


Fig. 9. Four algorithms comparison in F6 function.

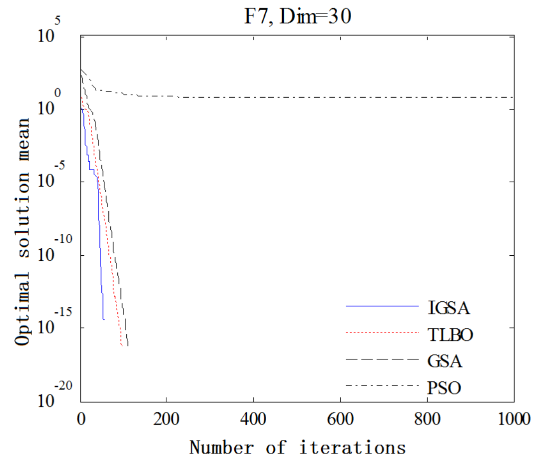


Fig. 10. Four algorithms comparison in F7 function.

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \vdots \\ \boldsymbol{\beta}_M^T \end{bmatrix}_{M \times n}, \boldsymbol{T} = \begin{bmatrix} \boldsymbol{t}_1^T \\ \vdots \\ \boldsymbol{t}_N^T \end{bmatrix}_{N \times n} \tag{13}$$

ELM is to find the least square solution $\widehat{\boldsymbol{\beta}}$, and its calculation formula is defined.

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{H}^{\dagger} \boldsymbol{T} \tag{14}$$

B. ELM Optimization Model

For the above model, through combining ELM optimization model, its process is shown in the figure 12.

V. Comparisons and Analysis of Experiments

IGSA-ELM prediction model is established to predict road congestion, and the optimal driving route is solved according to the proposed dynamic optimization algorithm. As mentioned earlier, the data set needed to predict road congestion is collected through on-board GPS data and transmitted to on-board information big data processing platform. For the process of large-scale on-board data transmission, the experimental implementation process is relatively difficult.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE III
VEHICLE GPS DATA TABLE

Vehicle ID	Date	Time	Longitude	Dimension	Position ID
4001	2014/08/3	06:01:18	30.583802	104.034407	86
482	2014/8/3	18:51:00	30.624811	104.136587	67
1	2014/8/3	21:18:15	30.624809	104.136612	67
1	2014/8/12	10:57:06	30.615417	104.040228	79
2824	2014/8/12	18:47:59	30.651394	103.984025	21
...					

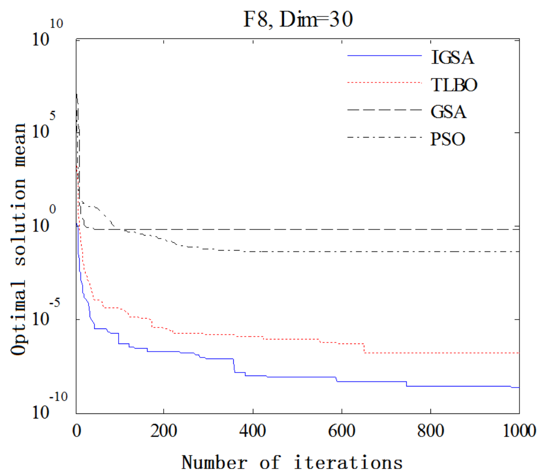


Fig. 11. Four algorithms comparison in F8 function.

