Forecasting Bus Passenger Flows by Using a Clustering-Based Support Vector Regression Approach

CHUAN LI[1], (Senior Member, IEEE), XIAODAN WANG[1], ZHIWEI CHENG[2], AND YUN BAI [1,3], (Member, IEEE)

[1]National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing 400067, China
[2]Chongqing CPI Zineng Science and Technology Company, Ltd., Chongqing 400050, China
[3]Centro de Electrónica, Optoelectronica e Telecomunicações (CEOT), Universidade do Algarve, 8005-139 Faro, Portugal

Corresponding author: Yun Bai (yunbai@foxmail.com)

ABSTRACT As a significant component of the intelligent transportation system, forecasting bus passenger flows plays a key role in resource allocation, network planning, and frequency setting. However, it remains challenging to recognize high fluctuations, nonlinearity, and periodicity of bus passenger flows due to varied destinations and departure times. For this reason, a novel forecasting model named as affinity propagation-based support vector regression (AP-SVR) is proposed based on clustering and nonlinear simulation. For the addressed approach, a clustering algorithm is first used to generate clustering-based intervals. A support vector regression (SVR) is then exploited to forecast the passenger flow for each cluster, with the use of particle swarm optimization (PSO) for obtaining the optimized parameters. Finally, the prediction results of the SVR are rearranged by chronological order rearrangement. The proposed model is tested using real bus passenger data from a bus line over four months. Experimental results demonstrate that the proposed model performs better than other peer models in terms of absolute percentage error and mean absolute percentage error. It is recommended that the deterministic clustering technique with stable cluster results (AP) can improve the forecasting performance significantly.

INDEX TERMS Affinity propagation, passenger flow, particle swarm optimization, forecasting, support vector regression.

I. INTRODUCTION

Forecasting bus passenger flows is an important component of the intelligent transportation system (ITS). It plays a key role in resource allocation, network planning, and frequency setting. As a result, it has attracted wide attention from researchers and engineers [1].

Forecasting models can generally be classified into three categories: parametric, non-parametric, and hybrid models [2], [3]. Main difference between parametric and non-parametric models lies in the functional dependency assumed between independent and dependent variables [4]. For parametric techniques, several methods have been used to

The associate editor coordinating the review of this manuscript and approving it for publication was Zhe Xiao.

forecast transportation demand, such as the Box-Jenkins [5], smoothing techniques [6], autoregressive integrated moving average (ARIMA) [4], grey forecasting [7], and state space models [8]. Of these models, the ARIMA model, which is a linear function of time-lagged variables and error terms, has been commonly used as early as 1970s [9]. However, bus passenger flows exhibit high fluctuations, non-linearity, and periodicity. Therefore, traditional parametric models may not be suitable for capturing the structure of non-linear flows because it assumes that the relationships between time-lagged variables are linear.

Within the non-parametric model, neural networks [10], k-nearest neighbors [11], Kalman filters [12], support vector regression (SVR) [13], [14], and other methods [15] were applied to predict passenger flow in transportation.

Neural networks are mainly expressed as the ability to describe the uncertain and complex nonlinearity of passenger flow time series [16]. They have great potential in forecasting nonlinearity and randomness. However, they are sensitive to parameter selection and prone to local minima and overfitting to some extent [17]. Unlike neural networks, the SVR implements the structural risk minimization principal, which seeks to minimize an upper bound of the generalization error rather than minimize the training error [18]. To some extent, it has the potential to overcome the inherent defects of neural works [19]. It also has the advantages of a global optimum, a simple structure, and a strong ability to promote small samples [20]. Therefore, it can be used to solve problems such as nonlinearity, small sample size, and a high number of dimensions and has been successfully applied to forecasting short-term traffic forecasting [21], complex motion of floating platforms [22], and electric load [23]. All the aforementioned cases have achieved good forecasting results.

Each model tends to have its own advantages and limitations for different applications [24]. Combining different models and retaining the advantages of each is the basic idea of the combination approach [25], [26]. To improve forecasting accuracy, the adoption of a hybrid model has become common [27]. Those hybrid models have achieved some good results when dealing with the forecasting problems.

For these models, it is challenging to recognize complex nonlinear passenger flows caused by emergencies, holiday trips, restricted traffic, and so on. To solve this problem, the data preprocessing for pattern recognition and feature learning maybe a way to reduce the complexity inherent in passenger flows [28]. There are two approaches: time series decomposition and cluster analysis [29], [30] to resolve the problem. For time series decomposition, the time series is viewed as a combination of quasi-periodic signals with noises. Many techniques such as wavelet analysis [16], singular spectrum analysis [31], and empirical mode decomposition [32] have been applied successfully. Meanwhile, cluster analysis attempts to classify data elements into clusters based on their similarity [33]. Differing from random time series analysis, similarity-dependent cluster analysis avoids resolution blurring and outliers in prediction models while considering only the similarity between data. Various similarity recognition methods, such as K-means method [34], Gaussian kernel-based fuzzy c-means clustering (KFCM) [35], and affinity propagation (AP) [36] clustering algorithm, have been proposed for clustering. Compared with K-means and KFCM methods, as an automatic clustering algorithm, AP algorithm does not require a specified number of cluster centers beforehand and can be used to obtain unchangeable clustering results with many tests. At same time, especially for large data, AP or AP-based methods can obtain better solutions with fewer errors and less time than previous algorithms [37]. Although it does not guarantee a global optimum, AP algorithm has been consistently superior to previous algorithms and has become an attractive clustering method [38].

Based on the preceding analyses, to adjust the fine characteristics of bus passenger flows, this paper introduces a novel forecasting model based on clustering and nonlinear simulation: an affinity propagation-based support vector regression (AP-SVR) model. First, the AP algorithm is used to partition the bus passenger flow and generate clustering-based intervals. Then, the SVR, with parameters optimized by partial swarm optimization (PSO) [39]–[41], is applied to fit and forecast the passenger flow for each cluster. Finally, the prediction results of the PSO-SVR are rearranged by chronological order rearrangement.

The rest of this paper is organized as follows. Section II describes the methodology in detail. Case study information, evaluation criteria, and model development are introduced in Section III. Section IV gives the experiment results and analysis of the proposed model. Conclusions are given in Section V.

