Short-Term Load Forecasting Using Artificial Neural Networks

Saeed M. Badran

Department of Electrical Engineering, Al-Baha University, Al-Baha, Saudi Arabia
sbadran@bu.edu.sa

Abstract: An accurate regional load forecasting is very important in improving management performance of Power Plant Generation. Various regional load forecasting methods have been developed for 24 hours ahead. There was a developed model based on Artificial Neural Networks (ANN^{s}) which had 24 output nodes. Other ANN^{s} model forecasted the peak and valley of the load and the result was used to forecast the load profile. A parallel architecture or topology used for the system with 24 separated ANN^{s}, which means that one model of ANN^{s} is used to forecast the load demand every hour during a day. The proposed Feed-Forward Neural Network principle is conducted to perform ANNs as a behavioral model for regional electricity system. Several data records such as hour, temperature and humidity data are used as the inputs for this model. In this paper, all raw data must be preprocessed first, before they are used as the training data. A behavioral model for Monday and Tuesday forecaster is developed in this paper based on statistical reason. Gradient Descent and Levenberg Marquardt training algorithms are involved in this ANNs behavioral model. The performance of each training algorithm is compared in visuals and numerical to validate the results. Finally, the results show that the ANNs model successfully predicts the load demand for dedicated regional electricity system.

[Saeed M. Badran. **Short-Term Load Forecasting Using Artificial Neural Networks.** Journal of American Science 2012;8(4):35-42]. (ISSN: 1545-1003). http://www.americanscience.org. 5

Keywords: Artificial neural networks, Forecasting, Multi-layer perceptron, Levenberg-Marquardt method.

I. Introduction

Artificial Neural Network (ANN) is an algorithm that imitates human being biological nervous systems. It has certain performance characteristics in common with biological neural networks [1]. The key element of this algorithm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working together to solve a specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist among neurons.

Neural network systems appear to be a recent development to solve problems in many areas. However, this field was established before the advent of computers, and has survived at least one major setback and several eras.

Historically, the first artificial neuron was designed and produced in 1943 by the neurophysiologist **Warren McCulloch and the logician Walter Pits** [2,3]. However, the technology available at that time did not allow them to go on much applicable research work.

Later on, many important advances have been boosted by the use of inexpensive computer emulations. Following an initial period of enthusiasm, the field survived a period of frustration and disrepute. During this period when funding and professional support was minimal, important advances were made by relatively few researchers. These pioneers were able to develop convincing technology which surpassed the limitations identified by **Minsky and Papert** [4]. Therein, they summed up a general feeling of frustration (against neural networks) among researchers, and thus most of them accepted without further analysis [5]. Currently, the neural network field enjoys a resurgence of interest and a corresponding increase in funding.

Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. This expert can then be used to provide projections given new situations of interest and answer "what if" questions. Other advantages include:

1. Adaptive learning: An ability to learn how to do tasks based on the data given for training or initial experience.
2. Self-Organization: An ANN can create its own organization or representation of the information it receives during learning time.
3. Real Time Operation: ANN computations may be carried out in parallel, and special hardware devices are being designed and manufactured which take advantage of this capability.
4. Fault Tolerance via Redundant Information Coding: Partial destruction of a network leads to the corresponding degradation of performance. However, some network capabilities may be retained even with major network damage.

A different approach is used by Neural Networks to problem solving than conventional computers. Conventional computers use an algorithmic approach i.e. the computer follows a set of instructions in order to solve a problem. Unless the specific steps that the computer needs to follow are known it cannot solve the problem. This restricts the problem solving capability of conventional computers to problems that we already understand and know how to solve. But computers would be so much more useful if they could do things that we don't exactly know how to do. Neural networks process information in a similar way the human brain does. The network is composed of a large number of highly interconnected processing elements (neurons) working in parallel to solve a specific problem. Neural networks learn by example. They cannot be programmed to perform a specific task. The examples must be selected carefully otherwise useful time is wasted or even worse, the network might be functioning incorrectly.

If the network is able to solve the problem by itself and the operation becomes unpredictable, then this can be a disadvantage of this method. On the other hand, it is common that a cognitive approach is used by computer to solve the problem. In this mechanism, problem must be known and stated in less unambiguous instructions. These algorithms are then converted to a high level language program and then compiled into machine code that the computer can understand. Artificial Neural networks and conventional computers are not in competition. They work together. There are tasks more suited to an algorithmic approach like arithmetic operations and tasks that are more suited to neural networks. Even more, a large number of tasks, require systems that use a combination of the two approaches (normally a conventional computer is used to supervise the neural network) in order to perform at maximum efficiency.

