공학 연구논문집 제29권 2호 pp. 87~99, 울산대학교.
Journal of Engineering Research Vol. 29, No. 2, pp. 87~99, 1998
University of Ulsan.

Daily Load Forecasting Using the Self-Organizing Map

Kab-Ju Hwang · Myung-Kook Yang · Sung-Woo Cho
Dept. of Electrical Engineering and Autamation.

<Abstract>

In this paper, a daily load forecasting algorithm using the self-organizing map(SOM) method is proposed and examined. SOM is a new, powerful software tool for the visualization of high-dimensional data. It requires less training time compared to other networks such as BP learning network, and moreover, its self organizing feature can amend the distorted data. The proposed algorithm analyzes the load patterns of the past couple of years and estimates future load demand by mapping the target day using SOM. KEPCO's hourly load record obtained between 1993 and 1995 is examined to investigate the efficiency of the proposed method. It is shown that the proposed algorithm provides better forecasting results than conventional exponential smoothing method.

1. INTRODUCTION

The accurate daily load forecasting is an important factor to ensure the economic and secure operation of power system. This provides reference data for basic operations functions such as thermal unit commitment, hydro-thermal coordination and security assessment.

Various load forecasting models are developed and examined to improve the accuracy of the prediction. A short-term load forecasting estimates the variation of load demand for a day or a week period. The load pattern of past days and the weather condition are major factors to determine the load demand of the next day. The time series forecasting model and the multiple regression model are well known as the daily load forecasting models.[1-2] The time series forecasting model starts from the temporal locality concept. It treats the load pattern as a time series signal and predicts the future load by using various time series analysis techniques. The multiple regression model requires several input parameters, such as weather condition, holiday distribution, etc. It estimates the future load demand by analyzing those various input parameters.

Statistic method is well known technique that has been used by many electric power companies to estimate the practical short-term load demand. Artificial intelligence(AI) technologies, such as artificial neural network(ANN) and fuzzy logic, are suitable to treat the non-linearity and uncertainty of the forecasting problems.[3-6] The ANN belongs to a class of data-driven approaches, as opposed to model-driven approaches. The ANN analyzes the available data and adapts itself to create required output. Back Propagation(BP) is one of the typical example of ANN. Multilayered BP neural network, however, not only consumes too much time for adaptation, but requests a output for every input data. It also includes the difficulty of partial data learning. On the other hand, the single layered self-organizing map is simple structured ANN, and has self-organize feature that can amend the distorted data. T. Baumann and A. J. Germond[7] presented a 2-step load forecasting method based on the Kohonen's self-organizing feature maps and the single layered linear delta-rule ANN.

In this paper, a practical daily load forecasting algorithm using the self-organizing map(SOM) method is proposed and examined. The SOM considers the past daily-load pattern and weather information as input parameters, and extracts the daily load pattern. It then estimates the next day's load demand using its own cognizance. The optimal size of ANN for KEPCO's daily load forecast is investigated via various simulations. KEPCO's hourly load record obtained between 1993 and 1995 is examined to verify the efficiency of the proposed technique. It is shown that the proposed algorithm provides stable and more accurate daily load estimation than the conventional smoothing method.

2. SELF-ORGANIZING MAP

In 1982, Teuvo Kohonen proposed a new neural network where the Self-organizing map(SOM) is implemented. It has a self-modeling characteristic that analyzes the arbitrary input data and generates the classified data groups.[8] SOM analyzes input data without any external guide and constructs the data cluster automatically according to the characteristic similarity of data.

Fig. 1 depicts the basic architecture of SOM. SOM consists of an input layer with n input nodes followed by a layer with [n×m] neurons, arranged in a two dimensional lattice. Each input is connected through a weight Wij to all neurons of the second layer.