II. METHODOLOGY
A. AFFINITY PROPAGATION
AP algorithm is an efficient clustering method introduced by Frey and Dueck [36] in 2007. It has been widely studied and applied in many fields because of its distinctive advantages [42]. AP is taken as an input for a collection passenger flow similarities between data points, where the similarity $s(i, k)$ indicates how well the data point with index k is suited to be the exemplar for data point i. Similarity between data x_i and x_k can be measured by the negative pairwise Euclidean distance:

$$s(i, k) = \begin{cases} -\|x_i - x_k\|^2 & i \neq k \\ p & i = k, \end{cases} \quad (1)$$

where p denotes "preference". Data points with a larger value of $s(k, k)$ are more likely to be chosen as exemplars [43].

There are two types of messages changing between data points: responsibility and availability. First, the responsibility $r(i, k)$ is sent from data point i to candidate exemplar point k. It reflects the accumulated evidence for how well-suited point k is to serve as the exemplar for point i. Second, the availability $a(i, k)$ is sent from candidate exemplar point k to point i. It reflects the accumulated evidence for how appropriate it would be for point i to choose point k as its exemplar [44].

Letting the availability $a(i, k) = 0$, the responsibility $r(i, k)$ are computed using the rule

$$r(i, k) \leftarrow s(i, k) - \max_{j, j \neq k}\{a(i, j) + s(i, j)\}), \quad (2)$$

where $s(i, j)$ is the similarity between points i and j for $i \neq j$. For $k = i$, $r(k, k)$ is set to the input preference that point k be chosen as an exemplar. Then $a(i, k)$ updates as

$$a(i, k) \leftarrow \min\left\{0, r(k, k) + \sum_{j, j \notin \{i, k\}} \max\{0, \ r(j, k)\}\right\}, i \neq k),$$

$$(3)$$

The "self-availability" $a(k, k)$ is updated differently:

$$a(k, k) \leftarrow \sum_{j, j \neq k} \max\{0, r(j, k)\}. \tag{4}$$

It should be noted that Eq. (4) is a special case of Eq. (3) by employing the self-availability concept. When updating the messages, it is important that they are damped to avoid numerical oscillations. The damping factor λ was introduced to avoid numerical oscillations when calculating $a(k, k)$ as shown in Eq. (4). Each message is set to λ times its value from the previous iteration plus 1-λ times its prescribed updated value, where the damping factor λ is between 0 and 1.

The responsibilities and availabilities are iterated until the cluster center remains unchanged for a user-set number of iterations. Then, clusters $e(k)$ are given by maximizing over the sum of responsibility and availability:

$$e(k) = \arg \max_j \{a(i, j) + r(i, j)\}. \tag{5}$$

For $e(k) = i$, point i is an exemplar or cluster center. If $e(k) \neq i$, then point k is an exemplar for point i.

B. SUPPORT VECTOR REGRESSION

SVR was developed for regression purposes by Vapnik *et al.* [45] in 1997. It has been used recently in most transportation fields and forecasting models with good accuracy. As an improved regression algorithm of the standard SVR, the least squares SVR (LSSVR) [46] uses the least squares linear system as a loss function to replace the quadratic programming question in the SVR. The changes greatly reduce the computational complexity and improve the speed of the algorithm [47]. In this study, a data-driven method is used to determine the general input structure of the LSSVR model $F(.)$ for the passenger flow prediction. The ith hour bus passenger flow can be associated with the historical data, i.e.,

$$\upsilon(i) = F[\upsilon(i - 1), \upsilon(i - 2), \cdots, \upsilon(i - n)]. \tag{6}$$

Considering a given sample set $T = \{(x_i, y_i) | i = n + 1, n + 2, \ldots, l\}$ with input data and output data, the LSSVR can be represented in the feature space as

$$y = w^T \varphi(x) + b, \tag{7}$$

where $\omega \in R^{nf}$, $b \in R$, and $\varphi(x)$ represent the nonlinear mapping function from the input space to high dimensional feature space. To solve the above problem, the objective function is obtained by minimizing

$$Z(\omega, \xi) = \frac{1}{2}\omega^T \omega + \frac{\gamma}{2} \sum_{i=n+1}^{l} \xi_i^2,$$

$$s.t. \; y_i = \omega^T \phi(x_i) + b + \xi_i, \quad i = n+1, n+2, \cdots, l, \tag{8}$$

where ξ_i is the error variable and γ the penalty coefficient. A Lagrange function can be defined as

$$L(\omega, b, \xi, \alpha) = Z(\omega, \xi) - \sum_{i=n+1}^{l} \alpha_i[\omega^T \phi(x_i) + b + \xi_i - y_i], \tag{9}$$

with $\alpha_i (i = 1, 2, 3, \ldots, l)$ are Lagrange multipliers. The conditions for optimality

$$\partial L / \partial \omega = 0 \Rightarrow \omega = \sum_{i=n+1}^{l} a_i \varphi(x_i) = 0,$$

$$\partial L / \partial b = 0 \Rightarrow \sum_{i=n+1}^{l} a_i = 0, \quad \partial L / \partial \xi_i = 0 \Rightarrow a_i = \gamma \xi_i,$$

$$\partial L / \partial a_i = 0 \Rightarrow \omega^T \varphi(x_i) + b + \xi_i - y_i = 0 \tag{10}$$

can be rewritten as the following set of linear equations eliminating ω, ξ

$$\begin{bmatrix} 0 & q^T \\ q & \Omega + \gamma^{-1}I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} o \\ y \end{bmatrix}$$

$$\Omega_{ij} = \varphi(x_i)^T \varphi(x_j) = k(x_i, x_j), \tag{11}$$

where q is a $l \times 1$ dimensional column vectors, $a = [a_1 a_2 \ldots a_l]^T$, $y = [y_1 y_2 \ldots y_l]^T$ and k is the kernel function that satisfies the Mercer condition. The solution to the problem can be obtained by

$$y = \sum_{i=n+1}^{l} a_i k(x_i, x_j) + b. \tag{12}$$

There are large variety of kernel functions with relevance to LSSVR, e.g., linear, polynomial, sigmoidal, and radial basis function (RBF). Since the reproducing kernel completely characterizes the hypothesis feature space, its choice has a crucial impact on the ability of the LSSVR [48]. The widely selected kernels for traffic flow prediction are RBF and polynomial (Poly) kernels, because of their good learning ability and good generalization ability, respectively [49]. They can be expresses as the following equations

$$k(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2), \tag{13}$$

$$k(x_i, x_j) = [x_i^T \times x_j + t]^q, \tag{14}$$

where Eq. (13) is RBF kernel function, and Eq. (14) represent Poly kernel, in which q determines the number of dimensions in the induced feature space, and t is the bias term in the kernel function.

