SHORT-TERM LOAD FORECASTING USING ARTIFICIAL NEURAL NETWORK TECHNIQUES: A CASE STUDY FOR REPUBLIC OF NORTH MACEDONIA

*Ana Kotevska, Nevenka Kiteva Rogleva**

Faculty of Electrical Engineering and Information Technologies –Skopje
North Macedonia

* Corresponding Author, e-mail: *nkiteva@feit.ukim.edu.mk*

Abstract: Modernization and liberalization of power system in North Macedonia offers an opportunity to supervise and regulate the power consumption and power grid. This paper proposes models for short-term load forecasting using artificial neural network in order to balance the demand and load requirements and to determine electricity price. Neural network approach has the advantage of learning directly from the historical data. This method uses multiple data points. In this research, it can be confirmed that the quality of the short term prediction depends on the size of the data set and the data transformation.

Key words: Artificial Neural Network (ANN), Short Term Load Forecasting (STLF), Back Propagation, Mean Absolute Percentage Error (MAPE).

1. INTRODUCTION

Applying intelligent surveillance techniques to various processes is a characteristic approach of the digital age. One example is the proposed in [1] system on the Internet of Things based secure application with IBM cloud platform. Most of the applications in this area are based on Artificial Neural Network (ANN) modelling [2] or use machine learning algorithms [3].

Artificial Neural Networks are able to recognize symbols, emotional expressions presented by graphical images, as well as more complex sequences of symbols and speech [3]. The paper [4] suggests models for short-term load forecasting (STLF) based on ANN techniques. ANNs are capable for modeling and adaptation. Furthermore, they do not require a functional link between the load data and the meteorological data. Load forecasting is crucial for the effective and efficient operation of any power system.

Power system planning ensures reliable operation of existing power grids. Reliability depends on load demand and load fluctuation, power generation units, load interconnections, transmission facilities, protection schemes etc. Distribution system operator (DSO)/Transmission system operator (TSO) must be able to predict the load flow at real time and in different locations of the power system in order to define generation scheduling, electricity transaction, etc. The aim of this paper is to confirm that the crucial impact for short-term load forecasting, has the model, the input parameters and their processing. For development of these models, it is necessary to have statistics knowledge, in order to understand and analyze the obtained values [5, 6] from the models and make necessary modifications in order to improve the model.

The paper is organized as follows. Section 2 includes a presentation and summary analysis of related work in the field of simulation of ANNs [3, 5]. In this section the backpropagation algorithm is explained along with the statistical tools. An ANN models for STLF are presented in Section 3. Three models are designed and implemented in Python [7]. The whole data set contains records of 2 years. Historical data set, weather data and type of month, day and time are considered. They are used to predict every load hour of the forecasted day. Temperature and humidity are the most frequently used meteorological predictors. Temperature is expected to have a significant impact on the consumption, regardless of if it occurs during peak or off-peak hours. Very often, temperature can explain more than 70% of the consumption variations [4].

Simulation results and related discussion are presented in Section 4. The results show that the size of the data set has a crucial role for STLF model accuracy.

2. ARTIFICIAL NEURAL NETWORKS

ANN has the advantage of learning directly from the historical data. Once the neural network is trained with historical data, predictions of future consumption can be made. Back propagation architecture is used for short-term load forecasting because of the: continuously valued functions and supervised learning [8, 11].

A representation of a typical multilayer neural network (Fig. 1) consists of three types of layers: input, output and hidden layer (1 or 2), which is located between the input and output layer.

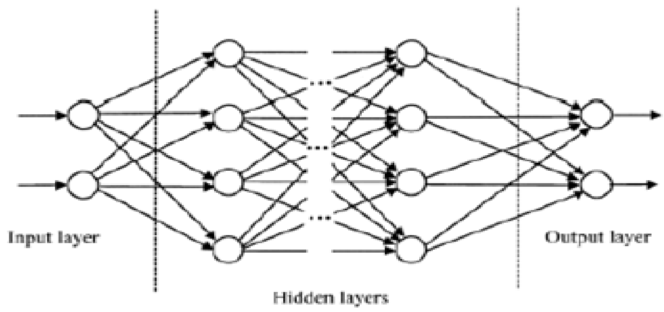

Figure 1. Multiple neural network

The information received through the input layer nodes is processed and delivered to the nodes of the hidden layers, where the information is further processed. The processed data is forwarded to the output layer nodes to predict the future power consumption. Each layer consists of unconnected neurons. The neurons from two adjacent layers are connected by the corresponding weights. Weights (w_1, w_2,..w_n) are optimized when back propagation algorithm is used [3]. Supervised learning involves comparing the network output (NO) with the desired output (DO) (eq.1) while changing the corresponding weights (w_i) to achieve an appropriate mapping [3]. This means that the output error is propagated back through the network, changing the internal weights along the way in order to simulate a lower error between the predictions and the actual data.

