An Integrated Machine Learning Model for Day-Ahead Electricity Price Forecasting

Shu Fan, James R. Liao, Kazuhiro Kaneko, Member, IEEE, and Luonan Chen, Senior Member, IEEE

Abstract—This paper proposes a novel and effective forecasting model based on the integrative machine learning technique possesses the following advantages:

1) It is well suited for capturing the dynamics of electricity price time series. The BCD classifier models the process generating each price series as an autoregressive model of order \( p \), say AR\((p)\), and then clusters those price time series with a high posterior probability of being generated by the same AR\((p)\) model \([14],[15]\). BCD is based on unsupervised learning, which has the ability to partition the space of input training data set into many subsets without prior knowledge about the classifying criteria. Compared to other clustering methods such as hierarchical clustering or Self Organizing Maps (SOM), BCD identifies the set of clusters with maximum posterior probability without requiring any prior input about the number of clusters, thereby avoiding the risk of overfitting.

2) It can fit the data well due to multiple local models. For each cluster of time series, we adapt 24 SVR for 24-hour predictions of a day. SVM or SVR (Support Vector Regression) is a new and powerful machine learning technique for data classification and regression based on recent advances in statistical learning theory \([16]\). Established on the unique theory of the structure risk minimization principle to estimate a

Index Terms—Electricity Price, Non-stationarity, Machine Learning.

I. INTRODUCTION

Electricity price forecasting has become increasingly important with the evolution of the electric power industry into an era of market economy due to deregulation and free competition \([1]\). Generators and loads rely on price forecasting information to develop their corresponding bidding strategies. If a generator has an accurate forecasting of the prices, it can develop a bidding strategy to maximize its profit. On the other hand, a load can make a plan to minimize its own electricity cost if an accurate price forecasting is available.

This paper focuses on forecasting the day-ahead hourly prices associated with a pool. Several approaches have been proposed in the recent years for such as problem, such as linear transfer function \([2]\), and Autoregressive Integrate Moving Average (ARIMA) models \([3]\); Artificial Neural Networks (ANNs) \([4] - [8]\); wavelet transform models \([9]\); machine learning techniques, \([10] - [12]\) and so on \([1], [13]\).

As indicated by many researchers, one key characteristic of electricity prices is their volatility and non-stationarity \([1], [8], [11], [12]\). The dynamics and properties of price series that frequently switch between different segments or regions are mainly due to discrete changes in competitors' strategies and multiple seasonality that generally give rise to piece-wise stationary time series. However, almost all the methods of time series analysis, both linear and nonlinear, assume some stationarity of the system under investigation. Therefore, when modeling time series of price, a key challenge is on how to handle the non-stationarity.

This paper proposes a novel and effective forecasting model by applying machine learning techniques with the emphasis on tackling the non-stationarity problem. Specifically, we adopt an integrated architecture based on two machine learning techniques: Bayesian Clustering by Dynamics (BCD) and Support Vector Machine (SVM). The basic procedure of the proposed model is stated as follows: firstly, a BCD classifier is used to identify the switching or piece-wise stationary dynamics for the input data and to partition the dataset into several subsets, so that the dynamics in each subset are similar; then groups of 24 SVMs are applied to respectively fit the hourly electricity price profile data in each partitioned subset by taking advantage of all past information and similar dynamic properties (e.g. piece-wise stationarity). After being trained, the forecasting model can predict the next-day electricity price with an acceptable level of accuracy on the specific subset in a 'first past the post' voting manner among the BCD and SVM, where the output of only one SVM model is used for the final forecast.

This forecasting model based on the integrative machine learning technique possesses the following advantages:

1) It is well suited for capturing the dynamics of electricity price time series. The BCD classifier models the process generating each price series as an autoregressive model of order \( p \), say AR\((p)\), and then clusters those price time series with a high posterior probability of being generated by the same AR\((p)\) model \([14],[15]\). BCD is based on unsupervised learning, which has the ability to partition the space of input training data set into many subsets without prior knowledge about the classifying criteria. Compared to other clustering methods such as hierarchical clustering or Self Organizing Maps (SOM), BCD identifies the set of clusters with maximum posterior probability without requiring any prior input about the number of clusters, thereby avoiding the risk of overfitting.

