ARPN Journal of Engineering and Applied Sciences

EFFECT OF INPUT VARIABLES SELECTION ON ENERGY DEMAND PREDICTION BASED ON INTELLIGENT HYBRID NEURAL NETWORKS

Badarul Islam, Zuhairi Baharudin and Perumal Nallagownden
Department of Electrical & Electronics Engineering Universiti Teknologi Petronas, Bandar Seri Iskandar, Perak, Malaysia
E-Mail: zuhairb@petronas.com.my

ABSTRACT

Numerous techniques have been applied by the researchers to predict the future electrical energy demand, which can be broadly categorized as parametric (statistical) and non-parametric (intelligent) techniques. The non-parametric or intelligent methods which are based on artificial intelligence are gaining a lot of attention during the recent past years. As compared to the other intelligent techniques, the Artificial Neural Network (ANN) has a tendency to map and memorize the non-linear relations between inputs and output variables. Because of this ability, they are extensively implemented in modern predictive model development. The efficacy of these models depends upon many factors such as, neural network architecture, type of training algorithm, input training and testing data set and initial values of synaptic weights. Among the others, the selection of most influential input variables has a critical effect on the forecast results. In this paper, the important issues related with the best input variable selection for a hybrid model is addressed. A hybrid approach that combines ANN and an evolutionary optimization technique, genetic algorithm (GA) is used for the development of a short term load forecast (STLF) model. GA and correlation analysis are used for the selection of the most influential input variables for the training and testing of the hybrid model. Multiple cases are developed using different optimally selected input variable vectors to train and test the back propagation neural network (BP-NN) and the hybrid model. The results show that hybrid forecast model provide better performance when it is trained and tested with optimally selected input variable vector (IV), containing historical load and meteorological variables. The proposed input variable selection approach not only improves that forecast accuracy but also reduces the computational efforts and training time of forecasting models.

Keywords: neural networks, hybrid, energy demand, prediction.

INTRODUCTION

Electric load forecast is the prediction of future load of power systems, which plays very important role in the energy management system to provide a better environment for future planning and decisions [1]. Electrical load forecasting is classified into short-term, medium-term and long-term forecasts with respect to planning horizon’s duration. Although, medium and long term load forecasts have their own benefits, yet short term load forecast is of utmost importance because it helps the operation managers in their day to day operational decisions, such as economic scheduling of generating capacity, scheduling of fuel purchase, estimation of peak demand and system security assessments [2-5]. ANN based model are most frequently used in the development of predictive models for short term load demand forecasts, as reported in the published work [6]. The enhancement is the forecast accuracy of these models is an important research issue that is concerned with the multiple factor [7]. However, the selection of most influential input variables and training data set for ANN is focused in this paper, because of its critical importance.

In this paper, the important issues related to the best input variable selection for ANN based STLF models are addressed. A new technique is proposed that integrates the GA and correlation analysis to establish the supremacy of the certain input variables over the others. In the next stage of experiments, the selected input variables are used to train a multilayer perceptron neural network (MLPNN) BP training technique. A Hybrid model is also developed, where the GA is combined with MLPNN to optimize the free parameters of neural networks. The optimally selected IVs are used to train the optimized hybrid model and the results are compared with the conventional BP training algorithm. The objective of this research is to find the most influential input variable vector for STLF.

Data analysis and pre-processing

The data related to historical load, electricity price and weather variables, sampled at the frequency of half hour for the previous five years of Australian grid is used in this research. A comprehensive quantitative analysis of the data is carried out to inspect various aspects of the data. The correlation coefficient analysis is conducted and the results are used in the proposed input variables selection technique.