$$Error = \frac{1}{2}\sum_{i=1}^{n}(DO_i - NO_i)^2 \tag{1}$$

ANN is a type of array that can perform non-linear inputs-output mapping. Input variables can be categorized into: historical load demand, temperatures, time of day, day of week, wind speed, humidity. Input variables classification depends mostly of the engineering assessment and experience [4].

So, one of the key features to design a good ANN architecture is to define the right input parameters. In STLF, these parameters can be categorized as consumption, weather and time information. Consumption information may include historical consumption. Weather information mostly includes temperature. Time information may include month, day and time of the day. For counties with low climatic difference, historical consumption and temperature are sufficient to make a good prediction of short-term consumption [8, 10].

Another issue is the determination of the optimal number of hidden neurons. Small data set means a lack of information and inaccurate predictions; large data results in a very long training process. However, the corresponding number of hidden neurons mainly depends on the data set and the number of input parameters [4]. There is no theoretical approach to calculate the right number of nodes in the hidden layers. This number can be determined using a similar approach as the training in epochs, meaning with examining the mean square error over a set of different numbers of nodes from the hidden layers, the one that gives the smallest error is chosen [9].

An important feature of ANN is the model training. In general, the error decreases after several training epochs, but there are some exception when error increases because the network begins to memorize the specifics of the training data. The epoch with the lowest validation error is usually chosen and this is achieved with a smaller number of epochs [12]. The mean absolute percentage error (MAPE) is used to determine the error.

The error of back propagation loop consists of two passes: forward and backward. When spreading forward, the input vector is applied to the network nodes, and its effect spreads through the network. Finally, a set of output values is obtained in response to the network. All weights on the network are adjusted according to the error correction rule, during the backward spreading. The weights are adjusted to bring the actual network response closer to the desired response [8, 11].

Training will stop when the error starts to increase. The parameters corresponding to the lowest error for the validation data are locked as last for the ANN model. There are several factors that can influence the structure of the ANN model, such as input variables, number of hidden neurons and hidden layers, interconnections, etc. Multiple ANN structures are usually tested using the training and validation data, and then the one with the lowest validation error is selected and the performance of the test data is reported.

3. ANN TECHNIQUES FOR STLF

This aim of this research is to create a model that can predict load demand 24 hours ahead and gained results to compare with the actual values recorded within the forecasted period. First, for each year datasets are queried from the database. The used variables are: hourly historical consumption, average daily temperature and type of hour, day and month. As a temperature representative in this study is used the average daily temperature in Skopje [7].

Data sets undergo a preprocessing (data cleaning and transforming) in order to eliminate noisy signals and then the descriptive statistics are calculated. The main problem in the simulation model is the quality of real data such as lack of values, due to SCADA or meteorological station problems. These missing data are determined based on some linear extrapolations.

3.1. ANN model

For this model is used 4-layer ANN back propagation algorithm: one input, two hidden and one output layer. Three models are created:

Model 1: 46 neurons are used in the input layer: the first 7 neurons indicate the type of day (Monday to Sunday), the next 24 neurons are the type of hour, the next 12 neurons are the type of month and 1 neuron is used for the consumption from the previous day, temperature from the previous day and temperature for the forecasted day. There are 2 hidden layers. The output layer contains 1 neuron, i.e. the forecasted hour.

Model 2: 46 neurons are used in the input layer: the first 7 neurons indicate the type of day (Monday to Sunday), the next 24 neurons are the type of hour, the next 12 neurons are the type of month and 1 neuron is used for the consumption from the previous week respectively for the same type of day, temperature from the previous week respectively for the same type of day and temperature for the forecasted day. There are 2 hidden layers. The output layer contains 1 neuron.

Model 3: 27 neurons are used in the input layer: the first neuron indicates the type of day (the day of the week is determined by a number in the range from 1 to 7, 1 neuron is used for the previous day's temperature, 1 neuron is used for the temperature of the forecasted day and the next 24 neurons represent the consumption of the previous day. There are 2 hidden layers. The output layer contains 24 neurons.

3.2. Data transformation

Some machine learning algorithms perform better if the data has a constant distribution. This can be achieved with two techniques: normalization and standardization.

