

Grid search of multilayer perceptron based on the walk-forward validation methodology

Tran Thanh Ngoc¹, Le Van Dai², Dang Thi Phuc³

^{1,2}Faculty of Electrical Engineering Technology, Industrial University of Ho Chi Minh City, Vietnam

³Faculty of Information Technology, Industrial University of Ho Chi Minh City, Vietnam

Article Info

Article history:

Received Jul 8, 2020

Revised Jul 27, 2020

Accepted Oct 15, 2020

Keywords:

Accuracy scores

Grid search model

Multilayer perceptron

Walk-forward validation

ABSTRACT

Multilayer perceptron neural network is one of the widely used method for load forecasting. There are hyperparameters which can be used to determine the network structure and used to train the multilayer perceptron neural network model. This paper aims to propose a framework for grid search model based on the walk-forward validation methodology. The training process will specify the optimal models which satisfy requirement for minimum of accuracy scores of root mean square error, mean absolute percentage error and mean absolute error. The testing process will evaluate the optimal models along with the other ones. The results indicated that the optimal models have accuracy scores near the minimum values. The US airline passenger and Ho Chi Minh city load demand data were used to verify the accuracy and reliability of the grid search framework.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Tran Thanh Ngoc

Faculty of Electrical Engineering Technology

Industrial University of Ho Chi Minh City

12 Nguyen Van Bao, Ward 4, Go Vap District, Ho Chi Minh City, Vietnam

Email: tranthanhngoc@iuh.edu.vn

1. INTRODUCTION

Load forecasting is extremely important for electrical system by means of being the key process used to determine the optimal schedule for electricity generation, it is pivotal to ensure the operational reliability of the power grid and load forecasting plays a leading role in energy trading and investment activities. However, it should be emphasized that load forecasting is a very challenging task owing to the complicated and unpredictable nature of human behaviour, causing difficulties in predicting future electricity demand. Therefore, the more reliable models are required to achieve better forecasting results [1-6]. Accordingly, several models have been utilized for load forecasting, including multiple regression, exponential smoothing, stochastic time series, fuzzy logic, neural networks and knowledge-based expert systems [7-21]. Among these models, the multilayer perceptron (MLP) belonging to the Neural has been widely used. MLP architecture is similar to that of biological neural networks consisting of one input layer, one or more hidden layers and one output layer [15-21].

Several hyperparameters of MLP structure and MLP training are also used which influences on the quality of MLP model such as batch size, training epochs, optimization algorithm, network weight initialization, activation function in the hidden layer, and the number of neurons in the hidden layer. Thus, finding out good hyperparameters is an important procedure when using MLP model for load forecasting [22-25]. Besides, in order to apply MLP model, the common way is to split the data into training and test sets, which are used to build the forecast model and to evaluate the accuracy of forecast values, respectively. In addition, the testing set is fixed during forecast operation. Unlike the traditional way, the walk-forward

validation (WV) methodology allows to get the best forecasts at each time step [25]. In present paper, the WV methodology for the MLP model is proposed, then applied for the grid search model of these hyperparameters. The US airline passenger and Ho Chi Minh city load demand data were analysed through training and testing processes to verify the accuracy and reliability of the grid search framework. Some of the accuracy scores such as root mean square error (RMSE), mean absolute percentage error (MAPE) and mean absolute error (MAE) were also used. In the training process, the optimal models with certain hyperparameters was obtained to satisfy the requirement for minimum of the accuracy scores. In the testing process, these optimal models will be compared to other ones to confirm the grid search process. The models are implemented using Keras library with Tensorflow as backend in the Python environment with Google colab, a free GPU on the cloud for running large scale machine learning projects [25-27].

This paper is organized as follows. In section 2, a short introduction to MLP modelling and WV methodology are presented along with proposing the grid search of MLP WV methodology. In section 3, we focus on experiments and analyse the results. The conclusions are given in section 4.

2. RESEARCH METHOD

2.1. MLP neural network

Artificial neural networks in general, or MLP network in particular are built based on mathematical models of the way brains are thought to work. A typical MLP network consisting of an input layer, one hidden layer, and output layer; and a first node in a hidden layer network are shown in Figures 1(a) and 1(b), respectively [18-20, 25-27].

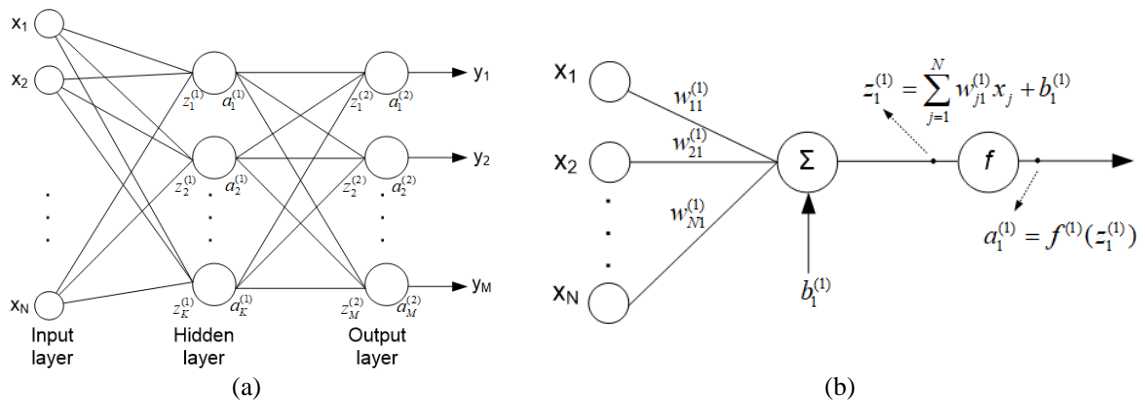


Figure 1. (a) A typical MLP network, (b) one node in hidden layer

The output of node i in hidden layer is obtained by:

