

# Load Forecasting Using Support Vector Machines: A Study on EUNITE Competition 2001

Bo-Juen Chen, Ming-Wei Chang, and Chih-Jen Lin, *Member, IEEE*

**Abstract**—Load forecasting is usually made by constructing models on relative information, such as climate and previous load demand data. In 2001, EUNITE network organized a competition aiming at mid-term load forecasting (predicting daily maximum load of the next 31 days). During the competition we proposed a support vector machine (SVM) model, which was the winning entry, to solve the problem. In this paper, we discuss in detail how SVM, a new learning technique, is successfully applied to load forecasting. In addition, motivated by the competition results and the approaches by other participants, more experiments and deeper analyses are conducted and presented here. Some important conclusions from the results are that temperature (or other types of climate information) might not be useful in such a mid-term load forecasting problem and that the introduction of time-series concept may improve the forecasting.

**Index Terms**—Load forecasting, regression, support vector machines, time series.

## I. INTRODUCTION

**E**LECTRICITY load forecasting has always been an important issue in the power industry. Load forecasting is usually made by constructing models on relative information, such as climate and previous load demand data. Such forecast is usually aimed at short-term prediction (e.g., [1]–[3] and references therein), like one-day ahead prediction, since longer period prediction (mid-term or long term) may not be reliable due to error propagation. Mid-term and long-term prediction on load demand, however, may still be very useful in some situations.

In 2001, EUNITE<sup>1</sup> network organized a competition on the similar problem. The goal is to predict daily load demand of a month. Given information includes the past two-year load demand data, the previous four-year daily temperature and the local holiday events. Dealing with such a mid-term forecast problem, reliable prediction and error propagation are both competitors' concern. During the competition, we proposed a model, which was the winning entry, to solve the problem. Though the competition has been closed, we find this topic interesting and useful. Moreover, we would like to figure out the performance of mid-term load forecasting with such limited information. In this paper, therefore, we present our approach and discuss more on this problem. The main technique used in our solution is support

vector machine (SVM) [4], a new machine learning method. To the best of our knowledge, this is the first work to successfully apply SVM on load forecasting. Some important conclusions from our experiments are that, temperature should not be considered in such a problem and the time-series modeling scheme is better. For the prediction of a 30-day period during which the temperature does not vary much, trying to predict the temperature and incorporate it into the model is not useful.

This paper is organized as follows. In Section II, we describe the goal of the competition task and the data provided. In addition, the analysis of data is also presented. Section III demonstrates the techniques we employed. Experiments and results of different models are showed in Section IV. Finally, the conclusion of our research and the comparison with other competitors are in Section V.

## II. DATA AND TASK DESCRIPTION

### A. Competition Task Description

The organizer of the EUNITE load competition provides competitors the following data.

- Electricity load demand recorded every half hour, from 1997 to 1998.
- Average daily temperature, from 1995 to 1998.
- Dates of holidays, from 1997 to 1999.

The task of competitors is to supply the prediction of maximum daily values of electrical loads for January 1999. Evaluation of submissions would mainly depend on the error metric, mean absolute percentage error (MAPE), of the results

$$\text{MAPE} = 100 \frac{\sum_{i=1}^n \left| \frac{L_i - \hat{L}_i}{L_i} \right|}{n}, \quad n = 31 \quad (1)$$

where  $L_i$  and  $\hat{L}_i$  are the real and the predicted value of maximum daily electrical load on the  $i$ th day of the year 1999 respectively, and  $n$  is the number of days in January 1999. The goal of the competition is to forecast electrical load with minimum MAPE.<sup>2</sup>

### B. Data Analysis

Before delving further into the solution we proposed, some observations about the data are examined first. Like many other literatures working on load forecasting, some relations between load demand and other information, such as climate or local

<sup>2</sup>Originally in the competition, there were two error metrics used. One is MAPE, and the other one, not really used, is the "maximal error"

$$\text{Maximal error} = \max |L_i - \hat{L}_i|, \quad i = 1, \dots, 31.$$

Manuscript received May 10, 2004. Paper no. TPWRS-00492-2002.

The authors are with the Department of Computer Science and Information Engineering, National Taiwan University, Taipei 106, Taiwan, R.O.C. (e-mail: cjlin@csie.ntu.edu.tw).

Digital Object Identifier 10.1109/TPWRS.2004.835679

<sup>1</sup>European Network on Intelligent TEchnologies for Smart Adaptive Systems [Online] Available: <http://www.eunite.org>. The competition page is <http://neuron.tuke.sk/competition/>.

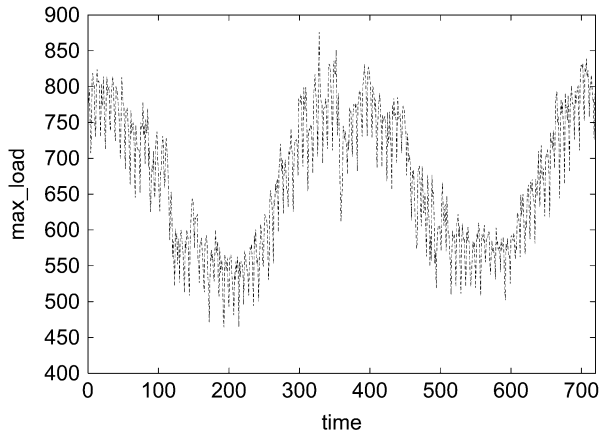


Fig. 1. Maximum daily load.

events, are also investigated. The following observations are somehow specific to the competition, yet they can be applied to general load forecasting.

1) *Properties of Load Demand*: Load demand data given are half-hour recorded. Fig. 1 gives a simple description of the maximum daily load demand from 1997 to 1998. By simple analysis, one can easily observe some properties of the load demand. First, the demand has some seasonal patterns: high demand for electricity in the winter while low demand in the summer. This pattern implies the relation between electricity usage and weather conditions in different seasons.

Moreover, if we scrutinize the data further, another load pattern could also be observed: a load periodicity exists in every week. Load demand in weekend is usually lower than that of weekdays (Monday through Friday). In addition, electricity demand on Saturday is a little higher than that on Sunday.

Further detailed examination of load data, such as daily patterns, can also be found, since the data set contains more details (half-hour recorded). However, since the aim of the competition is maximum values of daily load demand, this paper would, without loss of generality, focus on the maximum values only.

2) *Climate Influence*: In load forecasting, climate conditions have always played an important role. Previous work on short-term load forecasting [1]–[3] also indicate the relation between climate and load demand. Climate conditions considered may include temperature, humidity, illumination, and some special events like typhoon or sleet occurrences. These considerations may be regarded on different levels due to different localities. However, while forecasting load demand, more climate information usually give predictions more confidence.