- **Data normalization** is transformation of data from the original range so that all values are in the range between 0 and 1. It is necessary to calculate the minimum and maximum value of the monitored data (training data set).
- **Data standardization** involves preserving the values distribution. This technique is usually used when the input values are in different scales. Standardization assumes that the data fits into the Gaussian distribution. The mean and standard deviation of the monitored values (training data set) should be calculated.
- **Cyclical data characteristics**: One of the most common features in a database is the data cyclist. Usually this is the time data such as: months, days and hours, which occur in specific cycles. When using data with cyclic features, it is important how they will be represented so that the algorithm knows that these features occur in cycles, as it has a significant effect on the algorithm convergence rate. For example, if it is analyze an hour during a week, there is a cycle between 0 and 24, repeating 7 times. The problem with the presentation of cyclical data is the jump that occurs at the end of each day when the hourly value falls to 0. For example, from 22:00 to 23:00, an hour is passed and the absolute difference is the same one. But from 23:00 to 00:00, although the difference is one hour, the absolute difference is 23. To overcome this problem, sin and cos data transformation are used. It is necessary to use the both functions and unique values will be obtained. The benefit of this transformation is that the range of values is between [-1, 1], which will help to improve the neural network.

3.3. ANN models with transformed data

Based on the above data processing and their corresponding transformation, new models are created:

I. Using cyclic transformation and data normalization

Model N1: 9 neurons are used in the input layer: the first 2 neurons indicate the hour of the day, the next 2 neurons are the type of day and the next 2 neurons are the type of month, represented with sin and cos values according to the cyclic transformation. The next neurons for consumption from the previous day, temperature from the previous day and temperature for the forecasted day are normalized. The neural network has 2 hidden layers and the output layer has 1 neuron, i.e. the forecasted hour.

Model N2: 3 neurons are used in the input layer: 1 for consumption from the previous day, 1 for temperature from the previous day and 1 for the temperature for the forecasted day, which are normalized. There are 2 hidden layers. The output layer contains 1 neuron, i.e. the forecasted hour.

Model 3: In the input layer is used 1 neuron for the consumption from the previous day, which is normalized. There are 2 hidden layers. The output layer contains 1 neuron, i.e. the forecasted hour.

II. Using cyclic transformation and standardization of data

Model S1: 9 neurons are used in the input layer: the first 2 neurons indicate the hour of the day, the next 2 neurons are the type of day and the next 2 neurons are the type of month, represented with sin and cos values according to the cyclic transformation. The next neurons for consumption from the previous day, temperature from the previous day and temperature for the forecasted day are standardized. The neural network has 2 hidden layers and the output layer contains 1 neuron, i.e. the forecasted hour.

Model S2: In the input layer are used 3 neurons: 1 for consumption from the previous day, 1 for temperature from the previous day and 1 for the temperature for the forecasted day, which are standardized. There are 2 hidden layers. The output layer contains 1 neuron, i.e. the predicted hour.

Model S3: In the input layer is used 1 neuron for consumption from the previous day, which is standardized. There are 2 hidden layers. The output layer contains 1 neuron, i.e. the forecasted hour.

3.4. MAPE

Mean Absolute Percentage Error (MAPE) is used to measure the error in terms of percentage. It is calculated as the average of the absolute percentage error. It can be calculated as follows:

$$MAPE = \frac{\sum \frac{|Actual_t - Forecast_t|}{|Actual_t|}}{n} * 100 \qquad (2)$$

4. RESULTS

In this paper, the forecasted day is 09.09.2015 and the forecast accuracy is measured with MAPE.

4.1. ANN model

For Model 1 and Model 2, the training data set contains about 14,000 records, the validation data set around 1,000 records and the test data set is around 1,000 records. For Model 3, the training data set contains about 500 records, the validation data set around 100 records and the test data set is around 100 records. The models are developed in Python using the Keras library. The parameters of the models are explained below:

- The input layer receives 46 or 27 neurons (depending on the model) and gives 150 neurons to the output. The 'relu' function is used as an activation function.
- The first hidden layer gives 80 neurons and it is used 'relu' as activation function.
- The second hidden layer gives 30 neurons and it is used 'sigmoid' as activation function.
- The output layer gives 1 or 24 neurons (depending on the model) and as activation function is used 'linear' function.
- Before starting the training of the model, the learning process needs to be set with the 'compile' method. This method has three parameters: optimizer (in the model is used Adam algorithm), loss function (in the model is used MSE, the purpose is to try to minimize this loss) and a metric (in the model is used MAPE).
- The training of the model is done with the function 'fit', which has the training data set, the size of the series, the number of epochs and the validation data set.
- The 'evaluate' function calculates the loss based on the input data.
- The 'predict' function generates the forecasted values based on the input data.