There are two parameters (γ, σ^2) for the RBF based LSSVR and these parameters (γ, t, q) for the Poly based LSSVR to be determined. Considering the simple implementation (easy to realize and parameter determination), parallel processing, and global optimization, the PSO is applied to determine the parameters of the LSSVR model in this paper. For the standard PSO, it first initializes a set of random particles (usually $10 < N < 100$). Each particle has a position and a velocity for determining the direction and distance of the search. In addition, a fitness value is used to measure the merits of the particle based on the optimization problem. The velocity and position are determined by

$$v_{i,d}(t + 1) = \omega \times v_{i,d}(t) + c_1 \times r_1 \times (pbest_d(t) - x_{i,d}(t))$$
$$+ c_1 \times r_2 \times (gbest_d(t) - x_{i,d}(t)), \tag{15}$$

$$x_{i,d}(t+1) = x_{i,d}(t) + v_{i,d}(t+1), \qquad (16)$$

where the velocity is restricted to $[-v_{max}, v_{max}]$; r_1 and r_2 are random values within $[0\ 1]$; positive constants c_1 and c_2 are personal and social learning factors, respectively; ω is an inertia weight or constriction coefficient; $x_{i,d}$ and $v_{i,d}$ represent the ith particle position and velocity, respectively; and d is the dimension usually randomly initialized in a search space. Individual extreme $pbestd(t)$ is found by the particle itself, and global extreme $gbestd(t)$ is found by the whole population.

C. OVERVIEW OF THE PRESENT APPROACH

There are five steps in constructing the separation-fusion approach for forecasting bus passenger flows. These can be summarized as follows and are described in Fig. 1.
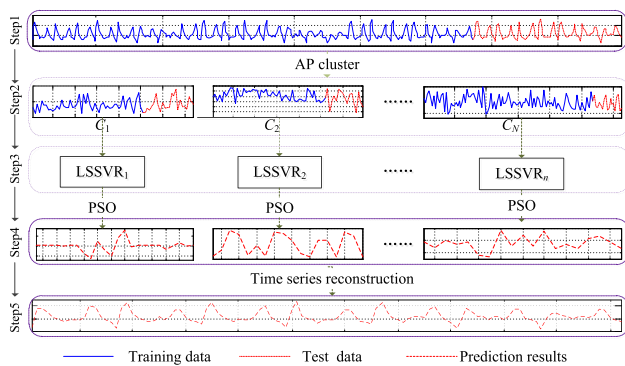

FIGURE 1. Flowchart of the present approach for short-term bus passenger flow forecasting.

Step 1: Collect the passenger flow dataset T partitioned into a training dataset (80%) and a test dataset (20%).

Step 2: Apply AP clustering method to divide T into C disjoined clusters, named as $C_1, C_2, \ldots, C_N$., where each cluster contains both training data as well as testing data.

Step 3: Testing data are predicted by LSSVR model optimized by the PSO for each disjoined cluster.

Step 4: In each cluster, prediction values $[\hat{p}_{c1}, \hat{p}_{c2}, \ldots, \hat{p}_{ck}]$ are generated based on LSSVR model.

Step 5: Rearrange these representations $[\hat{p}_{c1}, \hat{p}_{c2}, \ldots, \hat{p}_{ck}]$ in chronological order, and out final forecasting results.

III. CASE STUDY

A. DATA DESCRIPTION

The bus passenger flow data are collected from one bus service (line 15) in Guangzhou, China. The experimental data are obtained from the Guangdong Public Transport Big Data Contest, covering the period from August 25 to December 28, 2014 (126 days, 18 weeks). Every day, there are a total of 16-time intervals from 6:00 to 21:59. There are 2016 observations (126 days × 16 intervals). The hourly time series are plotted in Fig. 2.

As shown in Fig. 2, the passenger flow of the Guangzhou bus line is almost cyclical and similar every week, except
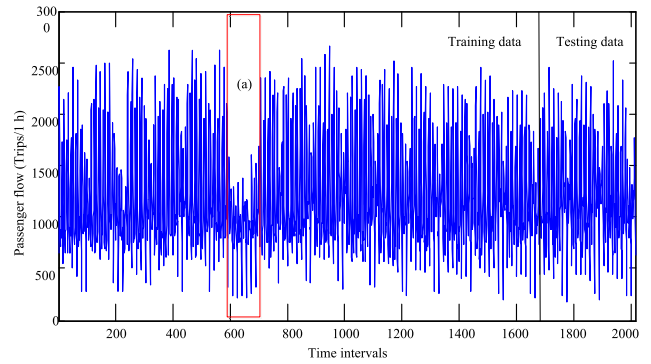

FIGURE 2. Bus passenger flow of line 15 from 25/8/2014 to 28/12/2014 in Guangzhou of China, where (a) represents 1/10/2014 to 7/10/2014 (National Day).
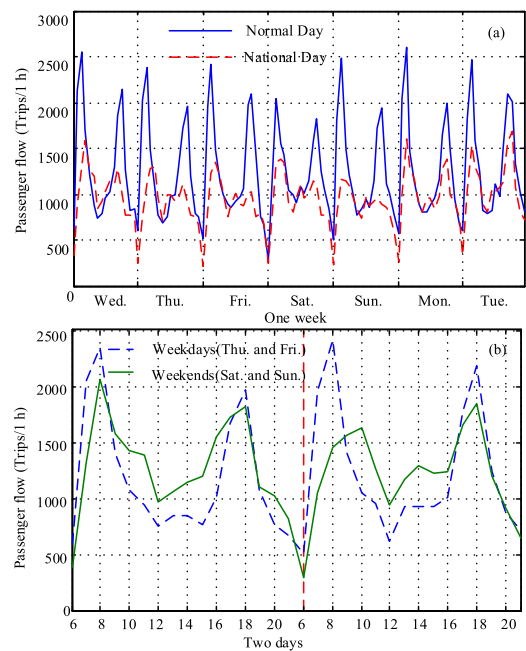

FIGURE 3. Passenger flow distribution of line 15: (a) one week, and (b) typical observed samples of two days.

for the abnormal time intervals during (a) Nation Day. The tendency is enlarged and can be seen clearly in Fig. 3(a). Passenger flow on National Day is significantly less than those on normal days. However, the tendency of passenger flow during one week (chosen randomly from the dataset) is the same as that on Nation Day. As such, it is considered in the proposed model. At the same time, the hourly relevant time series displays a similar regular trend and repeatable pattern from Monday to Sunday.