$$a_i^{(1)} = f^{(1)}(z_i^{(1)}) = f^{(1)}\left(\sum_{j=1}^N w_{ji}^{(1)} x_j + b_i^{(1)}\right), \forall i = 1, \dots, K \quad (1)$$

The i^{th} output of MLP is obtained by the same way:

$$y_i = a_i^{(2)} = f^{(2)}(z_i^{(2)}) = f^{(2)}\left(\sum_{j=1}^K w_{ji}^{(2)} x_j + b_i^{(2)}\right), \forall i = 1, \dots, M \quad (2)$$

where,

$x_j, j=1, \dots, N$ is the input; $y_j, j=1, \dots, M$ is the output

N is the dimension of input, M is the dimension of output, K is the number of node in the hidden layer.

w, b are the weight term and bias term of node; f is the active function of node.

The superscripts in the parameters $w^{(1)}, b^{(1)}, a^{(1)}, z^{(1)}, f^{(1)} \forall l = 1, 2$ indicate the values for the first layer (hidden layer) and the second layer (output layer). In the training process of MLP network, data comprising the input and output are fed into the network to construct an input-output mapping by adjusting

the weights and biases at each iteration based on an optimization problem which is solved to minimize a loss function. The loss function can be defined as [22, 28]:

$$L(\theta) = E(x, y; \theta) + R(\theta), \forall \theta = (w, b) \quad (3)$$

The error function E in the (3) is usually defined by the difference between the actual and desired output. The error function E of the MLP is normally chosen to be the mean square error, as given by:

$$E(x, y; \theta) = \frac{1}{M} \sum_{i=1}^M |y_i - \hat{y}_i|^2 \quad (4)$$

The regularization term R in (3) introduces a bias to reduce overfitting on the training data. Gradient descent (GD) algorithm is an iterative machine learning optimization algorithm to minimize the loss function. It is noticed that there are many gradient descent optimization algorithms as SGD, RMSprop, Adagrad, ...

2.2. Walk forward validation methodology

In load forecasting practice, the WFV methodology gives the forecasting model with the best opportunity to make good forecasts at each time step. The sequential operation of the WFV methodology is shown in Table 1. Firstly, using the history data (Months) for training, the forecasting model makes a load forecasting for the next month (Month1) and then stored or evaluated against the known value. Continuously, the training data is expanded to include the known value (Months + Month1), the forecasting model is updated and the next week is forecasted (Month2). The process is repeated to the end to ensure that the training data was updated with the unknown values at each step [25].

Table 1. The rolling of data in the Walk Forward Validation methodology

History data	Predictive data
[Months]	Month1
[Months + Month1]	Month2
[Months + Month1 + Month2]	Month3

2.3. The grid search of MLP WFV methodology

2.3.1. Tuning hyperparameters of MLP

There are many aspects of the MLP influence on the quality of the MLP network. They can be parameters of the data (such as features of input data, differencing of input data, ect.), or parameters of MLP configuration (such as number of hidden layers, number of nodes in each layers, activation functions, ect.), as well as parameters in the training process (such as optimization algorithms, batch, epoch, ect.). In our study, we focus on these hyperparameters to tune as follows [25].

a. Number of input (i)

The number of input is pre-specified by the available data. For the load forecasting, the data is usually considered in the form of a univariate electric load time series with N observations $\{X_{t1}, X_{t2}, \dots, X_{tN}\}$, while the number of input i represents the number of lag observation used as input and defined as the dimension of input features as well. For example, for a monthly load time series, the input and output of MLP model with number of input i equal to 3 are described in Table 2.

Table 2. The input and target values for times series

X : Inputs	Y : Output
month1, month2, month3	month4
month2, month3, month4	month5
month3, month4, month5	month6

The input variables of the MLP model are named by the lagged time series values (X_{t-i+1}, \dots, X_t), and the predicted output values are referred to the next values (X_{t+1}). Because of seasonality of load times series, the number of input i can be chosen as the multiples of period of time series. For instance, with the

period of hourly time series of 24, the period of weekly time series is 7, the period of yearly time series is 12, and so on.

b. Number of nodes (n)

In most situations, there is no way to determine the best number of hidden nodes. This is related to the fact that the utilization of too few nodes may lead to high error while too many nodes - overfit. In this study, we use MLP network with one hidden layer, and then tune the number of nodes in this hidden layer.

c. Number of batch (b)

The batch size is the number of samples that will be propagated through the network. For example, if the data have 1000 samples and the batch size equals to 100, the algorithm takes the first 100 samples from the data and trains the network. Next, it takes the second 100 samples and trains the network again. This procedure repeats several times until the algorithm propagates through all samples. Advantages of using a batch are related to requiring less memory as well as improving the training speed of networks owing to updating the weights after each propagation.

d. Number of epoch (e)