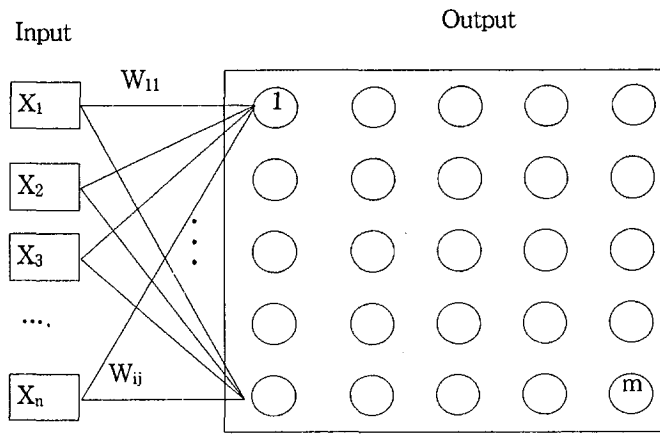


Fig. 1. Basic architecture of the SOM

For most ANN, the scalar product between input X and weight W result in output Y. Unlike other ANN, the basic operation of the SOM produces the Euclidian distance D_E which is calculated as

$$Y = D_E = \sum_{i=0}^{n-1} (X_{ij}(t) - W_{ij}(t))^2 \qquad (1)$$

where n is the number of input components.

An input vector X is presented to the map and the Euclidian distance between this input vector and all the weight vectors W_{ij} is calculated. Thus each neuron has an Euclidian distance. The neuron with the smallest distance is called winner. The weight vector of winner is consequently the nearest to the input vector X. Once the winner is determined, the unsupervised learning is processed by updating the weights of neurons located around winner CE. Fig. 2. depicts the selected area of neighborhood around the winner neuron.

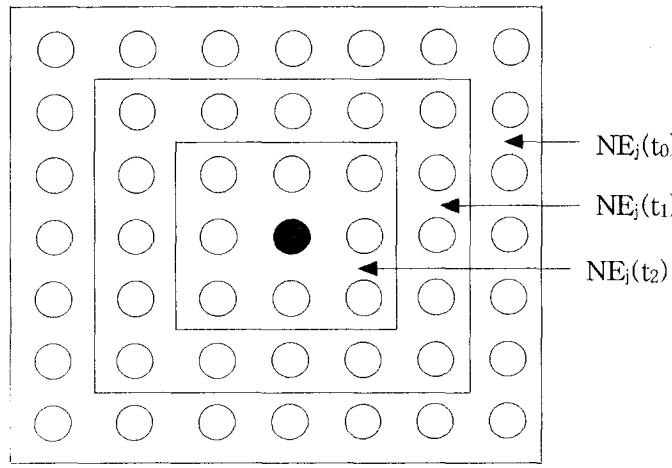


Fig. 2. Neighborhood around the winner neuron

The training of SOM is done by a continuous modification of the weights W_{ij}. This needs a repeated presentation of input vectors until the SOM in organized.

In the beginning of the learning process, the radius of neighborhood is initialized by big value. It then decreases as the learning step goes on. If the radius is zero, only the weight of winner neuron is updated. Update calculation is

$$W(t+1) = W(t) + \alpha(t)[X(t) - W(t)] \tag{2}$$

where $\alpha(t)$ is the learning gain. The range of $\alpha(t)$ that decreases as a function of the training iteration step is distributed between 0 and 1.

The choice of learning gain $\alpha(t)$ and the radius of neighborhood NE are dominant for a topologically correct organization of a SOM. The initial learning gain $\alpha(t_0)$ is adjusted between 0.2 and 0.5 and the radius of the neighborhood is initialized by half the size of the maximal dimension of the SOM.

The updating rate of both parameters is related to the tuning range. Less updating rate leads to better organization but needs more learning time.[7]

3. DAILY LOAD FORECASTING USING SELF-ORGANIZING MAP

3.1 Input data

A daily load pattern is changing periodically according to the following factors: a day of week, holiday distribution, and seasonal variation. The weather is another major source that affects the daily load pattern. Temperature, humidity, and intensity of