VOL. 10, NO 21, NOVEMBER, 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

Correlation analysis

Correlation analysis is the most frequently reported method to verify the significance of input variables used for prediction purpose in the literature [8]. It provides a comprehensive measure of the relationship between two variables. In general, the bigger correlation coefficient of the input variable with expected output values indicates a strong relationship between them [9]. The general form of the correlation coefficient R can be formulated as:

$$R = \frac{n\sum xy-(\sum x)(\sum y)}{\sqrt{n(\sum x^2)-(\sum x)^2}\sqrt{n(\sum y^2)-(\sum y)^2}} \quad (1)$$

where, n represents the number of pairs of data, x and y are the two variables, the value of R lies between -1 and +1. If the x and$y$ vectors have close relevance, then the value of R will be closer to 1, that indicates high correlation. The results of correlation analysis between load demand and input variables are illustrated Table-1.

Table-1. Correlation analysis results.

Input Variables	Correlation
Same time in the previous day load L(h-	0.9301
Same day and time in the previous week	0.9121
Same time, two days earlier load L(h-48)	0.8220
Same day, two hours earlier load L(H-2)	0.8818
Same time, previous day temperature	0.7909
Same time, previous week temperature	0.7017

The results of correlation analysis show that the load related variables have bigger coefficients as compared to weather variables. The load of the same day in the previous hour and same time in the previous hour are the most correlated variables. On the other hand temperature has a lesser correlation value. All the input variables are normalized before applying them to the model that results to restrict the input values in between 0 and 1.

Data normalization

All the input variables are normalized before applying them to the models. It is a transformation process that results to restrict the input values in a scale between 0 and 1 [10]. The data normalization improves the learning process of the neural network by providing easy to learn patterns and consequently, the performance of the MLPN based forecast models is enhanced [11]. The output of the network is converted back into the original format in a descaling process. The mathematical notations for data scaling and descaling are:

$$x_{(scl)} = \frac{[(x-x_{(min)})\times(x_{(max)}-x_{(min)})]}{(x_{(max)}-x_{(min)})} \quad (2)$$

$$y_{(dscl)} = [L_{(max)} - L_{(min)}] \times [y_0 + L_{(min)}] \quad (3)$$

where, $L_{(max)}$ and $L_{(min)}$ are the minimum and maximum values of the actual load and$y_0$is the normalized value of the neural network output. $y_{(dscl)}$is descaled value of the MLPNN output, which is equal to the predicted load.

Components of experimental techniques

MLPNN with BP training algorithm

MLPNN has been the most frequently used types of ANN architecture, which is composed of an input layer, one or more hidden layers and one output layer of neurons [12], as shown in Figure-1.

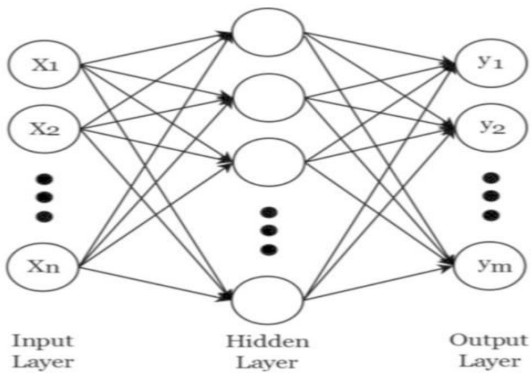

Figure-1. MLPNN with three layers.

In 1986, G.E. Rumelhart, proposed the BP algorithm for the training of neural networks. In the process of training, the inputs are linearly summed up after multiplying them by their respective connection weights and then passed through an activation function before passing them on to the next layer [13, 14]. The input and output equation of k_{th} neuron in this case can be formulated as:

$$u_k = \sum_{t=1}^{n} x_i w_i \quad (4)$$

where, u_k is the output of the linear summer, x_i is the i_{th} input and w_iIs its respective connection weight.

ANN learning is a process to extract the complex relationship between inputs and output of the network. In this relationship, an error is generated that is the difference between network output and target values. This error information is fed back to the network to update the weight values of network connections to minimize the error up to a certain acceptable threshold level. This error correction process generally deploys gradient descent algorithms [4],[15-18] to correct the error function. The

Back propagation learning algorithm uses the gradient descent method to update the weights and biases. For the network parameters, the partial derivative of the performance with respect to the weights and biases is calculated. Each node of the network is needed to be differentiated in accordance with back propagated error.

