Short-Term Load Forecasting Based on an Adaptive Hybrid Method

Shu Fan and Luonan Chen, *Senior Member, IEEE*

***Abstract*—This paper aims to develop a load forecasting method for short-term load forecasting, based on an adaptive two-stage hybrid network with self-organized map (SOM) and support vector machine (SVM). In the first stage, a SOM network is applied to cluster the input data set into several subsets in an unsupervised manner. Then, groups of 24 SVMs for the next day's load profile are used to fit the training data of each subset in the second stage in a supervised way. The proposed structure is robust with different data types and can deal well with the nonstationarity of load series. In particular, our method has the ability to adapt to different models automatically for the regular days and anomalous days at the same time. With the trained network, we can straightforwardly predict the next-day hourly electricity load. To confirm the effectiveness, the proposed model has been trained and tested on the data of the historical energy load from New York Independent System Operator.**

***Index Terms*—Adaptiveness, load forecast, nonstationarity, robustness, self-organizing map (SOM), support vector machine (SVM).**

I. Introduction

LOAD forecasting has always been a key instrument in power system operation. Many operating decisions are based on load forecasts, such as dispatch scheduling of generating capacity, reliability analysis, and maintenance plan for the generators. In particular, with the rise of deregulation and free competition of the electric power industry all around the world, load forecasting becomes more important than ever before. Load forecasts are vital for the energy transactions in competitive electricity markets. In addition, the accurate estimated load is key data for the electricity price forecast. Forecast errors have significant implications for profits, market shares, and, ultimately, shareholder value. However, the electric load is increasingly becoming difficult to forecast because of the variability and nonstationarity of load series that result from the dynamic bidding strategies of market players and price-dependent loads as well as time-varying prices. Therefore, more sophisticated forecasting tools with higher accuracy are necessary for the modern power system.

During the past years, a wide variety of techniques have been tried in the problem of load forecasting [1], most of which are based on time-series analysis. The time-series model mainly includes approaches based on statistical methods and artificial neural networks (ANNs). The statistical models are hard computing techniques based on an exact model of the system, which include moving average and exponential smoothing methods, linear regression models, stochastic process, data mining approach, autoregressive and moving averages (ARMA) models, Box–Jenkins methods, and Kalman filtering-based methods [2]–[8]. Basically, most of the statistical methods are based on linear analysis. However, the load series are usually nonlinear functions of the exogenous variables. Therefore, to incorporate the nonlinearity, the ANNs have received much more attention in solving the problem of load forecasting [9]–[12]. Neural networks have been shown to have the ability not only to learn the time-series load curves but also to model an unspecified nonlinear relationship between load and weather variables. Compared with the statistical models, the ANNs are based on a soft computing technique and do not require to explicitly model the underlying physical system. By simply learning historical samples, a mapping between the inputs and the load is reconstructed and then can be adopted for the prediction theoretically. ANN-based methods report a fairly good performances in forecasting. Recently, a new approach based on machine learning techniques and support vector machines (SVMs) has been used for load and electricity price forecasting or classification and achieved good performances [13], [14]. SVM or support vector regression (SVR) is a new and powerful machine learning technique for data classification and regression based on recent advances in statistical learning theory [15]. Established on the unique theory of the structure risk minimization principle to estimate a function by minimizing an upper bound of the generalization error, SVMs are shown to be very resistant to the over-fitting problem, eventually achieving a high generalization performance in solving forecasting problems of various time series [16], [17].

This paper focuses on short-term load forecasting (STLF), in particular on forecasting a day-ahead load profile, which means that for each day of the week, 24 load forecasts are computed. To develop such a forecasting system, there are three key problems that need to be deeply investigated and are stated as follows.

1) *Nonstationarity of load series.* As mentioned above, different from the traditional power system, an assumption of stationarity in competitive power markets may lead to an inaccurate prediction for load series. Specifically, the load series switch generally between different segments or regions mainly due to the change of consumption habits in the face of energy price amendments and multiple seasonality that generally give rise to piece-wise stationary time series. This fact also leads to gradual changes in the dependency between the input and the output variables. Therefore, when modeling the load series, it is important

Manuscript received June 20, 2005. Paper no. TPWRS-00360-2005.
The authors are with Osaka Sangyo University, 3-1-1 Nakagaito, Daito, Osaka 574-0013, Japan (e-mail: fanshu@hotmail.com; chen@elec.osaka-sandai.ac.jp).
Digital Object Identifier 10.1109/TPWRS.2005.860944

to capture the dynamic input–output relationship inherent in the data.

2) *Adaptiveness of the forecasting model.* Previous works have shown that the characteristics of load series between regular workdays and anomalous days, which include weekend, holidays, and days with anomalous events, are quite different [18], [19]. To achieve good forecasting results, the regular workdays and anomalous days should be treated with different schemes.
3) *Robustness of the forecasting tool.* According to our experiences, many exogenous variables besides the load itself, such as the day type, the hour, environment temperature, humidity, wind speed, etc., should be taken into account in the load forecasting. These variables are usually dealt with in a unity model, for instance, a neural network with multiple inputs and outputs. This usually makes the forecasting tools variable sensitive and system dependent. A model developed for one utility cannot be easily modified for another, and the installation at a new site is also a complex procedure requiring considerable time, which decreases the robustness of the forecasting model. A universal model with strong robustness will be very attractive to the power engineers.

Based on the above analysis, the purpose of this paper is to apply the new advances in machine learning technique to develop a load forecasting tool with the emphasis on tackling the above three key problems. A novel method for time-series analysis, which adopts a nonstationary model with two-stage adaptive hybrid architecture, is proposed in this paper. Specifically, a self-organizing map (SOM) network is used as a gating network to identify the switching or piece-wise stationary dynamics for the input training data set in the first stage. Since the SOM is an unsupervised network, it has the ability to partition the space of input training data set into many subsets without prior knowledge about the classifying criteria, where each subset may be considered as stationary time-series data. Then several groups of 24 SVMs are applied to, respectively, fit the hourly electricity load profile data in each partitioned subset by taking advantage of all past information and similar dynamic properties (e.g., piece-wise stationarity) in the second stage. In the first stage of the proposed model, the input data are also classified into two groups: regular days and anomalous days; then in the second stage, we use different feeders of SVM for the regular weekdays and anomalous days, which means the network can adapt to different models automatically and improve the forecasting accuracy for the anomalous days. To improve the robustness of the model, all the input variables are decomposed into two groups. Besides the load data itself, no other information except temperature data is used for the SVMs in the second stage. Moreover, other factors, such as day type, humidity, and wind speed are only used in the first stage so as to alleviate the sensitivity of the model to these variables and to increase the robustness of the model. In the mean time, the forecasting tool is able to accommodate new variables in the first stage when necessary. Finally, the next-day electricity load forecasting is conducted by the trained network with a satisfactory level of accuracy on the specific subset in a voting manner among the SOM and SVMs.