2) It can fit the data well due to multiple local models. For each cluster of time series, we adapt 24 SVR for 24-hour predictions of a day. SVM or SVR (Support Vector Regression) is a new and powerful machine learning technique for data classification and regression based on recent advances in statistical learning theory \([16]\). Established on the unique theory of the structure risk minimization principle to estimate a

S.Fan, K.Kaneko and L.Chen are with Osaka Sangyo University, 3-1-1 Nakagaito, Daito, Osaka, 574-0013, Japan (e-mail: fanshu@hotmail.com; k-kaneko@eic.osaka-sandai.ac.jp, chen@eic.osaka-sandai.ac.jp).

James R.Liao is with the Western Farmers Electric Cooperative, Anadarko, OK, 73005 USA (e-mail: j_liao@wfec.com)
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 [17],[18].

3) It is strongly robust. The proposed forecasting model uses different input variables for BCD and SVR respectively. All the input variables are decomposed into two groups: only price and load data are used for the SVMs; for BCD, besides price and load data, other factors, such as day type, weather factors and fuel price can be also used to capture the dynamics. Compared with a unity model dealing with all variables together, the approach will alleviate the sensitivity of the model to these variables and the user can intentionally split input variables using his knowledge and experience.

In the experiment, we adopt the price data of the New England Independent System Operator (ISO) to demonstrate the effectiveness of the learning and forecasting for the proposed method. To compare the forecasting accuracy, we also examine the model only with SVMs.

II. TASK DESCRIPTION AND PRICE DATA ANALYSIS

This paper uses the hourly electrical hub price series of New England Power Pool (NEPOOL) as a test example of our method [19]. In order to develop an appropriate model, we give a brief description of test system and examine the main characteristics of the hourly price series in this section.

The New England electricity market covers six-state region, including Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island and Vermont. The competitive electrical wholesale markets were implemented on December 1, 1998 [20]. All of New England has been treated as a single power market, with price set by an hourly bid process administered by the ISO New England. On March 1, 2003, the New England power market successfully implemented the Standard Market Design (SMD), a major redesign of the wholesale electricity market. The day-ahead market consists of 24 hourly auctions that take place simultaneously one day in advance. Hourly bids, expressed in $/MWh, are submitted on a day-ahead basis for the next 24 hours by the generators and loads. ISO New England then optimally schedules the generating units that will run on the following day by minimizing the total costs of the energy market. The market is settled after the fact on an hourly basis. All transactions are priced at the (ex-post) energy clearing price (ECP), which is determined at the intersection of aggregate supply and demand curves.

Fig.1 illustrates the latest electricity prices for the hub of NEPOOL from March 1, 2003 to July 31, 2005. According to Fig.1, it is clear that the price dynamics have multiple seasonal patterns, corresponding to a daily and weekly periodicity respectively, and are also influenced by calendar effect, i.e. weekend and holidays. These properties are just similar to those of load. However, in contrast to the load time series, there are several particular properties with price. The hourly price curve is widely divergent and fluctuates with a high frequency, and there are also a high percentage of abrupt changes or spikes in the price curve (mainly in periods of high demand). The price presents high volatility and non-constant mean. The abrupt changes and volatility of price can be reflected as a switch in the price series dynamics due to the discrete behaviors in competitors’ strategies. In other words, there exist different regimes in the price time series, which generally give rise to piece wise stationary dynamics.

III. METHOD AND THE LEARNING ALGORITHM

A. The Architecture of the Forecasting System

In this paper, a time series based nonlinear discrete-time dynamical model, is represented by (1) for the price forecasting.

\[
y(t + 1) = f(y(t),...,y(t - m + 1);D)
\]

where \(y(t)\) is a vector representing the daily electricity price profile at time \(t\), and \(m\) is the order of the dynamical system, which is a predetermined constant. \(D\) is a vector representing the control parameters of the dynamical system, such as weather conditions, day types, power plant mix, and fuel prices. Then, the task in price forecasting is to extrapolate past price behavior while taking into account the effect of other influencing factors.