The combined error (E) and gradient of error (E^p) for each point p in the network can be calculated as the following:

$$E = \sum_p E^p \quad (5)$$

The sum of all individual errors.E^pcan be calculated as:

$$E^p = \frac{1}{2}\sum(t_P^p - t_O^p)^2 \quad (6)$$

where, t_P^p and $t_O^p$are the desired and actual outputs for one neuron at point p.

Genetic algorithm (GA)

GA is a directed random search technique inspired by the living natural evolution process, proposed by John Holland [19]. The main idea is to design artificial systems, retaining the robustness and adoption properties of natural systems. Since the inception, these methodologies were then further improved by other researchers and are now widely used in various fields (business, science and engineering) to solve a variety of optimization problems. GA mimics the biological processes to perform a random search in a defined N-dimensional possible set of solutions [20]. For an optimization problem, it is needed to search and find the best solution in a given search space.

GAs work with a set of artificial elements called a population. An individual (string) is referred to as a chromosome, and a single bit in a string is called a gene [21]. GA generates a new population (called off springs) by applying the genetic operators to the chromosomes in the old population (called parents). Each iteration of genetic operations is referred to as a generation [17, 22]. The effective search for the optimal solution depends on the selection of population size, crossover rate, and mutation rate [23].

METHODOLOGY

In this section the proposed scheme of experiments is explained, which is highlighted in the Figure-2.

Selection of optimized input variable vector

The methodology proposed for the selection of most influential IVs, requires a sequence of tasks to be followed. The impact of correlation is used to analyze the relationship between meteorological variables, historical load data and calendar events on the load demand. Initially, for the input variable selection process, highly correlated historical load data, meteorological variables and other influencing time related variables are considered and then the GA is implemented for the final selection of IVs. The GA works on the principles of schema theorem, which can be derived as the following for this optimization problem.

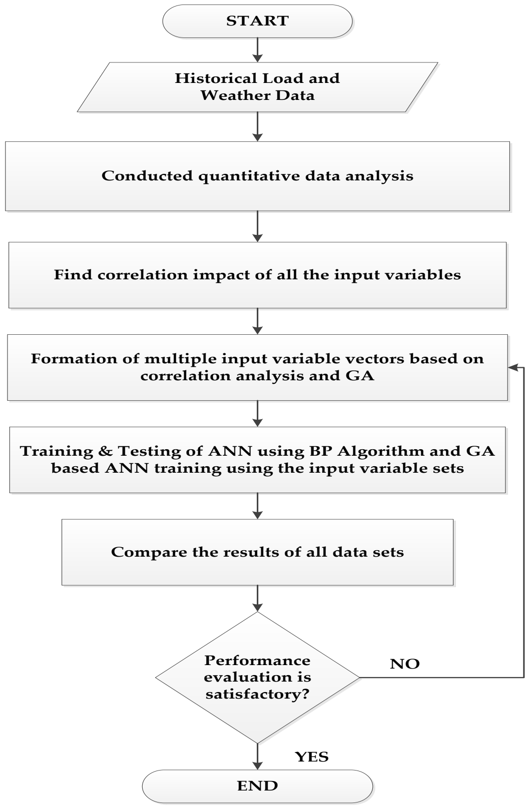

Figure-2. Flow diagram of research activities.

Let the total search space (set of possible chromosomes) isΩ, chromosome length (number of genes in each chromosome) is l. One chromosome and the population size are represented by x and S respectively.