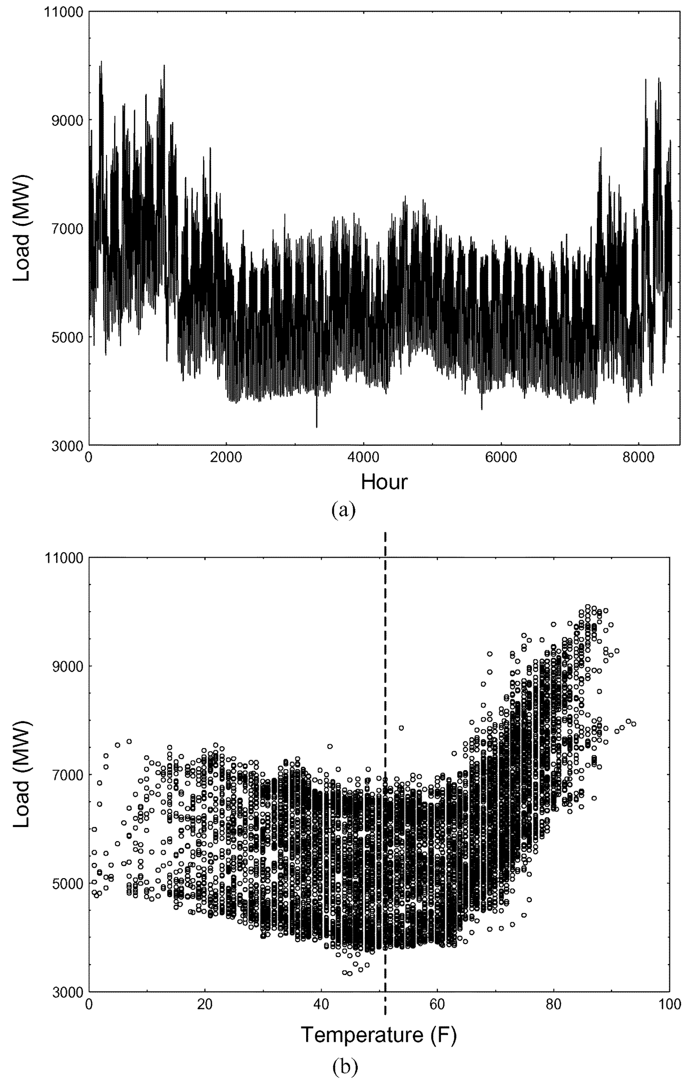

Fig. 1. Electricity demand in New York City from July 1, 2003 to June 30, 2004. (a) Hourly electricity demand. (b) Correlation between the electricity load and temperature at the same time point.

In the experiment, we adopt the load curve data of New York Independent System Operator (ISO) to verify the effectiveness of the learning and forecasting for the proposed methods. The criteria to compare the performance are the mean absolute percentage error (MAPE) and mean absolute error (MAE) in this paper, which indicate the accuracy of recall. In addition, for comparative study, we also examine the network only with SVMs. Detailed numerical results and their discussion for training errors are also given in the paper.

II. Task Description and Load Data Analysis

This paper uses the hourly electrical demand series of New York City as a test example of our method by comparing with the prediction of New York ISO [20]. In order to develop an appropriate model, we examine the main characteristics of the hourly load series in this section. Fig. 1 illustrates the electricity demand of New York City from July 1, 2003 to June 30, 2004.

According to Fig. 1(a), it is clear that the load dynamics have multiple seasonal patterns, corresponding to a daily and weekly periodicity, respectively, and are also influenced by calendar effect, i.e., weekends and holidays. Sometimes, the demand presents high volatility and nonconstant mean. It can be concluded that there exist different regimes in the load time series due to market and season effects, which generally give rise to piece-wise stationary dynamics. Based on such analysis, we use a hybrid model to classify the nonstationary price data set to several subsets with different characteristics, e.g., piece-wise stationary data subsets, on which a high accurate learning and prediction can be expected, compared with the conventional approaches.

We next analyze the correlation between load and temperature. It is well known that temperature information is very important for load forecasting. However, the correlation between the load and the temperature is not always constant or linear. For instance, the correlation coefficients between load and temperature of the same time point have usually various values for different seasons and different load forecasting problems. Such a fact explains why some studies said that loads are very sensitive to temperature, whereas others reveal that temperature does not strongly affect the load diagram [13]. Hence, it is necessary for us to analyze the inherent correlation between load and temperature for a specific system. Fig. 1(b) shows the correlation between load and temperature for New York City from July 1, 2003 to June 30, 2004.

According to Fig. 1(b), there exists approximately a piece-wise linear relationship of correlations between load and temperature with about 50° difference of their tangents. The correlation in each segment can be computed using the following expression:

$$\rho_{d,y} = \frac{\mathrm{Cov}(d,y)}{\sigma_d \sigma_y} \tag{1}$$

where $\mathrm{Cov}(d,y)$ is the covariance of loads d and y, and σ_d and σ_y are the standard deviations for d and y. $\rho_{d,y} = 1$ corresponds to a perfect linear correlation while an intermediate value describes partial correlations, and $\rho_{d,y} = 0$ represents no correlation at all.

The dotted line in Fig. 1(b) indicates the separate point between the two piece-wise segments, which is obtained by maximizing the two correlation coefficients on both segments. According to our computation, the separating point is approximately 55°, and $\rho_{d,y}$ on the two segments is -0.19 and 0.69, respectively. We use such information in the modeling of the load series in the next section.