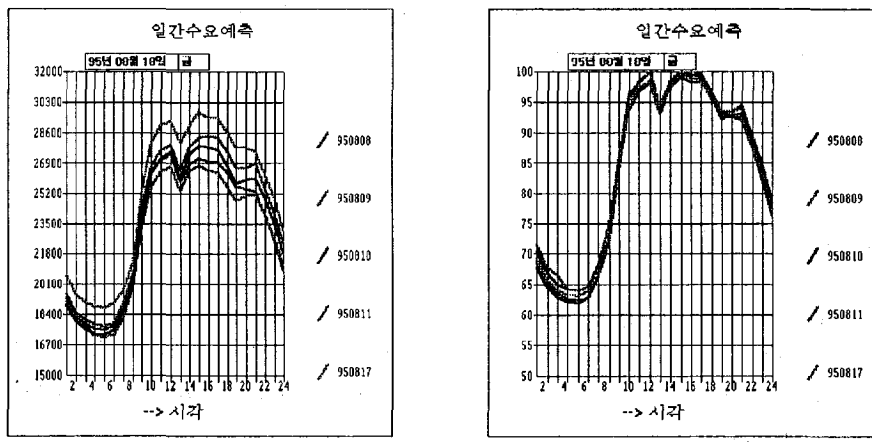
illumination are the main indexes of weather condition that can change the load demand. Since the temperature is the most important information among those weather variables, it is used most commonly in the load forecast.

In this paper, KEPCO's real load records obtained between 1993 and 1995, temperature, a day of week, special holiday distribution are used as an input data. Five major cities(Seoul, Pusan, Daejun, Daegu, Kwangju) that consume large portion of KEPCO electric power are selected as the sampling locations to measure the temperature.

The real load data are normalized and translated to the load demand as per unit value, Pu, before the neural network starts training. The per unit load demand is calculated as

$$P_U = \frac{P(d,t)}{P_{\max}(d)} \tag{3}$$

where $P(d,t)$ is the real power demand at hour t on day d, and $P_{\max}(d)$ is maximum power demand on day d. Fig. 3. presents the comparison of actual load pattern and normalized load pattern.



(a) Actual pattern (b) Normalized pattern

Fig. 3. Actual load pattern vs. normalized load pattern

3.2 Classification of Load Patterns

An appropriate classification of load pattern is an important issue to improve the accuracy of forecasting. Each output node of the SOM represents a group of classification strategy. Thus as the number of output nodes increases, the SOM provides a fine grain classification strategy that requires complicate analysis on input data. To figure out the optimal size of the output layer for KEPCO's load forecasting, various network sizes from

[3×3] to [17×17] are simulated. It has been shown that the small network, from [3×3] to [6×6], provides poor accuracy of load forecasting due to its coarse grain input analysis. Minimum network size for KEPCO load forecast is investigated as [7×7] that can classify the daily load pattern effectively. Table 1 presents the distribution of the daily load pattern assigned to each output node of [7×7] network. A typical hourly load patterns obtained using [7×7] network for special day, Sunday, Monday, Saturday, and other week days are depicted on Fig. 4. As the network size increases, the error range of the load estimation is getting smaller and smaller. The accuracy improvement, however, is saturated for greater than [10×10] network, i.e. the excessively fine grained input analysis does not always help to provide better load forecasts. In this paper, a [10×10] network is used to predict KEPCO's daily load forecast.

Day groups and weather are major factors that affect the load pattern. Day groups are classified into five categories: special day, Sunday, Monday, Saturday, and other week days. All day groups except special day group have their own codes. Special day load patterns are somewhat ambiguous. It varies according to the characteristic of the holiday. Thus the different day code assignment is required for each special day, such as 1225 for Christmas day, and 0815 for Thanksgiving day, etc. Moreover the accuracy of special day forecast is limited since the past load data of the specific holiday are not sufficient enough to provide a load pattern.

KEPCO's past data demonstrate that the load patterns of summer season and other seasons are quite different. This causes to classify the seasons into two groups: summer group, and the other three seasons group.

3.3 Learning of SOM

[10×10] output nodes analyze and categorize the input data according to the load pattern. Detail input data provided by fifty five input nodes are presented in Table 2.

In the beginning of the learning, all the output nodes are initialized as learning gain α= 0.5 and Euclidean distance D_E=7.