An integrated machine learning model is proposed to reconstruct the dynamics of electric price using the time series of its observables. The forecasting system is shown in Fig. 2. This model is based on an integrated architecture. Firstly, a BCD classifier is applied to cluster the input training data set into several subsets with similar dynamical properties in an unsupervised manner. Then, 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 similar property. Notice that there are 24 SVRs for each subset to train and predict 24-hour prices of next day respectively.

As indicated in Table I, the input data of BCD consist of the following elements: average daily price series of previous ten days \( P_0 \), maximal daily demand \( L \) and temperature \( W \), and daily fuel price \( F \). Specifically, the price time series \( P_0 \) are used as the primary input for BCD to cluster the day by dynamics, the other factors, including load, weather conditions and fuel prices, can be used to determine the prior probability for the BCD. Other factors can also be included, and the selection of input data mainly depends on the system to be studied.

For the SVM, in addition to the forecasted and actual load values, the input variables are the hourly price values of the last day available and the similar hours in the previous days or weeks. The input data of the SVM network is shown in Table II.

<table>
<thead>
<tr>
<th>Input</th>
<th>Variable name</th>
<th>Lagged value (hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-9</td>
<td>Hourly price ((L))</td>
<td>24, 25, 26, 48, 72, 96, 120, 144, 168</td>
</tr>
<tr>
<td>10-19</td>
<td>Hourly load ((T))</td>
<td>-1, 0, 1, 24, 48, 72, 96, 120, 144, 168</td>
</tr>
</tbody>
</table>

The inputs of training data consist of past hourly electricity price and the forecasted demand. In order to capture the time series style in price, we include the electricity price of the previous seven days around the predicting hour. Because our experiment shows that using load of the forecasting day only does not produce satisfactory forecasting results, load variables are added for every time point at which price was included.

B. Learning Algorithm: The BCD Classifier

Dynamics of electricity prices are non-stationary or approximately piece-wise stationary during a short period of time due to the switching nature related to multiple seasonality and discrete changes in participants’ strategies. In this paper, the BCD classifier is used as a gating network to identify the switching or piece-wise stationary dynamics for the input training data set firstly.

The clustering method implemented in BCD is based on a novel concept of similarity for time series: two time series are similar when they are generated by the same stochastic process. Therefore, the components of BCD are a model describing the dynamics of time series, a metric to decide when two time series are similar when they are generated by the same stochastic process, and a search procedure to efficiently explore the space of possible clustering models.

BCD models time series by autoregressive equations [15]. Let \( S_j = \{x_{j1}, \ldots, x_{jp}, \ldots, x_{jn}\} \) denote a stationary time series of continuous values. The series follows an autoregressive model of order \( p \), say AR\((p)\), if the value of the series at time \( t > p \) is a linear function of the values observed in the previous \( p \) steps. We can describe the model in a matrix form as

\[
x_j = X_j \beta_j + e_j
\] (2)
where \( x_j \) is the vector \( (x_{j(p+1)}, \ldots, x_{jn})^T \), \( X_j \) is the \((n-p) \times (p+1)\) regression matrix whose \( t \)th row is \((1, x_{j1}, \ldots, x_{j(p-1)})^T\) for \( t \geq p \), \( \beta_j \) is the vector of autoregressive coefficients and \( \varepsilon_j \) is the vector of uncorrelated errors that are assumed normally distributed with expected value \( E(\varepsilon_j) = 0 \) and variance \( V(\varepsilon_j) = \sigma_j^2 \), for any time point \( t \). Given the data, the model parameters can be estimated using standard Bayesian procedures, and details are described in [15].