The number of epochs is the number of times that the learning algorithm will work through all training dataset. Typically, the learning algorithm multiplies epochs until the values for weight and bias change insignificantly. An epoch is comprised of one or more batches.

e. Optimizer algorithms (o)

In neural networks, the most common optimization algorithms are GD. The Keras library provides some of optimization algorithms that usually include SGD (Stochastic gradient descent), RMSProp (Root Mean Square Propagation), Adagrad (Adaptive Gradient Algorithm), Adadelta, Adam, Adamax, Nadam.

f. Differencing data (d)

MLP can handle raw data with little pre-processing such as scaling and differencing. Nevertheless, when it comes to time series data, sometimes differencing the series can cause a problem easier than model. The differencing of time series data can be the first differencing or the seasonal differencing:

- The first differencing: $dy(t)=y(t)-y(t-1)$
- The seasonal differencing: $dy(t)=y(t)-y(t-d)$, where d is the seasonal period of time series data.

2.3.2. MLP WFV methodology

The framework of MLP WFV methodology which combines the MLP model and WFV methodology is shown in Figure 2 below and can be expressed by following steps:

- Step 1: Firstly, the data ($[y_1, y_2, \dots, y_{n-h}, y_{n-h+1}, y_{n-h+2}, \dots, y_n]$) is split into history (y_1, y_2, \dots, y_{n-h}) and testing data (y_{n-h+1}, \dots, y_n) using split1 function. The testing and history data have lengths of h and n-h, respectively, n is the number of original data.
- Step 2: The history data is then put into the diff function which makes differencing history data according to the values of different order **d**. The operation of diff function is shown in Figure 2.
- Step 3: Next, the split2 function will split the input data into the input component **X_train** and the target component **Y_train**. The first dimension of **X_train** is the number of lag observation used as input **i**, and the second dimension of **X_train** - the number of observation of data.
- Step 4: The determination, training and prediction procedures of MLP will be performed as shown in Figure 2. To build the MLP model, the the Keras library was used, and the MLP model has been defined as a sequence layer. The MLP consists of input layer with input dimension **i**, one hidden layer with the number of nodes **n**, and the optimization algorithms **o**. The numbers of epoch and batch in the traning process correspond to **e** and **b**. The output of MLP is the forecasting data.
- Step 5: The forecasting value from MLP model must be inverted by idiff function if the data was different from Step 2.
- Step 6: The known value y_{n-h+1} is added into the history data. Then, the circle from step 2 to step 5 is repeated. Besides, we repeat it after obtaining the last forecasting value of f_h .
- Final step: The accuracy score (AC) will be used to calculate the RMSE, MAPE and MAE values between the forecasting values (f_1 to f_h) and the testing data (y_{n-h+1} to y_n). The RMSE, MAPE and MAE values are obtained by (5) [29-33]:

$$RMSE = \sqrt{\frac{1}{h} \sum_{i=1}^h |y_i - f_i|^2}, MAPE = \frac{1}{h} \sum_{i=1}^h \left| \frac{y_i - f_i}{y_i} \right|, MAE = \frac{1}{h} \sum_{i=1}^h |y_i - f_i| \quad (5)$$

where y_i is the actual observed values; f_i is the forecasting values; h is the number of observed values.