Table 1. Typical load patterns assigned to output nodes

		Mn	Mn	Mn		St
Od,Mn	Od,Mn		Od,Mn	Mn	St	St
Od,Mn	Od,Mn		Od	Od,St	St	St
Od	Od,Mn	Od	Od			Sn
Od	Od	Od	Od	Od	Sp	Sp
Od	Od	Od		Sn	Sn	Sn
Od	Od	Od	Sn,Sp	Sn,Sp	Sn,Sp	Sn,Sp

Od : Ordinary Sn : Sunday Mn : Monday
St : Saturday Sp : Special day

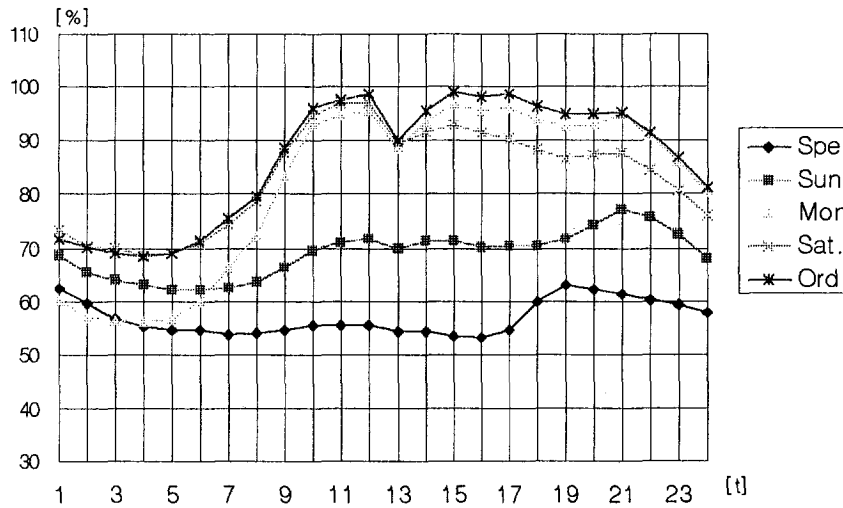


Fig. 4. Typical daily load patterns

Table 2. Definition of the inputs for learning
(1) Ordinary Day (Tue. - Fri.)

Inputs	**Description**
1 ~ 24	$\{(P(d\text{-}1,t) - P(d\text{-}2,t))/100,\ t = 1,24\}$
25	Weekday code
26	Special Day code
27 ~ 50	$\{Pu(d\text{-}1,t),\ t = 1,\ 24\}$
51 ~ 55	$\{Tmax(d,j),\ j = 1,\ 5\}$
Outputs	**Description**
1 ~ 24	$\{(F(d\text{-}1,t) - F(d\text{-}2,t))/100,\ t = 1,24\}$

(2)Week end (Sum., Mon., Sat.)

Inputs	Description
1 ~ 24	$\{Pu(d,h),\ t = 1,24\}$
25	Weekday code
26	Special day code
27 ~ 41	$\{Tmax(d,j), Tmin(d,j),\ Lave(d,j),\ j = 1,5\}$
42 ~ 50	NA
51 ~ 55	$\{Tmax(d,j) - Tmin(d,j),\ j = 1,5\}$
Outputs	**Description**
1 ~ 24	$\{Fu(d,h),\ t = 1,24\}$

j : weather station index(5 cities)
F : MW load forecast Fu : PU load forecast
$T_{max}(T_{min})$: maximum(minimum) temperature
L_{ave} : daily average luminous intensity

These initialization allows the same chance of training to every output nodes. As learning process goes on, the output nodes associated with similar load pattern reduce the corresponding Euclidean distances and eventually accomplish the grouping. Fig. 5 represents the flow of training process of the SOM.