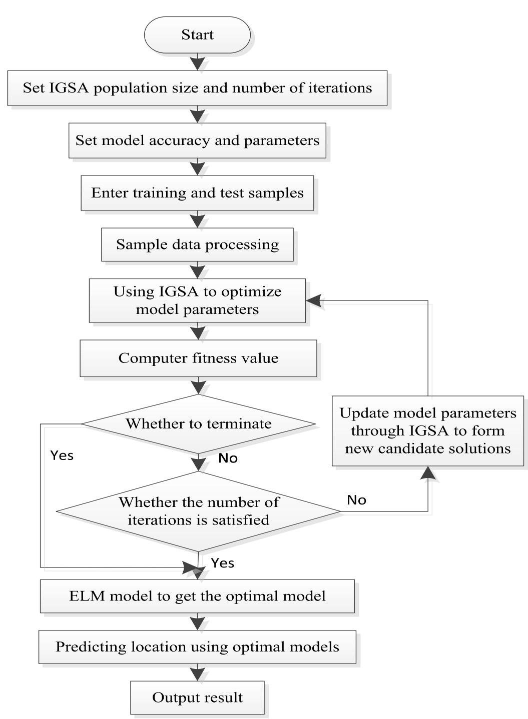


Fig. 12. Optimized parameters of ELM by IGSA.

Therefore, the process of data transmission between GPS data and the vehicle-mounted information big data processing platform is omitted, and the vehicle-mounted experimental

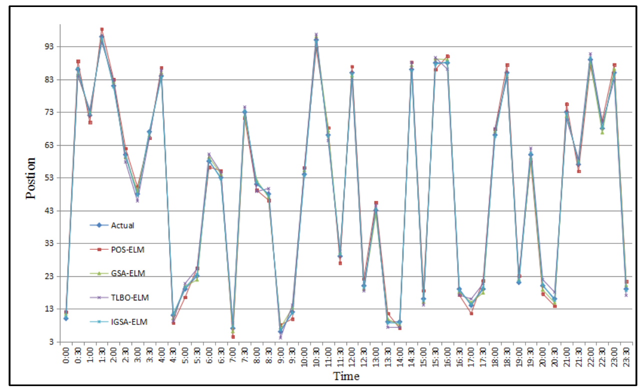


Fig. 13. Comparison of position prediction at different time points.

data are directly stored on the vehicle-mounted information big data processing platform, and then congestion prediction and road navigation calculation are carried out. This kind of experimental thinking is reasonable, because the focus of the road navigation problem studied in this paper is to process the received vehicle-mounted data, verify the rationality of IGSA-ELM prediction model and obtain the optimal position by using the prediction results.

An experimental platform for large data processing of vehicle-mounted information is constructed, and the average road speed data is selected as the training set for road congestion prediction, and a large data processing platform is built for experimental analysis. The IGSA-EML prediction model of each road is constructed on the big data vehicle platform, the model is trained with training sets of different time periods, the average speed of each road is predicted, and then the current optimal road is calculated according to the predicted congestion. Finally, the accuracy of IGSA-EML model in predicting road congestion is analyzed, and the optimized path solved by the improved algorithm is compared with the path solved by the traditional algorithm.

A. Data Characteristic Analysis

The data studied in this paper come from on-board GPS data. The data include: road number, vehicle ID, time, longitude, dimension, and speed and position number. The data time is 15,000 on-board GPS data information from August 1, 2014 to August 31, 2014. Its data format is as follows table III.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE IV
VEHICLE GPS ROAD STATISTICS

Road ID	Vehicle ID	Road_L(m)	Entry Time	Departure Time	Speed （km/h）	Road Conditions
101	11	325.6	06:01:18	06:02:23	32.6	2
482	15	1242.8	18:51:00	18:53:10	23.2	1
106	1003	1254.4	21:18:15	21:20:15	25.1	2
107	23	521.3	10:57:06	10:59:10	10.5	0
236	976	613.2	18:47:59	18:50:23	9.2	0
...						