ANN Architecture

A neuron model

An artificial neuron is the basic building block of any neural network architecture. It is considered as an information processing unit that is the fundamental ANN[$] operation. The three basic elements of the neuron model are:

1. A set of weights, each of which is characterized by a strength of its own. A signal x_i connected to neuron k is multiplied by the weight w_{ki}. The weight of an artificial neuron may lie in a range that includes negative as well as positive values.

2. An adder for summing the input signals, weighted by the respective weights of the neuron.

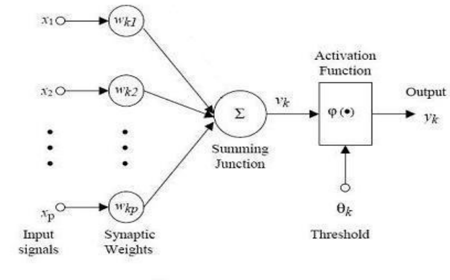

Fig. 1 A single neuron (k) model coupled with synaptic weights from other neurons w_{ki} (i,……,p)

$$V_k = \sum_{i=1}^{p} w_{ki} x_i \qquad (1)$$

$$y_k=\varphi(V_k+\theta_k) \qquad (2)$$

3. An activation function for limiting the amplitude of the output of a neuron. It is also referred to as transfer function which squashes the amplitude range of the output signal to some finite value. Two types of the sigmoid transfer (activation) functions are commonly used in ANN applications. First one is the logistic sigmoid [a'=log sig (n)]. The other function is [a'=tan sig (n)] that its value given at any arbitrary time instant (n) by equations (3) & (4) respectively. The graphical presentation of both functions is shown at Fig. 2.

$$y_k(n)=\varphi(V_k(n))= 1 /(1+e^{-\lambda v_k(n)}) \qquad (3)$$

$$y_k(n)=\varphi(V_k(n))=(1-e^{-\lambda v_k(n)})/(1+e^{-\lambda v_k(n)}) \qquad (4)$$

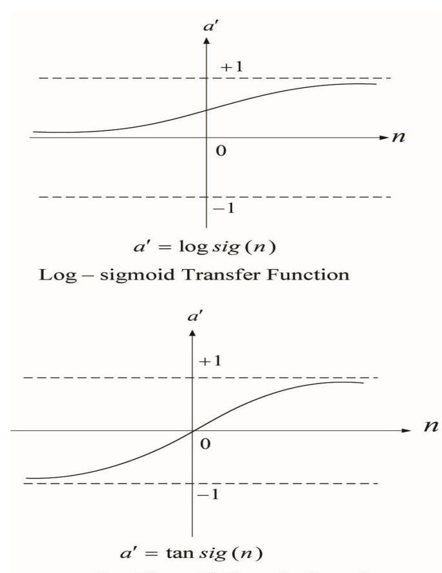

Fig. 2 Two types of transfer (activation) function [19].

A multilayer perceptron (MLP)

A generalized structure of MLP is shown at Fig.3. It consists of one input, two hidden, and one output layers. Furthermore this structure has three distinctive characteristics as follows:

1. The model of each neuron in the network includes a nonlinear activation (transfer) function. The logistic sigmoid function is commonly used which is defined as [log sig (n)] given by equation (3). Another commonly used is hyperbolic tangent sigmoid function defined as [tan sig (n)] given by equation (4).
   Noting that above nonlinearities of both activation (transfer) functions is important because otherwise the input- output relation of the network could be reduced to that of single layer perceptron.
2. The network contains one or more layers of hidden neurons that are not part of the input or output of the network. These hidden neurons enable the network to learn complex tasks.
3. The network exhibits a high degree of connectivity. A change in the connectivity of the network requires a change in the population of their weights.

Interestingly, the application of ANNs proved its suitability for load forecasting according to the following remarks:
1-ANNs are able to approximate numerically any continuous function to the desired accuracy. ANNs could be seen as multivariate, nonlinear and nonparametric methods.
2-ANNs considered as data-driven method, i.e., their models' parameters are mainly estimated on basis of preprocessed forecasting data.
3- Given a sample of input and output vectors, ANNs are able to automatically map the relationship between them.