To select the set of clusters, BCD uses a novel model-based Bayesian clustering procedure. A set of clusters \( C_1, \ldots, C_k, \ldots, C_m \), each consisting of \( m_k \) time series, is represented as a model \( M_{C_k} \). The time series assigned to each cluster are treated as independent realizations of the dynamic process represented by the cluster, which is described by an autoregressive equation. Each cluster \( C_k \) can be jointly modeled as

\[
x_k = X_k \beta_k + \varepsilon_k
\]

where the vector \( x_k \) and the matrix \( X_k \) are defined by stacking the \( m_k \) vectors \( x_{kj} \) and regression matrices \( X_{kj} \), one for each time series, as follows

\[
x_k = \begin{pmatrix} x_{k1} \\ \vdots \\ x_{km_k} \end{pmatrix}, \quad X_k = \begin{pmatrix} X_{k1} \\ \vdots \\ X_{km_k} \end{pmatrix}
\]

Given a set of possible clustering models, the task is to rank them according to their posterior probability. The posterior probability of the model \( M_{C_k} \) is computed by Bayes Theorem as

\[
P(M_{C_k} | x) \propto P(M_{C_k}) f(x | M_{C_k})
\]

where \( P(M_{C_k}) \) is the prior probability of \( M_{C_k} \) and \( f(x | M_{C_k}) \) is the marginal likelihood. Assuming independent uniform prior distributions on the model parameters and a symmetric Dirichlet distribution on the cluster probability \( p_k \), the marginal likelihood of each cluster model \( M_{C_k} \) can be easily computed in a closed form by solving the integral:

\[
f(x | M_{C_k}) = \int f(x | \theta_k) f(\theta_k) d\theta_k
\]

where \( \theta_k \) is the vector of parameters that describe the likelihood function, conditional on a clustering model \( M_{C_k} \), and \( f(\theta_k) \) is the prior density. In this way, each clustering model has an explicit probabilistic score and the model with maximum score can be found. In particular, \( f(x | M_{C_k}) \) can be computed as

\[
f(x | M_{C_k}) = \frac{\Gamma(1)}{\Gamma(1+m)} \prod_{k=1}^m \frac{\Gamma(m_k / m + m_k)}{\Gamma(m_k / m)} \frac{\text{RSS}(k)}{2} \left( \frac{n_k - q}{2} \right)
\]

\[
\times \left( \frac{n_k - q}{2} \right) ^{(q-n_k)/2} \left( \frac{n_k - q}{2} \right) ^{(q-n_k)/2} \frac{\text{det}(X_k^T X_k)}{(2\pi)^{n_k/2}}
\]

\[
where \( n_k \) is the dimension of the vector \( x_{kj} \) and \( \text{RSS}_k = x_k^T (I_{m_k} - X_k (X_k^T X_k)^{-1} X_k^T) x_k \) is the residual sum of squares in cluster \( C_k \). When all cluster models are \( a \ priori \) equally likely, the posterior probability \( P(M_{C_k} | x) \) is proportional to the marginal likelihood \( f(x | M_{C_k}) \), which becomes our probabilistic scoring metric.

As the number of clusters or subsets grows exponentially with the number of time series, BCD uses an agglomerative search strategy, which iteratively merges time series into clusters. The procedure starts by assuming that each of the \( m \) electricity price time series is generated by a different process. Thus, the initial model \( M_m \) consists of \( m \) clusters, one for each time series, with score \( f(x | M_m) \). The next step is the computation of the marginal likelihood of the \( m(m-1) \) models in which two of the \( m \) profiles are merged into one cluster. The model \( M_{m-1} \) with maximal marginal likelihood is chosen and the merging is rejected if \( f(x | M_m) \geq f(x | M_{m-1}) \) and the procedure stops. If \( f(x | M_m) < f(x | M_{m-1}) \), the merging is accepted and a cluster \( C_l \) merging the two time series is created. In such a way, the procedure is repeated on the new set of \( m-1 \) time series that consist of the remaining \( m-2 \) time series and the cluster profile.

Although the agglomerative strategy makes the search process feasible, the computational effort can be extremely demanding when the number of time series is large. To further reduce this effort, we use a heuristic strategy based on a measure of similarity between time series. The intuition behind this strategy is that the merging of two similar time series has better chances of increasing the marginal likelihood. The heuristic search starts by computing the \( m(m-1) \) pair-wise similarity measures of the time series and selects the model \( M_{m-1} \) in which the two closest time series are merged into one cluster. If \( f(x | M_m) < f(x | M_{m-1}) \), the two time series are merged into a single cluster, a profile of this cluster is computed by averaging the two observed time series, and the procedure is repeated on the new set of \( m-1 \) time series. If this merging is rejected, the procedure is repeated on the two time series with the second highest similarity until an acceptable merging is found. If no acceptable merging is found, the procedure stops. Note that the clustering procedure is actually performed on the posterior probability of the model and the similarity measure is only used to increase the speed of the search process and to limit the risk of falling into local maxima.