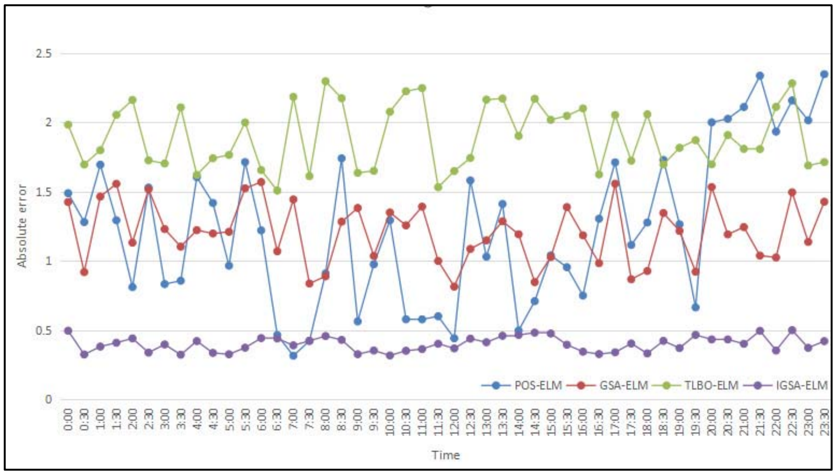


Fig. 14. Comparisons of absolute errors at different time points.

In the above table, the location number numbers the location information within a certain range of geographical locations (combining longitude and dimension). Generally, the longitude and dimension location numbers within the range of geographical locations with characteristics (e.g. Schools, hospitals,) are the same number. In this experiment, the position number is defined as about 90. Some position information with less longitude and dimension is deleted in this experiment and is not included in the experiment.

Through the adjustment of dimensions and geographical positions, the above data are sorted out, and the vehicle information in different time periods of different road sections is counted. Through the numbering of geographical road sections, the sorted data are described in the following table IV.

In practical application, the time interval for the release of dynamic traffic information is 5 minutes. The information evaluation of real-time road conditions can adopt three states of blockage, slow travel and ease, with 0, 1 and 2 respectively. The threshold intervals are respectively: blockage less than 12km/h, slow travel between 12 and 25km/h, and ease greater than 25km/h.

B. Prediction and Analysis of Vehicle Location Based on IGSA-EML Model

IGSA algorithm combined with Extreme Learning Machine (ELM) optimization algorithm is used to predict the vehicle communication position. The next position of the vehicle can be correctly identified, which is an excellent prediction model based on big data environment. Compared with other algorithm prediction methods, the effect is shown in the figure 13. The results show that IGSA-ELM algorithm has obvious advantages.

According to the prediction of different positions at different times from the positions in the above figure, GSA-ELM and PSO-ELM algorithms have relatively small errors, while TLBO-ELM algorithms have relatively large errors; especially the geographical location coordinates do not have specific positions (e.g. obvious markers such as schools and shops). As the number of such positions is relatively small, the training effect is not very ideal. Figure 14 is an absolute error predicted in one day by the IGSA-ELM algorithm.

The error described in this article is the absolute value of the actual value minus the predicted value. $Error = |actual\ value - predictedvalue|$; This can directly see the prediction accuracy. Figure 14 shows that the IGSA-ELM prediction error is basically about 0.5, indicating that the prediction accuracy is relatively high.

To embody the advantages of IGSA-ELM, We test the algorithm many times to verify the advantages of IGSA-ELM algorithm. Figure 15 below shows the average absolute error of IGSA-ELM algorithm. The average value of absolute error is mostly about 0.5, and the algorithm performance is relatively good.

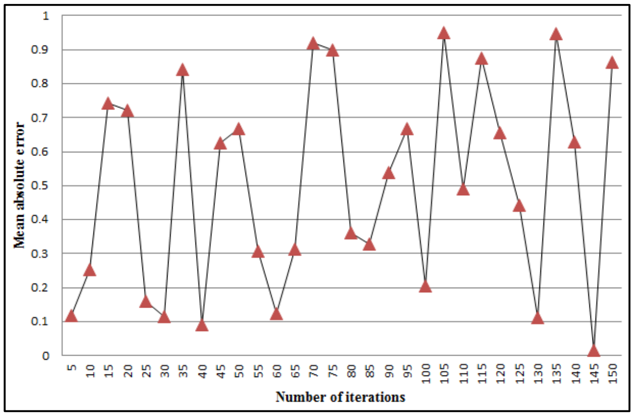


Fig. 15. Average number of iterations of IGSA-EML algorithm.