III. Method and the Learning Algorithm

A. Architecture of a Hybrid Network

In this paper, a time-series-based nonlinear discrete-time dynamical model is represented by (2) for the load forecasting

$$y(t+1) = f(y(t), \ldots, y(t-m+1); T) \tag{2}$$

where $y(t)$ is a vector representing the daily electricity load profile at time t, and m is the orders of the dynamical system, which

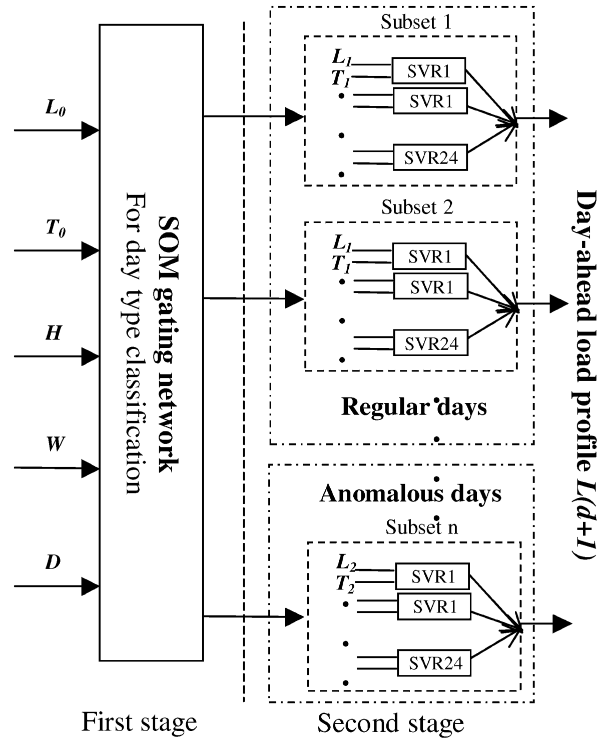

Fig. 2. Hybrid model for the electricity load forecasting.

is a predetermined constant. T is a vector representing the control parameters of the dynamical system, such as temperature, humidity, wind speed, and day types. Then, the task in STLF is to extrapolate past load behavior while taking into account the effect of other influencing factors.

An adaptive hybrid network is applied to reconstruct the dynamics of electric load consumption using the time series of its observables. The proposed network in this paper is shown in Fig. 2. This model is based on a two-stage architecture. In the first stage, a SOM network is applied to cluster the input training data set into several subsets with similar dynamical properties in an unsupervised manner. In this stage, we also classify the input data into two groups: regular days and anomalous days. Then, in the second stage, several groups of 24 SVMs (or exactly SVRs) are used to fit the training data in each subset in a supervised way. In other words, depending on their stationarity as well as other dynamic features, this method breaks the time series into different segments or subsets where data in the same segment can be modeled by the same SVR due to a similar property. Notice that there are 24 SVRs for each subset to train and predict 24-h loads of next day, respectively.

As shown in Fig. 2, the input variables are different for the SOM and SVM networks, which are given in Tables I–III.

As indicated in Table I, the input data of SOM consist of six elements: load vector L_0, temperature vector T_0, humidity scalar H, wind speed scalar W, and day type vector D, two binary code as the weekend and holiday indicator (1 for weekend and holiday, 0 for the other days). As mentioned in the previous section, we use a temperature sensitivity coefficient to represent the different correlation between load and temperature. If the temperature is larger than 55°, this coefficient is set as 1, otherwise 0. Considering the continuous changes of temperature

TABLE I
LIST OF INPUT DATA OF THE SOM NETWORK

Input	Variable	Detail description
1-2	Load vector L_0	Maximal daily load of one day ahead Average maximal daily load of the last week
3-4	Temperature vector T_0	Forecasted maximal temperature of the next day Average maximal temperature of previous two days
5	H	Forecasted maximal humidity of the next day
6	W	Forecasted average wind speed of the next day
7	T_c	Temperature sensitivity coefficient
8-9	Day type vector D	Weekend indicator Holiday indicator

TABLE II
LIST OF INPUT DATA OF THE SVM NETWORK FOR REGULAR DAYS

Input	Variable name	Lagged value (hours)
1-9	Hourly load (L_1)	24,25,26,48,72,96,120, 144,168
10-19	Hourly temperature (T_1)	0,1,2,24,48,72,96,120, 144,168

Assuming that the hour of load predication is at 0, the lag 0 represents the target instant, and the 24 lagged hours means the values that were measured 24 hours earlier than the hour of predication.

TABLE III
LIST OF INPUT DATA OF THE SVM NETWORK FOR ANOMALOUS DAYS

Input	Variable name		Lagged value (hours)
1-5	Hourly load		24,25,26,48,72
6-7	Hourly load of the previous Saturday	(L_2)	h,h-1*
8-9	Hourly load of the previous Sunday		h,h-1*
10-19	Hourly temperature (T_2)		0,1,2,24,48,72,96,120, 144,168

* h stand for the same clock with the target hour

and humidity in a short period, we apply the max value of these two elements to improve the sensitivity. On the contrary, since the changes of wind speed are usually stochastic, we use the average value. Hence, there are nine input variables in total for the SOM network, which are expected to identify the switching dynamics in load series.

For the SVM, in addition to the forecasted and actual temperature, the input variables are the hourly load values of the last day available and the similar hours in the previous days or weeks. In a real power system, there exist many anomalous conditions, such as weekend, holidays, sudden changes following switching operations or fault, sports events, etc. Such anomalous days represent a very hard task to load forecasting because these atypical load conditions are rare and quite different from regular workdays. Therefore, it is necessary for us to use different feeders for the regular days and anomalous days in the second stage. The proposed method is able to deal with regular days and anomalous days, respectively, and performs the load prediction under programmed grid switching operations. The input data of the SVM network for regular days is shown in Table II.

The inputs of training data for regular days consist of past hourly load demand and the forecasted temperature. In order to capture the time-series style in load, we include the electricity load of the previous seven days at the predicting hour. We use the temperature information of one and two hours earlier than the predication hour because temperature changes normally precede load changes. Because our experiment shows that only temperatures of the forecasting day included in the input set do not produce accurate forecasting results, temperature variables are added for every time point at which load was included.

Except weekends, how to handle holidays is also an important problem. Specifically, the input data for prediction of regular days may include data of anomalous days except for the weekend. For instance, when we forecast the load profile of January 2, the input data used for forecasting normally include data of January 1, which is a holiday. In this situation, the general load series are disturbed, which results in poor performances in prediction. Therefore, in this paper, the input of the SVM does not use the actual load data in the anomalous day but rather the forecasted load data for that day, as it was a regular day.

According to the historical load data, the same type of holiday showed a similar trend of load profile as in previous years. Previous works already show that different schemes should be developed for the regular days and anomalous days [18]. For instance, several studies conclude that the load diagram of a holiday has a strong connection with that of the two Saturdays before that day and the most recent diagram available, and holidays' forecasts should be assessed as a function of weekend behavior. Based on the above analysis, the input data of the SVM network for anomalous days is selected and shown in Table III.

As indicated in Table III, the difference of the input data for the SVM network between regular days and anomalous days is as follows. In addition to the most recent two days' load data, eight input variables, representing the load data around the predicting hour in the two weekends before the forecasting anomalous days, are used. All the input variables are selected based on the assumption that the load forecast is made at midnight for the next day with zero leading time, which means that all hourly loads from the previous day were known. On the other hand, if the forecast is performed before midnight, the inputs of SVMs will not include the diagram of the current day (lagged hour from 24 to 47) because they are not entirely available.

