A Data-Driven Fuzzy Information Granulation Approach for Freight Volume Forecasting

Shen Yin, *Senior Member, IEEE*, Yuchen Jiang, *Student Member, IEEE*, Yang Tian, and Okyay Kaynak, *Fellow, IEEE*

***Abstract*—The performance of the logistic system is one of the most important aspects in regional economy, and the freight volume is the biggest part of the logistic system. In this paper, an information granulation method is introduced to represent the freight volume in a fuzzy manner. After the characteristic features have been extracted from the raw time-series data and represented as information granules, the granules are modeled with the support vector machine (SVM). In consideration of both algorithm efficiency and prediction accuracy, an efficient version of SVM called least square (LS) SVM is employed and integrated with a parameter optimization algorithm, the particle swarm optimization. Simulation results on a real dataset illustrate the performance of the proposed method, and comparison studies are carried out with LS and partial LS-based methods.**

***Index Terms*—Data-driven, information granulation, particle swarm optimization (PSO) algorithm, prognosis, support vector machine (SVM).**

I. Introduction

WITH the rapid development of road transportation, the requirements of the freight traffic statistics for governmental transportation departments increase progressively. Some works have been published on this topic [1], [2]. It is also urgent that the government issues corresponding policies based on the traffic volume in the region. Road freight volume reflects the capacity and efficiency of the freight transport in the logistic system, and can also be used to estimate the role of road transportation in the transportation system. The statistics of freight volume are difficult to obtain, and traditional statistical methods cannot predict the real status of the transportation market accurately [3]. On the other hand, dynamic characteristics of the logistic systems are closely related with and affected by tons of factors in the real world, and could accordingly suffer from severe nonlinearity. Therefore, compared with modeling the traffic volume time series based on other factors, it is of great importance to study the volume prediction problem based on historical traffic data. In this paper, an information granulation method, which ensures that the granules contain enough information, is introduced to describe the freight volume in a fuzzy fashion. Information granulation can be either fuzzy or nonfuzzy. Nonfuzzy methods play an important role in many fields, including rough set quantizing, decision tree, and cluster analysis. However, as a matter of fact, human beings are more accustomed to learning and deducing in a fuzzy manner.

Research works focused on the freight volume forecasting have been carried out at home and abroad. Comparison studies among ordinary least square (LS) regression, principle component regression, and partial least squares (PLSs) regression approaches are reported in [4] on a China railway logistic dataset. The aforementioned algorithms are also tested in our work and the simulation results reveal the limitation of these statistical approaches due to system nonlinearity. In [5], a fuzzy-set time-series clustering method is proposed for dry cargo freight market modeling and prediction, based on the classification of historical occurrence with respect to the financial lead–lag relationship. The drawback therein lies in the negative effect of the response timing or equivalently the time lag. Moreover, with an alternative approach, passenger and freight traffic volume are forecasted with a combined model based on the variance reciprocal and optimal weighting to reduce the square error during the last period of time.

In recent years, a number of machine-learning approaches have been applied gradually to the industry and the economical fields [6]–[8], and one typical example is the artificial neural network (ANN) [9], [10]. Since ANN can imitate nonlinear functions, its prediction results are usually better than the counterparts obtained by the traditional methods [11]. However, ANN tends to fall into local optimal solutions instead of the global one [12]. Proposed in the 1990s, support vector machine (SVM) is a potential solution to this problem [13]. SVM is mainly used for classification and regression, and it has the unique advantage to deal with datasets with small samples sizes, nonlinearity, as well as high dimensionality [14]. Compared with ANN, SVM has better generalization ability and is more commonly used in the nonlinear regression and prediction problems [15]. Based on the original SVM approach, least

Manuscript received November 15, 2015; revised May 1, 2016 and August 9, 2016; accepted September 2, 2016. Date of publication September 27, 2016; date of current version January 10, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61304102, Grant 61472104, and Grant 61333012, and in part by the Fundamental Research Funds for the Central Universities under Grant HIT.BRETIV.201507.

S. Yin, Y. Jiang, and Y. Tian are with the Research Institute of Intelligent Control and Systems, Harbin Institute of Technology, Harbin 150001, China (e-mail: shen.yin2011@googlemail.com; yc.jiang2016@foxmail.com; yang.tian@gatech.edu).

O. Kaynak is with the Research Institute of Intelligent Control and Systems, Harbin Institute of Technology, Harbin 150001, China, and also with the Department of Electrical and Electronics Engineering, Bogazici University, Istanbul 34342, Turkey (e-mail: okyay.kaynak@boun.edu.tr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIE.2016.2613974

squares support vector machine (LSSVM) was proposed as an improved version. LSSVM changes the inequality constraints of SVM into equality constraints, and in the meantime transforms the original quadratic programming (QP) problem into linear equations. As a result, LSSVM reduces the computational complexity [16], [17]. Moreover, LSSVM involves less parameters than SVM, and can therefore solve the approximation problems easily. Furthermore, since parameters play a significant role in both modeling and generalization procedures, particle swarm optimization (PSO) algorithm is employed in this paper to select and optimize the parameters of LSSVM for more robust models [18].

The contributions of this work are as follows. First, an information granulation scheme is employed to construct three granules' time series from the original data. Compared with directly using the raw dataset to build the SVM models, the granules are equipped with more data information. The novel representation of the freight volume may bring benefits to the modeling phase, which is carefully discussed and verified in this paper. Second, LSSVM and PSO are combined, so as to simultaneously take algorithm efficiency and the prediction performance into consideration. Specifically, LSSVM acts as an alternative to the standard time-consuming SVM approach, and PSO functions as a parameter optimization method. Third, extensive comparison studies with traditional statistics approaches are conducted, which may serve as a reference to the future works in this area.

The organization of this paper is as follows. In the forthcoming section, the theoretical background and the mathematical models are introduced. Section III presents the technical core of a new combined prediction scheme. Section IV applies the new approach in the freight volume forecasting tasks, and comparison results are presented and discussed. Finally, Section V concludes this paper.

II. Related Work

A. Fuzzy Information Granulation

Information granulation refers to the process of dividing a holistic object into several parts, where each part can be recognized as an information granule. Prof. Zadeh thought of information granules as a combination of some elements, which are difficult to distinguish due to their characteristics [19]. Information granulation is the core of granular computing and computing with words. Its investigation objectives include information granule generation, expression, and linguistic explanation [20]. Information granulation has become one of the most important research branches in the artificial intelligence field, and is regarded as an effective tool for data analysis.

Although the concept of information granulation seems abstract, it is much easier to be understood through vivid examples, since information granulation is closely related to many aspects of human life. For instance, when the head is considered as the object, it contains information granules including eyes, ears, nose, mouth, cheek, and hair. These components not only constitute the system, but have individual functionality as well. Accordingly, these components are equipped with characteristic information and can in turn describe the whole system. Humans are accustomed to fuzzy concepts and fuzzy deduction, and obviously the nonfuzzy methods contradict this natural fact. For example, humans usually describe time intervals with "long" and "short." However, these concepts are equivocal and cannot be expressed accurately. As a consequence, fuzzy information granulation is introduced to solve this problem.