3.4 Forecast using Association

As the winner neuron is determined via training process, the input data patterns classified by the SOM are stored in the weighted connection. Once a training cycle is completed, the SOM starts using its own *associative effect* that can amend distorted data. The proposed daily load forecasting algorithm employs this *associative effect* of the SOM, i.e. the hourly load estimation is associated using the pre-known information of the target day. The associated information is a normalized load pattern by per unit domain. Thus the data conversion is required to obtain the actual load prediction. The real load forecast, $F(d,t)$, of the ordinary day and weekend is calculated respectively as

$$F(d,t) = F(d-1,t) + [P(d-1,t)-P(d-2,t)], \text{ for ordinary day} \quad (4)$$

$$F(d,t) = Pmax(d-1) \times RC(t) \times Pu(d,t), \text{ for weekend} \quad (5)$$

where F = estimated load[MW]
d = day index,
t = hour of day index
P = actual load[MW]
Pu = per unit load[PU]
$RC(t)$ = relative coefficient

The relative coefficient $RC(t)$ is defined to generate the stable prediction of weekend load. $RC(t)$ is given as

$$RC(t) = \frac{P(d,t) \quad \text{on foecasting weekend}}{P(d-1,t) \quad \text{on previous ordinary day}} \tag{6}$$

4. CASE STUDIES

KEPCO's real load records between 1990 and 1995 are simulated to examine the proposed algorithm. The SOM takes load records of past three years as input data for SOM adaptation and estimates the load demand of target date. For example, the load demand recorded between 1990 and 1992 is analyzed to predict the load of January 1993. Both the load record analyses and the future load forecasts are figured day by day and hour by hour. The forecast values are compared with the real load demand to evaluate the error rate of the proposed technique. Daily absolute average error rate is known as a main criteria to evaluate the performance of the forecasting system. The absolute average error rate is defined as

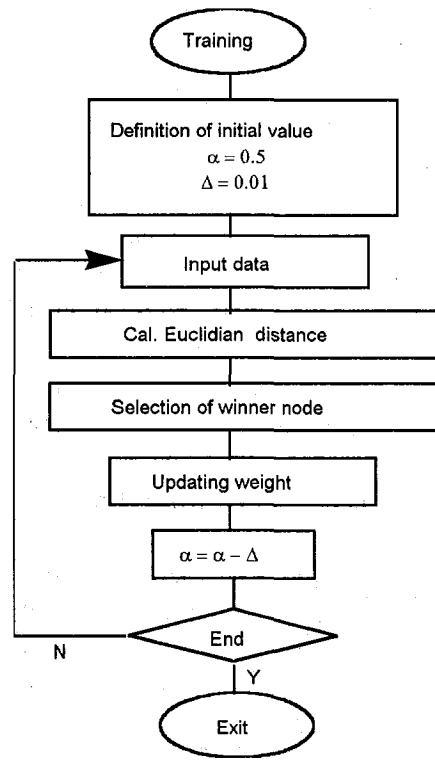


Fig. 5. Flow of training process of the SOM

$$\text{error rate} = \frac{|\text{measured value} - \text{forecast value}|}{\text{measure value}} \times 100[\%] \quad (7)$$

The average error rates of the summer season(from June to September) forecast for each day group are represented in Table 3. Table 4 shows the average error rates of fall, winter and spring seasons. It is shown that the error rate of the summer season forecast is bigger than the other season forecast. This phenomenon can be explained that the load pattern of the summer season is susceptible to the weather condition, i.e. the load demand varies in wide range according to the temperature in the summer season. Table 5 enumerates the monthly forecasting error.

Fig. 6 and Fig. 7 depict the error rate comparison of the proposed forecasting technique with the conventional exponential smoothing technique. KEPCO's historical data obtained between 1990 and 1992 are given to estimate the load demand of 1993. The forecast data are then compared to the real load data recorded in 1993 by KEPCO. It is investigated that the proposed method provides stable load prediction with much less error rate than the conventional exponential smoothing method. The average error rates of the ordinary day load forecast using proposed algorithm are 1.9% for summer season and 1.3% for the other seasons. The exponential smoothing method has been widely used for practical load forecast in electric power company. It estimates the load demand by simply smoothing(smoothing coefficient = 0.7) the past five load data which have the similar input patterns. The main drawback of the exponential smoothing method is that the weather sensitive load variation factor is not considered in the prediction. Unlike the exponential smoothing method, the SOM not only counts in the weather sensitive load variation factor but also utilizes its own associate function. These allow the SOM possible to provide more stable and accurate load forecast.