B. First Stage: SOM Gating Network

The dynamics of electricity loads are nonstationary or approximately piece-wise stationary during a short period due to the switching nature related to discrete changes in participants' strategies. In this paper, the SOM network is used as a gating network to identify the switching or piece-wise stationary dynamics for the input training data set in the first stage.

The SOM is an unsupervised neural model designed to build a representation of neighborhood relationships among vectors of an unlabeled data set [21], [22]. The structure of a two-dimension SOM network is illustrated in Fig. 3. The neurons in the SOM are put together in an output layer A in one-, two-, or even three-dimensional arrays. Each neuron i has a weight vector w with the same dimension of the input vector $X = (y(t),\ldots,y(t-m+1),D)$, which stores the information about the inputs and outputs of the mapping being studied. The network weights are trained according to a competitive–cooperative scheme in which the weight vectors of a winning neuron

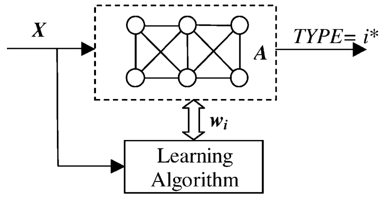

Fig. 3. Block diagram of a two-dimensional SOM network.

and its neighbors in the output array are updated after the presentation of an input vector.

The learning procedure consists of a series of training steps. During the training stage, the winning neuron $i*$ at step n is determined based on the input of the training sample X

$$i*(n) = \arg\min_{i\in A}\{\|X - w_i(n)\|\} \tag{3}$$

which means that neuron $i*$ has the minimal distance between the weight vector w_i and the input vector X among all neurons of the output layer A. The learning rules for updating the weights are stated as follows:

$$w_i(n+1) \leftarrow w_i(n) + \alpha(n)\cdot h_{ii*}(n)\cdot(X - w_i(n)) \tag{4}$$

where $\alpha(n)$ is the learning rate, and $h_{ii*}(n)$ is a Gaussian neighborhood function given by

$$h_{ii*}(n) = \exp\left(-\frac{\|r_i(n) - r_{i*}(n)\|}{2\sigma(n)^2}\right) \tag{5}$$

where $r_i(n)$ and $r_{i*}(n)$ are the locations of neurons i and $i*$ in the output array, respectively. Here, $\sigma(n)$ determines the size of the neighborhood region. The parameters of $\alpha(n)$ and $\sigma(n)$ are chosen fairly large at first for rapid adaptation of the neurons, and they decay quickly with time by

$$\alpha(n) = \alpha_0\left(\frac{\alpha_f}{\alpha_i}\right)^{\left(\frac{n}{N}\right)}, \quad \sigma(n) = \sigma_0\left(\frac{\sigma_f}{\sigma_i}\right)^{\left(\frac{n}{N}\right)} \tag{6}$$

where α_i and σ_i denote their initial values, and α_f and σ_f are the final ones. N is the number of total iterations. (6) is actually similar to the cooling schedule of simulated annealing process [22], [23].

After training, the SOM network can be used to identify the type of the input data subset as follows:

$$\text{TYPE} = i^* \tag{7}$$

with the number of the winning neuron $i*$ found in (3).

As the SOM is an unsupervised network, it is able to decompose the input training or testing data set into several subsets with specific characteristics in an unsupervised manner, such as season-dependent data or piece-wise stationary data, which subsequently are considered in next SVR training and forecasting.

Another benefit that can be obtained from the classification is the combination of the similar days. As stated above, the forecast results are not very satisfactory during anomalous periods characterized by exceptional circumstances. The reason is that these events represent week statistical samples for algorithms based on historical data learning. We do not have enough historical data for the learning of these anomalous days. For instance, besides a long time gap, there is only one example of Christmas day per year, and the corresponding load behavior varies depending on the day of the week and the nature of the load. However, if we can find some days similar to the anomalous day and combine some anomalous day into one group, the amount of the training samples for the anomalous day will be enlarged, and it is likely to have a better forecasting performance.

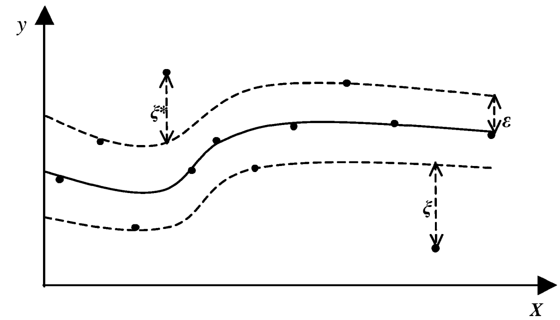

Fig. 4. ε-insensitive band for SVR.

C. Second Stage: The SVM Network

In this stage, several groups of SVRs are used for load time-series learning and prediction in each subset, and each SVR corresponds to one time instant. Supposing that we are given training data $(x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_n, y_n)$, where x_i are input patterns, and y_i are the associated output value of x_i, the support vector regression solves an optimization problem

$$\min_{\omega,b,\xi,\xi^*} \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$

subject to

$$\begin{aligned} y_i - (\omega^T\phi(x_i) + b) &\le \varepsilon + \xi_i^* \\ (\omega^T\phi(x_i) + b) - y_i &\le \varepsilon + \xi_i \\ \xi_i, \xi_i^* &\ge 0, i = 1, \ldots, n \end{aligned} \tag{8}$$

where x_i is mapped to a higher dimensional space by the function Φ, and ξ_i^* is slack variables of the upper training error (ξ_i is the lower) subject to the ε-insensitive tube $(\omega^T\phi(x_i)+b)-y_i \le \varepsilon$. The constant $C > 0$ determines the tradeoff between the flatness and losses. The parameters that control regression quality are the cost of error C, the width of the tube ε, and the mapping function Φ.

The constraints of (8) imply that we put most data x_i in the tube ε. If x_i is not in the tube, there is an error ξ_i or ξ_i* that we tend to minimize in the objective function. This can be seen in Fig. 4. SVR avoids under-fitting and over-fitting of the training data by minimizing the training error $C\sum_{i=1}^{n}(\xi_i+\xi_i^*)$ as well as the regularization term $\frac{1}{2}\omega^T\omega$. For traditional least-square regression, ε is always zero, and data are not mapped into higher dimensional spaces. Hence, SVR is a more general and flexible treatment on regression problems.