There are three information granulation models, including those based on fuzzy set, rough set, and quotient space. These models are related yet different. This paper utilizes only the fuzzy-set-based model, which can be represented by

$$g \triangleq (x \text{ is } G) \text{ is } \lambda \tag{1}$$

where x is a variable in the universe U, and G is a convex fuzzy set in the same universe, described by a membership function μ_G, and λ is the probability.

Prof. Zadeh thought that language is a rough granulation tool, which is very useful when people think and decide [19]. Language-based deduction and analysis involve computing with words, and they have normal rules similar to arithmetic ones. The only difference from the arithmetic rules is that fuzzy logic and relations are considered here.

Fuzzy information granulation of time series consists of two steps, namely, window generation and fuzzification. In the first step, time-series data are divided into a sequence of segments, which are recognized as operational windows. The reason to generate operational windows is that in most cases the data density has to be guaranteed identical to ensure the subsequent analysis is reasonable. After that, the data within the same window are fuzzified to generate granules. The second step is very critical. In order to retain information as much as possible, reasonable fuzzy sets should be established on operational windows. Given a time series X, it is possible to generate windows iteratively, so fuzzification in each window is first discussed. Fuzzification is to establish a fuzzy granule P or a fuzzy concept G, which can best represent the time series

$$g \triangleq (x \text{ is } G). \tag{2}$$

In this case, the core issue of fuzzy information granulation is to determine a membership function $A = \mu_G$. Different types of fuzzy concepts have different membership functions. There are some popular fuzzy forms, including triangle, trapezoid, and asymmetric Gauss, which are introduced in the following.

Triangle form:

$$A(x,a,m,b) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{m-a}, & a < x \le m \\ \dfrac{b-x}{b-m}, & m < x \le b \\ 0, & x \ge b \end{cases}. \tag{3}$$

Trapezoid form:

$$A(x,a,m,n,b) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{m-a}, & a < x \le m \\ 1, & m < x \le n \\ \dfrac{b-x}{b-n}, & n < x \le b \\ 0, & x \ge b \end{cases}. \tag{4}$$

Asymmetric Gauss form:

$$A(x, m, \sigma, \mu) = \begin{cases} \exp\left[-(x-m)/\sigma^2\right], & x \le m \\ \exp\left[-(x-m)/\mu^2\right], & x > m \end{cases}. \quad (5)$$

In this paper, the information granulation method proposed by Pedrycz is adopted [21]. The granules to be solved should satisfy that 1) the original data are fully expressed by granules; and (2) the granules possess some unique features and hold certain characteristics. In order to meet these requirements, a function of the membership function A is constructed to describe the quality of fuzzy information granulation

$$Q_A = \frac{M_A}{N_A} \quad (6)$$

where

$$M_A = \sum_{x \in X} A(x) \quad (7)$$

$$N_A = \mathtt{metric}\{\mathtt{support}(A)\} \quad (8)$$

satisfy the demand that M_A meets the first requirement and N_A meets the second one.

B. Support Vector Machine

SVM was proposed by Vapnik in 1995, mainly to address the problems of classification and regression [22]. It is established on statistical learning and belongs to the category of machine learning. According to the VC dimension theory and structural risk minimization principle, SVM aims to find the right balance between model complexity and learning ability, so as to suppress the occurrence of overfitting and underfitting. This characteristic enables SVM to display an excellent generalization performance. Different categories of samples are distinguished by a hyperplane in the feature space. In terms of the datasets with inseparable samples, slack variables and penalty factor are introduced. With the nonlinear projection by the kernel function, the inseparable samples in the low-dimensional feature space can be classified in the high-dimensional one [23].

SVM training is relatively time consuming in case of large-scale dataset, as it involves solving a QP problem. It is also a problem to select an appropriate kernel function for a given task. In addition, the performance of SVM depends heavily on parameter optimization [24].

With a set of training data (x_i, y_i), $i = 1, \ldots, l$, where x_i is the input samples and y_i is the corresponding category labels, the regression function is defined as

$$y = (w^* \cdot x_i) + c \quad (9)$$

where w^* is the weight vector and c is the bias term. These two parameters can be determined based on

$$\min_{w^*, c, \zeta^{(*)}} \frac{1}{2}\|w^*\|^2 + P\sum_{i=1}^{l}(\zeta_i + \zeta_i^*) \quad (10)$$

$$\mathtt{s.t.} \begin{cases} ((w^* \cdot x_i) + c) - y_i \le \varepsilon + \zeta_i, & i = 1, \ldots, l \\ y_i - ((w^* \cdot x_i) + c) \le \varepsilon + \zeta_i^*, & i = 1, \ldots, l \\ \zeta_i^* \ge 0, & i = 1, \ldots, l \end{cases} \quad (11)$$

where Pt is the penalty variable, ζ_i^* is the slack variable, and ε is the insensitive loss function.

The introduction of kernel functions, which should satisfy the Mercer condition, can map the samples from a lower feature dimension to a higher one despite that the nonlinear projection function is unknown. The nonlinear regression function is defined as

$$y = \sum_{i=1}^{l}(\overline{\alpha}_i' - \overline{\alpha}_i)K(x_i, x) + \overline{c} \quad (12)$$

$$\mathtt{s.t.} \begin{cases} \sum_{i=1}^{l}(\overline{\alpha}_i' - \overline{\alpha}_i) = 0 \\ 0 \le \overline{\alpha}_i', \overline{\alpha}_i \le P, i = 1, \ldots, l \end{cases} \quad (13)$$

where $\overline{\alpha}^{(')} = (\overline{\alpha}_1, \overline{\alpha}_1', \ldots, \overline{\alpha}_l, \overline{\alpha}_l')^T$ is the solution.

The performance of the SVM regression model depends on the values of related parameters. The value of P is very essential, as it is responsible for keeping the balance between the maximum of classification interval in the hyperplane and the minimum of training error, and also impacts on the fitting degree. Therefore, both P and the parameters of the kernel function should be tuned.

C. Least Squares Support Vector Machine

LSSVM was proposed to overcome the shortcomings of the original SVM. In the objective function, the square of error replaces the original slack variables. Thereby the QP problem is transformed into solving a set of linear equations, and analytical solutions can be achieved. Compared with the original SVM, LSSVM reduces the computational complexity and therefore improves the algorithm efficiency [25]. To be specific, the complexity of SVM is $O(\max(n, d), \min(n, d)^2)$, where n is the sample size and d is the feature dimensions [23], whereas the complexity of LSSVM is $O(n\log(n))$ [26]. However, this improvement is accompanied by the loss of sparsity in the solution. In SVM, only a portion of samples have nonzero Lagrange multipliers which participate in the construction of the SVM model. With LSSVM, the Lagrange multipliers of all data points are not zeros, which indicates that the model is constructed based on all the samples.