Taking our data for example, as we mentioned earlier, the load data have some seasonal variation, which indicates a great influence of climate conditions. A negative correlation between load demand and daily temperature can be easily observed in Fig. 2. The correlation coefficient between the maximum load and the temperature is  $-0.868$ . In our data set, it is clear to see that because of the heating use, higher temperature causes lower demands. Yet, unfortunately the only climate information provided in the competition is daily temperature. Such information limitation also affects solutions to such a problem.

There is another interesting observation: the temperature at December 31st, 1998 is the lowest from 1997 to 1998. This

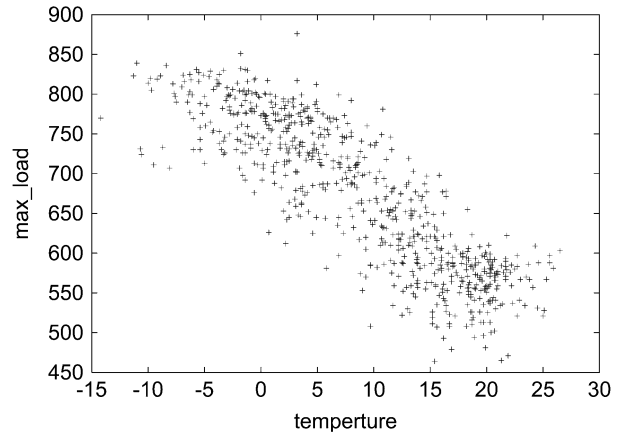


Fig. 2. Correlation between the maximum load and the temperature.

might imply the high uncertainty of the temperature and load of the incoming January 1999, and thus increase the difficulty of the load prediction.

3) *Holiday Effects*: Local events, including holidays and festivities, also affect the load demand. These events may lead to higher demand for extra usage of electricity, or otherwise. Influences of these events are usually local and highly depend on the customs of the area. From the two-year load data, it is easy to find out that the load usually lowers down on holidays. With further scrutiny, the load also depends on what holiday it is. On some major holidays such as Christmas or New Year, the demand for electricity may be affected more compared with other holidays.

### III. METHODS

SVM is a new and promising technique for data classification and regression [5]. In this section, we briefly introduce support vector regression (SVR) which can be used for time series prediction. Given training data  $(x_1, y_1), \dots, (x_l, y_l)$ , where  $x_i$  are input vectors and  $y_i$  are the associated output value of  $x_i$ , the support vector regression solves an optimization problem

$$\begin{aligned} \min_{w, b, \xi, \xi^*} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{subject to} \quad & y_i - (w^T \phi(x_i) + b) \leq \epsilon + \xi_i^*, \\ & (w^T \phi(x_i) + b) - y_i \leq \epsilon + \xi_i, \\ & \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, l \end{aligned} \quad (2)$$

where  $x_i$  is mapped to a higher dimensional space by the function  $\phi$ ,  $\xi_i^*$  is the upper training error ( $\xi_i$  is the lower) subject to the  $\epsilon$ -insensitive tube  $|y - (w^T \phi(x) + b)| \leq \epsilon$ . The parameters which control regression quality are the cost of error  $C$ , the width of the tube  $\epsilon$ , and the mapping function  $\phi$ .

The constraints of (2) imply that we would like to put most data  $x_i$  in the tube  $|y - (w^T \phi(x) + b)| \leq \epsilon$ . This situation can be clearly seen from Fig. 3. If  $x_i$  is not in the tube, there is an error  $\xi_i$  or  $\xi_i^*$  which we would like to minimize in the objective function. SVR avoids underfitting and overfitting the training data by minimizing the training error  $C \sum_{i=1}^l (\xi_i + \xi_i^*)$  as well as the regularization term  $(1/2)w^T w$ . For traditional least-square

regression  $\epsilon$  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  $\phi$  might map  $x_i$  to high or infinite dimensional space, we, instead of solving  $w$  for (2) in high dimension, deal with its dual problem

$$\min_{\alpha, \alpha^*} \frac{1}{2}(\alpha - \alpha^*)^T Q(\alpha - \alpha^*) + \epsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) \quad (3)$$

$$\text{subject to} \quad \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0, \\ 0 \leq \alpha_i, \quad \alpha_i^* \leq C, \quad i = 1, \dots, l$$

where  $Q_{ij} = \phi(x_i)^T \phi(x_j)$ . However, this inner product may be expensive to compute because  $\phi(x)$  has too many elements. Hence, we apply “kernel trick” to do the mapping implicitly. That is to employ some special forms which are inner products in higher space yet can be calculated in the original space. Some examples 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 very high dimensional space but can be computed efficiently.

For experiments in this paper, we use the software LIBSVM [6], which is a library for support vector machines, including the implementation of solving (3). While building SVM models, there are some parameters to select. Usually, we conduct cross-validation to choose suitable parameters. Details are in Section IV-B.

#### IV. EXPERIMENTS

Though the competition has been closed, the problem still presents an interesting issue in load forecasting. In order to study mid-term load forecasting more, we conduct some experiments on the competition problem. In this section, these experiments and results would be described.

##### A. Data Preparation

In Section III, we have described the SVM technique. We next need to prepare data sets to build SVM models. For that, we need to encode useful information into the data entries [i.e.,  $x_i$  in (2)]. In addition, different data encodings affect the selection of modeling schemes. Here we will discuss these issues in detail.

1) *Feature Selection*: Each component of the training data is called a feature (attribute). Here, we consider what kind of information should be included. Assuming that  $y_i$  is the load of the  $i$ th day, in general we incorporate information at the same day or earlier as features of  $x_i$ . There are a few choices for the feature:

a) *Basic Information: Calendar Attributes*: In Section II-B, we discuss the weekly periodicity of the load demand. Also, as we pointed out earlier, the load demand on holidays is lower than that on nonholidays. Therefore, encoding these information (weekdays and holidays) in the training

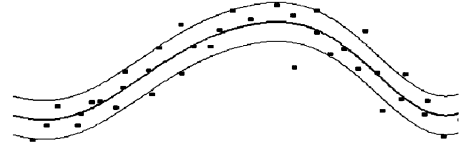


Fig. 3. SVR.

entries might be useful to model the problem. Actually, among literatures of load forecasting, many works (e.g., [3] and [7]) have used the calendar information (time, dates, or holidays).

b) *Temperature*: Another possible feature is the temperature data. This is quite a straightforward choice, since load demand and temperature have a causal relation in between. In most short-term load forecasting (STLF) works, meteorological information which includes temperature, wind speed, sky cover, etc., has been used to predict the load demand. However, to include the temperature in the training entries, there is one difficulty: in this competition, the real temperature data of January 1999 are not provided. In other words, for such mid-term load forecasting, temperature several weeks away are generally not available by weather forecasting. If we would like to encode the temperature in our training entries, we will also need to predict or estimate the temperature of January 1999. Yet, temperature forecast is not easy, especially with such limited data. The use of temperature, therefore, would be a dilemma.

