International Journal of Electrical and Power Engineering 2 (6): 393-397, 2008
ISSN: 1990-7958

Performance Evaluation of Statistical and Artificial Neural Network based Short Term Load Forecasting Techniques

[1]Ajay Shekhar Pandey, [2]D. Singh, [1]S.K.Sinha and [2]S.P. Singh
[1]Department of Electrical Engineering, Kamla Nehru Institute of Technology, Sultanpur, U.P, India
[2]Department of Electrical Engineering, Institute of Technology, Banaras Hindu University, Varanasi, U.P, India

Abstract: The performance evaluation of various short term load forecasting approaches has been made in this paper. The methods are selected to reflect the different categories of load forecasting approaches. These methods are Multiple Linear Regression and Time Series in Statistical/ Conventional approach, Feed Forward Neural Network in supervised Artificial Neural Network approach, Radial Basis Function Neural Network in Unsupervised/Supervised category and Numerical Taxonomy method in Self Organizing category. The performance is studied on the historical data of a Canadian utility. It is observed that the neural network based methods are quite accurate as compared to conventional statistical methods. Statistical methods are accurate only when the load behaviour is less erratic. The self-organizing Numerical Taxonomy method shows the best results.

Key words: Short term load forecasting, radial basis function, multiple linear regression, self-organizing, numerical taxonomy, back propagation

INTRODUCTION

Load Forecasting is an important function performed by utilities for Planning Operation and Control. The factors affecting the consumer's load include meteorological, climatic intensity, pricing schemes, tariff structures and many other factors. Short term forecasts; in particular have become increasingly important since, the rise of the competitive energy market. Short Term Load Forecasting (STLF) is primarily used for economic load dispatch, daily operation and control, system security and assurance of reliable power supply. The complex dependence of load on human behavior makes load forecasting, a tedious job. The estimated forecasts in the time range are important inputs for generator scheduling functions, power system security assessment and power system dispatcher. The importance of STLF has attracted power engineers and researchers for a long time. Vast and intriguing methods of estimation have been applied in short term electric load forecasting in recent times. These approaches range from well-established conventional forecasting methods to newer sophisticated methods, like supervised and unsupervised learning based neural networks. The methods relying on optimization tool and self–organization using pattern classification and matching are also used. However, there is a need to study and compare these approaches to come out with definite conclusions on their relative advantage and disadvantages. Supervised learning models have capability to learn from examples with inputs and corresponding outputs, which are presented to the network during the training stage. The unsupervised form of learning utilizes the pattern matching and classification methods to partition the n-dimensional input stage into various categories. In fuzzy based method the fuzzy is basically a tool for inference process rather than a full-fledged forecasting method. Self-organization is a system model, which finds a set of relevant independent variables and model parameters. Some approaches focus towards application of Probabilistic Neural Network (PNN) for developing optimal self-organizing forecasting structure, automated input variable and PNN parameter selection, which is a novel concept.

The performance evaluation based on comparison of accuracy in various forecasting approaches, Multiple Linear Regression and Time Series Formulation in conventional forecasting, Feed Forward Neural Networks in supervised learning approach, Radial Basis Function Neural Networks in unsupervised/supervised learning category and Numerical Taxonomy in Self-organizing

Corresponding Author: Ajay Shekhar Pandey, Department of Electrical Engineering, Kamla Nehru Institute of Technology, Sultanpur, U.P, India

category has been made in this study. The selection of methods is made to include contrasting approaches of short term load forecasting. The performance is studied on the historical data of a Canadian utility. This data includes hourly loads in temperature. The time series formulation includes the time delay variables along with the casual variables. Attempts have been made to develop sophisticated mathematical techniques like Auto Regressive Moving Averages (ARMA) State Space, Spectral decomposition and linear and nonlinear regression procedures. The study gives a brief description of these methods and studies the data requirements in relation to each other.