With a set of training data (x_i, y_i), $i = 1, \ldots, n$, the regression model of LSSVM can be denoted by

$$y = w^T\phi(x) + c \quad (14)$$

where $\phi(x)$ is a nonlinear transformation function, w denotes the weight coefficient, and c is the bias term. According to the empirical risk minimization principle, the cost function to be minimized is represented as

$$\min J(w, e_i) = \min\left(\frac{1}{2}w^T w + \gamma\frac{1}{2}\sum_{i=1}^{n} e_i^2\right) \quad (15)$$

$$\text{s.t.} \quad y_i = w^T\phi(x_i) + c + e_i \quad (16)$$

where e_i is a random error and γ is the regularization parameter. The constrained optimization problem is solved by constructing a Lagrange function

$$L(w,c,e_i,\bar{\alpha}_i)=J(w,e_i)-\sum_{i=1}^{n}\bar{\alpha}_i\{w^T\phi(x_i)+c+e_i-y_i\} \tag{17}$$

where $\bar{\alpha}_i$ is the Lagrange multiplier. The solution of (17) can be obtained by partially differentiating both sides with respect to w, c, e_i, and $\bar{\alpha}_i$, respectively.

$$\begin{cases}\frac{\partial L}{\partial w}=0 & \rightarrow \quad w=\sum_{i=1}^{n}\bar{\alpha}_i\phi(x_i)\\ \frac{\partial L}{\partial c}=0 & \rightarrow \quad \sum_{i=1}^{n}\bar{\alpha}_i=0\\ \frac{\partial L}{\partial e_i}=0 & \rightarrow \quad \bar{\alpha}_i=\gamma e_i\\ \frac{\partial L}{\partial \bar{\alpha}_i}=0 & \rightarrow \quad w^T\phi(x_i)+c+e_i-y_i=0\end{cases}. \tag{18}$$

Equation (18) can then be rewritten as

$$\begin{pmatrix}0 & \vec{1}^T\\ \vec{1} & \Omega+\gamma^{-1}I\end{pmatrix}\begin{pmatrix}c\\ \bar{\alpha}\end{pmatrix}=\begin{pmatrix}0\\ y\end{pmatrix}. \tag{19}$$

The parameters c and $\bar{\alpha}$ can be determined based on (19). With the introduction of the kernel function, the LSSVM model can be redefined as

$$y(x)=\sum_{i=1}^{n}\bar{\alpha}_iK(x,x_i)+c. \tag{20}$$

D. Particle Swarm Optimization Algorithm

The performance of LSSVM depends heavily on the choice of parameters. In order to develop a more robust model, parameter optimization methods and cross validation are used. In this paper, the PSO algorithm is introduced. Inspired by the observation of birds [29], Kennedy and Eberhart proposed the PSO algorithm in 1995. This algorithm simulates the process that a group of birds, which do not have a flying target at the beginning, finally fly to a particular habitat. In this process, a bird and the whole bird group can be abstracted into a particle and a particle swarm, respectively. Each particle has its own speed and position, and each position of each particle corresponds to a fitness degree. After setting up the initial position and velocity of every particle, the values of positions and velocities will be updated according to the fitness values until the particle swarm discovers and converges to the optimistic value of the fitness function. Intuitively, when one bird finds and lands in a habitat, this behavior will drive the other birds around it to fly to the habitat, and ultimately enable the whole flock to land in this habitat [27]–[29].

Compared with other evolutionary algorithms, PSO algorithm simulates a kind of group behaviors with consciousness. For one particle, the smaller the objective function value is, the better the corresponding fitness is. The movement directions of the particles are always toward the optimal direction of the group. PSO algorithm not only chooses the better one between offspring and parent, but also chooses the best one among all the historical values [30]. In this sense, although PSO may not converge to the global optimistic set of parameters, it does gradually enhance the performance of the models by repeatedly resetting the random seed points in the initialization procedure, with a cost of offline design efficiency, rather than the real-time capacity online [31], [32].

We use $Y_i=(y_{i1},y_{i2},\ldots,y_{in})$ and $\upsilon_i=(\upsilon_{i1},\upsilon_{i2},\ldots,\upsilon_{in})$ to denote the current location and the speed of the particle i, respectively. $L_i=(l_{i1},l_{i2},\ldots,l_{in})$ represents the best location that the particle i has ever experienced, which corresponds to the optimal fitness function value. A smaller target function value corresponds to a better fitness function value

$$L_i(t+1)=\begin{cases}L_i(t), & \text{if} \quad f(Y_i(\mathrm{t}+1))\geq f(L_i(t))\\ Y_i(t+1), & \text{if} \quad f(Y_i(t+1))<f(L_i(t))\end{cases} \tag{21}$$

where $f(Y)$ is the target function to be minimized. The current best location of the particle i can be obtained by the (21).

With s representing the particle number, the optimal location of the whole group is denoted as

$$\begin{aligned}L_g(t)\in\ &\{L_0(t),L_1(t),\ldots,L_s(t)\}|\, f(L_g(t))\\ &=\min\{f(L_0(t)),f(L_1(t)),\ldots,f(L_s(t))\}.\end{aligned} \tag{22}$$

The evolution equation of particles can be expressed as

$$\begin{aligned}\upsilon_{ij}(t+1)=\ &\upsilon_{ij}(t)+c_1\gamma_{1j}(t)(l_{ij}(t)-y_{ij}(t))\\ &+c_2\gamma_{2j}(t)(l_{gj}(t)-y_{ij}(t))\end{aligned} \tag{23}$$

$$y_{ij}(t+1)=\ y_{ij}(t)+\upsilon_{ij}(t+1) \tag{24}$$

where j represents the jth dimension of the particle, t stands for the tth generation of particle i, and c is a constant. γ_1 and γ_2 are two random numbers within the range [0, 1], and c_1 and c_2 are two acceleration constants.

III. Prediction Approach Based on Fuzzy Information Granulation With Optimized SVM

In this section, the technical core of a novel prediction approach is illustrated and summarized.

Generally, the prediction is based on the SVM model or the LSSVM model. The major difference from the existing methods lies in that the data used to train the models are not the raw time-series data with explicit physical meanings. Instead, an alternative representation in forms of information granules is fed into SVM or LSSVM. This indicates that

1) There are more than one time series available in both training and testing phases. They can be processed either separately or simultaneously, in which case the correlation relations can be investigated and utilized.
2) Different types of information granules can be calculated and adjusted by altering the fuzzy forms, e.g., the triangle form, the trapezoid form, and the asymmetric Gauss form, as it is shown in (3)–(5).
3) By thinking of the break points of the membership function as the upper and lower bound, a ribbon-shaped region instead of an exact value can be predicted. This benefits from the fuzzy concept and can be used for trend prediction and boundary violation alarm.

After the fuzzy information granules are generated, SVM models are to be trained. Since the parameters of SVM play an important role in model construction and have massive impact on regression accuracy and generalization capacity, a parameter optimization scheme called the grid search is introduced. The grid search includes three steps. First, multidimension grids of equal density are created, whose vertexes are collected as multivariate parameter sets to be chosen from. Second, cross validation is carried out under each parameter set. With K–fold cross validation or leave one out cross validation, samples' predicted error sum of squares (PRESS) is calculated for evaluation. Third, the optimal parameter set is selected as the one with the minimal PRESS value.

However, the optimization result of the grid search method is severely influenced by the grid size. Its performance can be improved accompanied by huge costs of efficiency. Specifically, the amounts of vertexes are in exponential growth relation with the parameter set dimension, and the computational load shoots up accordingly.