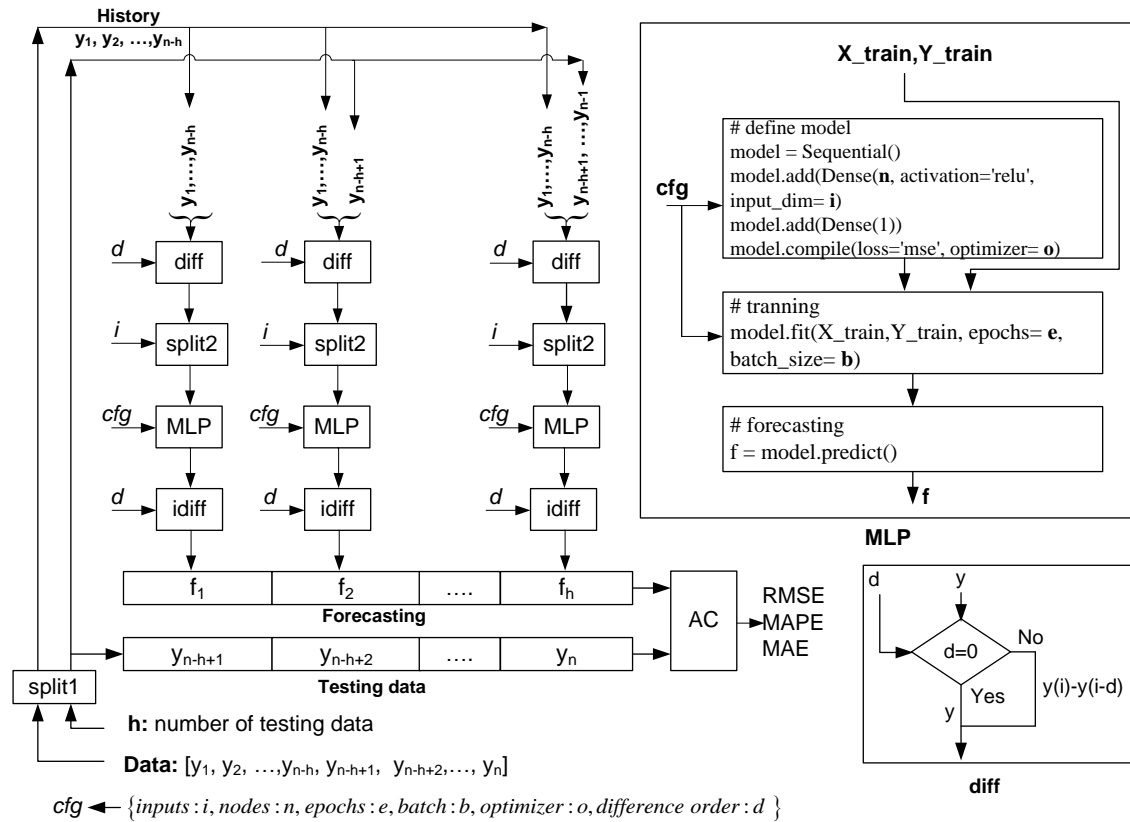


Figure 2. The framework for MLP WFV methodology

2.3.3. The grid search model of MLP WFV methodology

Based on the tuning hyperparameters and the model of MLP WFV methodology, the grid search model of MLP WFV methodology was established as shown in Figure 3. The training and testing processes have the same combination of tuning hyperparameters **cfg** and the same number of testing **h**. For the training data, we obtained the optimal models which satisfy minimum accuracy scores ($RMSE_{min}$, $MAPE_{min}$, MAE_{min}). In the testing process, these optimal models will be compared to all other models according to their accuracy scores in order to evaluate the reliability of the grid search model of MLP WFV methodology.

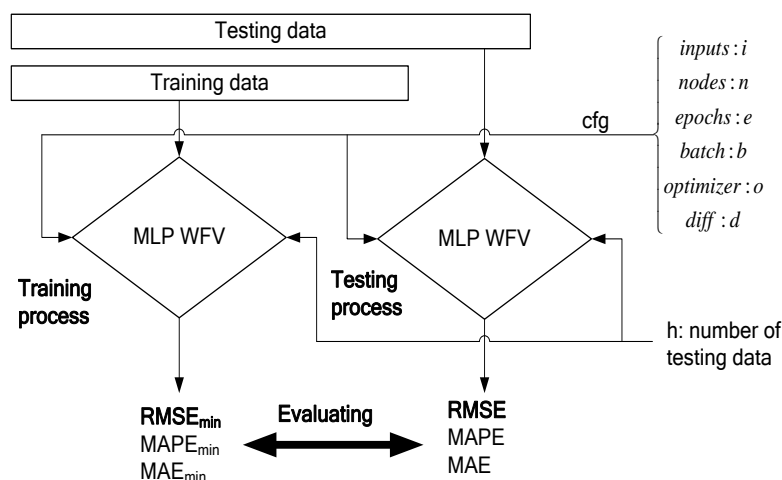


Figure 3. The grid search model of MLP WFV methodology

3. RESULTS AND DISCUSSIONS

3.1. Data description

In order to enhance the reliability of experiment results, the US airline passenger and electricity demand data of Hochiminh City, Vietnam were studied in our experiments. The US airline passenger dataset provides monthly totals in thousands of US airline passengers from 1949 to 1960 and has been used in many researches [34]. The Ho Chi Minh city load demand dataset provides daily max load demand in MW for three months from October 2018 to January 2019. The US airline passenger and Ho Chi Minh city load demand data are presented in Figure 4.

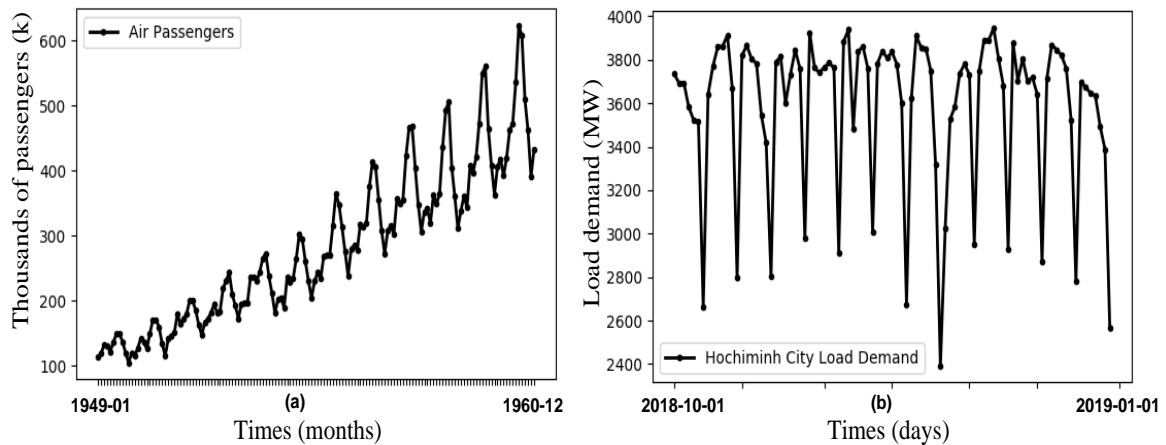


Figure 4. (a) The US airline passenger data, (b) Ho Chi Minh city load demand data

Setup values of the tuning hyperparameters for US airline passenger and Ho Chi Minh city load demand data are also listed in Table 3. Because of monthly seasonality of US airline passenger data, there is a numeric value of 12 for the values of the number of testing data **h**, the value of input **i** and the value of differencing **d**. It is the same for Ho Chi Minh city load demand data with a numeric value of 7 due to the weekly seasonality. The combination of all tuning hyperparameters gives 224 cases corresponding to 224 possible models of MLP WFV methodology.