Looking at typical observed samples for all time intervals on typical weekdays and weekends (shown in Fig. 3(b)), there are two obvious peaks in the morning and afternoon every day. However, unlike on weekdays, passenger flows which are severely affected by random factors, exhibit flexibility and variability on weekends. On weekdays, the peak hour periods are 7:00-9:00 and 17:00-19:00, and the off-peak periods are

10:00-16:00 and 20:00-21:00. On weekends, there are two peaks each day. However, the peak periods are different and later than those on weekdays (shown in Fig. 3(b)). Obviously, weekends are more suitable for people to travel.

B. PERFORMANCE CRITERIA

Two criteria, including the variance of absolute percentage error (VAPE) and mean absolute percentage error (MAPE), are applied to evaluate the prediction performance in this paper. The MAPE is used to measure the mean prediction accuracy, while the VAPE for presenting the prediction stability, described as

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{y_i}, \quad VAPE = Var(|y_i - \hat{y}_i|/y_i). \quad (17)$$

where y_i is the actual value and $\hat{y}_i$ the predicted value.

C. MODEL DEVELOPMENT

For AP algorithm, the similarity matrix is first computed as the input. It takes the time intervals and passenger flow to construct the $s(i, j)$ with 2016 points. An important parameter that decides the number of cluster centers, p is set as the diagonal value $(s(k, k))$. The default value for λ is 0.9 as suggested by [33].

For the SVR, there is not a universal way to determine the optimal model order n (as described in Eq. (6)). One may adopt a trial-and-error method and correlation analysis for this task. The maximum n value is set to be 8 with experience, and n is obtained when the minimum of the MAPE first appears. After determining n, PSO is designed to optimize the parameters of the LSSVR.

To testify the performance of the suggested prediction models, the collected 2016 data points are split into two sets: training and testing sets. The training set contains the first 1680 consecutive data points (the first 15 weeks, August 25 - December 8, 2014), and the testing set contains the remaining 336 data points (the remaining 3 weeks, December 8 - 28 2014). In the following study, the original time series and intervals are normalized as [0, 1].

IV. RESULTS AND DISCUSSION

A. CLUSTERING RESULTS

The bus passenger flows for all of the data points are clustered into 20 components $(C_1, C_2, \ldots, C_{20})$ by AP algorithm as shown in Fig. 4.

Fig. 4 proves that the peak periods are clustered into C_{15}, C_1, C_{18}, C_{10}, and C_{11} and that the non-peak periods are classified into the remaining clusters. It shows that the passenger flows in the same time intervals tend to be classified into different clusters except for C_4 and C_3 because AP considers the similarity between data points. The more the quantities of passenger flow are equal in the same time intervals, the easier they are to classify into the same cluster. When the distance between two-time intervals is more than 0.1, AP algorithm
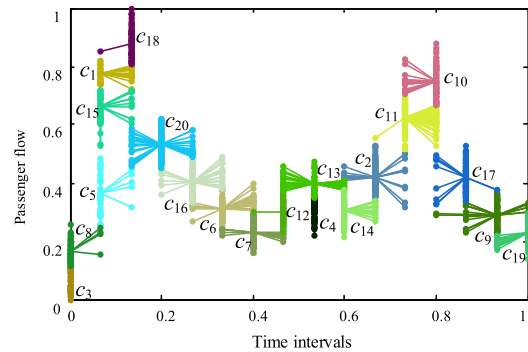

FIGURE 4. Clustering results of AP. Note that passenger flow and time intervals are normalized as [0, 1]. Each point stands for pair samples of $x_i = (x_{i1}, x_{i2})$, where x_{i1} is hourly time, and x_{i2} is passenger flow. Same color points belong to same cluster. Every cluster has a cluster center (exemplar), and remaining ones connect to it through a straight line.

tends to partition equal passenger flows into different clusters such as C_{11} and C_{20}.

B. FORECASTING RESULTS

After identifying patterns, SVR is designed to predict bus passenger flows. Table 1 lists the performance by using the proposed method with RBF kernel and Poly kernel for the passenger flow prediction, respectively.

TABLE 1. Prediction results based on RBF Kernel and poly Kernel.

Kernel	Model	Testing data	
		MAPE(%)	VAPE(%)
Poly kernel	LSSVR	16.94	13.91
	AP_SVR	8.28	10.13
Radial basis function	LSSVR	9.88	11.47
	AP_SVR	7.13	6.77

One can see that the RBF kernel has better prediction accuracy than the Poly kernel for the given case. It means that the RBF kernel is superior to the Poly kernel in the LSSVR model for the passenger flow prediction.

The best-fitted structure models based on RBF kernel for each model are identified according to Table 2, and an optimal n is also obtained toward the minimum MAPE for the testing data.

C. PATTERN RESULTS

It is hard to determine the type of kernel function for specific data patterns [14]. However, RBF kernel function is easier to implement and is capable of non-linearly mapping the training data into infinite dimensional space. Therefore, RBF kernel function is specified in this study.

After obtaining the prediction results of multiple clusters using SVR, the final results are rearranged by chronological order rearrangement. The pattern combination is plotted in Fig. 5. Note that the real line represents the records, and the dashed line means the forecasts.

As shown in Fig. 5, the proposed model captures the deviation between real values and prediction results.

TABLE 2. The specification of performances for input structures of the SVR in each cluster.

Mode	i-1	i-2	i-3	i-4	i-5	i-6	i-7	i-8	n (Optimal)
C_1	0.850	0.863	0.686	0.598	0.641	**0.517**	0.663	0.671	6
C_2	7.008	7.045	7.489	7.708	7.513	7.23	7.328	**6.872**	8
C_3	25.897	21.588	**17.862**	21.465	25.362	25.068	27.406	33.189	3
C_4	4.129	4.705	5.603	4.456	4.838	4.433	**3.728**	5.349	7
C_5	18.949	16.86	16.246	15.913	17.274	**15.826**	16.34	15.964	6
C_6	6.209	**6.058**	6.198	7.015	7.462	7.999	7.793	7.585	2
C_7	9.957	9.555	9.492	9.361	9.589	**9.335**	9.707	9.95	6
C_8	17.901	16.402	**14.628**	16.267	15.971	15.353	16.224	14.830	3
C_9	9.516	9.053	**8.693**	9.053	10.039	10.914	10.49	10.367	3
C_{10}	**4.75**	4.937	4.836	4.78	5.517	5.539	5.882	6.288	1
C_{11}	3.948	5.189	5.935	7.826	5.661	**3.774**	4.309	4.276	6
C_{12}	8.401	8.728	9.945	9.301	**8.371**	9.255	9.724	10.455	5
C_{13}	5.815	5.481	**5.402**	5.965	6.359	6.211	5.556	5.493	3
C_{14}	7.828	7.911	8.126	8.386	7.887	8.042	**7.804**	8.036	7
C_{15}	**6.986**	7.502	7.373	7.131	7.113	7.081	7.059	7.001	1
C_{16}	7.492	7.321	7.014	**6.599**	6.725	7.882	7.831	7.42	4
C_{17}	8.648	11.59	11.677	10.757	**7.428**	8.733	8.877	8.418	5
C_{18}	4.531	4.463	4.338	4.393	**4.217**	4.430	4.440	4.384	5
C_{19}	**9.233**	9.424	9.364	9.895	9.956	9.872	9.960	9.916	1
C_{20}	6.204	6.186	6.258	6.132	6.048	**5.908**	5.937	5.913	6

Note: Values in bold denote the performances of the optimal variables.