To improve the aforementioned defect, an intelligent algorithm called PSO is employed and investigated. As it is introduced in Section II, PSO is inspired by the group behavior of birds. Compared with the grid search method traversing through the hypergraph, PSO eliminates suboptimal solutions automatically in early stages, and can guarantee converging to a local optimal point. In this sense, equipped with more sparsity, PSO outweighs grid search in terms of both efficiency and accuracy.

On the other hand, standard SVM application is faced with the difficulty of solving the QP problem, where the matrix size is highly affected by the training sample size, and is time consuming in the big data cases. Suykens *et al*. proposed LSSVM by replacing the target function in the optimization problem with the two-norm form and transformed the inequality constraints in SVM into equality constraints. Subsequently, the optimization problem involved in LSSVM is a set of linear equations obtained by the Kuh–Tucker condition. In this paper, the grid search and PSO are also integrated with LSSVM for better parameter selection.

The complete procedure of the combined scheme—the "data-driven fuzzy information granulation based prediction approach" is simplified and summarized for readers' convenience. Refer to Tables I and II for the offline training phase and the online prediction phase, respectively.

TABLE I
ALGORITHM 1: DATA-DRIVEN FUZZY INFORMATION GRANULATION-BASED PREDICTION APPROACH: OFFLINE TRAINING PHASE

Step 1. Fuzzy information granulation
1) Choose a fuzzy form, e.g., the triangle form, the Trapezoid form or the asymmetric Gauss form, referring to (3), (4) or (5), respectively.
2) Run information granulation algorithm to calculate the granules' time series.
Step 2. Model construction
Train the SVM (or LSSVM model) with the granules' time series.
Step 3. Parameter optimization
Use the grid search or the PSO algorithm along with the predicted error sum of squares (PRESS) evaluation index to select the optimal set of parameters.
Step 4. Model verification (Optional)
1) Reconstruct the granules with the obtained SVM model (or LSSVM model).
2) Compare the reconstructed granules with the original ones.
3) If the reconstruction performance is acceptable, the offline training procedure is completed.
4) If the reconstruction performance is not acceptable, go to Step 1 and choose another fuzzy form.
Step 5. Refer to Table II for online prediction.

TABLE II
ALGORITHM 2: DATA-DRIVEN FUZZY INFORMATION GRANULATION-BASED PREDICTION APPROACH: ONLINE PREDICTION PHASE

Step 1. Collect a new observation from the time-series data.
Step 2. Fuzzy information granulation
1) Perform identical fuzzy granulation procedures with the same fuzzy form consistent with the Step 1 in Algorithm 1.
2) Calculate the granules' values for the new observation.
Step 3. Future value prediction
Carry out future value predictions based on the granules obtained in Step 2 and the SVM (or LSSVM model) established in the offline training phase.
Step 4. Go to Step 1 unless all predictions are performed.

IV. EXPERIMENTS

A. Fuzzy Information Granulation-Based SVM

Time-series data are extremely commonplace in reality. They are recorded and sorted in a set according to their temporal order. There usually exists certain significant information or patterns in the sequence, which can be mined and utilized by mathematical tools of time-series analysis. In this section, information granules are extracted from the raw time series, and then used for model construction. In addition, to test the validity and accuracy of SVM regression, the granules are in turn predicted by the model, which provides insights into the granule-based SVM model.

The experiments are conducted on the China highway transport freight traffic volume dataset (region monthly) from January 2001 to November 2014. The dataset was collected and sorted by the Economy Prediction System with data sources including the National Development and Reform Commission, Ministry of Transport, Civil Aviation Administration of China, State Post Bureau, Ministry of Industry and Information Technology, and National Bureau of Statistics. The first dimension (the rows of the matrix) denotes the national freight volume and the regional ones of 31 provinces. The second dimension (the columns of the matrix) denotes months throughout the timeline. Since the data are in an accumulative form from the beginning of each year, they are preprocessed to obtain the differentiated values for each month before the simulations. Moreover, it should be pointed out that there are some missing points in the dataset and the corresponding terms will not be selected for investigation.

It is not likely to make perfectly correct predictions on the basis of time series. Instead, an alternative—forecasting the tendency and range of variations, seems very promising. This part focuses on how to make regressions and predictions according to the recorded historical data.

To start with, the national monthly freight traffic data are taken into consideration. The time-series horizons in Figs. 1–4 are monthly recorded freight traffic volumes from January 2012 to April 2014, excluding the data points of each December. In the first step, fuzzy information granulation is carried out on

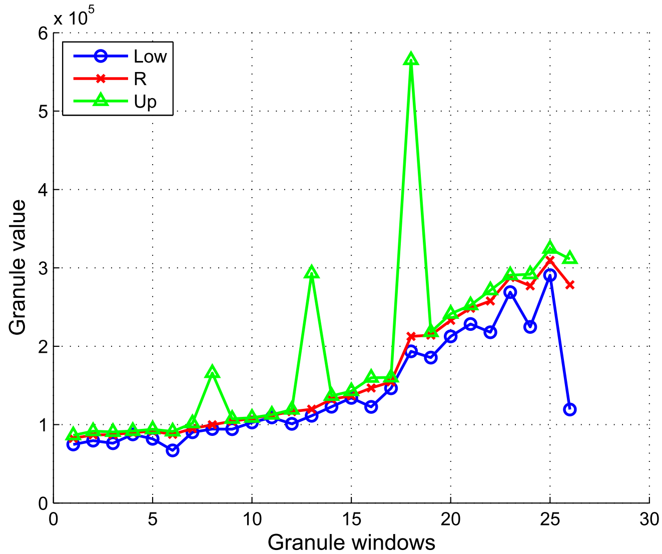

Fig. 1. Granules of national freight traffic.

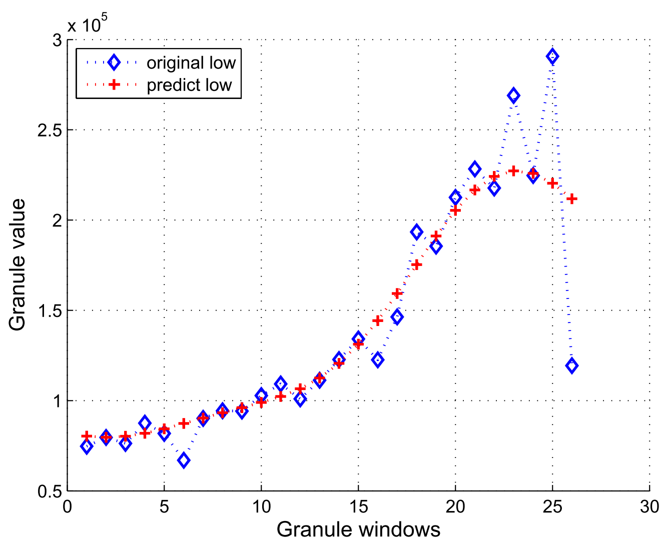

Fig. 2. low regression.