LOAD FORECASTING VIA STATISTICAL METHODS

The widely referred reviews Gross and Galiana (1987) and Moghram and Rahman (1989) are aimed at most of the traditional methods. The methods include Multiple Linear Regression, Auto Regressive Models, Stochastic Time Series, General Exponential Smoothing and others. A detailed review of these methods can be seen in ref (Papalexopoulos and Hasterberg, 1990; Haung 1997; Hagen and Behr, 1987; Christiaanse, 1988). These methods have sound mathematical background but they are generally based on statistically assumed model, which relates load to human behaviour and different environmental factors. The accuracy of such methods depends on the accuracy and relevance of the heuristically modeled load. Most of these methods are computationally involved, as the complexity of the method and order of the model increases. It is also important to note that programs based on these methods may require an overhaul as the time progresses and the load behavior undergoes a change.

A statistical approach emphasizes on general causality and time series formulation. These methods are also known as conventional or traditional forecasting methods. A rich literature is available on STLF using traditional statistical formulation. Several reviews and committee reports (IEEE Committee report, 1981; 1981) have been published in the past. These reports and reviews have covered the general forecasting philosophies and their application to short term load forecasting. Most of the methods reviewed are conventional forecasting method having established mathematical and statistical background and track record of usage by power utilities as well.

Multiple linear regression: Regression methods work with the historical data, extrapolating the past load patterns into future. The load is modeled in terms of explanatory variables such as weather and non-weather variables, which affect the load behaviour. Experience of the load behaviour helps in initial identification of the suspected influential variables.

Time series formulation: The time series formulation includes the time delay variables along with the causal variables. The formulation and method of identifying the coefficients are same as in case of Multiple Linear Regression. The time series formulation also includes the trend of the forecasted variables and therefore is more accurate.

LOAD FORECASTING VIA ARTIFICIAL NEURAL NETWORK

It is established that the load forecasting is a pattern matching and generalization problem. Artificial Neural Network (ANN) performs reasonably well for such problems. ANN models depend on mapping of input/output spaces and therefore are free from explicitly defined load models, from which the statistical methods suffer from. These methods used various network architectures and supervised as well as unsupervised learning rules. Though standard design procedures differ from author to author, networks are mostly designed on trial and error and engineering judgments. This is not a major obstacle, because these trial and error methods can be easily implemented. A large number of artificial neural network architectures based on Feed Forward Neural Network, Radial Basis Function Neural Network in Unsupervised/Supervised category, Numerical Taxonomy method in Self Organizing category (Peng *et al*., 1992; Yuan and Chien, 1991; Ranaweera *et al*., 1995; Singh and Malik, 1995; Singh and Singh, 2001; Singh *et al*., 2002) etc. for weekly as well as hourly forecasts are reported. Normally a neural network is trained to forecast the load up to predefined lead-time. It can be a lead of one day or a week ahead. Comparison of 2 schemes of the forecasts shows that the, day-ahead forecasts are more accurate than week-ahead forecasts.

Feed forward neural networks: The network (Singh and Malik, 1995) consists of hidden layers and an output layer. The number of hidden layer may be one or more. The neuron or the processing node receives a signal, which is sum of product of input and its connecting weight, called activation level. The final output is produced after passing this activation level through a predefined activation function. This phenomenon takes place at each neuron in a network. This is termed as forward pass. Design of Feed Forward Neural Network (FFNN) forecaster requires determination of proper inputs, detrimental to future loads, proper number of hidden layers and number of neurons in a layer. The FFNN has a remarkable property of learning through examples, a process called training.

In the multilayer network, the first set of neurons connecting to the inputs serve only as distribution points. They perform no input summation. The training instance set for the network must be presented many times in order for the interconnection weights between the neurons, to settle into a state for correct classification of input patterns. While, the network can recognize patterns similar to those they have learned, they do not have ability to recognize new patterns. This is true for all supervised learning networks. In order to recognize new patterns, the network needs to be retrained with these patterns along with previously known patterns. If only new patterns are provided for re-training, then old patterns may be forgotten. In this way learning is not incremental over time. This is a major limitation for supervised learning networks. Another limitation is that the back propagation network is prone to local minima, just like any other gradient descent algorithm.