The error rates of the special day forecasts are distributed on a wide range. In the case of special day, the input data are not sufficient enough to figure out the load pattern using SOM. Only one load datum is available for a specific special day from the one year load record. To improve the special day load forecast, one can either provide more special day load data or apply other forecasting approaches.

Table 3. Forecasting average error[%] of summer season

year / pattern	1993	1994	1995	Average
Ordinary Day	1.6	1.6	1.9	1.7
Sunday	1.5	2.6	2.5	2.2
Monday	1.4	2.1	1.9	1.8
Sturday	1.5	1.6	2.3	1.8
Average	1.5	2.0	2.1	1.9

Table 4. Forecasting average error[%] of ordinary season

pattern \ year	1993	1994	1995	Average
Ordinary Day	1.0	1.1	1.0	1.0
Sunday	1.5	1.4	1.7	1.5
Monday	1.3	1.4	1.6	1.4
Sturday	1.4	1.1	1.4	1.3
Average	1.3	1.3	1.4	1.3

Table 5. Forecasting average error[%] of monthly base

	Monday			Ord. days			Saturday			Sunday		
mo	93	94	95	93	94	95	93	94	95	93	94	95
1	1.6	1.3	1.4	1.1	0.8	0.7	1.5	0.9	1.0	2.2	0.9	1.5
2	1.2	1.1	1.6	0.9	1.0	0.9	1.1	1.0	0.8	1.7	0.9	1.4
3	0.9	0.9	1.4	1.0	1.1	1.0	1.1	1.0	1.5	1.2	1.2	1.5
4	0.8	1.6	1.6	1.1	1.0	1.1	1.1	0.9	1.1	1.2	1.2	1.5
5	1.3	1.4	1.4	1.0	1.2	1.1	3.0	2.5	1.7	2.4	1.8	1.9
6	1.3	1.9	1.0	1.4	1.3	1.1	0.7	1.3	1.8	1.4	1.3	1.8
7	1.5	2.7	2.5	1.4	1.8	1.8	2.1	2.2	2.4	1.4	2.6	2.0
8	2.2	2.3	2.2	2.1	1.8	3.0	1.7	1.2	2.7	2.0	2.5	3.0
9	1.2	1.6	1.6	1.0	1.6	2.0	2.0	1.2	1.2	1.0	2.6	2.5
10	1.5	1.2	1.6	1.3	1.1	0.8	1.3	0.9	2.3	1.0	2.3	2.1
11	1.4	1.2	1.5	1.0	0.9	0.9	0.8	0.6	1.7	1.8	1.7	1.9
12	1.4	1.6	1.6	1.0	0.9	0.9	1.1	0.5	1.3	1.1	1.3	1.6
Ave.	1.4	1.6	1.6	1.2	1.2	1.3	1.4	1.2	1.6	1.5	1.6	1.9

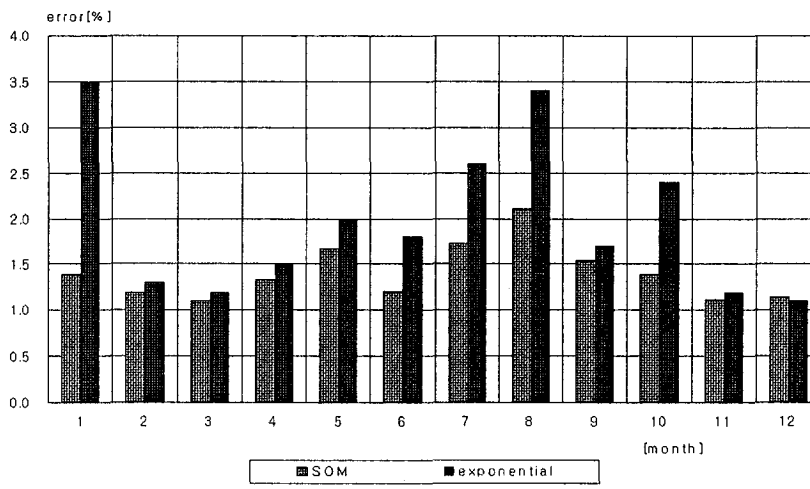


Fig. 6. Average error rate of monthly forecasting in 1993
SOM method vs. Exponential smoothing method