Since Φ might map x_i to a high or infinite dimensional space, instead of solving ω for (8) in a high dimension, we deal with its dual problem

$$\min_{\alpha,\alpha*}\frac{1}{2}(\alpha-\alpha*)^TQ(\alpha-\alpha*)+\varepsilon\sum_{i=1}^{n}(\alpha_i+\alpha_i*)+\sum_{i=1}^{n}(\alpha_i-\alpha_i*)$$

subject to

$$\sum_{i=1}^{n}(\alpha_i-\alpha_i*)=0$$
$$0\le\alpha_i,\quad \alpha_i*\le C,\quad i=1,\ldots,n \qquad (9)$$

where $Q_{\mathrm{ij}}=\phi(x_i)^T\phi(x_j)$. However, this inner product may be expensive to compute because $\phi(x)$ has too many elements. Hence, we apply a "kernel trick" to do the mapping implicitly. That is, to employ some special forms, inner products in a higher space yet can be calculated in the original space. Typical examples for the kernel functions are polynomial kernel $\phi(x_i)^T\phi(x_j)=(\gamma x_1^T x_2+c_0)^d$ and RBF kernel $\phi(x_i)^T\phi(x_j)=e^{-\gamma(x_1-x_2)^2}$. They are inner products in a very high dimensional space (or infinite dimensional space) but can be computed efficiently by the kernel trick, even without knowing $\phi(x)$.

As each data subset classified from the SOM is considered to be approximately stationary, 24 SVRs are applied to, respectively, fit the hourly electricity load profile data by taking advantage of all past information and similar dynamic properties (e.g., piece-wise stationarity). The next-day electricity load forecasting is conducted by the trained network with an acceptable level of accuracy in a voting manner in the SOM and SVRs. For numerical experiments in this paper, we use the software LIBSVM [24], which is a library for support vector machines, including the efficient implementation of solving (9).

IV. Numerical Experiments

A. Data Collection and Preprocess

The daily electricity load in New York City and weather data observed at Central Park have been considered for the paper. Two typical months have been selected to forecast and validate the performance of the proposed model. The first one corresponds to January 2004, which is a winter month with high demand. The second one corresponds to July 2004, which is a summer month also with high demand. The hourly data used to forecast the first month are from January 1, 2003 to December 31, 2003 (365 days). The hourly data used to forecast the second month are from July 1, 2003 to June 30, 2004 (366 days). Because New York City had experienced an electric power blackout on August 14, 2003, data of one week, including the outage day and six following days, are eliminated from the training samples. A few missing load and temperature data were filled in by interpolating between neighboring values.

Basically, large data sets are effective antidotes to fight against over-fitting. However, increasing the size of the data set by aggregating data way back in the past may not be feasible, because the load series show a very clear upward trend and load patterns have varied slightly year in and year out. Considering that the work in this paper is for academic research, we select training data of one year for simplicity. However, we think more training samples, for example, two years, should be used in real application.

The test sets are completely separate from the training sets and are not used during the learning procedure. Because it is clear that a larger forecasting lead-time does not necessarily imply a larger forecasting error, which depends on the data variability for the different periods.

As stated above, the training patterns for SOM and SVM are different. Training patterns of SOM only include input data, while that of SVM consists of the input–output pairs, and the output y of the SVM training pattern is the actual hourly load. Once trained, the proposed network is then used to predict the hourly load of the next day.

Input variables, load, and the other variables are all scaled, respectively, in our program. This procedure can set the input variables with a uniform range, which can avoid numerical problems and make every input variable be treated with the same importance.

Before the training, we still need to preprocess the training samples. As indicated in Fig. 1, there are some extremely high or low electricity demands in the training data. Some of them are caused by implicit reasons such as high temperature, urgent maintenance of generator, etc. However, some others cannot be explained with assessable reasons, which can be considered noise in the training samples. The noise in the data could lead to the over-fitting and under-fitting problem. To improve the accuracy of training, we remove the noise data in the training sample according to the correlation relationship between load and temperature.

We have already computed that the historical correlations of the two sets of training data are -0.19 and 0.69, respectively. After removing the unreasonable ones from the training data, we can obtain the higher demand-load correlations of -0.26 and 0.75, respectively, which can enhance the training accuracy due to the approximately linear relationship between demand and temperature.

B. Algorithm for Learning

A cross-validation procedure is used in the learning procedure for the proposed architecture, which is outlined as follows.

1) Classify the entire training data set into two groups: training set used for updating the network parameters and verification set for testing the performance.
2) Preset the minimum number $N_{\min}$ and max number $N_{\max}$ of the neural units in the SOM network, and set $N_{\min}$ as the initial number of neural units.
3) Train the SOM network using the training data set, and classify the neural unit into two groups: regular days and anomalous days.

4) In each subset of the input space, train the SVM to fit the data subset according to (9).
5) Forecast the load profile using the data set for verification, and calculate the MAPE.
6) Increase the number of the neural units in the SOM, and repeat step 1) to step 5) until the number reaches $N_{\max}$.
7) By comparing the MAPE for a different neural unit of the SOM network, the network parameters at the minimum of the MAPE are used as the final ones.

For different systems, the main difference may only lie in the number of the classifications. Therefore, we can easily apply the proposed model to different power systems.

C. Implementation

In the training process for SOM, there are several parameters, which may influence the performance of the proposed model. First, we need to select $N_{\min}$ and $N_{\max}$, which determine the minimum and maximum number the neural units of the SOM, and then decide the number of training data sets in each subset, namely, the number of load states to be partitioned. Basically, a bigger number of states give the model with higher classification capabilities but also with a higher risk of over-fitting and high computational effort. Moreover, a model with too many states is also more difficult to interpret. On the other hand, a smaller number of states may lead to under-fitting problem.

For the training of SOM, there are two parameters α and σ to be selected. In this paper, the initial and final values of α are set as $\alpha_i = 1$ and $\alpha_f = 0.01$. The initial value of σ is chosen as $\sigma_i = 2$, and the final value is $\sigma_f = 0.1$.

On the other hand, for SVM models, there are two key parameters, which are the cost of error C and γ in the RBF function. To decide the proper parameters, each data subset is divided into two subsets: training subset and validation subset. The training data subset is used for updating the parameters of the network, whereas the validation data subset is adopted to monitor the training performance in the training procedure. Based on this partition, we conduct cross-validation to choose suitable parameters, by the following test process.

1) Identify the type of the test day according to the information of the test day and previous days, using the gating SOM network.
2) Use the corresponding SVM network to output the forecasting load.

D. Numerical Results

The criteria to compare the performance are the MAE and MAPE in this paper, which indicate the accuracy of recall.

MAE is defined as follows:

$$\mathrm{MAE} = \sum_{i=1}^{n} (|y_{ai} - y_{fi}|)/n \tag{10}$$

where y_{ai} is the actual value, y_{fi} is the forecast value, and n is the total number of value predicted.