The tendency of those curves shows that the forecasting results fit the actual values well, except for the rush hour (marked with arrowhead) on weekends. The passenger flow on weekends is more flexible and changeable than on weekdays due to the severe effect of random factors.

To further analyze the superiority of the proposed approach, the aforementioned MAPE and VAPE were employed as the statistical variables for the test. It is calculated that MAPE value is 7.13% and VAPE value is 6.77% for the forecasting results, indicating high accuracy and stability (detailed comparisons are shown in Table 4).

According to both quantitative and qualitative analyses, the proposed model is suitable for capturing the structure of non-linear passenger flow well. The performance can be attributed to the similarity feature observation.

D. COMPARISONS WITH PEER METHODS

To highlight the forecasting performance of the proposed model, four models are employed for comparisons using the same dataset, including three non-parametric models (KFCM-based SVR, SVR, and BPNN models) and a parametric model (seasonal ARIMA (SARIMA) model). Structures and parameters of these models are listed in Table 3.

TABLE 3. Structure and parameters of different models for comparisons.

|	Model Structure and parameters for comparisons
KFCM-based SVR	The number of clusters is 7 with Xie-Beni validity index [50] 0.11369. Optimal model order $n = 8$.
BPNN	Input nodes = 8, hidden layer = 7, output nodes = 1
SARIMA	Autoregressive degree (AR) = 2, differencing degree = 0, moving average degree (MA) = 1, seasonality autoregressive degree (SAR) = 2, seasonality differencing degree = 1, seasonality moving average degree (SMA) = 16, and seasonality = 16.
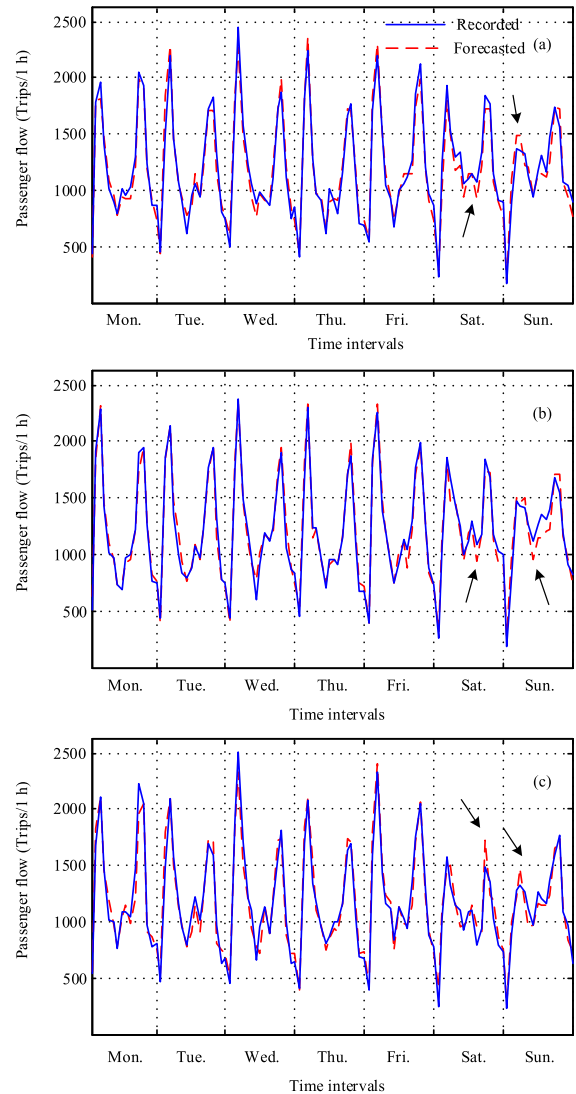
FIGURE 5. Forecasting results by using the present AP-SVR model. The testing data is split three weeks, where (a) is first week, (b) second week, and (c) third week.

TABLE 4. Performance criteria of different forecasting methods.

Models	MAPE (%)	VAPE (%)
AP-SVR	7.13	6.77
LSSVR	9.88	11.47
KFCM-SVR	11.61	14.18
BPNN	11.85	14.38
SARIMA	12.57	15.55

Note that the structures of the models are determined by the experiments (except for the SARIMA model, which is determined by Ref. [4]). Their training processes and parameters are the same as those of the proposed model.

KFCM model uses a new kernel-induced distance to replace Euclidean distance in the FCM model. It uses a nonlinear function to map into a feature space that may have more dimensions. Thus, the original linear indivisible sample
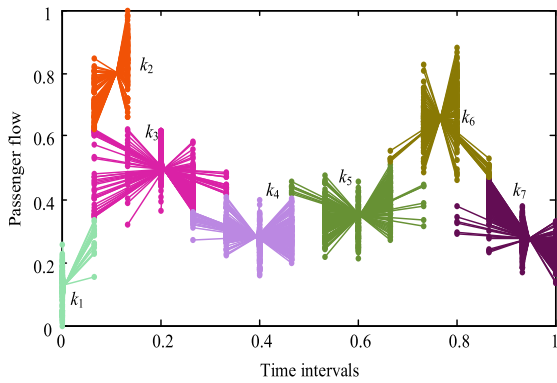
FIGURE 6. KFCM clustering results.

points become linearly separable in the nuclear space [31]. The clustering results of the KFCM model are shown in Fig. 6. Note that the description of Fig. 6 is consistent with the description of Fig. 4. Comparing Figs. 4 and 6 show that the clustering result of AP algorithm is more detained than that of KFCM model. Using the peak period as an example, there are five clusters ($C_{15}, C_1, C_{18}, C_{10}, C_{11}$) in AP algorithm and two clusters (K_2, K_6) in KFCM model. Quantitative evaluation results with MAPE are 6.981%, 0.291%, 3.869%, 4.599%, and 3.534% for $C_{15}, C_1, C_{18}, C_{10}$, and C_{11}, respectively, and 8.574% and 10.966% for K_2 and K_6, respectively. As such, the peak periods with more clusters are stable and suitable for forecasting models. At the same time, the cluster center of AP exists in the real data. However, KFCM model depends on a partition matrix. Comprehensively, AP is an automatic clustering algorithm. The number of clusters in the algorithm is obtained from the real-valued messages between data points. On the contrary, KFCM model comes from the lowest Xie-Beni Validity Index [50].