Table 3. The values of tuning hyperparameters

Items	US airline passenger data	Ho Chi Minh city load demand data
number of validation data h	12	7
input i	1, 12	1, 7
nodes n	50, 100	50, 100
epochs e	10, 100	10, 100
batch b	1, 10	1, 10
differencing d	0, 12	0, 7
optimizer o	'SGD', 'RMSprop', 'Adagrad', 'Adadelata', 'Adam', 'Adamax', 'Nadam'	'SGD', 'RMSprop', 'Adagrad', 'Adadelata', 'Adam', 'Adamax', 'Nadam'
Number of possible models	224	224

3.2. The results in case of US airline passenger data

Table 4 shows the results of training and testing processes. As obviously seen that we have the same optimal models determined by the minimum of accuracy scores RMSE, MAPE and MAE in training process. For testing process, the column 'Optimal' shows the accuracy scores for the optimal model, and the columns 'Min', 'Average' and 'Max' - the min, the average and the max values for 224 possible models that can be generated. Figure 5(a) gives the forecasting and testing series of the optimal model, while Figure 5(b), 5(c) the forecast and testing series for the model of minimum and maximum accuracy scores in the testing process. Figure 6 indicates the distribution of accuracy scores for testing process. Figure 6(a) presents the box plot for RMSE component with the first column for the distribution of all the 224 possible models and the second column for the optimal model. The same distributed data are plotted in Figures 6(b) and 6(c).

Table 4. The results of training and testing process in case of US airline passenger data

The accuracy scores	The optimal models of training process	The accuracy scores of testing process			
		Optimal	Min	Average	Max
RMSE (k)	[1, 100, 100, 1, 12, 'Nadam']	20.71	15.39	31.16	64.37
MAPE (%)	[1, 100, 100, 1, 12, 'Nadam']	3.34	2.60	5.42	13.22
MAE (k)	[1, 100, 100, 1, 12, 'Nadam']	15.55	13.74	25.65	57.92

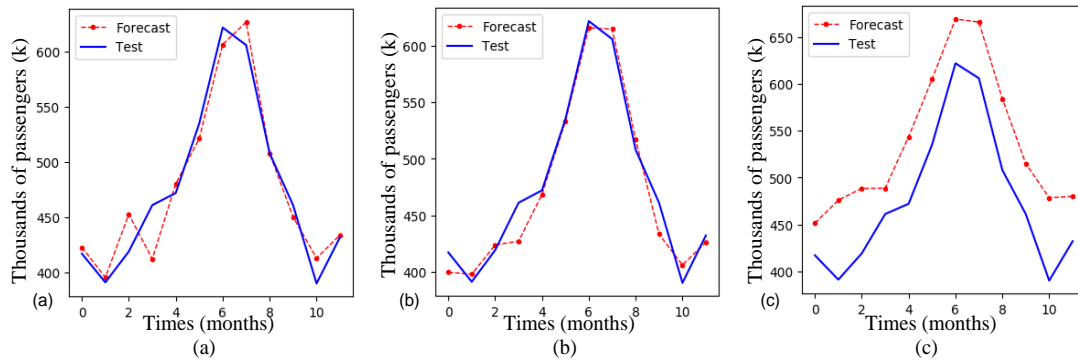


Figure 5. The forecast and testing series for testing process in case of US airline passenger data, (a) optimal model, (b) minimum model, (c) maximum model

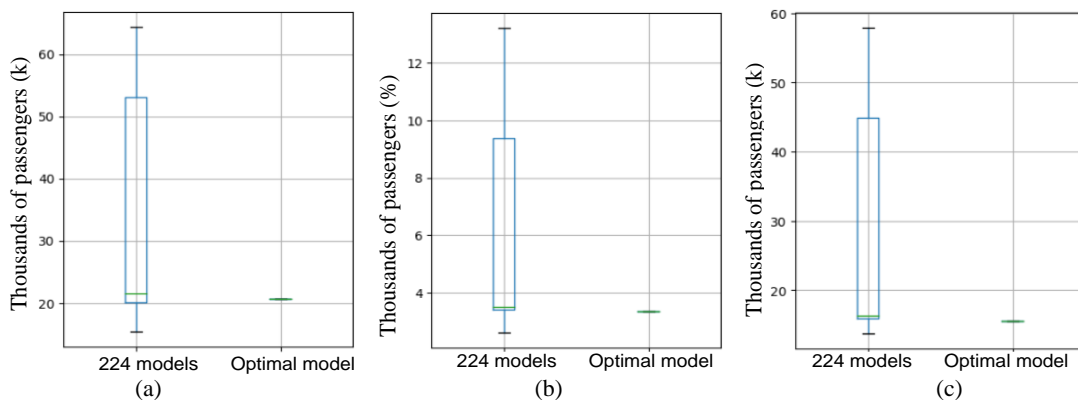


Figure 6. The box plot of accuracy scores for testing process in case of US airline passenger data, (a) RMSE, (b) MAPE, (c) MAE