The number of instances the string x appears in the population:

$$S = m(x) \quad (7)$$

The H (schema) is a string equal to l in size, where H can have all the possible alphabets in the string. The number of strings in the population that belong to a schema H can be is represented:

$$m(H) = \sum_{x \in \Omega} m(x) \quad (8)$$

ARPN Journal of Engineering and Applied Sciences

Here x is a member of the population (vector $x \epsilon \Omega$).

Let population at a certain time t is S(t) and number of strings in the population belonging to schema H is $m(H,t)$, the average fitness of all the strings representing schema H is defined as:

$$f(H,t) = \frac{\sum_{x \epsilon \Omega} f(x).m(x)}{m(H,t)} \tag{9}$$

$f(H,t)$ is referred as an average payoff function of the schema. The fitness of string x if equal to f(x), then, according to the standard roulette wheel selection mechanism the expected no. of selections of x is given by:

$$E(x) = f(x)/f'(x) \tag{10}$$

Here $f'(x)$, stands for the average fitness of all the strings in the population at time t. The expected number of strings belonging to the schema H in the next population can be described as:

$$m(H,t+1) = m(H,t).\frac{f(H,t)}{f\prime(t)} \tag{11}$$

This equation represents that the schema with above average fitness values $\left(\text{where } f(H,t) \geq f'(H)\right)$, will be reproduced in the new generation and those having a value less than average fitness $\left(f(H,t) < f'(H)\right)$, will not be reproduced in the next generation and eventually die off.

It is to be noted that no genetic operator is applied to the selected strings before they are inserted in the new population.

When $a = \frac{f(H,t)}{f\prime(t)}$ is relatively constant, the Equation. 11 can be approximated by a linear difference equation of the form:

$$m(H,t+1) = a.m(H,t) \tag{12}$$

The final solution is given by:

$$m(H,t) = m(H,0).a^1 \tag{13}$$

In this case strings that belong to an above average schema will grow exponentially, whereas the strings belonging to the below average schema will decline exponentially. However, this imposed condition of the constant value of a will not prevail for multiple generations, because an increase of above average strings and decrease in below average strings will result in the form of an increase in the average fitness value$f'(t)$. The $f(H,t)$ will remain almost constant. After the initial exponential growth of above average schemata, a slow increase and lastly an almost constant number of strings belonging to schemata H would be the result, when the value of a would be near 1. In this situation the above average schema H has spread over the whole population in a way that every member of the generation belongs to this schema. This makes $f(H,t)$ equals to $f'(t)$ and a equals to 1.

The steps involved in the selection process of IV are as the following:

Step 1

A chromosome is defined on the first step of the GA. A fixed length chromosome equal to 8 is implemented. The value of each genome in the defined chromosome is the index of input variables. These chromosomes correspond to the possible solutions in the selection process. Once the individual chromosome is defined, generate an initial population S composed of n chromosomes.

Step 2

The fitness function is developed that maximizes the correlation coefficient and minimizes the mean square error (MSE) between actual and forecast load. The defined fitness function is called maximum correlation value and minimum error (mxRmnE) as shown in Equation. 17. Mean square error (Equation. 17) is used as a performance index in this case.

Step 3

The roulette wheel selection method with a single point crossover and multipoint mutation is implemented in this technique. The mutation operator is defined as given in Equation. 14. This operator complements the values in the genome to avoid from the local minima [24].

In the first step of the GA, the chromosome length and index of input variables are defined. Real coded GA is implemented with a fixed length chromosome i.e. 8 and the value of each genome in a particular chromosome is the index of the input variable. In the second step, the GA is used for the selections of the final data set. The initial parameters of GA are set as follows:

Population size: $P = 60$

Individual coding length: $L = 65$

Maximum number of iterations: $I_{max} = 150$

Mutation probability: $P_m = 0.001$

Crossover probability: $P_c = 0.8$

The fitness function used by GA is represented by the Equation. 17.

$$mut(S,k), S_k = K - S_k \tag{14}$$

$$E = \sum (A_t - P_t)^2 \tag{15}$$

$$mxRmnE(S,p), = \frac{1}{|S|} \sum_{x_i \epsilon S} \frac{1}{\sum (A_t - P_t)^2} (x_i, p).R(x_i) \tag{16}$$

$$\text{fintness function} = \frac{R}{1+E} \tag{17}$$

ARPN Journal of Engineering and Applied Sciences

where, S is a subset of input variables, R is the correlation value and p is target class of the prediction variable. A_t and P_t are the actual and predicted load values. Whereas, k is mutation point and K is the length of input dimension.