c) *Time Series Style or Not*: Besides the weekdays, holidays and temperature, there is another information we consider to encode as the attributes: the past load demand. That is to introduce the concept of time-series into our models. To be more precise, if  $y_i$  is the target value for prediction, the vector  $x_i$  includes several previous target values  $y_{i-1}, \dots, y_{i-\Delta}$  as attributes. In the training phase all  $y_i$  are known but for future prediction,  $y_{i-1}, \dots, y_{i-\Delta}$  can be values from previous predictions. For example, after obtaining an approximate load of January 1, 1999, if  $\Delta = 7$ , it is used with loads of December 26–31, 1998 for predicting that of January 2. We continue this way until finding an approximate load of January 31. An earlier example using SVR for time series prediction is [8]. As we know, the past load demand could affect and imply the future load demand. Therefore, considering to include such information in the models might help forecast the load demand. In fact, in some load forecasting works, time series models have been explored [2], [9].

2) *Data Segmentation*: Besides the features choices, we may consider only a subset of load data, since Section II also shows the seasonal pattern. Previous works [1], [2], [7] on load forecasting also propose models built on data of different seasons. This inspires us to do some analyzes for the data segmentation.

Usually, people model time-series data by using the formulation

$$y_t = f(x_t), \quad \forall t = \Delta + 1, \dots, l$$

where  $x_t = (y_{t-1}, y_{t-2}, \dots, y_{t-\Delta})$ , and  $\Delta$  is the embedding dimension. However, this formulation is not suitable for a nonstationary time series, because the characteristic of the time series

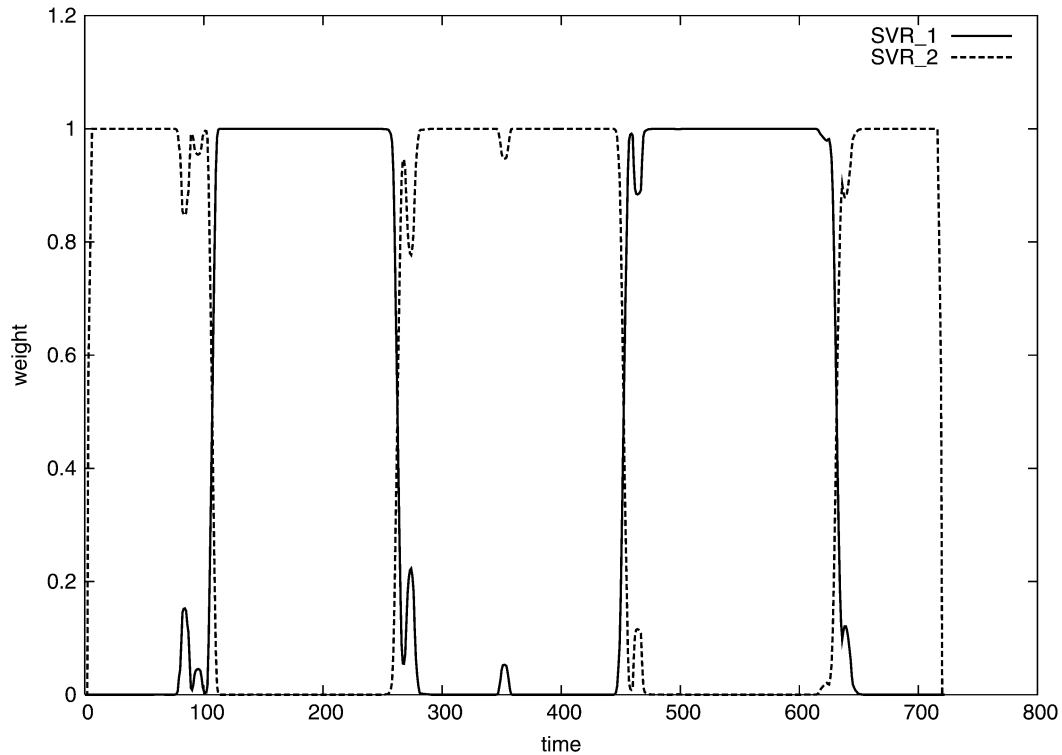


Fig. 4. Unsupervised segmentation for EUNITE data.

may change with time. For such a time series which alternate in time, we can consider a mixture model where

$$y_t = f_{i(t)}(x_t), \quad \forall t = \Delta + 1, \dots, l.$$

Note that the formulation allows different characteristic functions in different time. Given the series  $y_t, t = 1, \dots, l$ , the task is to identify all  $i(t)$ . We call this unsupervised segmentation where earlier work can be found in, for example, [10]. In other words, the method breaks the series into different segments where points in the same segment can be modeled by the same  $f_i$ . Recently, [11] states a similar framework using SVR with different parameters tunings. At any time point  $t$ , these methods consider different weights representing the probability that  $y_t$  belongs to corresponding functions. The sum of weight at any given time point  $t$  is always fixed to one. The weights are iteratively updated until one weight is close to one but others are close to zero. That means eventually  $y_t$  is associated to one particular time series.

Now, we can consider the loads as a time series. First, we linearly scale all load data to  $[0, 1]$ . Then we can get time series style data by incorporating load of last seven days and weekday information to attributes. After that, we follow the framework of [11] to analyze the data. We consider two possible time series so at each time point there are two weights. The experimental result is in Fig. 4. The x-axis indicates days from January 1997 to December 1998 and the y-axis indicates the weights of two time series. Interestingly, “winter” and “summer” data are automatic separated without any seasonal information. The figure shows that the loads in the summer and in the winter have different characteristics.

Unsupervised data segmentation has been very useful for time series prediction (e.g., [10]). If the training data are associated with different time series, it is better to consider only data segments related to the same series of the last segment. Now the objective of the competition is to predict the load demand of January 1999, so we consider to use only the winter segment for training. Here, we choose January to March and October to December to be our “winter” period, as the analytical result indicates. That is, the “winter” data set would contain half of the data in 1997 and 1998. Also, we further extract data of January and February to form another possible training data set. This data set is much smaller than the “winter” one, and it would focus more on the load pattern in the period of our target concern.

3) *Data Representation:* After selecting useful information and proper data segments for encoding, we can prepare several combinations of training data sets. In these data sets, we encode a training entry [i.e.,  $x_i$  in (2)] for the particular  $i$ th day, as follows:

(calendar, temperature (optional), past load (optional.))