3.3. The results in case of Ho Chi Minh city load demand data

Table 5 shows the results of training and testing processes. Obviously, the minimum of RMSE, MAPE and MAE gives the different optimal models in the training process. For testing process, the column 'Optimal' shows the accuracy scores for the optimal model, while the column 'Min', 'Average' and 'Max' - the min, the average and the max values for 224 possible models that can be generated. Figure 7(a) gives the forecasting and testing series of the optimal model of RMSE, while Figures 7(b), 7(c) the forecast and testing series for the model of minimum and maximum accuracy scores in the testing process. Figure 8 indicates the distribution of accuracy scores for testing process. Figure 8(a) presents the box plot of RMSE component with the first column for the distribution of all the 224 possible models and the second column for the optimal model. The same distributed data are plotted in Figure 8(b) and 8(c).

Table 5. The results of training and testing process in case of Ho Chi Minh city load demand data

The accuracy scores	The optimal models of training process	The accuracy scores of testing			
		Optimal	Min	Average	Max
RMSE (MW)	[7, 100, 10, 1, 7, 'Adamax']	123.01	103.31	256.67	740.59
MAPE (%)	[7, 100, 100, 10, 7, 'Adamax']	2.83	2.49	6.03	18.59
MAE (MW)	[7, 100, 10, 10, 7, 'Adamax']	123.62	80.77	192.95	659.24

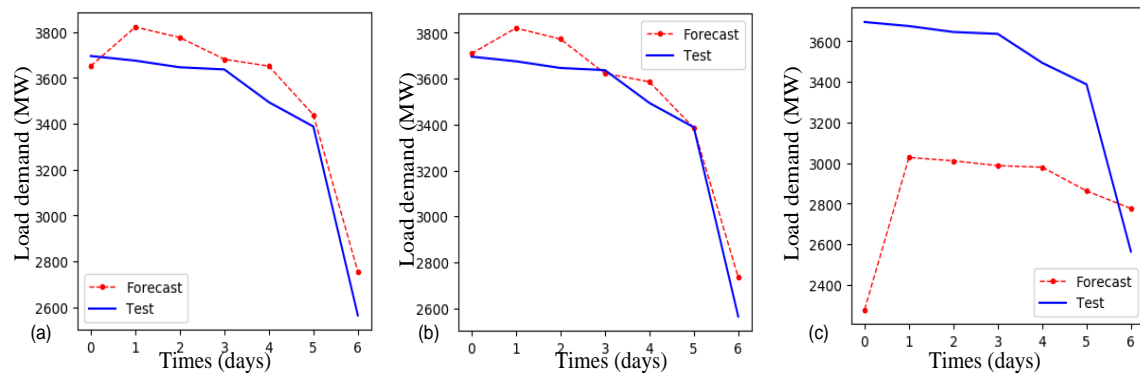


Figure 7. The forecast and testing series for testing process in case of Ho Chi Minh city load demand data, (a) optimal model, (b) minimum model, (c) maximum model

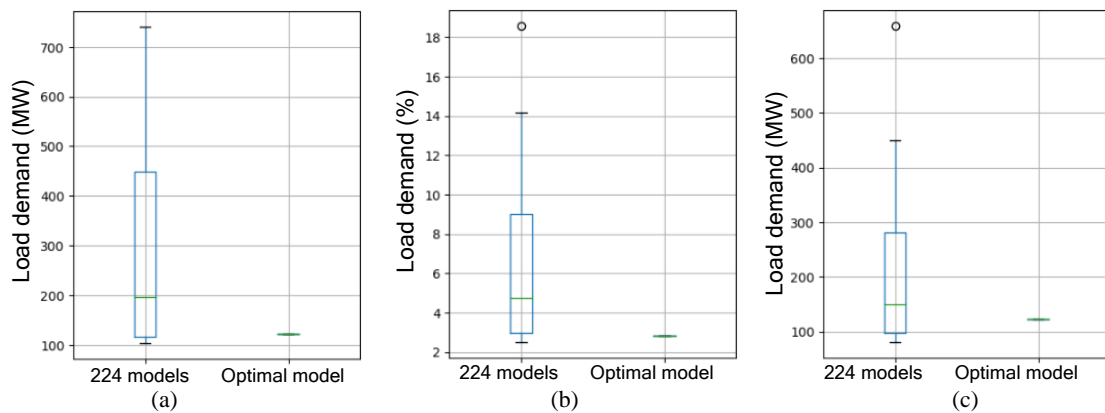


Figure 8. The box plot of accuracy scores for testing process in case of Ho Chi Minh city load demand data, (a) RMSE, (b) MAPE, (c) MAE

3.4. Evaluation

As described above, the optimal models in training process were determined by minimizing accuracy scores such as RMSE, MAPE and MAE. The analysis of results listed in Table 4 and Table 5 indicated the existence of the optimal model that satisfies the criterion of minimum of accuracy scores RMSE, MAPE, MAE. In the case of US airline passenger data, there was the unique model that satisfies all RMSE, MAPE and MAE. In the case of Ho Chi Minh city load demand data, three different optimal models were found to satisfy correspondingly RMSE, MAPE and MAE.