On Table 1 are given the calculated MAPE for each model and it can be concluded that Model 2 has the lowest MAPE, which is 3.21% (model is correct about 96.79%) [11]. Figure 2 shows the graph that compares the forecast and actual values with ANN models 1 and 2.

Table 1. MAPE for ANN models

Models	***Model 1***	***Model 2***	***Model 3***
MAPE	*4.93%*	*3.21%*	*4.70%*

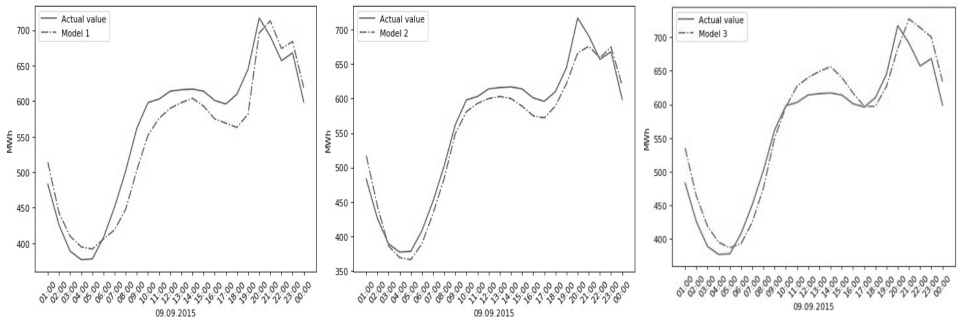

Figure 2. Comparison between the forecast and real values with ANN models

Table 1 MAPE for ANN models with normalized data

Models	***Model N1***	***Model N2***	***Model N3***
MAPE	*8.17%*	*6.58%*	*4.96%*

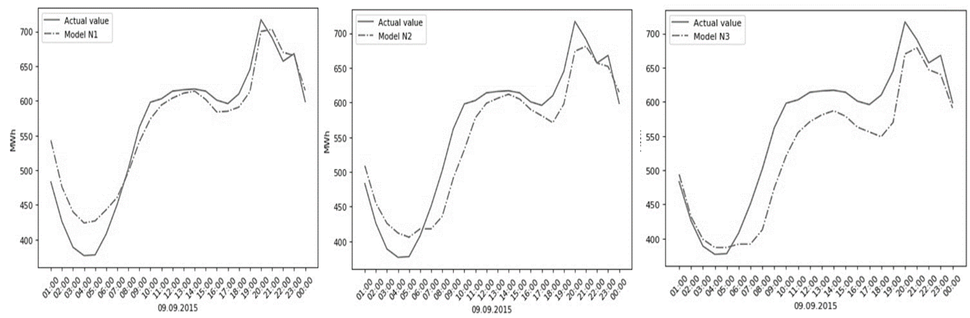

Figure 3. Comparison between the forecast and real values with normalized data

From the graph, it can be noticed that Model N1 gives good forecast values from 06:00 to 00:00 and Model N3 gives good forecast values from 01:00 to 06:00. So as a future work, combination of these two models can be analyzed.

Normalization is done for each data set separately (training, validation and test). The results of the respective models are given on Table 3, i.e. the calculated MARE for each model.