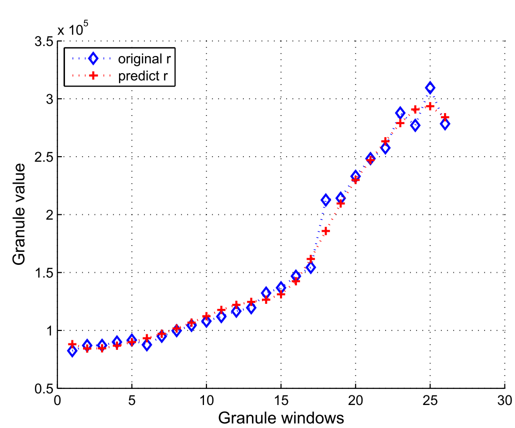

Fig. 3. R regression.

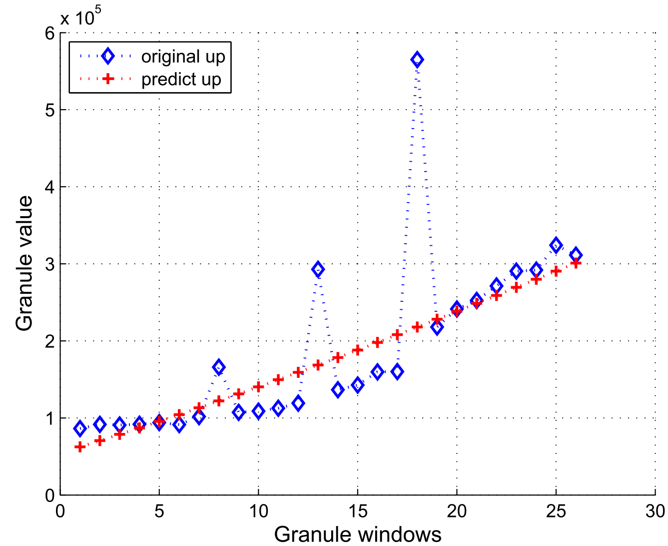

Fig. 4. up regression.

TABLE III
RESULTS OF SVM BASED ON GRANULES

Variables	P	λ	RMSE ($\times 10^4$ tons)	SCC	Prediction ($\times 10^4$ tons)
low	1.41421	0.015625	25613	0.842308	201070
R	4	0.0625	8041	0.988869	263660
up	32	0.0000863	76843	0.535468	311600

the freight volume data, to obtain three granules' time series, which are subsequently used to train the SVM model. This model can be used to forecast the tendency and the range of the national freight traffic variation since May 2014. Fuzzy information granulation is realized via corresponding toolbox. Triangle fuzzy granules are selected and each granule window contains five months. The outputs of fuzzy information granulation are three vectors, denoted as low, R, and up. These three vectors are related to the three parameters—a, m, and b, in triangle fuzzy granules. Here, low represents the lower limit of the original data variation, R represents the general level, and up represents the upper limit. The fuzzy information granule results of national freight traffic are shown in Fig. 1.

Afterwards, SVM is utilized for regression and prediction of the granules. The three outputs, namely, low, R, and up, are treated as training dataset for model development. The data are normalized to the real numbers ranging from 1 to 5. After that, a simple grid search method is used to optimize the parameters of SVM. It is worth to mention that PSO is also implemented in place of grid search for parameter optimization purpose in the subsequent study. Table III presents the results of simulations, including optimized parameters P and λ, root-mean-squared error (RMSE), squared correlation coefficient (SCC), and prediction values. Figs. 2–4 display the regression results of these three variables. The final prediction results are reported in Fig. 5. The algorithm running time for national monthly freight traffic data regression is 4.11 s (simulations are run on a

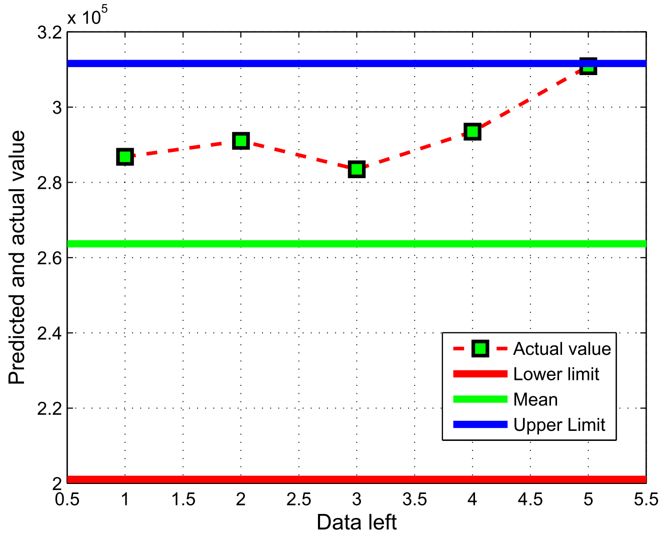

Fig. 5. Prediction for data left.

computer with Intel Core i5-2430M, 2.40 GHz CPU and 4 GB memory).

The curves of the original data in Figs. 2 and 4 are far from being smooth, resulting in great difficulty of accurate fitting with SVM. Comparatively, the curve in Fig. 3 is more smooth and SVM can accomplish the fitting easily. This observation can be justified by the statistical results as well. In fact, the SCCs of low and up are less than 0.85, while that of R is close to 0.99. The reason behind this phenomenon is that the freight volume value of every December is missing, and that there are step changes in different years' traffic data. These characteristics are reflected by low and up, which is the fundamental cause of difficulty in SVM fitting. Fortunately, some important information is reserved in R, which makes it possible to realize further prediction.

As is shown in Fig. 5, the actual traffic is within the predicted range between R and up. This figure illustrates that when combined with fuzzy information granulation, SVM can yield extraordinary effects. In addition, the last set of parameters of the training dataset is $[low, R, up] = [211770, 284120, 301000]$ ($\times 10^4$ tons), and the predicted counterpart is $[low, R, up] = [201070, 263660, 311600]$ ($\times 10^4$ tons).

Apart from simulations with the national data, the cases of several provinces are also taken into consideration. The results are similar to the national one, as is shown in Table IV.

B. Freight Traffic Volume Forecasting

In this section, the freight traffic volumes of 14 provinces are used to forecast the whole country's total freight traffic volume. SVM, LSSVM, and PSO-LSSVM models are trained with this dataset and then used for prediction. The training data are the monthly freight traffic volumes of the nation and 14 provinces from January 2001 to July 2011, whose granules are used to build the models. The horizon of prediction is the months from August 2011 to May 2013, excluding the December of each year. The forecasting results with SVM, LSSVM, and PSO-LSSVM are shown in Figs. 6–8, respectively. The prediction accuracy of these methods are measured by the RMSE, and the computational efficiency is evaluated by the running time. In these three figures, the blue curves represent the real data and the red ones represent the predicted data.

As shown in Fig. 6, while the SVM prediction results reveal the actual trend, there are big deviations on some points. The RMSE of the entire dataset is 5502.64 ($\times 10^4$ tons) and the program's running time is 0.3410 s. Fig. 7 indicates that LSSVM also predicts the variation trend accurately. However, the RMSE of LSSVM is 10193.13 ($\times 10^4$ tons), greater than that of traditional SVM, and the running time is 2.047 s. Because the parameter optimization process is added into the LSSVM program, it is natural that the running time of LSSVM is longer than that of SVM (the default parameter optimization method for LSSVM is grid search). If the parameter optimization process is removed from the program, the LSSVM running time is reduced to 0.29 s, smaller than that of SVM. This means that LSSVM is more computationally efficient than SVM. Fig. 8 reports the predication results of PSO-LSSVM, which is a combination of PSO and LSSVM.