Forecasting results for different models are compared in Fig. 7 using the testing data of the first week. Comparing with AP-SVR and SVR model, KFCM-SVR model performances poor quality both in forecasting peak value and hush hour at weekends due to scant cluster analysis. Fig. 7 shows that the SARIMA model has limitations in predicting nonlinearity and randomness. Comprehensively, the AP-based SVR has the best performance among these models.

It is notes that MAPE is the mean value of all the absolute percentage errors (APEs). Fig. 8 shows the contour map of the forecast APEs for different models. The AR-SVR model provides the best prediction. Generally, the highest APEs for the four models occur in the early morning period (6:00-7:00) on weekends, especially for the SARIMA model (APE higher than 90%). The passenger flow fluctuates greatly on weekends due to the destinations and departure times. It is hard to capture all of the different features for different intervals, especially for the transition periods. Comparatively, the AP-SVR model performs better than the other models for this period. On weekends, other than the early morning period, the times 15:00 and 20:00 exhibit higher APEs. Unlike the weekdays, these two time intervals represent peak hours for
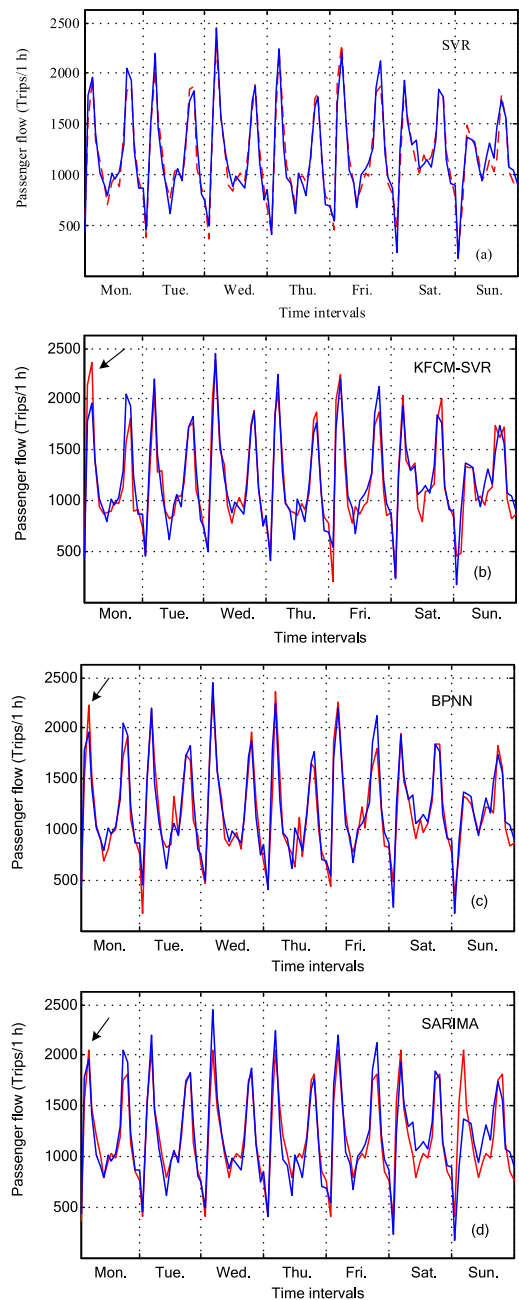

FIGURE 7. Forecasted results for the testing data of the first week by using: (a) KFCM-based SVR, (b) SVR, (c) BPNN, and (d) SARIMA models, respectively, where the blue line represent the real value, and the red line is the forecasted data.

travelers heading out and back. For weekdays, during the 6:00-9:00 period, the prediction APEs are under 12%, except for Tuesday in the AR-SVR and BPNN models, Thursday in the KFCM-SVR model, Tuesday and Wednesday in the LSSVR model, and Monday in the SARIMA model. This may be due to the relatively stable recurrent demand for the AM peak period. The APEs of the four models during the 12:00-14:00 period and at 20:00 are higher than during the other time intervals (except for 6:00-9:00). Fig. 3(b) shows that 12:00-14:00 is rush hour. At 20:00, the passenger
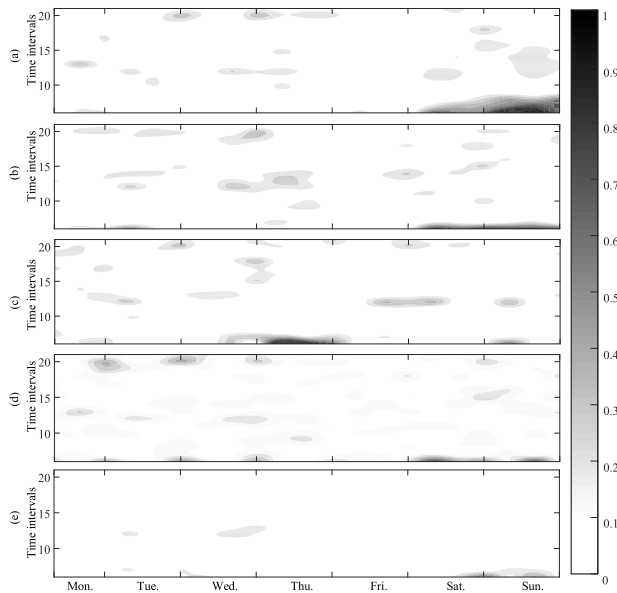
FIGURE 8. Contour map of the forecast APE for different models. (a) SARIMA model, (b) BPNN model, (c) KFCM-SVR model, (d) LSSVR, and (e) AP-SVR model.

flow decreases after the peak period. Comprehensively, the AP-SVR model can deal with this rush hour more intelligently.