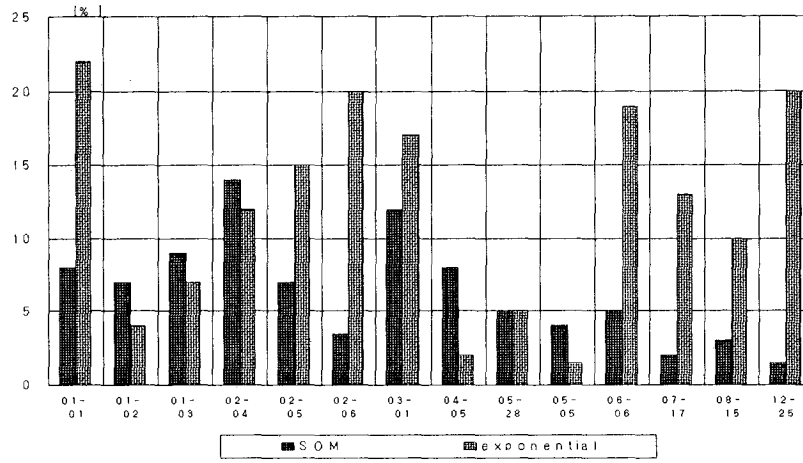


Fig. 7. Average error rate for special day forecasting in 1993
SOM method vs. Exponential smoothing method

5. CONCLUSIONS

A Daily load forecasting algorithm using the self-organizing map(SOM) is proposed and examined. The algorithm analyzes the load patterns of the past few years using SOM and estimates the future day's load by merging the former load pattern and the expected information of the target day. Various network sizes are simulated to find the optimal size of SOM for KEPCO's daily load forecast. It has been investigated that the [10×10] network provides satisfactory forecasting accuracy with reasonable network cost. The SOM analyzes the KEPCO's hourly load data recorded between Jan. 1990 - Dec. 1992, and Dec. 1992 - Nov. 1995 and predicts the load demands for Jan. 1993, and Dec. 1995, respectively. A day of the week(Sunday, Monday, Saturday, and other days), special holidays, and seasons (summer, and other seasons) are selected as classification categories of the input data. The simulation results show that the proposed algorithm provides stable load forecast for the ordinary day's load demand with limited error rate: 1.9% for summer season and 1.3% for fall, winter, and spring seasons.

REFERENCES

[1] IEEE Committee Report, "Load Forecasting Bibliography Phase I", IEEE Trans. on Power Apparatus and Systems, vol. PAS-99, pp. 53-58, 1980

[2] W.R.Christiaanse, "Short term Load Forecasting Using General Exponential Smoothing", IEEE Trans. on Power Apparatus and Systems, Vol. PAS-90, No.2, March/April 1972, pp. 900-910

[3] S.Rahman, R.Bhatnagar, "An Expert System Based Algorithm for Short Term Load Forecast," IEEE Trans. on Power Systems, vol.3, no.2, pp. 392-399, May 1988

[4] K.Y.Lee, Y.T.Cha, J.H.Park, "Short - Term Load Forecasting Using An Artificial Neural Network," IEEE Trans. on Power Systems, vol.7, No.1, pp.124-132, Feb. 1992

[5] K.H.Kim, J.K.Park, K.J Hwang, S.H.Kim, "Implementation of Hybrid Short-Term Load Forecasting System Using Artificial Neural Networks and Fuzzy Expert Systems", IEEE Trans. on power Systems, Vol. 10, No. 3, August 1995

[6] "Next day peak load forecasting using an artificial neural network", CRIEPY Report. No. Y92007, 1993

[7] T. Bauman, A. J. Germond, "Application of the Kohonen Network to Short-Term Load Forecasting", IEEE Trans. PWRS, pp. 407-412, 1993

[8] T. Kohonen, E. Oja, O. Simula, A. Visa, J. Kangas, "Engineering Application of the Self-Organizing Map", Proc. IEEE, Vol. 84, No. 10, pp. 1358-1384, Oct. 1996