Similarity measures of two time series implemented in BCD include Euclidean distance, correlation and Kullback–Leiber distance. In the numerical experiments, we have tried different distances and finally adopted the Euclidean distance of two time series \( S_i = \{x_{ij}, \ldots, x_{im}\} \) and \( S_j = \{x_{ij}, \ldots, x_{jn}\} \), computed as

\[
D(S_i, S_j) = \sqrt{\sum_{t=1}^{n} (x_{it} - x_{jt})^2}
\]

C. Learning Algorithm: The SVM Network

In this stage, several groups of SVRs are used for price time series learning and prediction in each subset, and each SVR corresponds to one time instant. Suppose that we are given training data \( (x_{ij}, y_i)(x_{ij}, y_j) \ldots (x_{ij}, y_n) \), 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
\[ \min_{\omega, b, \varepsilon} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^{n} (\varepsilon_i + \varepsilon_i^*) \]
\[ \text{Subject to } y_i - (\omega^T \varphi(x_i) + b) \leq \varepsilon + \varepsilon_i^*, \]
\[ (\omega^T \varphi(x_i) + b) - y_i \leq \varepsilon + \varepsilon_i^*, \]
\[ \varepsilon_i, \varepsilon_i^* \geq 0, \quad i = 1, \ldots, n \]

where \( x_i \) is mapped to a higher dimensional space by the function \( \varphi \), and \( \varepsilon_i^* \) is slack variables of the upper training error (\( \varepsilon_i \) is the lower) subject to the \( \varepsilon \)-insensitive tube \((\omega^T \varphi(x_i) + b) - y_i \leq \varepsilon \). The constant \( C > 0 \) determines the trade off between the flatness and losses. The parameters which control regression quality are the cost of error \( C \), the width of the tube \( \varepsilon \), and the mapping function \( \varphi \).

The constraints of (7) imply that we put most data \( x_i \) in the tube \( \varepsilon \). If \( x_i \) is not in the tube, there is an error \( \varepsilon_i \) or \( \varepsilon_i^* \) which we tend to minimize in the objective function. SVR avoids under-fitting and over-fitting of the training data by minimizing the training error \( C \sum_{i=1}^{n} (\varepsilon_i + \varepsilon_i^*) \) as well as the regularization term \( \omega^T \omega / 2 \). For traditional least-square regression, \( \varepsilon \) 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 \( \varphi \) might map \( x_i \) to a high or infinite dimensional space, instead of solving \( \omega \) for (7) in a high dimension, we deal with its dual problem, which can be derived using the Lagrange theory.

\[ \max_{\alpha, \alpha^*} \left\{ \frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i - \alpha_j^*) \alpha_j^* Q(x_i, x_j) \right\} - \varepsilon \sum_{i=1}^{n} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \]
\[ \text{Subject to } \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) = 0 \]
\[ 0 \leq \alpha_i, \alpha_i^* \leq C, \quad i = 1, \ldots, n \]

where \( Q_g = \varphi(x_i)^T \varphi(x_j) \), \( \alpha_i \) and \( \alpha_i^* \) are the Lagrange multipliers. However, this inner product may be expensive to compute because \( \varphi(x) \) has too many elements. Hence, we apply “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 \( \varphi(x_i)^T \varphi(x_j) = (x_i^T x_j + c_0)^d \) and RBF kernel \( \varphi(x_i)^T \varphi(x_j) = e^{-\gamma(x_i^T x_j)} \). Here, \( \gamma \), \( c_0 \), and \( d \) are kernel parameters. 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 \( \varphi(x) \).