MAPE is given as follows:

$$\mathrm{MAPE} = \sum_{i=1}^{n} (|y_{ai} - y_{fi}| * 100/y_{ai})/n. \tag{11}$$

TABLE IV
RESULTS FOR THE ALL DAYS OF EACH MONTH

	Winter month		Summer month	
	MAE	MAPE	MAE	MAPE
ISO	163.05	2.86	245.47	3.55
Single SVM	143.21	2.38	207.74	3.03
Hybrid network	106.97	1.82	162.20	2.29

TABLE V
RESULTS FOR THE NORMAL DAYS (WORK DAYS)

	Winter month		Summer month	
	MAE	MAPE	MAE	MAPE
ISO	147.97	2.49	239.37	3.44
Single SVM	152.64	2.41	185.49	2.54
Hybrid network	91.83	1.49	155.17	2.11

TABLE VI
RESULTS FOR ANOMALOUS DAYS (INCLUDING WEEKENDS AND HOLIDAYS)

	Winter month		Summer month	
	MAE	MAPE	MAE	MAPE
ISO	190.47	3.52	236.59	3.80
Single SVM	126.05	2.32	254.47	4.07
Hybrid network	134.51	2.42	176.94	2.68

For comparative study, numerical simulations comparing with other methods are also conducted. First, we calculate the MAE and MAPE of the forecasting load published by New York ISO in the same period. At the same time, to verify the effectiveness of the hybrid structure, a model using SVM network without the gating stage of SOM is also built, and the performance of such a model is studied.

The hourly MAE and MAPE for the two test months with the three different methods are shown in Tables IV–VI. Among them, Table IV shows the results of the total days for the two months, whereas Tables V and VI show the results for regular days (work days) and anomalous days (including weekends and holidays), respectively. From the three tables, clearly the hybrid network outperforms all others in almost all the situations. The only exception is that a single SVM gives higher precision for the forecast of the anomalous days in January 2004. We think that the reason is the lack of training samples in the situation of hybrid network for anomalous days. However, the overall accuracy of the proposed method for normal days is much better, as shown in Tables IV–VI.

Figs. 5 and 6 show the forecasted results as well as the real load demand of January 2004 and July 2004. To avoid confusion, the data markers are set at every 12 h in the two figures.

To further study the adaptiveness of the proposed method, we also compare the forecasted results for the anomalous with SVM for a normal day and SVM specialized for anomalous day, respectively, as shown in Table VII. Fig. 7 illustrates the predicted values and the actual load of a long weekend in July 2004, which includes July 3, 4 (weekend), and 5 (holiday). It can be seen that the accuracy is improved for anomalous when a specialized model has been applied.

The MAE and MAPE of the whole test data for the 24 h are summarized in Table VIII. The overall MAPE of the hybrid net-

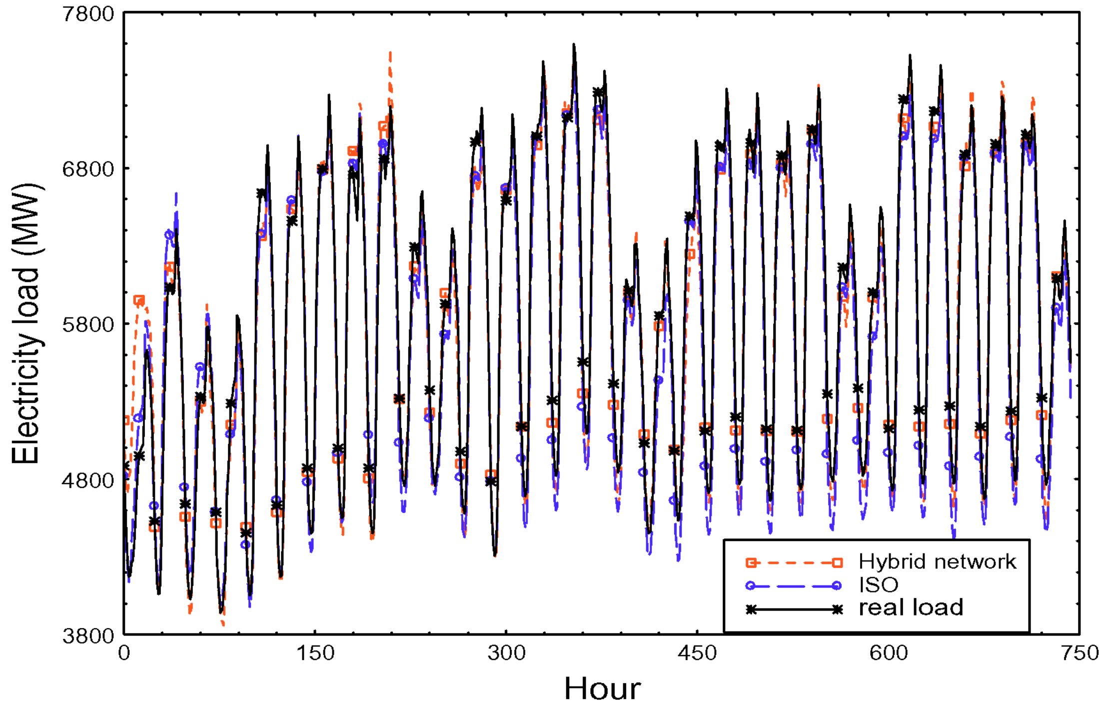

Fig. 5. Predicting and real electricity load of January 2004. (Color version available online at http://www.ieeexplore.ieee.org.)