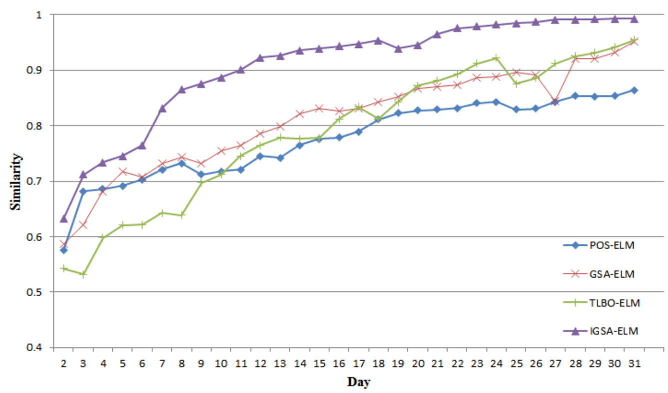


Fig. 16. Comparison of prediction similarity for different days.

KL divergence is called relative entropy and relative probability distribution. In vehicles v_i, and v_j, the probability of predicting the target geographic location is P_i and P_j. The relative entropy p_{ik} of the vehicle v_i for the predicted target position p_k is defined as flow.

$$p_{ik} = \frac{F_{ik}+1}{\sum_{k=1}^{N}(F_{ik}+1)} \tag{15}$$

$$KL(P_i|P_j) = \sum_{k=1}^{N} p_{ik}|\log\frac{p_{ik}}{p_{jk}}| \tag{16}$$

$$Sim^{KL}(v_i, v_j) = 1 - \frac{KL(P_i|P_j)}{\max\{KL\,(P_i|P_x) : u_x \in U\}} \tag{17}$$

When the actual position of the vehicle is the same as the path or preference of the predicted position, the result of the above formula is 0, otherwise it is the value of [0, 1].

The following experiments predict the vehicle location information for different days, and IGSA-ELM has the most ideal prediction effect. From August 1, 2014 to August 31, 2014, the vehicle positions for 31 days are predicted and analyzed, and the figure is a 31-day prediction effect diagram as figure 16.

On the whole, IGSA-ELM algorithm predicted that the similarity reached above 0.9 on the 12th day, while other algorithms had to reach above 20 to reach above 0.9. Among several algorithms, the overall similarity of POS-ELM algorithm is relatively low, with the highest value reaching 0.88. However, the overall similarity between GSA-ELM and TLBO-ELM is relatively close, but the overall number of days above 0.9 is relatively small.

VI. Conclusion

In order to accurately predict the vehicle position, an improved universal gravitation optimization algorithm (IGSA) combined with limit learning machine (ELM) is proposed to predict the vehicle position, and the traffic characteristics of vehicle position are affected by many factors, and the influence relationship is complex. IGSA is used to pre-select ELM model parameters to improve the accuracy and generalization of the prediction model. Taking the 31-day vehicle GPS data of a certain city as the test object, the prediction model of IGSA-ELM's geographical position is established. Through training and testing the vehicle GPS data as training samples and testing samples, the model is trained and tested. The simulation experiment shows that the vehicle GPS position prediction model of IGSA-ELM has better accuracy and strong generalization ability, and has higher practical value.

References

[1] V. C. MahaVishnu, M. Rajalakshmi, and R. Nedunchezhianm, "Intelligent traffic video surveillance and accident detection system with dynamic traffic signal control," *Cluster Comput.*, vol. 21, no. 4, pp. 1–13, Jun. 2017.

[2] W. Jiang *et al.*, "The impact of the biomass crop assistance program on the United States forest products market: An application of the global forest products model," *Forests*, vol. 10, no. 3, p. 215, Feb. 2019.

[3] J. Liu, J. Zhao, and Z. Zhu, "On the number of spanning trees and normalized Laplacian of linear octagonal–quadrilateral networks," *Int. J. Quantum Chem.*, vol. 119, no. 17, Sep. 2019, Art. no. e25971.

[4] J.-B. Liu, J. Zhao, and Z.-Q. Cai, "On the generalized adjacency, Laplacian and signless Laplacian spectra of the weighted edge corona networks," *Phys. A, Stat. Mech. Appl.*, vol. 540, Feb. 2020, Art. no. 123073.