PSO algorithm serves as an enhancement tool for LSSVM to select optimal SVM-model-related parameters to increase prediction accuracy. With the addition of PSO, the running time increases greatly and reaches 172.455 s. At the same time, the prediction accuracy is improved compared with LSSVM. There are minor differences between Figs. 7 and 8. The prediction RMSEs are 10193.13 ($\times 10^4$ tons) and 9801.22 ($\times 10^4$ tons), respectively. The reason why the curves are so similar is that both LSSVM and PSO-LSSVM go through the parameter optimization procedures, i.e., the grid search and the PSO, respectively. Although the grid search method is not as efficient and "intelligent" as the PSO, it can find the optimal LSSVM parameters and achieve similar prediction accuracy to PSO.

It should be pointed out that PSO is only utilized in the offline training phase, rather than the online phase. In other words, introducing PSO has only positive effects for the online prediction results. Although parameters selected by PSO may differ from time to time, suboptimal ones can be obtained by running PSO with multiple seed points at a cost of a slightly longer training time. The termination criterion in this simulation includes limiting the iteration cycles by 200 and checking if the squared error of the objective function values in two consecutive iterations is smaller than a given threshold.

While both SVM and LSSVM can be used to predict the trend of the freight traffic volume, Figs. 6 and 7 show that the SVM prediction results are more accurate. The original LSSVM without parameter optimization has higher computation efficiency than SVM, and this improvement in efficiency results from the fact that LSSVM gets rid of the QP problem in SVM. As shown in Fig. 8, the LSSVM prediction accuracy can be improved by means of parameter optimization with PSO. However, the prediction accuracy increase is accompanied by a significant increase of computation load. To summarize, a reasonable compromise should be achieved between efficiency and accuracy. On one hand, from the perspective of efficiency, LSSVM outweighs SVM. With acceptable accuracy, SVM involves more parameters than LSSVM and is less efficient than LSSVM. On the other hand, when accuracy is in focus, SVM has smaller

TABLE IV
PREDICTION FOR SEVERAL PROVINCES

Provinces	Actual Freight Traffic ($\times 10^4$ tons)					Prediction Range ($\times 10^4$ tons)	Operation Time
Hebei	15700	15172	17540	17355	20540	[12461, 14197, 21566]	3.602917 s
Jilin	3773	3853	40.13	4306	4466	[2356.4, 4116.3, 4654.6]	3.568265 s
Liaoning	17249	18329	17101	17805	19360	[11180, 15791, 19156]	3.487026 s
Inner Mongolia	12767	12193	11019	11084	11585	[7655.2, 10436, 13657]	3.382921 s

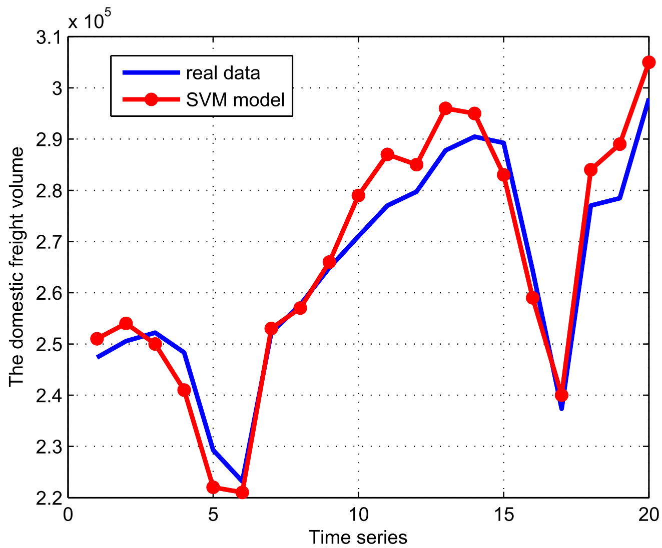

Fig. 6. Forecasting results by SVM.

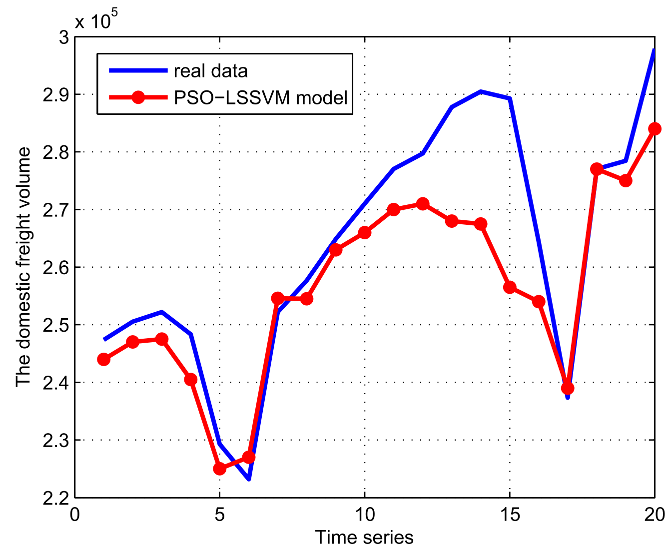

Fig. 8. Forecasting results by PSO-LSSVM.

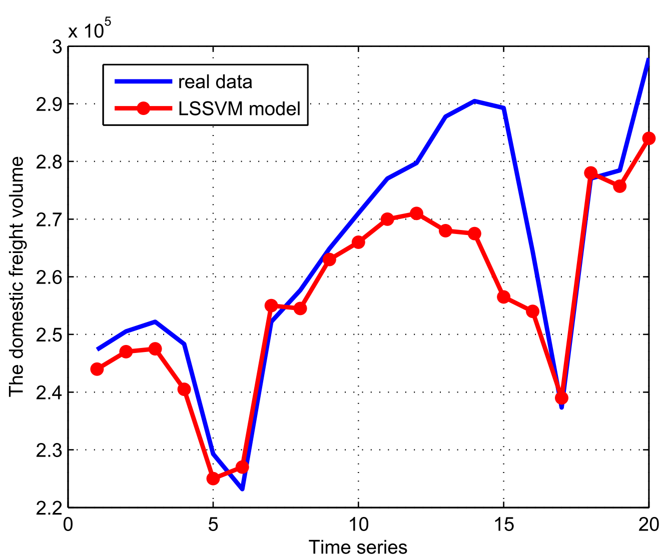

Fig. 7. Forecasting results by LSSVM.

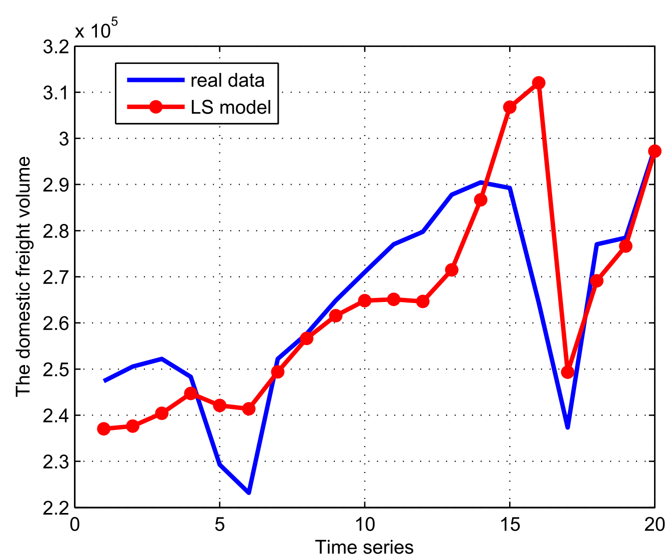

Fig. 9. Forecasting results by least squares.