Here we use seven binaries to encode calendar information which includes weekdays, weekends, and holidays, where six are for weekdays and weekends, and the other one for holidays. The six binaries stand for Monday to Saturday respectively and Sunday is represented as all six attributes are set to zero. Also, one numerical attribute is used for normalized temperature data, if the temperature is encoded. As for the past load, if encoded, we use seven numerics for the past seven daily maximum loads. The reason for using “seven” instead of other numbers is the complexity of model selection. We will elaborate more on this

later. Finally, for such an entry, its target value  $[y_i$  in (2)] is assigned to the maximum load demand of the  $i$ th day.

Yet, for the models built with temperature information, the lack of temperature of January 1999 will pose a problem in the predicting phase. As we mentioned before, this may lead us to the prediction or estimation of the temperature in 1999 and that will be a more difficult problem, especially with only daily temperature of the past four years (1995–1998). A straightforward idea is to use the average of the past temperature data for the estimation. That is, the temperature of each day in January 1999 is estimated by averaging the past January daily temperature data (from 1995 to 1998). Competitors like [12]–[14] used the averaged temperature in their models. Also, [15] uses temperature of the other cities close to Slovakia for the estimation. In their report, the temperature for January 1999 was calculated through a linear combination of that of the other three cities. However, for an application in real world, the actual temperature information of January 1999 at any place is not available since the prediction would be made at December 31, 1998. Therefore, the formulation in [15] might be helpful for inspecting the influence of temperature on the load prediction, yet impractical in use. For experiments here, these two estimations for temperature will be employed in the predicting phase, provided the temperature is encoded in our training data sets.

There is one remark about the testing data of January 1999. In this period, there are two holidays, January 1 and 6. However, in the data entries prepared for the prediction, we remove the holiday flag of these two entries. In other words, we treat them as nonholidays. The reason is that our models cannot learn well about the load demand on holidays. The number of holidays in the training data is too small to provide enough information. Moreover, for the time-series-based approach, inaccurate prediction at one day could affect the succeeding forecasting.

## B. Implementation and Results

With different schemes of model construction, a series of experiments are conducted. MAPE's of different model schemes are the main concern of our comparison.

Upon the data sets we prepare, SVM models are built for load forecasting. When training an SVM model, there are some parameters to choose. They would influence the performance of an SVM model. Therefore, in order to get a “good” model, these parameters need to be selected properly. Some important ones are:

- 1) cost of error  $C$ ;
- 2) the width of the  $\epsilon$ -insensitive tube;
- 3) the mapping function  $\phi$ ;
- 4) load of how many previous days included for one training data.

In our experiments, for each training data we simply include the maximum load of the previous seven days. For the parameter  $\epsilon$ , after a few trials we fix it to 0.5. In addition, we consider only the radial basis function (RBF) kernel, as we mention in Section III. The RBF function has the property that  $K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}$ . Note that  $\gamma$  is a parameter associated with the RBF function which has to be tuned. Thus, parameters left are  $C$  and  $\gamma$  and the search space is reduced.

TABLE I  
MAPE USING DIFFERENT DATA PREPARATION

(seg., TS style)	W/out T	W/ avg. T	W/ 3c T	W/ real T
(Winter, yes)	1.95%	3.14%	2.71%	2.70%
(Jan-Feb, yes)	2.54%	3.21%	2.93%	2.96%
(Winter, no)	2.86%	3.48%	2.97%	3.10%

As searching for the proper parameters, we need to access the performance of models. To do this, usually we divide the training data into two sets. One of them is used to train a model while the other, called the validation set, is used for evaluating the model. According to their performance on the validation set, we try to infer the proper values of  $C$  and  $\gamma$ . Here, due to the different characteristics of the data encoding schemes, we employ two procedures for the validation.

For time-series-based approaches, we respectively extract the data entries of January 1997 and 1998 to form the validation set and evaluate the models on them. The performance is decided by averaging the errors of these two validations. As for the non-time-series models, we simply conduct tenfold cross validation to infer the parameters. That is, we randomly divide the training set into ten sets. Using each set as a validation set, we then train a model on the rest. The performance of a model would be the average of the ten validating predictions.

With this procedure, proper  $C$  and  $\gamma$  are selected to build a model for the future prediction. We then evaluate its performance by forecasting the load demand for January 1999. To estimate the efficiency of model training, we also measure the time consumed to build a model. On a machine with Windows 2000 Professional, CPU 598 MHz, and 256-MB RAM, the time required to train a time-series model ranges from 0.3 to 0.5 s for winter data set and 0.02 s for January-February data. For a non-time-series data set, it takes about 0.15 s to build a model.

Now we are ready to present experimental results. Table I contains the prediction errors generated by different data encodings and segmentations. In the table, the first column shows the data segments used and if the past load demand is encoded. Then the next four columns indicate the predictions with or without the temperature (T): “avg. T” for average temperature, “3c T” for the estimation derived from the three other cities’ data, and “real T” for the real temperature of January 1999.

1) *Time-Series Models With Winter Data*: In Table I, it can be observed that using the “winter” data along with the past loads (time-series information), the model (with  $C = 2^{12}$ ,  $\gamma = 2^{-4}$ ) built without temperature outperforms all others. In fact, the winning entry in the competition is generated by such a model<sup>3</sup>. The MAPE of using the temperature of three other cities is smaller than that using the average temperature. Moreover, after the competition is closed, we get the real temperature data of January 1999. We employ the temperature-incorporated models on data sets with the real temperature. The result is shown in the last column of Table I. Its performance, compared with those two using different temperature estimations, is not better. That means, even assuming

<sup>3</sup>During the competition, our validation procedure is different from what we employ in this paper. Thus, the actual parameters that generate the winning entry is not the same as those we present here. However, the modeling scheme is the same.