The process flow diagram of the proposed technique is shown in Figure-3. In Figure-4, the input variable vector that produced the best performance for the forecast models is depicted. This vector is composed of; day and time indicators; dry bulb temperature of the same time in the previous day *T(w, d-1, h)*, dew point of the same time in the previous day *D(w,d-1,h),* electrical load of the same time in the previous day *L(w, d-1, h),* load of the same day and time in the previous week *L(w-1, d, h)*, load of the previous day minus one hours *L (w, d, h-1)* and load of the previous day minus two hours *L (w, d, h-2)*.

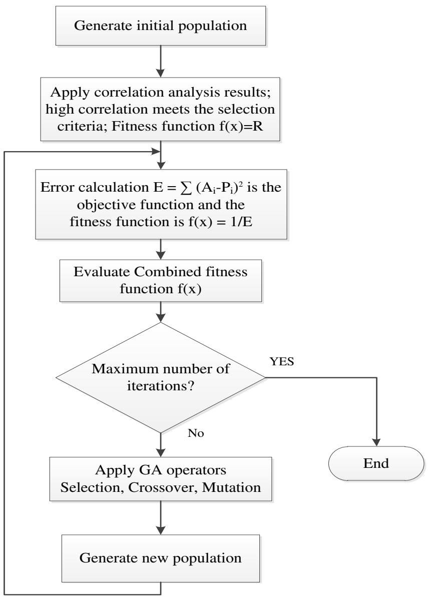

Figure-3. Flow diagram of input variable selection technique.

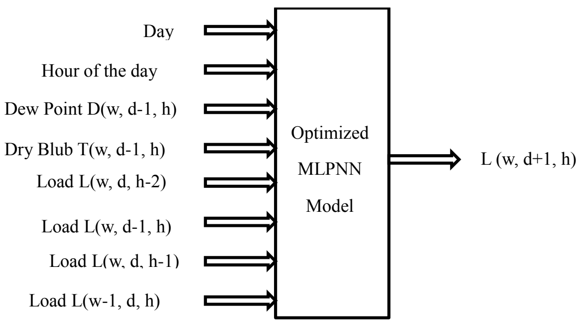

Figure-4. Optimally selected input variable vector with historical load and meteorological variables.

Performance analysis

To analyze the performance and accuracy of optimized prediction models multiple evaluation measures are implemented. The coefficient of determination, R^2, is a useful performance measure as it returns the variance of one variable that is predictable from the other variable. The value of R^2 lies in between 0 and 1 and prediction accuracy is supposed to be high as its value approaches to 1. This useful performance indicator can be determined by using following expression;

$$R^2 = \frac{\left[\sum_{t=1}^{N}(P_t-\bar{P})(A_t-\bar{A})\right]^2}{\left[\sum_{t=1}^{N}(P_t-\bar{P})^2 \sum_{t=1}^{N}(A_t-\bar{A})^2\right]} \quad (18)$$

The performance can be measured in terms of root mean square error (RMSE), mean absolute percentage error (MAPE), mean square error (MSE) and mean absolute error (MAE), which are computed as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(A_t - P_t)^2} \quad (19)$$

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\frac{|A_t-P_t|}{|A_t|}100\% \quad (20)$$

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(A_t - P_t)^2 \quad (21)$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|A_t - P_t| \quad (22)$$

where, A_t and P_t are the actual and predicted values at time point *t.*

The MAPE is considered as a benchmark performance index due to its stable performance that resolves the inconsistency problem in the predicted results [25, 26].