By comparing Fig. 5(a) with Fig. 7, one can barely identify (but not very clear) the good result of the present AR-SVR. For this reason, the aforementioned MAPE and VAPE were again employed as the statistical variables for the comparison. Table 4 shows the corresponding prediction performances using different models. It shows that the addressed AP-SVR model performs well in terms of accuracy because it has the lowest MAPE value (7.13%), whereas the SARIMA model performs worst because it has the highest MAPE value (12.57%). Obviously, the forecasting accuracy of the AP-SVR model, considering the similarity of data points and presents data in a finite and small number of oscillatory modes, outperforms the other models. Compared with the LSSVR model, the KFCM-SVR model is limited in pattern identification. Additionally, the VAPE values of the four models are shown in Table 4. The AP-SVR model presents the best forecasting stability because it has the smallest VAPE value. The MAPE and VAPE values of the SARIMA model for the test dataset are 12.57% and 15.55%, respectively. This model has the worst forecasting accuracy and stability among all of the models considered in this paper. Obviously, the SARIMA model does not predict the passenger flow well. The results indicate that it is not suitable for constructing a forecasting model for non-linear passenger flow due to its assumption of linearity.

V. CONCLUSION

A novel AP-SVR approach has been proposed to forecast bus passenger flows. The dataset was first segmented into different clusters by AP algorithm. Subsequently, SVR was optimized by PSO for forecasting passenger flows for each cluster. Finally, the results of all SVRs were rearranged with chronological order. Taking bus line 15 in Guangzhou, China as a case study, this paper evaluated AP-SVR model. According to both quantitative and qualitative analyses, the proposed model is superior to the comparison models. The addressed model weakens the non-linear characteristics of bus passenger flows to some extent due to its similarity clustering. Pattern recognition and feature learning are capable of improving the performance accuracy and stability. However, the model does not comprehensively forecast disruptions caused by special events and the external environment, such as weather and temperature. These factors and additional forecasting capabilities should be investigated in future.

REFERENCES

[1] 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.

[2] F. Schimbinschi, L. Moreira-Matias, V. X. Nguyen, and J. Bailey, "Topology-regularized universal vector autoregression for traffic forecasting in large urban areas," *Expert Syst. Appl.*, vol. 82, pp. 301–316, Oct. 2017.

[3] C. Li, Y. Tao, W. Ao, S. Yang, and Y. Bai, "Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition," *Energy*, vol. 165, pp. 1220–1227, Dec. 2018.

[4] M. Milenković, L. Švadlenka, V. Melichar, N. Bojović, and Z. Avramović, "SARIMA modelling approach for railway passenger flow forecasting," *Transport*, vol. 33, no. 5, pp. 1113–1120, 2018.

[5] S. Anvari, S. Tuna, M. Canci, and M. Turkay, "Automated Box-Jenkins forecasting tool with an application for passenger demand in urban rail systems," *J. Adv. Transp.*, vol. 50, no. 1, pp. 25–49, Jan. 2016.

[6] A. Samagaio and M. Wolters, "Comparative analysis of government forecasts for the Lisbon Airport," *J. Air Transp. Manage.*, vol. 16, no. 4, pp. 213–217, Jul. 2010.

[7] X. Xiao, J. Yang, S. Mao, and J. Wen, "An improved seasonal rolling grey forecasting model using a cycle truncation accumulated generating operation for traffic flow," *Appl. Math. Model.*, vol. 51, pp. 386–404, Nov. 2017.

[8] A. Stathopoulos and M. G. Karlaftis, "A multivariate state space approach for urban traffic flow modeling and prediction," *Transp. Res. C, Emerg. Technol.*, vol. 11, no. 2, pp. 121–135, Apr. 2003.

[9] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, to be published.

[10] F. Moretti, S. Pizzuti, S. Panzieri, and M. Annunziato, "Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling," *Neurocomputing*, vol. 167, pp. 3–7, Nov. 2015.

[11] B. Sun, W. Cheng, P. Goswami, and G. Bai, "Short-term traffic forecasting using self-adjusting k-nearest neighbours," *IET Intell. Transp. Syst.*, vol. 12, no. 1, pp. 41–48, Feb. 2018.

[12] J. Guo, W. Huang, and B. M. Williams, "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 50–64, Jun. 2014.

[13] J. Guo, Z. Liu, W. Huang, Y. Wei, and J. Cao, "Short-term traffic flow prediction using fuzzy information granulation approach under different time intervals," *IET Intell. Transp. Syst.*, vol. 12, no. 2, pp. 143–150, Mar. 2018.

[14] Y. Sun, B. Leng, and W. Guan, "A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system," *Neurocomputing*, vol. 166, pp. 109–121, Oct. 2015.

[15] L. Li, X. Chen, and L. Zhang, "Multimodel ensemble for freeway traffic state estimations," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 3, pp. 1323–1336, Jun. 2014.

[16] Y. Li, X. Wang, S. Sun, X. Ma, and G. Lu, "Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks," *Transp. Res. C, Emerg. Technol.*, vol. 77, pp. 306–328, Apr. 2017.

[17] G. Xie, S. Wang, Y. Zhao, and K. K. Lai, "Hybrid approaches based on LSSVR model for container throughput forecasting: A comparative study," *Appl. Soft Comput.*, vol. 13, no. 5, pp. 2232–2241, May 2013.

[18] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, and R. E. Vásquez, "Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis," *Neurocomputing*, vol. 168, pp. 119–127, Nov. 2015.

[19] M.-L. Huang, "Intersection traffic flow forecasting based on *v*-GSVR with a new hybrid evolutionary algorithm," *Neurocomputing*, vol. 147, pp. 343–349, Jan. 2015.

[20] K. Wang, W. Zhu, and P. Zhong, "Robust support vector regression with generalized loss function and applications," *Neural Process Lett*, vol. 41, no. 1, pp. 89–106, Feb. 2015.

[21] W. Hu, L. Yan, K. Liu, and H. Wang, "A short-term traffic flow forecasting method based on the hybrid PSO-SVR," *Neural Process. Lett.*, vol. 43, pp. 155–172, Feb. 2016.

[22] W.-C. Hong, M.-W. Li, J. Geng, and Y. Zhang, "Novel chaotic bat algorithm for forecasting complex motion of floating platforms," *Appl. Math. Model.*, vol. 72, pp. 425–443, Aug. 2019.

[23] Z. Zhang and W.-C. Hong, "Electric load forecasting by complete ensemble empirical mode decomposition adaptive noise and support vector regression with quantum-based dragonfly algorithm," *Nonlinear Dyn.*, vol. 98, no. 2, pp. 1107–1136, 2019.

[24] T.-H. Tsai, C.-K. Lee, and C.-H. Wei, "Neural network based temporal feature models for short-term railway passenger demand forecasting," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3728–3736, Mar. 2009.

[25] M. Khashei and M. Bijari, "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 2664–2675, Mar. 2011.

[26] T. Pant, C. Han, and H. Wang, "Examination of errors of table integration in flamelet/progress variable modeling of a turbulent non-premixed jet flame," *Appl. Math. Model.*, vol. 72, pp. 369–384, Aug. 2019.