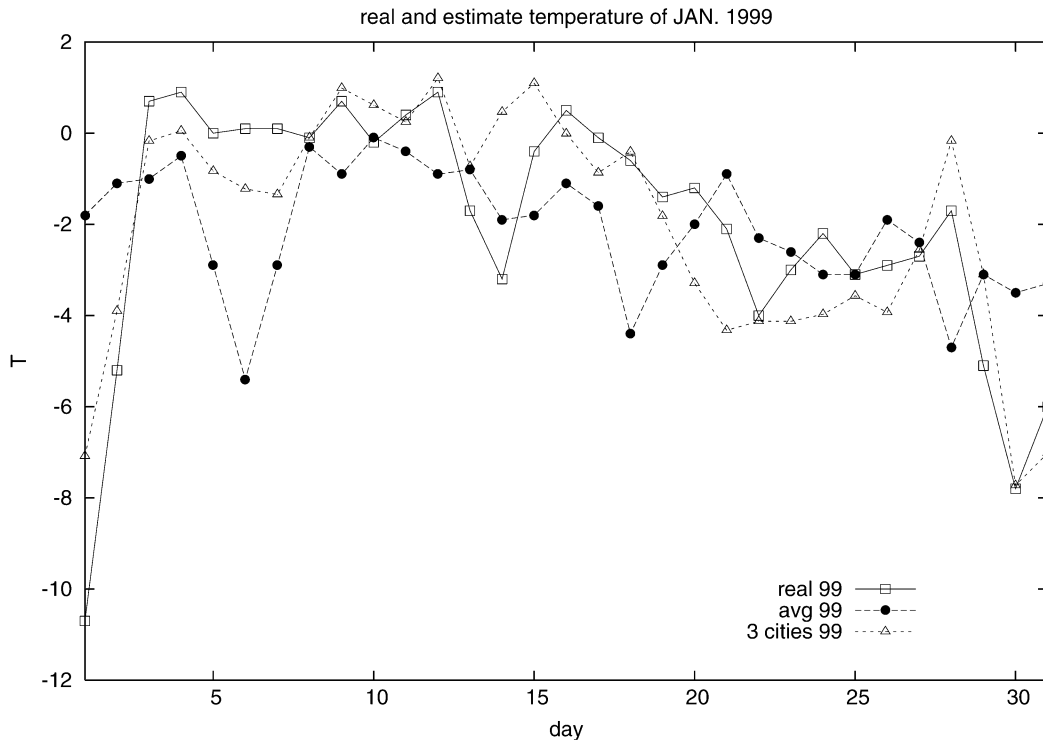


Fig. 5. Estimates and real temperature in Jan. 1999.

the real temperature is known, the forecasting result is still not satisfactory.

2) *Time-Series Models With January-February Data:* With the training set containing only data of January and February, the model built without temperature also performs better than all others built with temperature. This is the same as using the “winter” data segment.

3) *Non-Time-Series Models With Winter Data:* Besides time-series-based approaches, we build models without taking the past load demand as attributes. The training set of these models is limited to the “winter” segment. In Table I, we show the test errors of the forecast with or without temperature information. Just like the aforementioned experimental results, the MAPE of prediction on the data set without temperature is better than that with. Actually as one can see in Fig. 8, the forecast generated by the model using only calendar attributes is the same weekly. That is because besides the calendar attributes, the data set does not provide any other information for the model.

4) *Remarks:* From the results, an interesting observation can be made: models built without the temperature generally perform better than those built with. This renders the issue about the usage of temperature information. As we mentioned earlier, temperature information is important for load forecasting. However, models constructed with such information may be really sensitive to the temperature and thus the estimation for the temperature in 1999 would surely affect the performance of the models.

In Fig. 5, the real temperature data and the two estimations are shown. Figs. 6–8 plot the predicted values as well as the real load demand of January 1999. Roughly the higher the temperature is, the lower the demand is. Interestingly, the model using real temperature cannot predict the load demand more precisely

than using the estimation coming from the data of the three other cities. This result somehow implies the difficulty of incorporation of the temperature. To further inspect the influence of temperature information, we also conduct some simple experiments with another part of the data set. Targeting at predicting the loads in August 1998, we use another period of the data to build two different time-series models, one with temperature and the other without. The resulting MAPEs are both 2.47%. Note that for the temperature-included model, we already use the real temperature of August 1998, which, again, is not available for practical use. This confirms the futility of temperature information in such a mid-term load forecasting problem. In fact, as we compute the correlation coefficient between the maximum load and the temperature for each month instead of the whole two year, we find they are variant (ranging from  $-0.64$  to  $0.32$ ). This also indicates the fuzzy correlation between the load and the temperature in a shorter period.

Models built on data segments of January and February perform not as well as those built on the “winter” segments. We think the main reason is that, with the limited data given (only two years), such data segments cannot provide enough information for models compared with the “winter” segments that contains more entries. Therefore, though data set containing data only from January and February may represent the period of the prediction better, due to the limited data, models built on such segments are somehow not as competitive as those built with more information.

For comparison, models without time-series information are also built. The performance of such models is less competitive. This shows that models built without the past loads may not be able to learn the tendency of the load demand. Moreover, if the temperature is used to build the model, it would be the main attribute that affects the predicted values. Due to the limited data

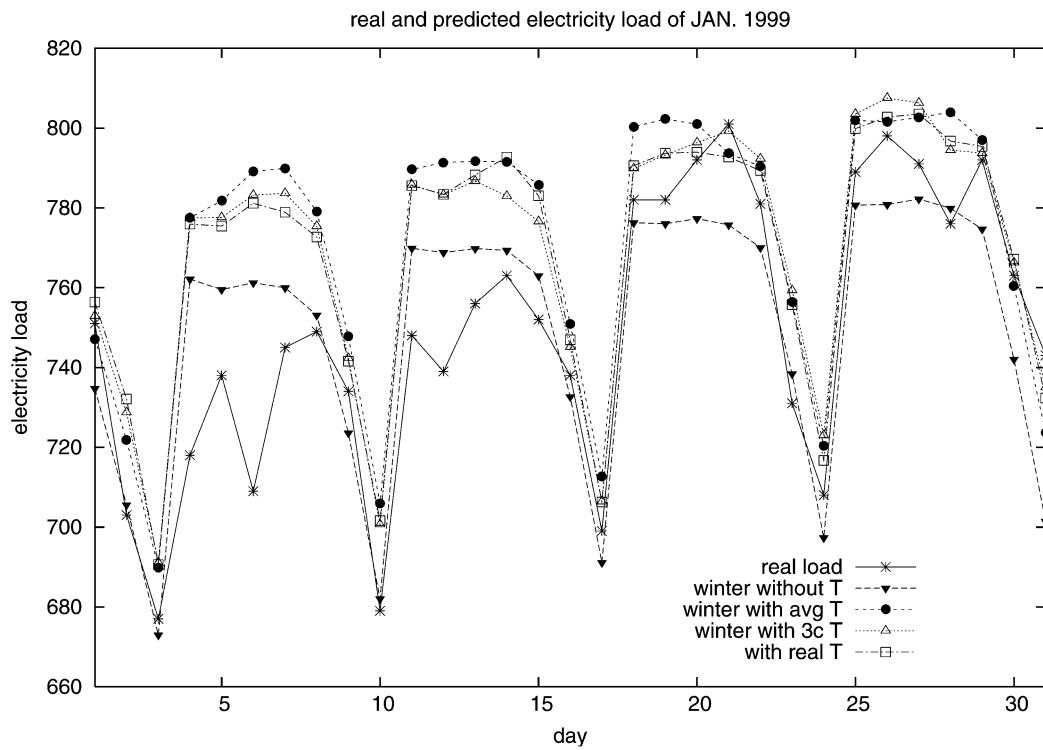


Fig. 6. Estimates and real electricity load in January 1999, with “winter” data.