This network consists of a set of input units, a set of output units and one or more layers of intermediate units. These intermediate unit layers are called hidden unit layers. The learning rule is a generalization of the Delta rule for multi-layer networks. It carries out a minimization of the mean square error E which is now a function of both weight matrices w and W. Since, the weight errors are successively back propagated from output layer to hidden layer, this algorithm is known as 'Error Back –Propagation'. The main difficulties with these networks are that they sometimes get stuck into local minimum and also their convergence is slow. The slow convergence can be overcome partially by using momentum terms. It learns through training examples that are similar to human expert forecasters.

Radial basis function neural networks: In Radial Basis Function (RBF) network (Singh and Singh, 2001) the input selection and network parameter tuning for optimal performance of the network is demonstrated. The selection of inputs is based on the performance of the network. The method does not suffer from convergence problem which is common in feed forward and recurrent network. A Radial Basis Function (RBF) Network consists of two layers, a hidden layer with nonlinear neurons and an output layer with linear neurons. Thus the transformation from the input space to the hidden unit space is non-linear whereas the transformation from the hidden unit space to the output space is linear. The basis functions in the hidden layer produce a localized response to the input i.e. each hidden unit has a localized receptive field. The basis function can be viewed as the activation function in the hidden layer. The most common basis function chosen is a Gaussian function. The outputs of the hidden unit lie between 0 and 1. Closer the input to the center of the Gaussian, larger will be the response of the node; because the node produces an identical output for inputs with equal distance from the center of the Gaussian, it is called a radial basis.

The output units form a linear combination of the nonlinear basis functions and thus the overall network performs a nonlinear transformation of the input. The normalization factor represents a measure of the spread of the data in the cluster associated with the hidden unit. The average distance between the cluster center and the training instances in that cluster commonly determines it. Designing an RBF network involves choosing the centers from the data points. A radial basis neuron receives the vector distance between its weight vector (cluster center) 'W' and the input vector divided by the spread constant factor, unlike sum of product of the inputs and respective synaptic weights in case of feed forward network.

The RBF neural network generalizes on the basis of pattern matching. The different patterns are stored in a network in form of cluster centers of the neurons of the hidden units. The number of neuron determines the number of cluster centers that are stored in the network. If the numbers of neurons are as large as number of training patterns, permissible maximum number of neurons, all the input patterns will be recognized as separate cluster center, thus it acts like a memory. In case the number of neurons is less than the training patterns, the network will group the similar inputs patterns a single cluster. Thus it will act like a generalizer. The response of particular hidden layer node is maximum (i.e. 1) when the incoming pattern matches the cluster center of the neuron perfectly and the response decays monotonically as the input patterns mismatches the cluster center. The rate of decay can be small or large depending on the value of the spread. Neurons with large spread will generalize more, as it will be giving same responses (closer to 1) even for the wide variation in the input pattern and the cluster centers, whereas a small spread will reduce the generalization property and work as a memory. Therefore, spread is an important parameter and depends on the nature of input pattern space.

The values of the different parameters of the RBF networks are determined during training. These parameters are spread, cluster centers and weights and biases of the linear layer. The number of neurons for the network and spread is determined through experimentation with a large number of combinations of spread and number of neuron. The best combination is one, which produces minimum Sum Squared Error (SSE), on the testing data. The optimal set of parameters is determined through following the above procedure.

Numerical taxonomy (clustering): In Numerical Taxonomy Method (Singh *et al.*, 2002) the advantage of distance metric and clustering based neural network learning is obvious from speed and accuracy. The method is formulated as an optimization problem using Mean Absolute Percentage Error (MAPE) as an objective function. The data matrix and the cluster size are the optimization variables. In this research a clustering based formulation of STLF is demonstrated, which do not require training and uses less number of inputs as compared to neural network forecasters.

A clustering based formulation of STLF does not require training and uses less number of independent variables/ inputs as compared to neural network forecasters. Once the inputs are decided, the model parameters are easily found. Since, the method does not involve any training process, the continuous monitoring and trial with different parameter settings are not required. The method of clustering is well known for data partitioning and analysis in market prediction and classification in biology. The data partitioning approach is normally used to classify the data according to an objective function. The data are made to self-organize into clusters/partitions by finding similarities among the variables. Normally, this requires very simple computations of averaging and Euclidean distance. Generally the clustering technique is used for partitioning the data matrix to minimize the certain objective function among the data itself i.e. the problem of clustering can be addressed as the optimal partitioning of the data matrix such that the MAPE is minimum for the given data matrix.