[5] T.-Y. Qian, B. Liu, L. Hong, and Z.-N. You, "Time and location aware points of interest recommendation in location-based social networks," *J. Comput. Sci. Technol.*, vol. 33, no. 6, pp. 1219–1230, Nov. 2018.

[6] R. W. Scholz and Y. M. Lu, "Detection of dynamic activity patterns at a collective level from large-, vol. trajectory, data," *Int. J. Geograph. Inf. Sci.*, vol. 28, no. 5, pp. 946–963, Jan. 2014.

[7] G. Technitis, W. Othman, K. Safi, and R. Weibel, "From A to B, randomly: A point-to-point random trajectory generator for animal movement," *Int. J. Geograph. Inf. Sci.*, vol. 29, no. 6, pp. 912–934, Mar. 2015.

[8] J. Li, Q. Qin, J. Han, L.-A. Tang, and K. H. Lei, "Mining trajectory data and geotagged data in social media for road map inference," *Trans. GIS*, vol. 19, no. 1, pp. 1–18, Feb. 2015.

[9] S. Qiao, T. Li, J. Peng, and J. Qiu, "Parallel sequential pattern mining of massive trajectory data," *Int. J. Comput. Intell. Syst.*, vol. 3, no. 3, pp. 343–356, Sep. 2010.

[10] G. Khoshamooz and M. Taleai, "Multi-domain user-generated content based model to enrich road network data for multi-criteria route planning," *Geograph. Anal.*, vol. 49, no. 3, pp. 239–267, Apr. 2017.

[11] Y. Dong and D. Pi, "Novel privacy-preserving algorithm based on frequent path for trajectory data publishing," *Knowl.-Based Syst.*, vol. 148, pp. 55–65, May 2018.

[12] A. K. Laha and S. Putatunda, "Real time location prediction with taxi-GPS data streams," *Transp. Res. C, Emerg. Technol.*, vol. 92, pp. 298–322, Jul. 2018.

[13] F. Dantas Nobre Neto, C. D. S. Baptista, and C. E. C. Campelo, "Combining Markov model and prediction by partial matching compression technique for route and destination prediction," *Knowl.-Based Syst.*, vol. 154, pp. 81–92, Aug. 2018.

[14] Z. Liu, L. Hu, C. Wu, Y. Ding, and J. Zhao, "A novel trajectory similarity-based approach for location prediction," *Int. J. Distrib. Sensor Netw.*, vol. 12, no. 11, pp. 1–13, Nov. 2016.

[15] N. Zhang, H. Chen, X. Chen, and J. Chen, "Forecasting public transit use by crowdsensing and semantic trajectory mining: Case studies," *Int. J. Geo-Inf.*, vol. 5, no. 10, pp. 1–13. Sep. 2016.

[16] M. Ni, Q. He, and J. Gao, "Forecasting the subway passenger flow under event occurrences with social media," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 6, pp. 1623–1632, Jun. 2017.

[17] J. Ding, "Trajectory mining, representation and privacy protection," in *Proc. 2nd ACM SIGSPATIAL Workshop.*, Bellevue, WA, USA, Nov. 2015, pp. 1–4.

[18] P. Geetha and E. Ramaraj, "Tree based space partition of trajectory pattern mining for frequent item sets," *Austral. J. Basic Appl. Sci.*, vol. 10, no. 2, pp. 250–261, Jan. 2016.

[19] W. Yu, "Discovering frequent movement paths from taxi trajectory data using spatially embedded networks and association rules," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 1–12, Mar. 2019.

[20] L. Liu, C. Andris, and C. Ratti, "Uncovering cabdrivers' behavior patterns from their digital traces," *Comput., Environ. Urban Syst.*, vol. 34, no. 6, pp. 541–548, Nov. 2010.

[21] S. Liu, Y. Liu, L. M. Ni, J. Fan, and M. Li, "Towards mobility-based clustering," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Washington, DC, USA, 2010, pp. 919–928.