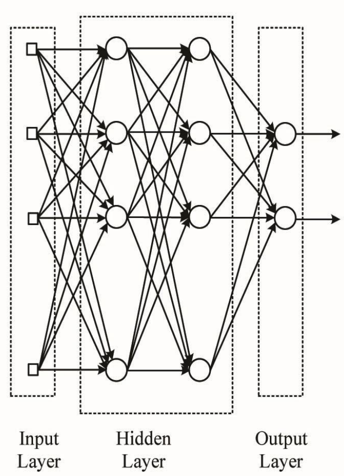

Fig. 3 A generalized structure of MLP consists of one input, two hidden, and one output layers.

II- Load forecasting

System load forecasting is an essential operation and important component in power system control centers for modern energy management systems. For many reasons which are in line with economic consideration, such as several decisions in unit commitment, scheduling of spinning reserve capacity and planning device maintenance, many market operators take the advantage of load forecasting in their operation to obtain system security and reliability in the highest possibility. For instance, an over-prediction of load to meet security requirements could involve the start up of too many power plant generation units. Cause of this imprecision will add the operating costs, as market operators must allocate such demand [6,7]. Load forecasting can maintain different planning objectives and assist system operators to meet some critical conditions in the short, medium and long term. It can be classified into four types depending on the future time window of the forecasting task [8]:

1) Long-term Load Forecasting (LTLF)
More than a year; used to identify needs for major generation planning and investment, since large power plants may take a decade to become available due to the challenging project requirements and the needs to design, finance and build them; In the deregulated economy, decisions on capital expenditures based on long-term forecasting are also more important than in a non-deregulated economy when rate increases could be necessary by capital expenditure projects [9].

2) Medium-term Load Forecasting (MTLF)
Medium-term forecasts the load in the range of weeks to a year and is used to ensure the security and capacity constraint.

3) Short-Term Load Forecasting (STLF)
Short Term Forecasting is for the range of one hour to one week ahead and is used to assist planning and market participants. It can help to estimate load flows and to make decisions that can prevent overloading. Appropriate implementations of such decisions lead to the improvement of network consistency and to the reduced occurrences of equipment failures and blackouts. Weather factors that include temperature, humidity, precipitation, wind speed, cloud cover, light intensity etc. often affect consumer's usage of some appliances such as space heater, water heater and air conditioner. Forecasted weather parameters are the most important factors in short-term load forecasts. Various weather variables could be considered for load forecasting. Humidity is also an important factor, because it affects human comfort. Temperature and humidity are the most commonly used factors in load

predictors. That is why THI is sometimes employed as an affecting factor of load forecasting. Fig. 4 presents schematically (Input/output) forecasting relation. So, it illustrates inputs parameters such as past 24 hours load, temperature, humidity ,wind speed, season(month) and day of the week to forecast 24 hours ahead load demands(output)

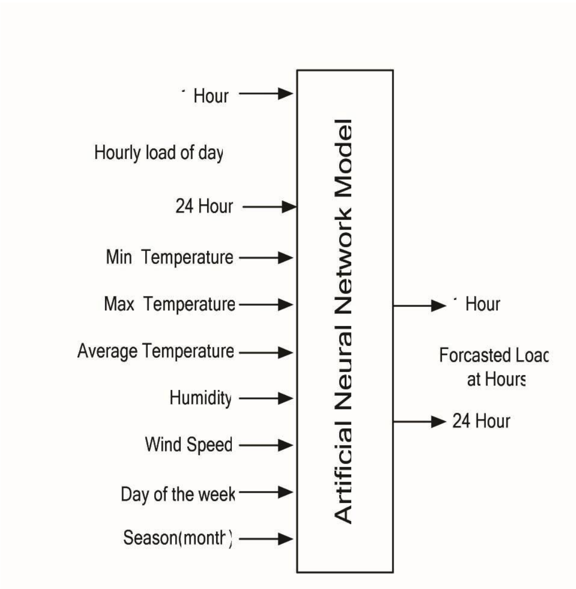

Fig.4 Input-output schematic diagram of a forecasting system

4) Very Short-Term Load Forecasting (VSTLF)

Very Short Term Forecasting is done to forecast the hours and minutes ahead. It is used to assist trading and eventually dispatch.