RESULTS AND DISCUSSIONS

Five year historical data of system load and four meteorological variables of the State of Victoria and New South Wales of Australia (from year 2006 to 2010) at half hour sampling frequency are deployed in the experimentation.

BP based MLPNN forecast results using the different combinations of input variables of half hourly recorded samples for 24 hours of a day are presented below in graphical form in Figure-5. The graph shows a plot of both actual and forecast loads in MW against half hourly data samples. The respective error measures are shown in Table-2. The minimum error is recorded when an optimally selected input variable vector, based on load and weather related variables is applied for the training and testing of MLPNN-BP. The input variable combinations are further used for the training of hybrid models based on MLPNN-GA. This model is also trained and tested using load related variables, weather related variables and combination of load and weather related

VOL. 10, NO 21, NOVEMBER, 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

variables. The results of these experiments on the bases of forecast error are also shown in Table-2 and Figure-6.

Table-2. Performance evaluation of MLPNN-BP and MLPNN-GA based on optimally selected input variable vectors.

Models	Input Variable Vectors	MAE	MSE	MAPE	RMSE	R^2
ANN-BP	Load & Weather	156.210	40061.39	3.840	200.112	0.9829
	Load Related	160.866	42518.79	3.924	206.163	0.9818
	Weather Related	922.342	1252712.6	10.913	1118.672	0.4644
ANN-GA	Load & Weather	63.617	15656.35	1.755	77.254	0.9969
	Load Related	125.465	29390.03	2.474	147.172	0.9866
	Weather Related	760.855	928054.66	9.117	963.564	0.6028

The results of the hybrid model based on optimally selected input variable vector show its superiority over the conventional BP training algorithm as all the error measures produced lesser error. These results clearly show that the proposed input variable selection method produced better forecast accuracy for STLF models.

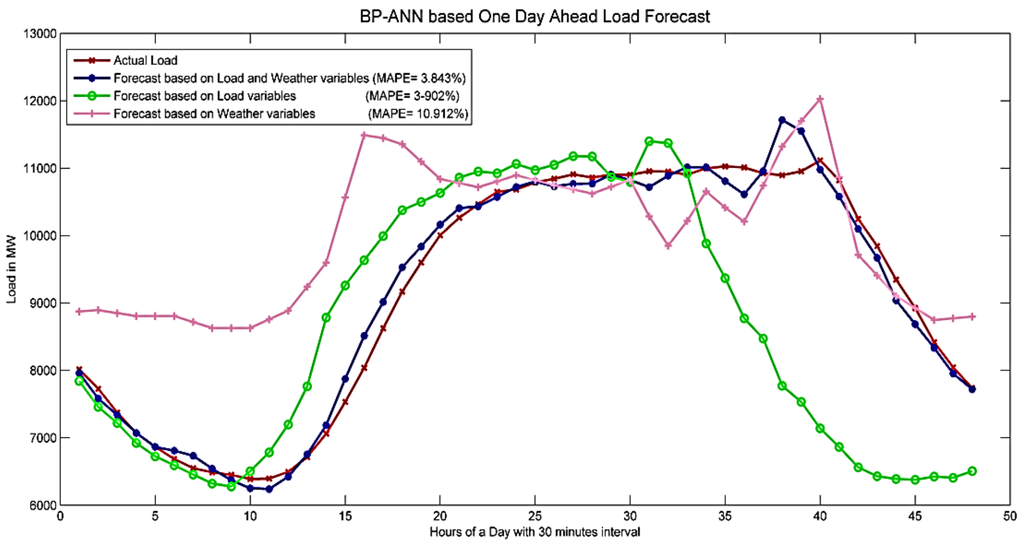

Figure-5. MLPNN-BP based forecast using optimally selected input variable vectors.