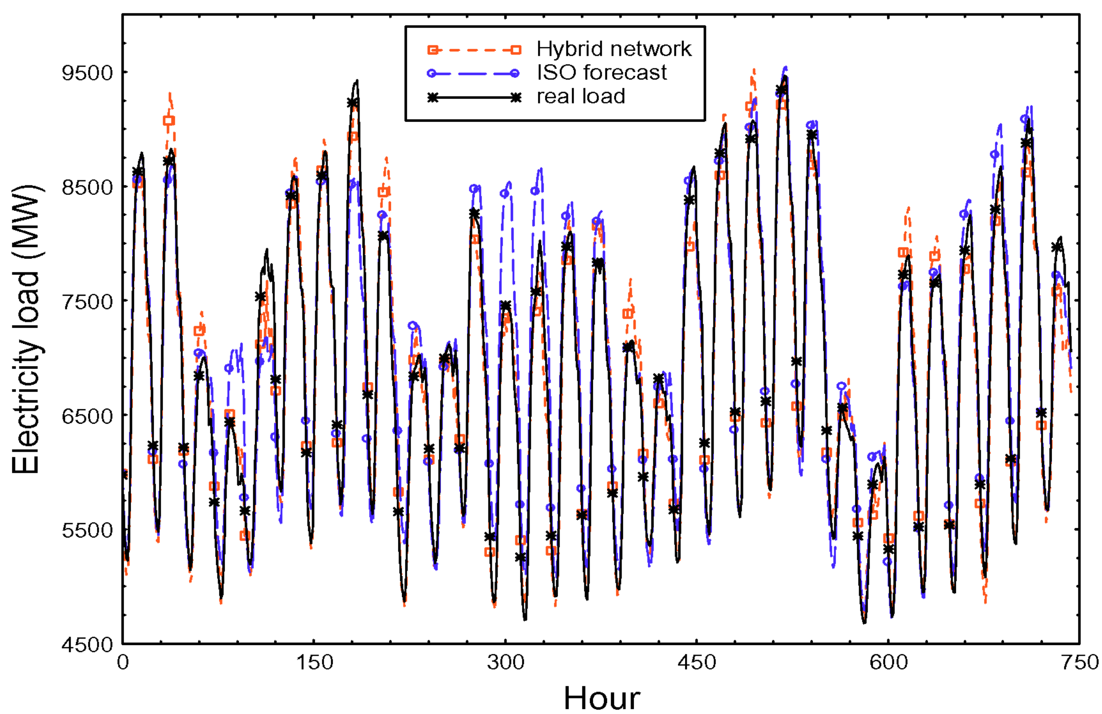

Fig. 6. Predicted and real electricity load of July 2004. (Color version available online at http://www.ieeexplore.ieee.org.)

work is 2.06, which is much better than the other methods and can be considered high in terms of accuracy.

In addition to superior and satisfactory performance, the results of the proposed model are good in a global sense, even with limited samples of only one year. When more training samples are incorporated in our method, a higher accuracy of prediction can be expected.

All the numerical studies have been run on a PC with 1 GB of RAM memory and 2.79-GHz clock frequency. The running time, including training and forecasting, is generally under 7 min for each case.

TABLE VII
RESULTS FOR ANOMALOUS DAYS WITH DIFFERENT MODELS

	Winter month		Summer month	
	MAE	MAPE	MAE	MAPE
SVM for a normal day	143.87	2.57	208.71	3.21
SVM specialized for anomalous day	134.51	2.42	176.94	2.68

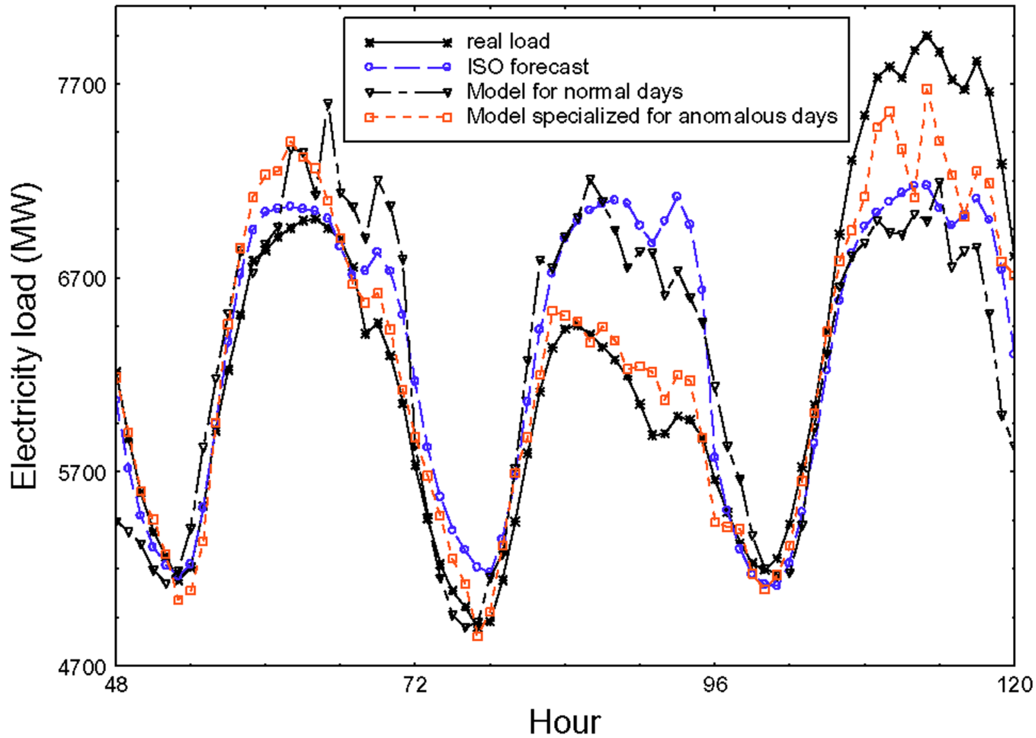

Fig. 7. Forecasted results of a long weekend in July 2004. (Color version available online at http://www.ieeexplore.ieee.org.)

TABLE VIII
COMPARISON FOR THE PROPOSED METHOD, SINGLE SVM, AND ISO

Hour	ISO		Single SVM		Hybrid network	
	MAE (MW)	MAPE (%)	MAE (MW)	MAPE (%)	MAE (MW)	MAPE (%)
1	210.69	3.80	94.83	1.67	94.57	1.67
2	197.03	3.76	94.72	1.77	91.83	1.72
3	188.63	3.74	89.80	1.74	77.49	1.53
4	180.73	3.66	84.18	1.66	77.73	1.57
5	176.48	3.58	95.53	1.90	77.71	1.58
6	168.23	3.32	117.75	2.30	67.45	1.34
7	174.34	3.22	159.22	2.95	103.23	1.91
8	183.13	3.13	234.20	4.02	127.06	2.20
9	187.02	3.01	240.07	3.87	125.77	2.06
10	184.87	2.81	236.33	3.60	128.93	1.97
11	191.92	2.80	229.04	3.37	141.13	2.07
12	194.23	2.77	224.46	3.21	145.92	2.07
13	202.58	2.86	228.62	3.23	171.50	2.37
14	212.69	2.97	242.58	3.40	174.50	2.39
15	215.58	3.01	248.38	3.48	192.31	2.65
16	223.08	3.11	247.06	3.43	183.66	2.50
17	227.68	3.14	228.78	3.13	202.01	2.70
18	223.73	3.05	223.75	3.03	179.81	2.42
19	222.69	3.11	188.71	2.63	168.88	2.34
20	202.26	2.91	169.94	2.45	163.06	2.33
21	205.26	3.03	154.56	2.26	153.32	2.23
22	219.24	3.29	137.11	2.02	139.46	2.05
23	215.71	3.36	129.41	1.99	129.19	1.96
24	210.47	3.50	112.34	1.83	113.49	1.81
Avg	200.76	3.21	175.47	2.71	134.58	2.06

V. CONCLUSION AND DISCUSSION

In this paper, an adaptive hybrid model based on the SOM and SVR network has been developed to forecast electricity load. The SOM gating network in the proposed method can identify the state of load and cluster the input data set into several subsets in an unsupervised manner. Then, in each subset, different SVMs have been used to fit the input data belonging to a different market state in a supervised way. The proposed method was applied to the prediction of the next day's load profile in New York City, which demonstrates the effectiveness and efficiency of the learning and prediction in contrast to others.