ANNs are able to perform a nonlinear mapping of the load demand electricity series, which allows the extraction of more complex relationships. These characteristics often make it possible to obtain more precise forecasts [5, 18]. Most of the literature surveyed use multilayer perceptron (MLP) that might be classified into two groups, according to the number of output nodes. The first group is MLP that has several output nodes to forecast a sequence of hourly loads, typically 24 nodes, to forecast next day's 24 hourly loads (this is called the load profile) [10- 13]. The second group is the ones that has only one output node, used to forecast next hour's load, next day's peak load or next day's total load [6, 14- 17].

This paper will focus on Very Short-Term Load Demand Forecasting (VSTLF). Basically, the Artificial Neural Network (ANN) that will be used act as a behavioral model. Based on previous data such as Time, Temperature, Humidity and Load Demand as the output, a behavioral model using ANN will be constructed. Once the behavioral model is concluded it will be used as Load Demand Forecaster.

III. Characteristics of Electrical Loading

Some parameters such as time, random effects and anomalous days are the main factors affecting the VSTLF. *Time:* Electricity demand at the day differs from the demand at the night, and demand during weekdays differs from the demand during weekends. However, all these differences have a cyclic nature, as the electricity demand on the same weekday and time but on a different date is likely to have the same value. Any shift to and from daylight saving time and the start of school year also changes the previous load profiles. *Random effects:* A power system is continuously subjected to random disturbances and transient phenomena. In addition to a large number of very small disturbances, there are large load variations caused by devices such as steel mills, synchrotrons and wind tunnels, as their hours of operations are usually unknown to utility dispatchers. There are also certain events such as strikes, shutdown of industrial facilities and special television programs whose occurrences are not known *a priori* but affect the load. *Irregular days:* These days include public holidays, consecutive holidays, and days preceding and following the holidays, days with extreme weather or sudden weather change and special event days [8].

A good VSLTF system should fulfill the requirement of high accuracy and speed. However, there are several challenges. First, it is not clear how to select the best prediction algorithm and the best feature set. Good feature selection is the key to the success of a prediction algorithm. It is needed to reduce the number of features by selecting the most informative and discarding the irrelevant features. Second, over fitting is a common problem in load prediction, especially for the predominantly used neural network-based prediction algorithms. It means that the error on the training data (the historical data used to build the prediction model) is low but the error on the new data is high. Thirdly, the neural network algorithms have many parameters that require manual tuning and greatly influence their performance.

In this paper, a Multi Layer Feed Forward Neural Network (MLFF) will be used**.** In MLFF network the neurons are arranged in layers and only neurons in adjacent layers are connected. It has a minimum of three layers of neurons; (i) the input layer, (ii) the middle or hidden layer(s), (iii) the output layer. The information propagation is only in the forward direction (input to output) and there are no feedback loops.

Figure 5 shows the architecture (network topology) of Neural Network that will be used in these experiments. The number of neurons at each layer is determined later depending on the experimental condition.

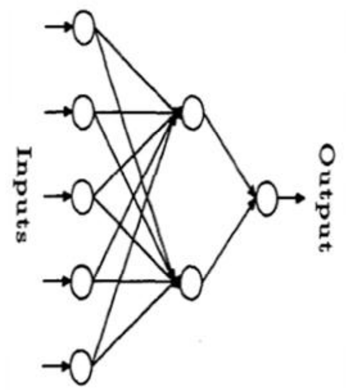

Fig. 5 Architecture of Feed-forward Neural Network (5 inputs, 1 output, and 2 hidden neurons)

In order to obtain bounded output from neurons a sigmoidal activation function is chosen where output is limited to [0, 1] for the input range of [-∞,∞]. MLFF network is trained using back propagation algorithm. The back propagation algorithm adjusts the weights using the relationship given below:

$$w_{ij}(new) = w_{ij}(old)\eta\delta_i O_j + \alpha(\Delta w_{ij}(old)) \quad (5)$$

Where

$$\delta_i = (t_i - O_i)O_i(1 - O_i) \quad (6)$$

for output layer neurons

$$\delta_i = O_i(1 - O_i)\sum_{l=1}^{n_o} w_{li}\delta_l \quad (7)$$

for hidden layer neurons

t_i is the target for i[th] output neuron.
O_i is the output of i[th] neuron.
n_o is the number of output neuron.
η is the learning coefficient.
α is the momentum factor.