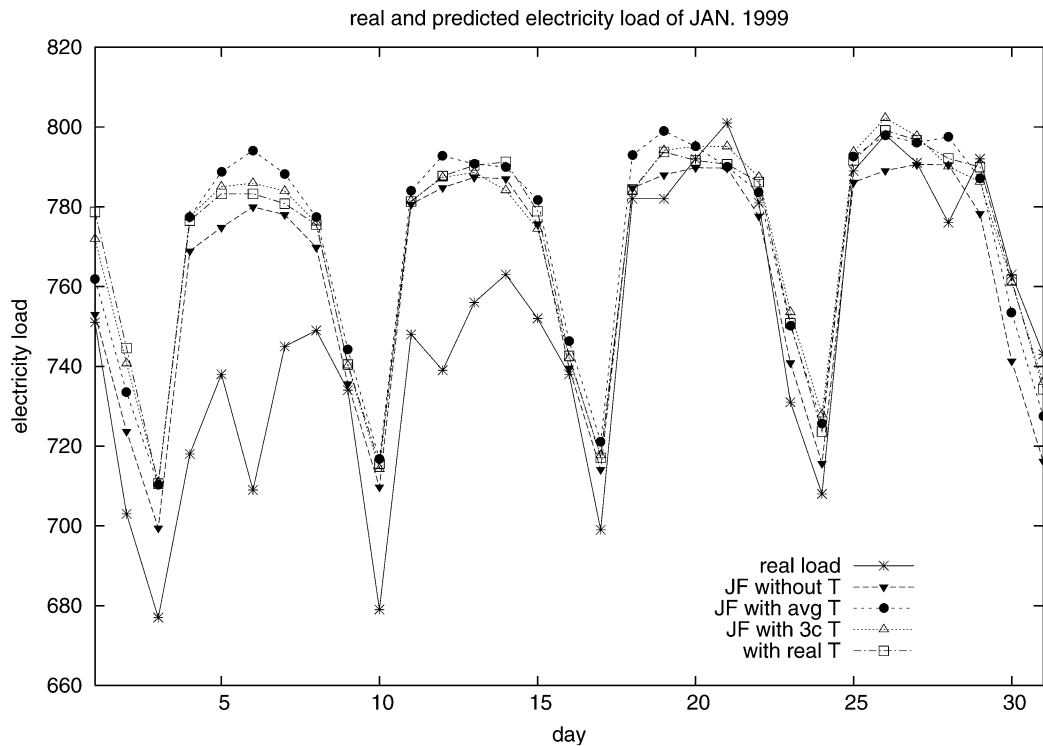


Fig. 7. Estimates and real electricity load in January 1999, with January-February data.

and the fuzzy correlation between the temperature and the load demand, such incorporation of temperature information could introduce higher variance into the model and result in unreliable prediction. That is why even using the real temperature, the error is higher than that of using only calendar information.

Recall that the temperature of the last day in 1998 is the lowest of the whole data set, but the load is not extremely high. More-

over, the correlation coefficient between the load and the temperature in December 1998 is 0.092. All these could make the forecast difficult for temperature-related models; however, with the time-series information, the models could at least catch the load tendency, and thus give reasonable predictions following the trend of December 1998. That is why here the time-series-based models perform better than nontime-series-based ones.

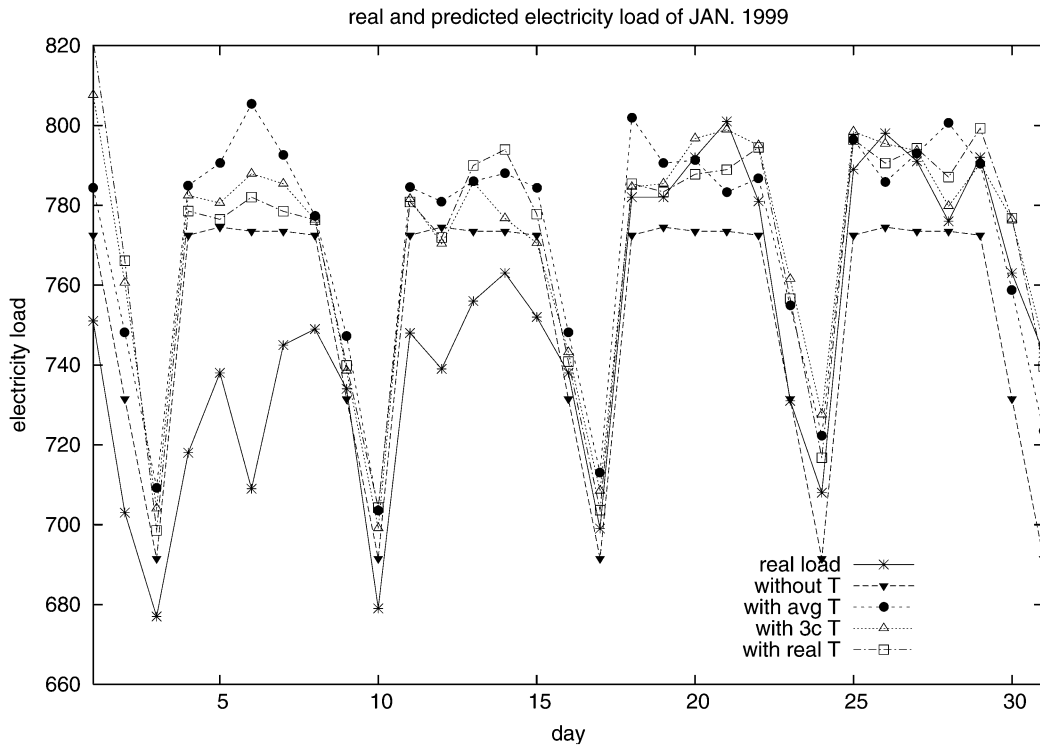


Fig. 8. Estimates and real electricity load in January 1999, with nontime-series model.

## V. DISCUSSIONS AND CONCLUSIONS

### A. Comparisons

1) *With Local Models:* Besides SVM, during the competition we also tried another method, local models. Local modeling is a simple analysis scheme for time-series data. With stationary time-series data, local models can also have good performance [16]. They generate predictions by finding segments of the time series that closely resemble the segment of the points immediately preceding the point to be predicted. Then the prediction is usually the average of elements that occurred immediately after these similar segments of points.