The optimal model obtained in the training process does not ensure completely the best results in the testing process. Let's analyse the Table 4 in the case of US airline passenger data for the optimal model. When using the optimal model for the testing process, the accuracy scores of RMSE, MAPE and MAE correspond to 20.71k, 3.34% and 15.55k. Meanwhile, there are the other MLP models that give better results with minimum values of the RMSE, MAPE, MAE are 15.39 k, 2.60%, 13.74 k, respectively. However, in comparison with the average values of all the models (the RMSE, MAPE, MAE values of 31.16 k, 5.42%, 25.65 k, respectively) and with the max values of all the models (the RMSE, MAPE, MAE values of 64.37 k, 13.22%, 57.92 k, respectively), the accuracy scores of the optimal model are too small. Besides, analysing the boxplot of accuracy scores shown in Figure 6 clearly indicates that the accuracy scores of the optimal model are nearly the same as the minimum values of all the other models. Moreover, the forecast values shown in Figure 5(a) are well consistent with the testing values for the optimal model. We have the same results for the case of Ho Chi Minh city load demand data. These results clearly show that the optimal model obtained in the training process by applying the grid search model of MLP WFV methodology will give good values in testing process.

Noted that the characteristics of times series in cases of US airline passenger data and Ho Chi Minh city load demand data are different from each other. Indeed, the time horizon and value vertical in case of US airline passenger data are built by scales of months and thousands of passengers, respectively. Meanwhile,

for data of Hochiminh city load demand, daily and MW scales were used. Besides, the seasonality in cases of US airline passenger data is 12 (monthly) and of Ho Chi Minh city load demand data-7 (daily). The results in both cases showed good values that are promising for applying the grid search model of MLP WFV methodology at any times series.

4. CONCLUSION

Based on the MLP structure and Walk-Forward Validation Methodology, a framework for grid search model of MLP WFV methodology was proposed. In the training process, the minimum of accuracy scores of RMSE, MAPE, MAE was applied to specify the optimal models. In the testing process, the accuracy scores have been used to compare the optimal model with all other ones. Both the US airline passenger and Ho Chi Minh city load demand data were used for the analysis. The results indicated the existence of the optimal model that satisfies requirement of minimum for accuracy scores. In the testing process, the accuracy scores of the optimal model gave the good values close to the minima which were much smaller than the average values. In addition to that the max values were comparable to those of all other models. In this regard, the positive results obtained in this study suggest an effective way for load demand forecasting.

REFERENCES

- [1] W. Li, et al., "A novel hybrid model based on extreme learning machine, k-nearest neighbor regression and wavelet denoising applied to short-term electric load forecasting," *Energies*, vol. 10, no. 5, pp. 1-16, 2017.
- [2] A. Kaysal, et al., "Artificial Neural Networks and Adaptive Neuro-Fuzzy Inference Systems Approaches to Forecast the Electricity Data for Load Demand, an Analysis of Dinar District Case," *ISMSIT 2018-2nd International Symposium on Multidisciplinary Studies and Innovative Technologies, Proceedings*, 2018, pp. 1-6.
- [3] S. Netsanet, et al., "Short term load forecasting using wavelet augmented non-linear autoregressive neural networks: A single customer level perspective," *IEEE 3rd International Conference on Big Data Analysis (ICBDA 2018)*, 2018, pp. 407-411.
- [4] J. W. Taylor and P. E. McSharry, "Short-term load forecasting methods: an evaluation based on European data," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 2213-2219, 2007.
- [5] P. E. McSharry, et al., "Probabilistic forecasts of the magnitude and timing of peak electricity demand," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 1166-1172, 2005.
- [6] E. Gonzalez-Romera, et al., "Monthly electric energy demand forecasting based on trend extraction," *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1946-1953, 2006.
- [7] T. Hong, et al., "A naïve multiple linear regression benchmark for short term load forecasting," *IEEE Power and Energy Society General Meeting*, vol. 2, pp. 1-6, 2011.
- [8] T. Hong, et al., "Modeling and forecasting hourly electric load by multiple linear regression with interactions," *IEEE PES General Meeting (PES 2010)*, pp. 1-8, 2010.
- [9] A. Karim and S. Ariffin, "Electricity load forecasting in UTP using moving averages and exponential smoothing techniques," *Applied Mathematical Sciences*, vol. 7, no. 77-80, pp. 4003-4014, 2013.
- [10] R. Goeb, et al., "Electrical load forecasting by exponential smoothing with covariates," *Applied Stochastic Models in Business and Industry*, vol. 29, no. 6, pp. 629-645, 2013.
- [11] M. A. Acquah, et al., "Real-Time Peak Control algorithm using Stochastic Optimization," *IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV*, 2017, pp. 1-6.
- [12] S. Ahmadi, et al., "A fuzzy inference model for short-term load forecasting," *2nd Iranian Conference on Renewable Energy and Distributed Generation (ICREDG 2012)*, vol. 37, 2012, pp. 39-44.
- [13] T. Y. Kurniawan, "Electrical load time series data forecasting using Interval Type-2 Fuzzy Logic system," *Proceedings 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT 2010)*, vol. 5, 2010, pp. 527-531.
- [14] S. Rahman and O. Hazim, "A Generalized Knowledge-Based Short-Term Load-Forecasting Technique," *IEEE Transactions on Power Systems*, vol. 8, no. 2, pp. 508-514, 1993.
- [15] D. C. Park, M. a. El-Sharkawi, R. J. Marks, L. E. Atlas, and M. J. Damborg, "Electric load forecasting using an artificial neural network," *IEEE Trans. Power Syst.*, vol. 6, no. 2, pp. 442-449, 1991.
- [16] Z. Saad et al., "Short-term load forecasting of 415V, 11kV and 33kV electrical systems using MLP network," *Proceeding 2017 Int. Conf. Robot. Autom. Sci. ICORAS 2017*, vol. 2018, 2018, pp. 1-5.
- [17] S. Govender and K. A. Folly, "Short-Term Load Forecasting using Artificial Neural Networks and Multiple Linear Regression," *IEEE PES/IAS PowerAfrica Conf. Power Econ. Energy Innov. Africa, PowerAfrica 2019*, no. 1, pp. 273-278, 2019.
- [18] L. Hernández-Callejo, et al., "Artificial neural networks for short-term load forecasting in microgrids environment," *Energy*, vol. 75, pp. 252-264, 2014.
- [19] A. Khotanzad, et al., "ANNSTLF-A neural-network-based electric load forecasting system," *IEEE Transactions on Neural Networks*, vol. 8, no. 4, pp. 835-846, 1997.