IV. Behavioral model using Neural Network

Dataset is fundamental and it is unwise to use all the available examples in the training set to estimate the neural network parameters. As shown in Figure 6, a dataset of observations is divided into two sub-data sets: training and forecasting set. The forecasting dataset is not used during the neural network training procedure.

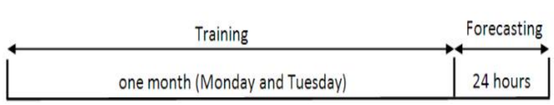

Fig. 6 Division of dataset into two parts

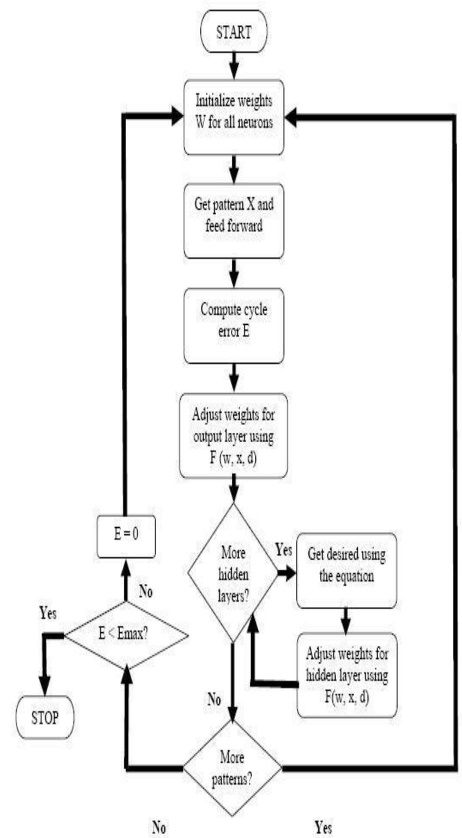

Fig. 7 : Flowchart showing working of the back-propagation algorithm.

The size of the training dataset is very important for good performance because the neural network obtains information from the training set. If the training dataset does not cover the full range of operating conditions, the model may perform badly when deployed. Under no circumstances should the training set be less than the number of weights in the neural network. A good size of the training dataset is ten times the number of weights in the network, with the lower limit being set around three times the number of weights. The size of the neural network topology should also be carefully selected. If the number of layers or the size of each layer is too small, the network does not have enough degree of freedom to classify the data to approximate the function, and the performance suffers. Vice versa, if the size of the network is too large, performance may also suffer.

The generalization error can be decomposed into the sum of the bias squared plus the variance. A model which is too simple, or too inflexible, will have a large bias or under fitting, while one which has too much flexibility in relation to the particular dataset will have large variance or over-fitting. The best generalization is obtained when we have the best compromise between the conflicting requirements of small bias and small variance.

The purpose of building a behavioral model using artificial neural networks is to make a good prediction for new inputs, not to learn an exact representation of the training data itself.

V. Computer based Experiments

- **Preprocessing Data**

From the previous section, it is known that time, random effects and anomalous days are some of the main factors influencing the VSTLF. In order to achieve a good behavioral model and an accurate Load Demand forecasting, it is necessary to do some preprocessing data. First, we must look for the dynamical behavior in our raw data. Variable selection plays a critical role in building a good forecasting model. It is important to first analyze the data to ensure all essential variables are included.

In this paper, hour, temperature and humidity will be used as an input variable to the model in order to produce or forecast Load Demand electricity. After using some statistical computation, it is decided that only Monday and Tuesday electricity demand will be used as data for this experiment, both for the Training and Forecasting steps, because the behavior of Load Demand electricity for these days are nearly the same.

- **Training**

Training procedure is the first step in building a behavioral model. Inputs for the behavior model are hour, temperature and humidity. Meanwhile, the target or desired output is Load Demand electricity. So there are three inputs and one output. As stated in the subsection before, the number of dataset for the training step is 192 whereas 24 dataset is prepared for evaluating the prediction's capability.

Figures 8 and 9 show the Load Demand electricity for Training and Forecast procedure. These figures only depict the Load Demand electricity. Input for the Load Demand is a vector that consists of hour, temperature and humidity data.