 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

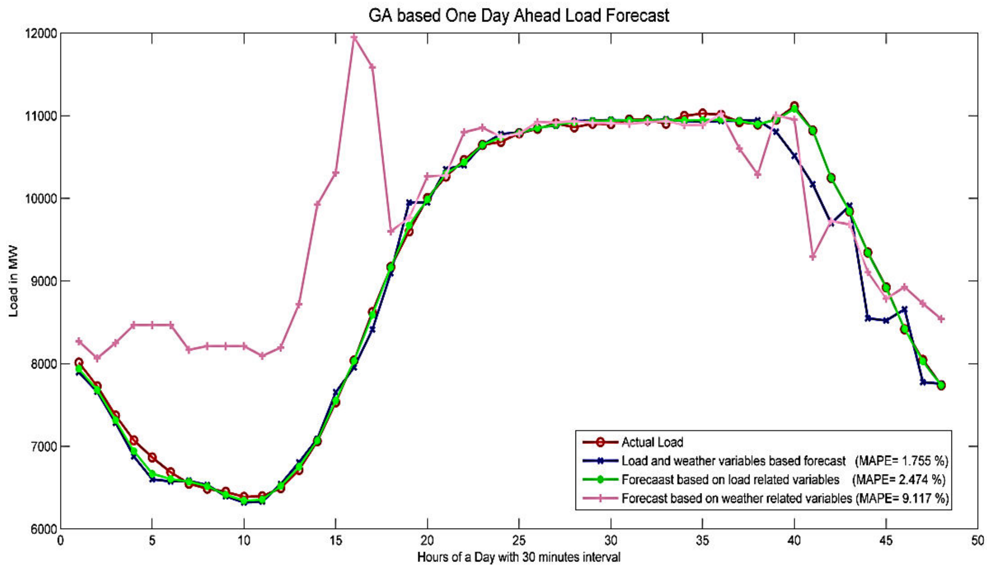

Figure-6. MLPNN-GA based forecast using optimally selected input variable vectors.

CONCLUSIONS

The optimized hybrid prediction model with a new data pre-processing framework is proposed in this paper to enhance the forecast accuracy. This method integrates the correlation analysis and genetic algorithm for the selection of most appropriate input variables vector for MLPNN training and testing. The hybrid STLF model is developed by combining the multilayer perceptron neural networks and optimization techniques genetic algorithm. This model is trained and tested by the optimally selected input variable vectors and the results are compared with the conventional BP training algorithm. The obtained simulation results show that, the performance of the optimized MLPNN-GA model outperformed the results of the conventional BP training algorithm. In particular, when this model is trained with optimally selected input vector, containing the combination of historical load and metrological variables produced the best results.

ACKNOWLEDGEMENT

The authors wish to thank University Technology PETRONAS, Malaysia for providing the funding to conduct this research.

REFERENCES

[1] Gross, G. and F.D. Galiana, Short-term load forecasting. Proceedings of the IEEE, 1987. 75(12): p. 1558-1573.

[2] Alfares, H.K. and M. Nazeeruddin, Electric load forecasting: literature survey and classification of methods. International Journal of Engineering Sciences and Emerging Technologies, Feb 2012. ISSN: 2231 - 6604 Volume 1, Issue 2, pp: 97-107 ©IJESET.

[3] Papalexopoulos, A.D., S. Hao, and T.M. Peng. Short-term system load forecasting using an artificial neural network. In Neural Networks to Power Systems, 1993. ANNPS '93. Proceedings of the Second International Forum on Applications of. 1993.

[4] Adepoju, G.A., Application of Neural Network to Load Forecasting in Nigerian Electrical Power System. The Practical Journal of Science and Technology, 2007. Volume 8.

[5] Mrs. J. p. Rothe, D., A. K. Wadhwani, Short Term Load Forecasting Using Multi Parameter Regression. International Journal of Computer Science and Information Security (IJCSIS), vol 6(2).