TEST SYSTEM

The study is made on the data of a Canadian utility, TransAlta, Alberta, Canada. The data consists of hourly-integrated load data and temperature data of two important cities (load centers) namely Edmonton and Calgary. There is wide variation in temperatures for these 2 load centers. The load may vary from 6190 MW H to 4427 MW H in winters.

RESULTS AND DISCUSSION

The performance is studied on the historical data of a Canadian utility. This data includes hourly loads and temperatures (T1 and T2) of two major load centres. The forecasting models discussed in earlier sections were designed for comparison. The data of three weeks were taken for fixing the parameters of the various models (historical data). The forecaster was used to forecast the hourly load up to a week. The models were designed to give the optimum performance. The parameters of the models were finalized after several trial and error efforts to give the optimum performance. The models designed had the following parameters which are also shown in the Table 1-3. The MLR had two temperature (T1 and T2) as the independent variables. The time series formulation had T1 and T2 at the forecast hour as exogenous variables where as, load at hour k-1 and k-2 was taken as the time series inputs, where k is the hour of forecast. There were separate models for forecasting weekdays and weekend days.

The FFNN had 35 input variable containing loads at k-1, k-2, k-3, k-25, k-26, k-27, k-168, k-169, k-170 as load variables, temperature of both places at k, k-1, k-2, k-3 along with 07 input for day type. The networks with 10 hidden neurons were selected on basis of trial and error.

Table 1: Forecast error (MAPE) for a selected week in winter

Day	MLR	Time series	FFNN 35×10×1	RBFNN 6×5×1 3×3×1	NT
Mon.	1.5307	1.0164	1.1672	1.0776	1.1194
Tues.	2.5946	1.4431	1.0621	1.0727	1.3289
Wed.	2.5572	1.4522	1.8903	1.1105	1.6555
Thu.	0.8748	1.2026	1.6119	0.7494	0.8902
Fri.	1.2807	1.2488	1.4074	1.1171	1.3887
Sat.	3.6593	2.0467	0.9937	1.6459	1.0929
Sun.	2.5460	1.4010	1.2432	1.5838	0.8067
Average	2.1490	1.4015	1.3393	1.1939	1.1832

Table 2: Forecast error (MAPE) for a selected week in spring

Day	MLR	Time series	FFNN 35×10×1	RBFNN 6×5×1 3×3×1	NT
Mon.	1.3874	0.9597	0.7371	1.0856	1.2167
Tues.	0.9410	0.6830	0.6662	0.7082	0.6759
Wed.	0.7514	1.4804	1.2875	0.9606	0.5802
Thu.	2.4449	1.1851	2.4804	2.2876	2.3343
Fri.	1.4025	0.9368	1.3930	1.1114	1.4699
Sat.	1.4023	1.3493	0.9916	0.7726	1.1042
Sun.	1.2428	2.1244	1.4455	1.7412	1.3413
Average	1.3675	1.2455	1.2854	1.2310	1.2461

Table 3: Forecast error (MAPE) for a selected week in summer

Day	MLR	Time series	FFNN 35×10×1	RBFNN 6×5×1 3×3×1	NT
Mon.	0.8501	0.7641	1.6191	1.2466	0.9938
Tues.	0.6679	0.6653	0.7331	2.2017	0.9387
Wed.	0.9014	1.0485	0.9248	0.8057	1.2389
Thu.	1.0093	1.3438	1.0421	1.2365	0.6789
Fri.	0.8507	1.5750	1.3168	0.9062	0.8277
Sat.	1.1833	1.8962	1.0007	1.0312	0.8950
Sun.	1.4736	1.1163	0.9046	1.1475	1.2590
Average	0.9909	1.2013	1.0773	1.2246	0.9788

The RBFNN model consisted of 06 inputs and five hidden layer neurons for weekday forecasting and 03 inputs and 03 hidden layer neurons for weekend day forecasting. The numerical taxonomy method used loads at k-1, k-2 and temperatures at the kth h as input for week days architecture, whereas only load at k-1 hour is used for weekend days as input.