[27] Y. Zhang and Y. Liu, "Analysis of peak and non-peak traffic forecasts using combined models," *J. Adv. Transp.*, vol. 45, no. 1, pp. 21–37, Jan. 2011.

[28] A. Roy, R. M. Cruz, R. Sabourin, and G. D. Cavalcanti, "A study on combining dynamic selection and data preprocessing for imbalance learning," *Neurocomputing*, vol. 286, pp. 179–192, Apr. 2018.

[29] Y. Bai, Z. Sun, B. Zeng, J. Deng, and C. Li, "A multi-pattern deep fusion model for short-term bus passenger flow forecasting," *Appl. Soft Comput.*, vol. 58, pp. 669–680, Sep. 2017.

[30] C. Li, M. Cerrada, D. Cabrera, R. V. Sanchez, F. Pacheco, G. Ulutagay, and J. Valente De Oliveira, "A comparison of fuzzy clustering algorithms for bearing fault diagnosis," *J. Intell. Fuzzy Syst.*, vol. 34, no. 6, pp. 3565–3580, Jun. 2018.

[31] Y. Xiao, J. J. Liu, Y. Hu, Y. Wang, K. K. Lai, and S. Wang, "A neuro-fuzzy combination model based on singular spectrum analysis for air transport demand forecasting," *J. Air Transp. Manage.*, vol. 39, pp. 1–11, Jul. 2014.

[32] C.-F. Chen, M.-C. Lai, and C.-C. Yeh, "Forecasting tourism demand based on empirical mode decomposition and neural network," *Knowl.-Based Syst.*, vol. 26, pp. 281–287, Feb. 2012.

[33] E. Lord, M. Willems, F.-J. Lapointe, and V. Makarenkov, "Using the stability of objects to determine the number of clusters in datasets," *Inf. Sci.*, vol. 393, pp. 29–46, Jul. 2017.

[34] L. He and H. Zhang, "Kernel K-means sampling for Nyström approximation," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2108–2120, May 2018.

[35] R. Kuo, T. Lin, F. Zulvia, and C. Tsai, "A hybrid metaheuristic and Kernel intuitionistic fuzzy c-means algorithm for cluster analysis," *Appl. Soft Comput.*, vol. 67, pp. 299–308, Jun. 2018.

[36] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.

[37] P. Li, H. Ji, B. Wang, Z. Huang, and H. Li, "Adjustable preference affinity propagation clustering," *Pattern Recognit. Lett.*, vol. 85, pp. 72–78, Jan. 2017.

[38] F. Shang, L. C. Jiao, J. Shi, F. Wang, and M. Gong, "Fast affinity propagation clustering: A multilevel approach," *Pattern Recognit.*, vol. 45, no. 1, pp. 474–486, Jan. 2012.

[39] Y. Bai, P. Wang, C. Li, J. Xie, and Y. Wang, "A multi-scale relevance vector regression approach for daily urban water demand forecasting," *J. Hydrol.*, vol. 517, pp. 236–245, Sep. 2014.

[40] J. Guo, Z. Sun, H. Tang, X. Jia, S. Wang, X. Yan, G. Ye, and G. Wu, "Hybrid optimization algorithm of particle swarm optimization and cuckoo search for preventive maintenance period optimization," *Discrete Dyn. Nature Soc.*, vol. 2016, Feb. 2016, Art. no. 1516271.

[41] H. Kundra and H. Sadawarti, "Hybrid algorithm of cuckoo search and particle swarm optimization for natural terrain feature extraction," *Res. J. Inf. Technol.*, vol. 7, no. 1, pp. 58–69, Jan. 2015.

[42] W. Hang, F.-L. Chung, and S. Wang, "Transfer affinity propagation-based clustering," *Inf. Sci.*, vol. 348, pp. 337–356, Jun. 2016.

[43] G. Gan, Y. Zhang, and D. K. Dey, "Clustering by propagating probabilities between data points," *Appl. Soft Comput.*, vol. 41, pp. 390–399, Apr. 2016.

[44] M. Mézard, "Where are the exemplars?" *Science*, vol. 315, no. 5814, pp. 949–951, Feb. 2007.

[45] V. Vapnik, S. Golowic, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," in *Proc. Adv. Neural Inf. Process Syst.*, vol. 9, Jan. 1997, pp. 281–287.

[46] J. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, pp. 293–300, Jun. 1999.

[47] Y. Bai, Z. Sun, B. Zeng, J. Long, L. Li, J. V. De Oliveira, and C. Li, "A comparison of dimension reduction techniques for support vector machine modeling of multi-parameter manufacturing quality prediction," *J. Intell. Manuf.*, vol. 30, no. 5, pp. 2245–2256, Jun. 2019.

[48] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, no. 3, pp. 657–682, Mar. 2014.

[49] X. Feng, X. Ling, H. Zheng, Z. Chen, and Y. Xu, "Adaptive multi-kernel SVM with spatial–temporal correlation for short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2001–2013, Jun. 2019.

[50] X. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841–847, Aug. 1991.

CHUAN LI (Senior Member, IEEE) received the Ph.D. degree from Chongqing University, China, in 2007. He has been successively a Postdoctoral Fellow with the University of Ottawa, Canada, a Research Professor with Korea University, South Korea, and a Senior Research Associate with the City University of Hong Kong, China. He is currently a Professor with Chongqing Technology and Business University, China. His research interests include prognostics and health management, and intelligent systems.
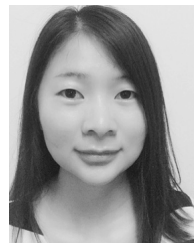
XIAODAN WANG received the B.S. degree from the Henan University of Chinese Medicine, China, in 2013. She is currently pursuing the M.S. degree with Chongqing Technology and Business University, China. Her research interests include traffic intelligence prediction and machine learning.

ZHIWEI CHENG received the B.S. degree in engineering from the Chongqing University of Arts and Sciences, in 2013. He is currently pursuing the M.S. degree with Chongqing University, China. He is also a Data Analyst with Chongqing CPI Zineng Science and Technology Company, Ltd. His research interests include machinery healthy maintenance, artificial intelligence, and data mining.

YUN BAI (Member, IEEE) received the Ph.D. degree from Chongqing University, China, in 2014. He is currently a Postdoctoral Researcher with the University of Algarve, Portugal. He is also an Assistant Professor with Chongqing Technology and Business University, China. His current research interests include intelligent system management, modeling, and forecasting.

• • •