[6] Morinigo-Sotelo, D., *et al*. Short-term hourly load forecasting of a hospital using an artificial neural network. in International Conference on Renewable Energies and Power Quality.

[7] Srinivasan, D., Evolving artificial neural networks for short term load forecasting. Neuro computing, 1998. 23(1-3): p. 265-276.

[8] Yang, K.-h., G.-L. Shan and L.-L. Zhao. Correlation coefficient method for support vector machine input samples. In International Conference on Machine Learning and Cybernetics. IEEE.

[9] Wu, J., *et al.*, Short term load forecasting technique based on the seasonal exponential adjustment method and the regression model. Energy Conversion and Management, 2013. 70(0): p. 1-9.

[10] Xiao, L., *et al.*, A hybrid model based on data preprocessing for electrical power forecasting. International Journal of Electrical Power and Energy Systems, 2015. 64(0): p. 311-327.

[11] De Aquino, R.R., *et al.*, Combining artificial neural networks and heuristic rules in a hybrid intelligent load forecast system, in Artificial Neural Networks–ICANN 2006. 2006, Springer. p. 757-766.

[12] De Aquino, R.R., *et al.,* Using genetic algorithm to develop a neural-network-based load forecasting, in Artificial Neural Networks–ICANN 2007. 2007, Springer. p. 738-747.

[13] Jun, C., *et al.* Application of BP Neural Networks based on genetic simulated annealing algorithm for short term electricity price forecasting. In Advances in Electrical Engineering (ICAE).

[14] Guo, Z.-h., *et al.,* A case study on a hybrid wind speed forecasting method using BP neural network. Knowledge-Based Systems, 2011. 24(7): p. 1048-1056.

[15] Nahi Kandil, R.W., An efficient approach for short term load forecasting using artificial neural networks. ELSEVIER Electrical Power and Energy Systems.

[16] J. Nuno Fidalgo, M.A.M., Forecasting Portugal Global Load with Artificial Neural Networks. J. Marques de Sá *et al*. (Eds.): ICANN 2007, Part II, LNCS 4669, pp. 728-737 Springer-Verlag Berlin Heidelberg.

[17] Yang Zhangang, C.Y., K.W. Eric Cheng, Genetic Algorithm-Based RBF Neural Network Load Forecasting Model.

[18] Nayak, S.C., B.B. Misra, and H.S. Behera. Index prediction with neuro-genetic hybrid network: A comparative analysis of performance. In Computing, Communication and Applications (ICCCA), 2012 International Conference.

[19] Holland, J.H., Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. 1992: MIT press.

[20] De Menezes, L.M. and N.Y. Nikolaev, Forecasting with genetically programmed polynomial neural networks. International Journal of Forecasting, 2006. 22(2): p. 249-265.

[21] Nayak, S., B. Misra, and H. Behera. Index prediction with neuro-genetic hybrid network: A comparative analysis of performance. In Computing, Communication and Applications (ICCCA), 2012 International Conference on. 2012. IEEE.

[22] Ahmadizar, F., *et al.*, Artificial neural network development by means of a novel combination of grammatical evolution and genetic algorithm. Engineering Applications of Artificial Intelligence, 2015. 39(0): p. 1-13.

[23] Shapiro, A.F., The merging of neural networks, fuzzy logic, and genetic algorithms. Insurance: Mathematics and Economics, 2002. 31(1): p. 115-131.

[24] Satpathy, H., Real-coded GA for parameter optimization in short-term load forecasting. Artificial Neural Nets Problem Solving Methods, 2003: p. 1051-1051.

[25] Ghayekhloo, M., M. Menhaj, and M. Ghofrani, A hybrid short-term load forecasting with a new data preprocessing framework. Electric Power Systems Research, 2015. 119: p. 138-148.

[26] Wang, J.-J., *et al.*, Stock index forecasting based on a hybrid model. Omega, 2012. 40(6): p. 758-766.