- [20] D. P. Yuan Octavia, A. N. Afandi, and H. Putranto, "Power demand forecasting considering actual peak load periods using artificial neural network," *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, vol. 2018, 2018, pp. 198–203.
- [21] M. Zou, et al., "Comparison of Three Methods for a Weather Based Day-Ahead Load Forecasting," *Proc. 2019 IEEE PES Innov. Smart Grid Technol. Eur. ISGT-Europe*, 2019, pp. 1-5.
- [22] A. Burkov, "The Hundred-Page Machine Learning," *Andriy Burkov*, 2019.
- [23] D. Rothman, "Artificial intelligence by example : develop machine intelligence from scratch using real artificial intelligence use cases," *Packt Publishing Ltd*, 2018.
- [24] F. M. Bianchi, et al., "An overview and comparative analysis of Recurrent Neural Networks for Short Term Load Forecasting," *Springer Briefs in Computer Science*, 2017.
- [25] Jason Brownlee, "Deep Learning for Time Series Forecasting," *Machine Learning Mastery Pty. Ltd.: Vermont Victoria, Australia*, 2018.
- [26] J. Moolayil, "Learn Keras for Deep Neural Networks: A Fast-Track Approach to Modern Deep Learning with Python," *Apress (DE)*, 2019.
- [27] Jason Brownlee, "Deep Learning With Python," *Machine Learning Mastery Pty. Ltd.: Vermont Victoria, Australia*, 2019.
- [28] H. Daumé, "A Course in Machine Learning," *Todo*, 2012.
- [29] J. Moon, et al., "A Short-Term Electric Load Forecasting Scheme Using 2-Stage Predictive Analytics," *Proceedings IEEE Int. Conference on Big Data and Smart Computing (BigComp 2018)*, 2018, pp. 219-226.
- [30] Ngoc T. T. and Dai L. V., "Grid search of exponential smoothing method: A case study of Ho Chi Minh City load demand," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 19, no. 3, pp. 1121-1130, 2020.
- [31] H. Kusdarwati and S. Handoyo, "System for prediction of non stationary time series based on the wavelet radial bases function neural network model," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 4, pp. 2327-2337, 2018.
- [32] I. Mado, A. Soeprijanto, S. Suhartono, "Applying of Double Seasonal ARIMA Model for Electrical Power Demand Forecasting at PT. PLN Gresik Indonesia," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 6, pp. 4892-4901, 2018.
- [33] A. Özdemir, et al., "Performance evaluation of different classification techniques using different datasets," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 5, pp. 3584-3590, 2019.
- [34] R. J. Hyndman and G. Athanasopoulos, "Forecasting: principles and practice," Monash University, Australia 2019. [Online]. Available: OTexts.org/fpp/.

BIOGRAPHIES OF AUTHORS



Tran Thanh Ngoc received the B.Sc. and M.Sc. degrees from the Ho Chi Minh University of Technology, Vietnam in 2001 and 2006, respectively. He received his Ph.D. degree in Information and electricity engineering from Hunan University, Changsha, China in 2012. Currently, he is a Lecturer with the Faculty of Electrical Engineering Technology, Industry University of Ho Chi Minh City, Vietnam. His main research interests include HVDC transmission systems, active filters, power quality and load forecasting



Le Van Dai received MSc degree in Electrical Engineering from the Ho Chi Minh City University of Technology, Vietnam in 2008. He received his Ph.D. degree in Information and electricity engineering from Hunan University, Changsha, China in 2016. Since 2011, he is a lecturer of the Faculty of Electrical Engineering Technology, Industry University of Ho Chi Minh City, Vietnam. His research interests include FACTS control, solar PV energy conversion, control of wind power generation, and grid integration



Dang Thi Phuc received the Specialist degree from the Lomonosov Moscow State University, Moscow, Russia in 2008. In 2018 she received Ph.D. degree in System analysis, Control and Information Processing from Peoples' Friendship University of Russia, Moscow, Russia. Since 2018, she is a lecturer of Faculty of information Technology, Industry University of Ho Chi Minh City, Vietnam. Her research interests include Machine Learning, Computer Vision, NLP, Deep Learning, Operator network.