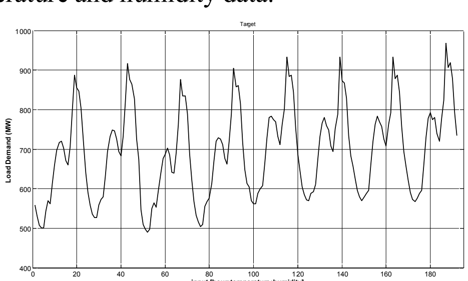

Fig. 8 Load Demand graph for Training step

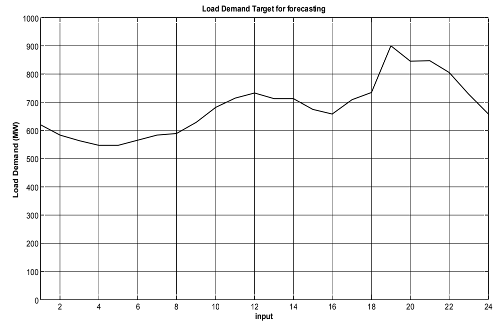

Fig. 9 Load Demand graph for Forecast step

At this training step, gradient descent and Levenberg Marquardt method are used to train the Artificial Neural Network [20]. Some experiments are done to determine which training method is better. The input structure and number of hidden layer that yield the best performance for the model has 1 output node, 2 hidden layers 10 and 15 neurons, and 3 inputs. Figures 8 and 9 show the results for this training step.

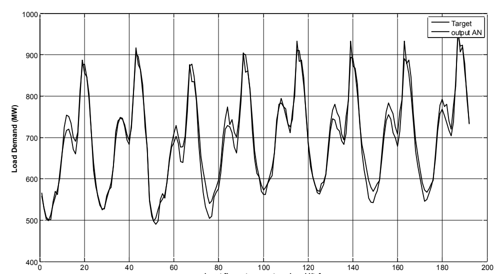

Fig. 10 Training results using gradient descent method. 4000 Iteration, Index Performance: 0.0530

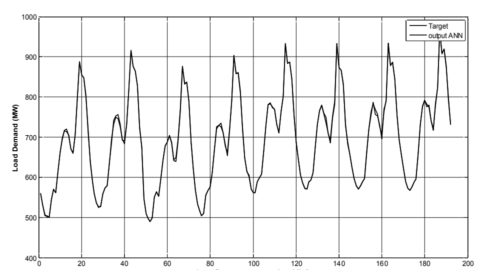

Fig. 11 Training results using Levenberg Marquardt method, 148 iteration, Index Performance: 0.000990

Two training methods are compared in terms of iteration and index performance parameter. It is concluded that Levenberg-Marquardt method is a powerful tool for neural network modeling. Therefore at the forecasting step, this method will be used as an update weight value.

- **Forecasting**

Based on the experiments, a network topology that is constructed by 3 inputs, 2 hidden layers 10 and 15 neurons and 1 output using Levenberg Marquardt method gives the best results during training [18]. The most suitable topology for 24 hours ahead forecasting of Load Demand electricity is constructed by 3 inputs, 2 hidden layers and 4 neurons and 1 output.

As stated, a dataset which consists of 24 vectors; hour, temperature and humidity; will be used as an input data. However, further study needs to be conducted to analyze the relation of topology during training and prediction process.

In order to measure the accuracy of forecasted Load Demand electricity 24 hours ahead, Mean Average Percentage Error (MAPE) will be used, which is defined as:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t-\hat{y}_t|}{y_t} \times 100\% \quad (8)$$

Where, y_t is actual load, $\hat{y}_t$ is predicted load and n is number of the forecasted target. Figure 12 shows two graphics of Load Demand electricity. The dashed line is a graphic for Real Power Load Demand electricity. The solid line is a graphic for forecasted Load Demand. It is shown that forecasted Load Demand is almost the same as real power Load Demand with MAPE equals to 3.5704%.

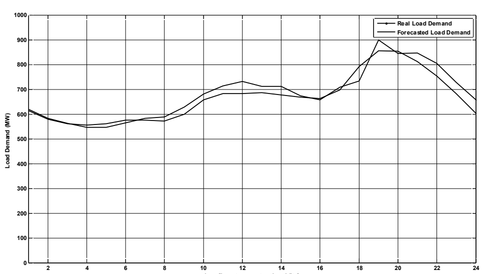

Fig. 12 Real and Forecasted Load Demand

, MAPE = 3.5704 %

VI. Conclusion

The result of MLP network model used for one day ahead short term load forecasting shows that MLP network has a good performance. a reasonable prediction accuracy was achieved for this model. Its forecasting reliabilities were evaluated by computing the mean absolute error between the exact and predicted values.