Practically we have to decide the length of each segment and the number of similar segments. Here we simply take loads of seven days as a segment and select the closest segment. That is, if we would like to predict the load  $L_t$  on date  $t$ , we compute the distances between segment  $(L_{t-7}, L_{t-6}, \dots, L_{t-1})$  and other segments  $(L_{k-7}, L_{k-6}, \dots, L_{k-1})$ , for  $k = 8, \dots, t-1$ . After finding the closest segment with  $\hat{k}$ , we use the load  $L_{\hat{k}}$  as the prediction. In addition, we take the information of weekdays into consideration since there is a weekly pattern in the load demand. That means we choose the closest segment with  $\hat{k}$  so that the weekdays of date  $\hat{k}$  and  $t$  are the same. Then, we take the load on date  $\hat{k}$ ,  $L_{\hat{k}}$ , as the predicted value.

Using the local-model scheme described above, we tried to forecast the load in January 1999. However, the result is not promising. The MAPE of the predictions is 8.81%. Actually, some competitor [17] also used local models to do the forecast.

2) *With Neural Network Models:* Among competitors, neural network models are quite popularly employed. Here we also employ the technique to conduct comparisons.

The implementation of neural network we choose to work with is *nnet*, which is wrapped in the package *VR* [18] for the statistical computing environment *R* [19]. There are also several tunable parameters while training a neural network, such as *decay rate for weight*, *number of units in the hidden layer*, and etc. Hence in our experiments, we try to tune these parameters along with the two data sets of winter, with or without temperature. The determination for the values of parameters is similar to what we used for SVM models. Different values for parameters are used to build models and then validate on data sets of Januaries in 1997 and 1998. With the chosen parameters we finally train models and use them to forecast loads in January 1999.

The models with tuned parameters, however, do not perform well on the problem. Using temperature information, the built model predicts the loads of January 1999 with MAPE reaching to 5.73%, while the model without temperature produces results with 7.8% for the error metric. Generally, the neural networks we tune here do not perform well. There is one extreme case that the tuned model generates *a single value* for all the prediction with a MAPE of 3.79%. That single value, in a way, is a kind of average the model predicts for this problem. Considering the difficulties of parameter selection and insufficient proficiency in neural networks, we then decide to simply use the default values to train our models. Surprisingly, the results generated by these models are better. The MAPE is 3.63% for the data set with temperature and 4.94% for that without.

Overall, the results produced by neural networks are not satisfactory at all. However, these results indicate more about the incapability of the neural networks due to the difficulty of choosing parameters.



3) *With Other Competitors:* In the competition, many different methods and models are proposed to solve the problem. Reports submitted by all competitors are published in [20].

First, the data segments used for training are different among the competitors. In our method, we select the winter segment as our training base. Some competitors like [12], and [21]–[23], also select particular segments rather than the whole two-year data. [12] even uses only January and February as such a choice could focus on the target period. Comparing all these competition entries, we find the selection of data segments may improve performance of models, since the load demand appears to have different distribution in different seasons.

As for choosing the features for models, the causal relation between temperature and load demand is most used. Many competitors use the temperature, along with some other information, to build causal models while some also incorporate with time-series analysis (e.g., [22], [24] and [17]). The techniques competitors used for causal models include neural networks ([21], [13], [25] and [14]), adaptive logic network [12], fuzzy rule-based models [26], self-organizing maps combined with ARIMA models [27], clustering [28], etc. However, the use of causal relation between temperature and load demand, would involve the inference for the temperature of January 1999.

In fact, some competitors do try to predict or infer the weather condition of January 1999 ([29], [21] and [15]), yet, with the limited data, forecasting the temperature may even be a more difficult problem. Without good forecast for the temperature, the model might end up with imprecise prediction, even though it may perform well in the training phase.

Holiday is another issue. For most competitors, load forecasting on holidays is not easy to handle. The difficulty is that the historical load data on holidays are not enough, and thus make the behavior of load demand on these special days hard to model. In our work, we do include holiday information in the training phase. However, while predicting the load in 1999, we remove the holiday flags on January 1 and 6. Other competitors treat some holidays as Saturday or Sunday [21], [12], build another model for holidays [13], or manually adjust the predicted result [22].

## B. Conclusion

Among these different proposed solutions, we find that choosing appropriate data segments seems to enhance the model performance. Furthermore, models built with continuous estimation of the future temperature, may be inaccurate in the end. In other words, including unprecise information causes higher variance on prediction so a conservative approach using only available correct information is sometimes recommended for such mid-term load forecasting. We have also demonstrated a validation procedure to judge whether temperature should be incorporated or not. As the experimental results show, the introduction of the time-series attributes also gives models better information to forecast load demand more precisely. Actually, these considerations are what we were upon to propose the time-series based, winter-data only, without temperature information, model which generated the winning entry.

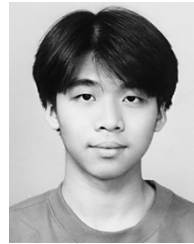
## ACKNOWLEDGMENT

The authors express their appreciation to EUNITE for having organized such an interesting competition which inspires them to conduct research on load forecasting.