Table 2 MAPE for ANN models with standardized data

Models	***Model S1***	***Model S2***	***Model S3***
MAPE	*8.47%*	*7.49%*	*5.38%*

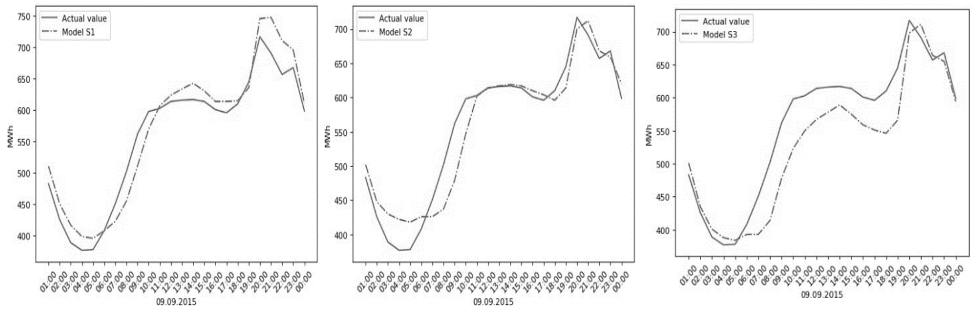

Figure 4. Comparison between the forecast and real values with standardized data

4. CONCLUSION

The short-term load forecasting using artificial neural network techniques has been applied on two years load data from power system operators in Republic of North Macedonia. This paper gives a comparative study of different models with artificial neural networks. For development of these models, it is required a solid understanding of how the artificial neural network works, determining and adjusting the parameters, meaning how many hidden layers are needed or how many neurons, which algorithm is best to use, etc. Neural networks are using learning method, meaning their output is not always the same, although the same input data and parameters are used. The network is learning better if there is more data, such as type of hour, day (within the week or if it is working, non-working or a holiday), month, previous consumption, previous temperature and forecasted temperature (or other weather information). Model 2 has shown better results than the other presented ANN models, if they are compared in terms of error measurements with mean absolute percentage error of having 3.21%. The size of the data set has a crucial role, since it has shown a reduction of the error measurement, having the mean absolute percentage error of 2.64%. This study presents usable and accurate models for short-term load forecasting which is very important for the operation of any power system.

REFERENCES

[1] I. Adamopoulos, A. Ilias, Ch. Makris, Y. C. Stamatiou. Intelligent surveillance systems on the Internet of Things based on secure applications with the IBM cloud platform. *International Journal on Information Technologies and Security*, vol. 15, no. 2, 2023, pp. 59-74. DOI: https://doi.org/10.59035/XVRS3592

[2] Ukov, T., Tsochev, G. Machine learning algorithm for intelligent bots in multiplayer video game: A case study. *International Journal on Information Technologies and Security*, vol. 14, no. 4, 2022, pp. 67-178.

[3] Yoana Ivanova. Optimizing simulation models of an artificial neural network for digital recognition, *International Journal on Information Technologies and Security*, No. 4 (vol. 13), 2021, pp. 59-70.

[4] Y. Rui, A.A. El-Keib. *A review of ANN-based Short-Term Load Forecasting models*, Tuscaloosa, AL 35487.

[5] Romansky, R. Evaluation of experimental data from monitoring and simulation of network communication parameters. *International Journal on Information Technologies and Security*, vol. 14, no. 2, 2022, pp. 75-86.

[6] Romansky, R. Statistical analysis of empirical network traffic data from program monitoring. *International Journal on Information Technologies and Security*, No. 3 (vol. 14), 2022, pp. 15-24.

[7] Kotevska Ana. Analysis and models for Short Term Load Forecasting in Republic of North Macedonia (master thesis).

[8] Manoj Kumar. *Short-Term Load Forecasting using Artificial Neural Network techniques*, 2012.

[9] Muhammad Buhari, Sanusi Sani Adamu. Short-Term Load Forecasting using Artificial Neural Network, *International MultiConference of Engineers and Computer Scientists*, March 2012, vol. 1, 2012.

[10] Heiko Hahn, Silja Meyer-Nieberg, Stefan Pickl. Electric load forecasting methods: Tools for decision making. *European Journal of Operational Research*, Elsevier, vol. 199, no. 3, 2009, pp. 902-907.

[11] Swaroop R., Hussein A. Abdulqader. Load forecasting for power system planning and operation using Artificial Neural Network at Al Batinah region Oman. *Journal of Engineering Science and Technology*, vol. 7, no. 4, 2012, pp. 498-504.

[12] Swasti R.Khuntia, Jose Luis Rueda, Mart Meijden. Neural Network based load forecasting and error implication for Short term horizon, *Conference Paper*, July 2016, DOI:10.1109/IJCNN.2016.7727854

Information about the authors:

Ana Kotevska – Works as a Scrum Master on Agile transformation in the healthcare domain. The scientific activities currently focus on agile methodology, agile transformations, agile metrics for predictability, change management and leadership.

Nevenka Kiteva Rogleva – Associate Professor in Dept. of Power Plants and Switchgears at Faculty of Electrical Engineering and Information Technologies, Ss. Cyril and Methodius University in Skopje, North Macedonia, areas of scientific research - power plants, project management, risk management, agile methodology.

Remark:

Manuscript received on 30 May 2023 and is accepted after double-blind reviewing to take part in the 37th International Conference on Information Technologies (InfoTech-2023), IEEE conference, Rec. # 58664, Section E: “Technologies of System Design” and has not been published in full text elsewhere.