As each data subset classified from the BCD is considered to be approximately stationary, 24 SVRs are applied to respectively fit the hourly electricity price profile data by taking advantage of all past information and similar dynamic properties (e.g. piece-wise stationarity). The next-day electricity price forecasting is conducted by the trained network with an acceptable level of accuracy in a voting manner in the BCD and SVRs. For numerical experiments in this paper, we use the software LIBSVM [21], which is a library for support vector machines with an efficient implementation of solving (8).

IV. NUMERICAL EXPERIMENTS

A. Data Collection and Preprocess

The hourly electrical prices and load data in New England have been considered for the study [19]. The SMD market period of 3/2003-present is studied in the experiments. We use the hub price of New England market as forecasting object. Two months have been selected to forecast and validate the performance of the proposed model. The first month corresponds to May 2005, and the latter month corresponds to August 2005. The hourly data used to forecast the three months are from March 1, 2003 to April 30, 2005; and March 1, 2003 to July 31, 2005.

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

B. Learning Procedure and Implementation

The learning procedure of the proposed architecture is outlined as follows.

1) Classify the entire training data set into two groups: a training set used for updating the network parameters and a verification set for testing the performance.

2) Determine prior probability of the BCD classifier according to the control variables;

3) Classify the day type using BCD according to the dynamics of price time series;

4) In each subset of the input space, train the SVM to fit the data subset according to (8);

5) Forecast the price profile using the data set for verification and calculate the MAPE;

6) Tune the parameters of the model and repeat step 1) to step 5) until satisfactory results are obtained;

7) Select the network parameters at the minimum of the MAPE as the final ones.

After the training procedure is finished, the following test process is applied to verify the proposed model.

1) Identify the type of the test day according to the information of the test day and previous days, using the BCD classifier.

2) Use the corresponding SVM network to output the forecasting price.
C. Numerical Results

The criteria to compare the performance are the mean absolute error (MAE) and mean absolute percentage error (MAPE) in this paper, which indicate the accuracy of recall.

Numerical results with the proposed method are presented. To demonstrate the effectiveness of the integrated structure, a SVM network without the BCD classifier is also adopted for forecasting. The MAEs and MAPEs for the two test months are shown in Tables III.

<table>
<thead>
<tr>
<th>Table III</th>
<th>Forecasting results of May and August 2005</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single SVM</td>
</tr>
<tr>
<td></td>
<td>MAE</td>
</tr>
<tr>
<td>May 2005</td>
<td>3.78</td>
</tr>
<tr>
<td>August 2005</td>
<td>8.04</td>
</tr>
<tr>
<td>Average</td>
<td>5.91</td>
</tr>
</tbody>
</table>

Clearly the integrated machine learning model outperforms single SVM in all the situations. The variety of error indices for the 24-hour is smooth, and indicate stability of the proposed model. Especially, the average MAPE for the two months in the new SMD market is 6.08%, which is considered high in terms of accuracy.

Fig. 3 shows the forecasting and the actual price for ten days in August. It can be seen that the proposed model well predicts the trend of the price curve generally. Since the SMD market has been implemented for two and a half years, only limited data have been available for learning so far. When more training samples become available and are included in training in the future, it is possible to get more accurate forecasting.

REFERENCES

Shu Fan received his B.S., M.S., and Ph.D. degrees in Department of Electrical Engineering, from Huazhong University of Science and Technology (HUST), in 1995, 2000 and 2004 respectively. Presently he is carrying out postdoctoral research in Osaka Sangyo University. His research interests are power system control, high-power power electronics and hybrid intelligent forecasting system.

James R. Liao (M’89) received the M.S. degree from the University of Missouri, Rolla, in 1980, and the Ph.D. degree from the University of Oklahoma, Norman, in 1992, both in electrical engineering. Since 1980, he has been with the Western Farmers Electric Cooperative, Anadarko, OK. He was a Transmission/Generation Systems Analyst from 1980 to 1985 and an EMS System Software Engineer from 1985 to 1999. Since 1999, he has been Principal Operations Engineer. Dr. Liao is a Registered Professional Engineer in the State of Oklahoma.

Luonan Chen (M’94, SM’98) received the B.E. degree from HUST, 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 of 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.