RMSE values and performs better than both LSSVM and PSO-LSSVM.

In order to evaluate the performance of the proposed methods, the traditional multivariate statistics-based approaches including LS regression and PLS are tested on the identical dataset for the purpose of future freight volume values and trend prediction. Considering that the number of training samples is far from adequate for LS regression, to avoid overfitting, the order of LS model is selected as 2, and in PLS the latent space is span with one latent variable. The LS approach is a linear model in nature, which indicates that any nonlinearity in the dataset may trigger huge performance loss of generalization. As shown in Figs. 9

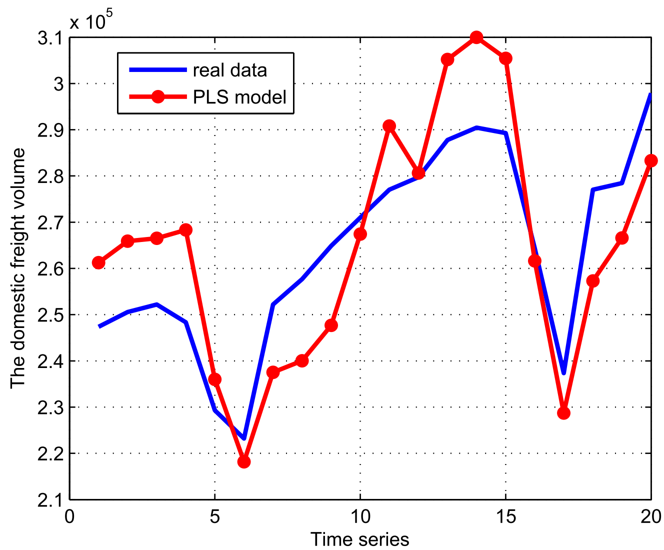

Fig. 10. Forecasting results by partial least squares.

TABLE V
RMSEs FOR FORECAST ACCURACY COMPARISON

Algorithms	LS	PLS	SVM	LSSVM	PSO-LSSVM
RMSE ($\times 10^4$ tons)	14843.33	13961.89	5502.64	10193.13	9801.22
MAPE ($\times 10^4$)	418.24	478.67	194.15	296.63	299.05
ESS ($\times 10^8$ tons^2)	0.983	1.268	1.162	0.61	0.627

and 10, both prediction results of LS and PLS reflect the true circumstances roughly. The LS method displays a variation with less fluctuation, especially during the time interval 1 to 13. Also, it shows some trend tracking capability in the rest seven sample points. However, the defect is also obvious. The RMSE of the LS approach is 14843.33 ($\times 10^4$), which is possibly worse than acceptable. As for the PLS-based scheme, its fluctuation is more severe than that of LS. Generally, PLS shows its effectiveness in estimating the trend of freight volume variation. To be more specific, PLS reaches local minima at the 6th and 17th sample, and shoots up to a local maximum at the 14th sample, coinciding with the real circumstances. However, the RMSE of PLS is also larger than that of the proposed methods. All the RMSE, mean absolute percentage error (MAPE) values, and Explained Sum of Squares (ESS) values are summarized for algorithms' accuracy comparison (refer to Table V).

V. CONCLUSION

In this paper, an information granulation method was presented to represent the freight volume in a fuzzy manner. By combining fuzzy information granulation and SVM, we were able to predict the range and tendency of freight traffic volume. In the simulations we found that in a short term, the real freight traffic volume lies exactly in the range predicted by our method. This means that SVM can yield extraordinary prediction results in combination with fuzzy information granulation. Furthermore, we studied the performance of SVM, LSSVM, and PSO-LSSVM in predicting the freight traffic volume, and carry out comparison studies with statistical methods including LSs and PLS. As a result, it is found that all these methods can be used to forecast the trend of freight traffic volume, although their prediction accuracy and computation efficiency are different. SVM generates the best prediction accuracy with a relatively low computation load. By transforming the QP problem in SVM to a set of linear equations, LSSVM reduces the computation load in comparison with SVM. However, this improvement in efficiency is accompanied by the decrease of prediction accuracy. By applying PSO algorithm to optimize related parameters, PSO-LSSVM improves the prediction accuracy compared with LSSVM. Overall, the newly proposed fuzzy granulation integrated scheme shows its priority over the basic methods when applied to the raw time-series data.

REFERENCES

[1] F. Pereira, C. Antoniou, J. Fargas, and M. Ben-Akiva, "A metamodel for estimating error bounds in real-time traffic prediction systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 3, pp. 1310–1322, Feb. 2014.
[2] K. Halawa, M. Bazan, P. Ciskowski, T. Janiczek, P. Kozaczewski, and A. Rusiecki, "Road traffic predictions across major city intersections using multilayer perceptrons and data from multiple intersections located in various places," *IET Intell. Transport Syst.*, vol. 10, no. 7, pp. 469–475, May 2016.
[3] S. Yin, L. Liu, and J. Hou, "A multivariate statistical combination forecasting method for product quality evaluation," *Inf. Sci.*, vol. 355–356, pp. 229–236, Aug. 2016.
[4] Y. Yang and C. Yu, "Prediction models based on multivariate statistical methods and their application for predicting railway freight volume," *Neurocomputing*, vol. 158, pp. 210–215, Jun. 2015.
[5] D. Okan, "A multivariate model of fuzzy integrated logical forecasting method and multiplicative time series clustering: A model of time-varying volatility for dry cargo freight market," *Expert Syst. Appl.*, vol. 39, pp. 4135–4142, Mar. 2012.
[6] G. Wang and S. Yin, "Quality-related fault detection approach based on orthogonal signal correction and modified PLS," *IEEE Trans. Ind. Informat.*, vol. 11, no. 2, pp. 398–405, Jan. 2015.
[7] S. Yin, X. Zhu, J. Qiu, and H. Gao, "State estimation in nonlinear system using sequential evolutionary filter," *IEEE Trans. Ind. Electron.*, vol. 63, no. 6, pp. 3786–3794, Jan. 2016.
[8] S. Yin, G. Wang, and H. Gao, "Data-driven process monitoring based on modified orthogonal projections to latent structures," *IEEE Trans. Control Syst. Technol.*, vol. 24, no. 4, pp. 1480–1487, Jul. 2016.
[9] Z. Hou and S. Jin, "Data-driven model-free adaptive control for a class of MIMO nonlinear discrete-time systems," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2173–2188, Dec. 2011.
[10] S. Maiti, V. Verma, C. Chakraborty, and Y. Hori, "An adaptive speed sensorless induction motor drive with artificial neural network for stability enhancement," *IEEE Trans. Ind. Informat.*, vol. 8, no. 4, pp. 757–766, Nov. 2012.
[11] A. Chen, M. Leung, and H. Daouk, "Application of neural networks to an emerging financial market: Forecasting and trading the Taiwan Stock Index," *Comput. Oper. Res.*, vol. 30, no. 6, pp. 901–923, May 2003.
[12] R. Balabin, E. Lomakina, and R. Safieva, "Neural network (ANN) approach to biodiesel analysis: Analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (NIR) spectroscopy," *Fuel*, vol. 90, no. 5, pp. 2007–2015, May 2011.
[13] F. Wu, S. Yin, and H. Karimi, "Fault detection and diagnosis in process data using support vector machines," *J. Appl. Math.*, vol. 2014, Mar. 2014, Art. ID 732104.
[14] S. Gupta, R. Kambli, S. Wagh, and F. Kazi, "Support-vector-machine-based proactive cascade prediction in smart grid using probabilistic framework," *IEEE Trans. Ind. Electron.*, vol. 62, no. 4, pp. 2478–2486, Apr. 2015.