[22] G. Andrienko and N. Andrienko, "Poster: Dynamic time transformation for interpreting clusters of trajectories with space-time cube," in *Proc. IEEE Symp. Vis. Analytics Sci. Technol.*, Salt Lake City, UT, USA, Oct. 2010, pp. 24–29.

[23] L. Zheng *et al.*, "Spatial–temporal travel pattern mining using massive taxi trajectory data," *Phys. A, Stat. Mech. Appl.*, vol. 501, pp. 24–41, Jul. 2018.

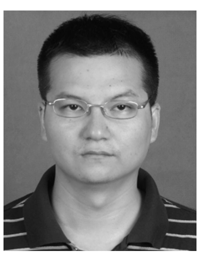
[24] W. Jiang and L. Zhang, "The impact of the transportation network companies on the taxi industry: Evidence from Beijing's GPS taxi trajectory data," *IEEE Access*, vol. 6, pp. 12438–12450, 2018.

[25] H. Ji-hua, H. Ze, and D. Jun, "A hierarchical path planning method using the experience of taxi drivers," *Proc.—Social Behav. Sci.*, vol. 96, pp. 1898–1909, Nov. 2013.

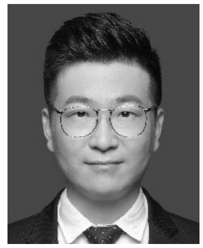
[26] B. Y. Chen, W. H. K. Lam, Q. Li, A. Sumalee, and K. Yan, "Shortest path finding problem in stochastic time-dependent road networks with stochastic first-in-first-out property," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1907–1917, Dec. 2013.

[27] Y. Wang, Y. Zheng, and Y. Xue, "Travel time estimation of a path using sparse trajectories," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Apr. 2014, pp. 25–34.

[28] L. X. Pang, S. Chawla, W. Liu, and Y. Zheng, "On detection of emerging anomalous traffic patterns using GPS data," *Data Knowl. Eng.*, vol. 87, pp. 357–373, Sep. 2013.



Guobin Chen was born in Yancheng, Jiangsu, China, in 1982. He received the master's degree in computer software and theory from Chongqing Normal University. His research directions are software architecture, web services, and intelligent information processing. Since 2010, he has been with the Rongzhi College of Chongqing Technology and Business University. He is currently an Associate Professor. He has published more than 20 articles. His research interests include D2D networks, big data, and machine learning. He has presided over the Chongqing Municipal Education Commission Science and Technology Project, the Chongqing Science and Technology Innovation Project, and participated in the National Natural Science Foundation and the Chongqing Municipal Science and Technology Commission Project.



Lukun Wang (Member, IEEE) received the B.S. degree from the Shandong University of Science and Technology in 2006, the M.S. degree from the Dalian University of Technology in 2009, and the Ph.D. degree in computer science from the Ocean University of China in 2016. He is currently an Assistant Professor with the College of Intelligent Equipment, Shandong University of Science and Technology. His current research interests include artificial intelligence, image processing and pattern recognition, machine learning, control theory and applications, and sensor networks.

Muhammad Alam (Senior Member, IEEE) received the Ph.D. degree in computer science from the University of Aveiro, Portugal. He is with the Department of Computer and Software Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, China. He has authored/edited five books and over 100 publications in refereed journals and conferences as well as book chapters. His research interests include the IoT, blockchain, and 5G networks. He is a Senior Member of the IEEE IES. He is also a TPC member and a reviewer for a number of reputed conferences, journals, and magazines.

Mohamed Elhoseny (Member, IEEE) is currently an Assistant Professor with the Faculty of Computers and Information, Mansoura University, where he is also the Director of the Distributed Sensing and Intelligent Systems Laboratory. He has been appointed as an ACM Distinguished Speaker from 2019 to 2022. He has authored/coauthored over 85 ISI journal articles in high-ranked and prestigious journals, such as the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, the IEEE TRANSACTIONS ON RELIABILITY, *Future Generation Computer Systems*, and *Neural Computing and Applications*. He has authored/edited conference proceedings, book chapters, and ten books. His research interests include smart cities, network security, artificial intelligence, the Internet of Things, and intelligent systems.