The proposed model has three notable advantages. First, it has the ability to tackle with the nonstationarity in the electricity load time series, which becomes increasingly significant due to the market effects. Second, it can treat regular days and anomalous days with different schemes under programmed grid switching operations. Last, it has strong robustness and can be easily modified for different power systems or markets.

Although the forecasting of electricity load had reached a comfortable state of performance with errors of 1%-3% in the past years [1], this problem is still such a difficult task that a comprehensive and general solution is far from easy. This can also be explained by the large amount of publications on load forecasting in the past years. For a specific system, the best performances can be achieved only if a deep investigation of the inherent characteristics for the system has been carried out. For the proposed method, different sets of feeders for the SOM network should be selected according to characteristics of the system, because the sensitivities of the input variables may vary when the system changes. For instance, the humidity will be very important for the load forecasting in tropic areas and summer. However, in temperate zone and autumn, the load may not be sensitive to the humidity. In the future, we will further incorporate operational and market factors in our model to improve the prediction accuracy.

REFERENCES

[1] H. S. Hippert, C. E. Pedreira, and R. Castro, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 44–55, Feb. 2001.
[2] T. Haida and S. Muto, "Regression based peak load forecasting using a transformation technique," *IEEE Trans. Power Syst.*, vol. 9, no. 4, pp. 1788–1794, Nov. 1994.
[3] A. D. Papalexopoulos and T. C. Hesterberg, "A regression-based approach to short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 5, no. 4, pp. 1535–1550, Nov. 1990.
[4] S. Rahman and O. Hazim, "A generalized knowledge-based short term load-forecasting technique," *IEEE Trans. Power Syst.*, vol. 8, no. 2, pp. 508–514, May 1993.
[5] S. J. Huang and K. R. Shih, "Short-term load forecasting via ARMA model identification including non-Gaussian process considerations," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 673–679, May 2003.
[6] H. Wu and C. Lu, "A data mining approach for spatial modeling in small area load forecast," *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 516–521, May 2003.
[7] G. E. P. Box and G. M. Jenkins, *Time Series Analysis—Forecasting and Control*. San Francisco, CA: Holden-Day, 1976.
[8] D. G. Infield and D. C. Hill, "Optimal smoothing for trend removal in short term electricity demand forecasting," *IEEE Trans. Power Syst.*, vol. 13, no. 3, pp. 1115–1120, Aug. 1998.
[9] A. Khotanzad, R. C. Hwang, A. Abaye, and D. Maratukulam, "An adaptive modular artificial neural network hourly load forecaster and its implementation at electric utilities," *IEEE Trans. Power Syst.*, vol. 10, no. 3, pp. 1716–1722, Aug. 1995.
[10] T. Czernichow, A. Piras, K. Imhof, P. Caire, Y. Jaccard, B. Dorizzi, and A. Germond, "Short term electrical load forecasting with artificial neural networks," *Eng. Intell. Syst.*, vol. 2, pp. 85–99, 1996.
[11] L. M. Saini and M. K. Soni, "Artificial neural network-based peak load forecasting using conjugate gradient methods," *IEEE Trans. Power Syst.*, vol. 12, no. 3, pp. 907–912, Aug. 2002.
[12] S. Fan, C. X. Mao, and L. N. Chen, "Peak load forecasting using the self-organizing map," in *Advances in Neural Network-ISNN 2005*. New York: Springer-Verlag, 2005, pt. III, pp. 640–649.
[13] B.-J. Chen, M.-W. Chang, and C.-J. Lin, "Load forecasting using support vector machines: A study on EUNITE competition 2001," *IEEE Trans. Power Syst.*, vol. 19, no. 4, pp. 1821–1830, Nov. 2004.
[14] S. Fan, C. X. Mao, and L. N. Chen, "Next-day electricity price forecasting using a hybrid network," *IEEE Trans. Power Syst.*, submitted for publication.
[15] C. Cortes and V. Vapnik, "Support-vector network," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.
[16] N. Cristianini and J. Shawe-Tylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[17] K.-R. Müller, A. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik, "Predicting time series with support vector machines," in *Advances in Kernel Methods—Support Vector Learning*, B. Schölkopf, C. J. C. Burges, and A. J. Smola, Eds. Cambridge, MA: MIT Press, 1999, pp. 243–254.
[18] K. B. Song, Y. S. Baek, D. H. Hong, and G. Jang, "Short-term load forecasting for the holidays using fuzzy linear regression method," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 96–101, Feb. 2005.
[19] J. N. Fidalgo and J. A. Pecas Lopes, "Load forecasting performance enhancement when facing anomalous events," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 408–415, Feb. 2005.
[20] New York Independent System Operator. [Online]. Available: http://www.nyiso.com
[21] T. Kohonen, *Self-Organizing Maps*. Berlin, Germany: Springer-Verlag, 1997.
[22] G. A. Barreto and A. F. R. Araujo, "Identification and control of dynamical systems using the self-organizing map," *IEEE Trans. Neural Netw.*, vol. 15, no. 5, pp. 1244–1259, Sep. 2004.
[23] L. Chen and K. Aihara, "Chaotic simulated annealing by a neural network model with transient chaos," *Neural Netw.*, vol. 8, pp. 915–930, 1995.
[24] C.-C. Chang and C.-J. Lin. (2001) LIBSVM: A Library for Support Vector Machines. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Shu Fan received the B.S., M.S., and Ph.D. degrees from Huazhong University of Science and Technology, Wuhan, China, in 1995, 2000, and 2004, respectively. Presently, he is carrying out postdoctoral research at Osaka Sangyo University, Osaka, Japan.

His research interests are power system operation and intelligent systems.

Luonan Chen (M'94–SM'98) received the B.E. degree from Huazhong University of Science and Technology, Wuhan, China, in 1984 and the M.E. and Ph.D. degrees from Tohoku University, Sendai, Japan, in 1988 and 1991, respectively.

Since 1997, he has been a faculty member with Osaka Sangyo University, Osaka, Japan, where he is currently a Professor with the Department of Electrical Engineering and Electronics. His fields of interest are nonlinear dynamics and optimization in power systems.