The results suggest that the Levenberg-Marquardt method is a powerful tool for neural network modeling. ANN model with the developed structure can perform good prediction with least error and finally this neural network could be an important tool for short term load forecasting.

The developed neural network model is used for the prediction of hourly load demand and the forecasting results are practically acceptable with MAPE 3.5704%.

VII. References

[1] Haykin, S., Neural Networks (1999): A Comprehensive Foundation, 2nd Ed., Prentice-Hall, Englewood Cliffs, NJ,.

[2] http://www.i-programmer.info/babbages-bag/325-mcculloch-pitts-neural-networks.html

[3] Gualtiero Piccinini (2004): The First Computational Theory of Mind and Brain: A Close Look at McCulloch and Pitts' Logical Calculus of Ideas Immanent in Nervous Activity. Synthese, 141 (2):175-215.

[4] Marvin L. Minsky and Seymour A.(1988): Papert. Perceptrons: An Introduction to Computational Geometry, Expanded Edition. Cambridge, MA: MIT Press, 1988. 292pp.

[5] HINTON, G. E. (1984): Distributed Representations.CMU-CS-84-157, ittsburgh, PA: Carnegie-Mellon University, Computer Science Department.

[6] Gross and F.D. Galiana(1987):Short term load forecasting. IEEE Proceedings, vol. 75, no.(12): , pp. 1558-1573, Dec..

[7] Bunn D. W. (1982): Short term forecasting: a review of procedures in the electricity supply industry. Journal of the Operational Research Society, vol. 33, pp.: 533-545.

[8] Setiawan A, Koprinska I, and Vassilios G. Agelidis (2009):Very Short-Term Electricity Load Demand Forecasting Using Support Vector Regression. Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, June 14-19,

[9] Mandal J.K., A.K. Sinha(1995): Artificial Neural Network Based Hourly Load Forecasting for Decentralized Load Management", IEEE Catalogue,.

[10] Lee, K.Y. Cha, Y.T. and Park, J.H. (1992): Short term load forecasting using an artificial neural network. IEEE Transaction on Power Systems, 7(1): pp. 124-132.

[11] Kiartzis, S.J. *et al.* (1997):Short term load forecasting in an autonomous power system using artificial neural networks. IEEE Transactions on Power Systems, 12(4): pp. 1591-1595.

[12] Chow, T.W.S. and Leung, C.T. (1996): Neural network based short term load forecasting using weather compensation. IEEE Trans on Power System, 11(4): pp. 1736-1742.

[13] Mohammed, O. *et al.* (1995): Practical experiences with an adaptive neural network short term load forecasting system. IEEE Transaction on Power Systems, 10(1): pp. 254-262.

[14] Park, D.C. *et al.* (1991): Electric load forecasting using an artificial neural network. IEEE Transaction on Power Systems. 16 (May): pp. 442-449.
[15] Ho, K.L. Hsu, Y.Y. and Yang, C.C. (1992): Short term load forecasting using a multilayer neural network with an adaptive learning algorithm. IEEE Transaction on Power Systems, 7(1): pp. 141-149.
[16] Chen, Shin Tzo.Yu, David C. Moghaddamjo, A.R. (1992): Weather sensitive short term load forecasting using non-fully connected artificial neural network". IEEE Transaction on Power System, 7(3): pp. 1098-1105.
[17] Drezga, I. and Rahman, S. (1998): Input variable selection for ANN based short term load forecasting. IEEE Transactions on Power Systems,13(4): pp. 1238-1244.

3/1/2012

[18] Othman M. F., T. Andromeda, Saeed Badran (2009):Application of Artificial Neural Network for Load Demand Forecasting based on Hour, Temperature and Humidity Data. EEIES 2009, Penang, Malyasia.
[19] Mohsen H., and Yazdan S. (2007): Artificial Neural Network Approach for Short Term Load Forecasting for Illam Region. published at World Academy of Science, Engineering and Technology ,28.
[20] Amir Abolfazl Suratgar, Mohammad Bagher Tavakoli, and Abbas Hoseinabadi (2005): Modified Levenberg-Marquardt Method for Neural Networks Training. published at World Academy of Science, Engineering and Technology-6 -.

3/1/2012