[15] H. Yoon, S. Jun, Y. Hyun, G. Bae, and K. Lee, "A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer," *J. Hydrol.*, vol. 396, no. 1, pp. 128–138, Nov. 2011.
[16] P. Wong, Q. Xu, C. Vong, and H. Wong, "Rate-dependent hysteresis modeling and control of a piezostage using online support vector machine and relevance vector machine," *IEEE Trans. Ind. Electron.*, vol. 59, no. 4, pp. 1988–2001, Aug. 2012.
[17] F. Chauchard, R. Cogdill, S. Roussel, J. Roger, and V. Bellon-Maurel, "Application of LS-SVM to non-linear phenomena in NIR spectroscopy: Development of a robust and portable sensor for acidity prediction in grapes," *Chemometr. Intell. Lab. Syst.*, vol. 71, no. 2, pp. 141–150, May 2004.
[18] F. Silva, R. Cruz, and A. D'Assuncao, "Blending PSO and ANN for optimal design of FSS filters with Koch island patch elements," *IEEE Trans. Magn.*, vol. 46, no. 8, pp. 3010–3013, Jul. 2010.
[19] L. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets Syst.*, vol. 90, no. 2, pp. 111–127, Sep. 1997.
[20] Y. Li and F. Yu, "Optimized fuzzy information granulation based machine learning classification," in *Proc. 7th IEEE Int. Conf. Fuzzy Syst. Knowl. Discovery*, Sep. 2010, vol. 1, pp. 259–263.
[21] W. Pedrycz and G. Vukovich, "Feature analysis through information granulation and fuzzy sets," *Pattern Recognit.*, vol. 35, no. 4, pp. 825–834, Apr. 2002.
[22] C. Cortes and V. Vapnik, "Support vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[23] O. Chapelle, "Training a support vector machine in the primal," *Neural Comput.*, vol. 19, no. 5, pp. 1155–1178, May 2007.
[24] X. Liang, H. Zhang, J. Xiao, and Y. Chen, "Improving option price forecasts with neuralnetworks and support vector regressions," *Neurocomputing*, vol. 72, nos. 13–15, pp. 3055–3065, Aug. 2009.
[25] J. Suykens, T. Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*. Singapore: World Scientific, 2002.
[26] L. Ding and S. Liao, "Approximate model selection for large scale LSSVM," in *Proc. Asian Conf. Mach. Learn.*, 2011, pp. 165–180.
[27] C. Huang and F. Dun, "A distributed PSO-SVM hybrid system with feature selection and parameter optimization," *Appl. Soft Comput.*, vol. 8, no. 4, pp. 1381–1391, Sep. 2008.
[28] J. Napoles *et al.*, "Selective harmonic mitigation technique for cascaded h-bridge converters with nonequal dc link voltages," *IEEE Trans. Ind. Electron.*, vol. 60, no. 5, pp. 1963–1971, May 2013.
[29] E. Romero-Cadaval, G. Spagnuolo, L. G. Franquelo, C. A. Ramos-Paja, T. Suntio, and W. M. Xiao, "Grid-connected photovoltaic generation plants components and operation," *IEEE Ind. Electron. Mag.*, vol. 7, no. 3, pp. 6–20, Sep. 2013.
[30] Y. Da and G. Xiurun, "An improved PSO-based ANN with simulated annealing technique," *Neurocomputing*, vol. 63, pp. 527–533, Jan. 2005.
[31] M. Clerc and J. Kennedy, "The particle swarm—Explosion, stability, and convergence in a multidimensional complex space," *IEEE Trans. Evol. Comput.*, vol. 6, no. 1, pp. 58–73, Jan. 2000.
[32] I. C. Trelea, "The particle swarm optimization algorithm: Convergence analysis and parameter selection," *Inf. Process. Lett.*, vol. 85, no. 6, pp. 317–325, Mar. 2003.

Shen Yin (M'12–SM'15) received the B.E. degree in automation from Harbin Institute of Technology, Harbin, China, in 2004, and the M.Sc. degree in control and information systems and the Ph.D. degree in electrical engineering and information technology from the University of Duisburg-Essen, Duisburg, Germany.

He is currently a Professor with the Research Center of Intelligent Control and Systems, Harbin Institute of Technology. His research interests include model-based and data-driven fault diagnosis and prognosis in process control and their applications, fault-tolerant control, and big data focused on industrial electronics applications.

Yuchen Jiang (S'16) received the B.E. (first-class honors) degree in automation from Harbin Institute of Technology, Harbin, China, in 2016, where he is currently working toward the Ph.D. degree in the Center of Intelligent Control and Systems.

He participated in an exchange program at the University of New South Wales, Sydney, Australia, from July to December 2014. His research interests include data-driven fault diagnosis, process monitoring, and artificial intelligence.

Yang Tian was born in Shanxi, China, in 1993. He received the B.E. degree in automation from the State Key Laboratory of Robotics and Systems, Harbin Institute of Technology, Harbin, China, in 2015. He is currently working toward the Ph.D. degree at the Georgia Institute of Technology, Atlanta, GA, USA.

His research interests include applications of learning machines such as neural networks and support vector machines, focusing on improvements by means of data preprocessing algorithms.

Okyay Kaynak (M'80–SM'90–F'03) received the B.Sc. (first-class honors) and Ph.D. degrees in electronic and electrical engineering from the University of Birmingham, Birmingham, U.K., in 1969 and 1972, respectively.

From 1972 to 1979, he held various positions in industry. In 1979, he joined the Department of Electrical and Electronics Engineering, Bogazici University, Istanbul, Turkey, where he is currently a Full Professor, holding the UNESCO Chair on Mechatronics. He has held long-term (near to or more than a year) Visiting Professor/Scholar positions with various institutions in Japan, Germany, the U.S., and Singapore. His current research interests include intelligent control and mechatronics. He is the author of three and the editor of five books. In addition, he is the author or coauthor of close to 350 papers that have appeared in various journals and conference proceedings.

Prof. Kaynak is active in international organizations, has served on many committees of the IEEE, and was the President of the IEEE Industrial Electronics Society during 2002–2003. He is currently on the IEEE MGA Board.