## REFERENCES

- [1] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Trans. Power Syst.*, vol. 16, pp. 44–55, Feb. 2001.
- [2] P. Murto, "Neural network models for short-term load forecasting," Master's thesis, Dept. Eng. Phys. Math., Helsinki Univ. Technology, Helsinki, Finland, 1998.
- [3] K. Jabbour, J. F. V. Riveros, D. Landsbergen, and W. Meyer, "Alfa: Automated load forecasting assistant," *IEEE Trans. Power Syst.*, vol. 3, pp. 908–914, Aug. 1988.
- [4] C. Cortes and V. Vapnik, "Support-vector network," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.
- [5] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
- [6] 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>
- [7] S. Rahman and O. Hazim, "A generalized knowledge-based short-term load-forecasting technique," *IEEE Trans. Power Syst.*, vol. 8, pp. 508–514, May 1993.
- [8] 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.
- [9] N. Amjadi, "Short-term hourly load forecasting using time-series modeling with peak load estimation capability," *IEEE Trans. Power Syst.*, vol. 16, pp. 498–505, Aug. 2001.
- [10] K. Pawelzik, J. Kohlmorgen, and K.-R. Müller. (1996) Annealed competition of experts for a segmentation and classification of switching dynamics. Neural Computation. [Online]. Available: <http://citeseer.nj.nec.com/pawelzik96annealed.html>
- [11] M.-W. Chang, C.-J. Lin, and R. C. Weng, "Analysis of switching dynamics with competing support vector machines," in *Proc. IJCNN, 2002*, pp. 2387–2392.
- [12] D. Esp, "Adaptive logic networks for east slovakian electrical load forecasting," in *Electricity Load Forecast Using Intelligent Technologies*, P. Sinčák, J. Strackeljan, M. Kolcun, D. Novotný, and P. Szathmáry, Eds: EUNITE: The European Network on Intelligent Technologies for Smart Adaptive Systems, 2002, pp. 55–74.
- [13] E. Castillo, B. Guijarro, and A. Alonso, "Electricity load forecast using functional networks," in *Electricity Load Forecast Using Intelligent Technologies*, P. Sinčák, J. Strackeljan, M. Kolcun, D. Novotný, and P. Szathmáry, Eds: EUNITE: The European Network on Intelligent Technologies for Smart Adaptive Systems, 2002, pp. 75–84.
- [14] I. King and J. Tindle, "Storage of half hourly electric metering data and forecasting with artificial neural network technology," in *Electricity Load Forecast Using Intelligent Technologies*, P. Sinčák, J. Strackeljan, M. Kolcun, D. Novotný, and P. Szathmáry, Eds: EUNITE: The European Network on Intelligent Technologies for Smart Adaptive Systems, 2002, pp. 93–106.
- [15] W. Kowalczyk, "Averaging and data enrichment: Two approaches to electricity load forecasting," in *Electricity Load Forecast Using Intelligent Technologies*, P. Sinčák, J. Strackeljan, M. Kolcun, D. Novotný, and P. Szathmáry, Eds: EUNITE: The European Network on Intelligent Technologies for Smart Adaptive Systems, 2002, pp. 209–218.
- [16] J. McNames, J. A. K. Suykens, and J. Vandewalle, "Winning entry of the K. U. leuven time series prediction competition," *Int. J. Bifurc. Chaos*, vol. 9, no. 8, pp. 1485–1500, 1999.
- [17] G. Bontempi, "Eunite world-wide competition: Research report," in *Electricity Load Forecast Using Intelligent Technologies*, P. Sinčák, J. Strackeljan, M. Kolcun, D. Novotný, and P. Szathmáry, Eds: EUNITE: The European Network on Intelligent Technologies for Smart Adaptive Systems, 2002, pp. 33–40.
- [18] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, 4th ed. New York: Springer-Verlag, 2002.
- [19] R Development Core Team, R Foundation for Statistical Computing. (2003) R: A Language and Environment for Statistical Computing, Vienna, Austria. [Online]. Available: <http://www.R-project.org>

- [20] P. Sinčák, J. Strackeljan, M. Kolcun, D. Novotný, and P. Szathmáry, Eds., *Electricity Load Forecast Using Intelligent Technologies*: EUNITE: The European Network on Intelligent Technologies for Smart Adaptive Systems, 2002.
- [21] A. Lewandowski, F. Sandner, and P. Protzel, "Prediction of electricity load by modeling the temperature dependencies," in *Electricity Load Forecast Using Intelligent Technologies*, P. Sinčák, J. Strackeljan, M. Kolcun, D. Novotný, and P. Szathmáry, Eds: EUNITE: The European Network on Intelligent Technologies for Smart Adaptive Systems, 2002, pp. 107–114.
- [22] E. Pelikán, "Middle-term electrical load forecasting by time series decomposition," in *Electricity Load Forecast Using Intelligent Technologies*, P. Sinčák, J. Strackeljan, M. Kolcun, D. Novotný, and P. Szathmáry, Eds: EUNITE: The European Network on Intelligent Technologies for Smart Adaptive Systems, 2002, pp. 167–176.
- [23] W. Brockmann and S. Kuthe, "Different models to forecast electricity loads," in *Electricity Load Forecast Using Intelligent Technologies*, P. Sinčák, J. Strackeljan, M. Kolcun, D. Novotný, and P. Szathmáry, Eds: EUNITE: The European Network on Intelligent Technologies for Smart Adaptive Systems, 2002, pp. 41–54.
- [24] F. Rivieccio, "SVM for an electricity load forecast problem," in *Electricity Load Forecast Using Intelligent Technologies*, P. Sinčák, J. Strackeljan, M. Kolcun, D. Novotný, and P. Szathmáry, Eds: EUNITE: The European Network on Intelligent Technologies for Smart Adaptive Systems, 2002, pp. 177–186.
- [25] D. Živčák, "Electricity load forecasting using ANN," in *Electricity Load Forecast Using Intelligent Technologies*, P. Sinčák, J. Strackeljan, M. Kolcun, D. Novotný, and P. Szathmáry, Eds: EUNITE: The European Network on Intelligent Technologies for Smart Adaptive Systems, 2002, pp. 219–231.
- [26] R. C. Weizenegger, "Maximum electricity load problem," in *Electricity Load Forecast Using Intelligent Technologies*, P. Sinčák, J. Strackeljan, M. Kolcun, D. Novotný, and P. Szathmáry, Eds: EUNITE: The European Network on Intelligent Technologies for Smart Adaptive Systems, 2002, pp. 187–208.
- [27] G. A. Ivakhnenko, "Inductive self-organizing algorithm for maximum electrical load prediction," in *Electricity Load Forecast Using Intelligent Technologies*, P. Sinčák, J. Strackeljan, M. Kolcun, D. Novotný, and P. Szathmáry, Eds: EUNITE: The European Network on Intelligent Technologies for Smart Adaptive Systems, 2002, pp. 85–92.
- [28] Z. Boger. (2001) Electricity Load Forecast Using Artificial Neural Networks Clustering. [Online]. Available: <http://neuron.tuke.sk/competition/>
- [29] A. Lotfi, "Application of learning fuzzy inference systems in electricity load forecast," in *Electricity Load Forecast Using Intelligent Technologies*, P. Sinčák, J. Strackeljan, M. Kolcun, D. Novotný, and P. Szathmáry, Eds: EUNITE: The European Network on Intelligent Technologies for Smart Adaptive Systems, 2002, pp. 123–130.



**Bo-Juen Chen** received the B.S. degree in computer science and information engineering in 2001 and the M.S. degree in computer science and information engineering in 2003, both from National Taiwan University, Taipei, Taiwan, R.O.C.

His research interests include data mining, machine learning, and related applications.



**Ming-Wei Chang** received the B.S. degree in 2001 and the M.S. degree in 2003, both in computer science and information engineering, from National Taiwan University, Taipei, Taiwan, R.O.C..

His research interests include machine learning, numerical optimization, and algorithm analysis.



**Chih-Jen Lin** (S'91–M'98) received the B.S. degree in mathematics from National Taiwan University, Taipei, Taiwan, R.O.C., in 1993 and the M.S. and Ph.D. degrees from the Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, in 1998.

He is currently an Associate Professor in the Department of Computer Science and Information Engineering, National Taiwan University. His research interests include machine learning, numerical optimization, and applications of operations research.