The forecast error in MAPE for a week in winter is given in Table 1. A comparative chart showing accuracy of different approaches for winter season has been shown in Table 1. It is observed that the non-conventional methods such as FFNN, RBFNN and NT methods are quite accurate as compared to MLR and Time Series methods. Forecast errors for the representative week in spring are given in Table 2. A comparative chart showing accuracy of different approaches for spring season. It is observed that the MLR is quite accurate. However, RBFNN, FFNN, NT and Time Series are equally good and comparable.

The forecast errors and a comparative chart showing accuracy of different approaches for a week in summer season are shown in Table 3, respectively. It is observed that in this case numerical taxonomy method is quite accurate and MLR is comparable. However, the other models are less accurate.

From the above results it is observed that the conventional methods (MLR and Time Series) are accurate where the load behaviour is less erratic. During summers the loads are much more predictable than those in winters.

CONCLUSION

Performance evaluation on the basis of forecasting accuracy (MAPE) between two statistical and three artificial neural networks based short term load forecasting approaches has been made in this study. Also a study of the data requirements for different methods in relation to each other is made. It is observed that the conventional statistical methods such as MLR and Time Series performance are comparable and in some cases even it is better. Season in which data is less erratic, it is better but the accuracy deteriorates for other seasons. However, the neural network approaches performs consistently in all seasons. The MLR and Time Series methods use fewer amounts of data for fixing the parameters as compared to Neural Network methods. However, the neural network methods have the ability to map different type of data such as day type and hour of the day in their models. The self- organizing (NT) method shows the best results.

ACKNOWLEDGEMENT

The authors would like to thank TransAlta, Alberta, Canada for providing the load data used in the studies.

REFERENCES

Christiaanse, W.R., 1988. Short Term Load Forecasting using general exponential smoothing. IEEE Trans. Power Appar. Syst., PAS, 3: 900-911.

Gross, G. and F.D. Galiana, 1987. Short Term Load Forecasting. Proc. IEE., 75 (12): 1558-1578.

Hagen, M.T. and S.M. Behr, 1987. The time series approach to time series load forecasting. IEEE Trans. Power Syst. PRWS, 2 (3): 785-791.

Haung, S.R., 1997. Short Term Load Forecasting using threshold autoregressive models. IEE. Proc. C, 144 (5): 477-481.

IEEE Committee Report, 1980. Load Forecasting Bibliography, Phase I. IEEE. Trans. Power Appl. Syst., 99 (1): 53-58.

IEEE Committee Report, 1981. Load Forecasting Bibliography, Phase II. IEEE. Trans. Power Appl. Syst., 100 (7): 3217-3220.

Moghram, I. and S. Rahman, 1989. Analysis and evaluation of 5 short term load forecast techniques. IEEE. Trans. Power Syst., 4 (4): 1484-1491.

Papalexopoulos, A.D. and T. Hasterberg, 1990. A Regression based Approach to Short Term System Load Forecast. IEEE. Trans. Power Syst., 5 (4): 1535-1544.

Peng, T.M., N.F. Hubele and G. Karady, 1992. Advancement in the application of Neural Networks for short term load forecasting. IEEE. Trans. Power Syst., 7 (1): 250-257.

Ranaweera, D.K., N.F. Hubele and A.D. Papalexopoulos, 1995. Application of Radial Basis Function Neural Network Model for Short Term Load Forecasting. IEE. Proc. C, 142 (1).

Singh, D. and S.P. Singh, 2001. Self selecting neural network for short-term load forecasting. J. Elec. Power Component Syst., 29: 117-130.

Singh, D., S.P. Singh and O.P. Malik, 2002. Numerical Taxonomy Method for STLF. J. Elec. Power Component Syst., 30 (5): 443-456.

Singh, S.P. and O.P. Malik, 1995. Single ANN architecture for short term load forecasting for all seasons. Int. J. Eng. Intelligent Syst., 3 (4): 249-254.

Yuan-yih Hsu and Chien-chuen Yang, 1991. Design of artificial neural network for short term load forecasting. Part1: Self-organizing feature maps for day type identification. IEE. Proc. C, 138 (5): 407-413.