Artificial neural network based short term electrical load forecasting

Oon Yi Her, Mohd Saiful Azimi Mahmud, Mohamad Shukri Zainal Abidin, Razman Ayop, Salinda Buyamin
School of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia Skudai, Johor Bahru, Malaysia

ABSTRACT

In power generation, a 24-hour load profile can vary significantly throughout the day. Therefore, power generation must be adjusted to reduce money loss due to excess generation. This paper presents a short-term load forecasting (STLF) system design using artificial neural network (ANN). As ANN come in many different configurations, this paper analyzes the best ANN configuration via trial-and-error method. To train the ANN, historical load data from 2016 to 2018 of power south energy cooperative (AEC) is used. A simple feedforward ANN type with one hidden layer is implemented, where 48 neurons are used at the input layer. For hidden layer, an arbitrary 50 neurons are chosen and 24 neurons at output layer are used to generate a day ahead 24-hour load profile. To measure the best activation function for SLTF application, four non-linear activation functions will be tested and the best activation function is used to create two and three hidden layer ANN architecture. Finally, the performance of the two new networks will be compared against one hidden layer model. From the obtained result, the best performing model is found as two hidden layers ANN with Tanh as its hidden layer activation function with 8.9% of testing mean absolute percentage error (MAPE).

Keywords:
Artificial neural network
Electrical energy
Load forecasting
Power distribution
Short term

1. INTRODUCTION

Electrical energy is a very crucial resource of modern human society as it powered various important industries that satisfy human needs. However, electricity is difficult to be produced and distributed especially in large scale area. In addition to that, electrical energy storage system does not have widespread implementation due to economic reasoning as most of the generated electricity must be consumed immediately. Therefore, to ensure that the power system distribution operation to be running smoothly, an efficient load forecasting system is required [1]. In general, electrical load forecasting is divided into three types which are short-term load forecasting (STLF), medium-term load forecasting (MTLF) and long-term load forecasting [2]. Each type is divided based on different forecasting ranges. In designing load forecasting system, an inaccurate forecast will cause a mismatch between demand and generation of electrical power and eventually it will result in significant amount of money loss. Thus, the selection of load forecasting method needs to be chosen based on its application to solve specific load forecasting type.

Load forecasting methods can be categorized in three major groups which are traditional forecasting technique, Modified forecasting technique and soft computing technique. All types of load forecasting methods have their own advantage and disadvantage and their usage is dependent on the load pattern, type of
model inputs and forecasting time horizon. In traditional forecasting method, conventional mathematical techniques such as regression [3], [4], multiple-regression [5] and exponential smoothing [6], [7] is usually used. From all available traditional forecasting methods, the multiple-regression technique is the most popular and has been widely used to forecast the load that are affected by numerous factors from meteorological effects, electricity prices, economic growth and others.

The modified traditional forecasting method is designed by modifying the traditional methods to enable the automatic parameter correction of forecasting model under changing environmental conditions. Some of the techniques used in modified traditional forecasting are adaptive load forecasting [8], [9], stochastic time series [10] and support vector machine [11], [12]. By comparing all methods in this category, the adaptive load forecasting method has the most advantages as the demand forecasting model parameters are automatically corrected to keep track of the changing load conditions, thus enable the prediction system to be used on-line.

Recently, soft computing techniques have been emerging as a flexible approach to forecast the electrical load in power system. This technique mimics the human reasoning system to employ the ability to produce mode of reasoning that is approximate rather than accurate. This method using algorithms such as fuzzy logic [13]–[15], artificial neural network (ANN) [16]–[19] and evolutionary algorithms such as genetic algorithm [20]–[22] and particle swarm optimizations [23]–[25]. In soft computing method, each factor affecting the forecast is considered as a cost and the method will exploit all possibilities to find the potential solution based on the computed costs.

Each algorithm in soft computing techniques has its own advantages and disadvantages. In fuzzy logic based method, the knowledge must be adapted accurately using fuzzy rules as the quality of the forecasting system will be mainly affected by the fuzzy rules. In ANN based methods, the selection of the parameters needed in training the models must be carefully chosen as each parameter will affect the performance of the forecasting system. Finally, solution generated by the evolutionary algorithm usually fall into local minimum thus creating low quality electrical load prediction for power systems. Therefore, the implementation of soft computing technique needs a careful design process to create an efficient load forecasting system.

As ANN comes with different configuration, this paper proposed analysis on the effect of ANN parameters towards short-term load forecasting system. In this paper, multiple layers feed forward network will be used which will predict the electrical usage a day ahead in 24-hours using historical data, day of week, week of month and month of year as the inputs. The analysis of different number of hidden layers and activation function types are conducted in this paper to find the most optimized parameters in short team load forecasting.

2. RESEARCH METHOD

Figure 1 shows the overall process for short term load forecasting. In this paper, a densely connected feedforward ANN with backpropagation learning algorithm will be used to implement the short-term load forecast. Based on Figure 1, the process of SLTF starts with the data initialization where the historical data were loaded. Then, the data was preprocessed where encoding process is executed in this stage. After the date has been loaded, the model will be trained where different activation function will be used to find the best activation function for the model. Finally, the hidden layer experiment is conducted where different number of hidden layers will be tested to analyze the effect of number of hidden layers towards the prediction quality. In this paper, 48 input neurons will be used and 50 hidden neurons were arbitrary chosen to forecast the electrical load. In the output layer, 24 neurons were selected and mean absolute error is used as the loss function which will be optimized using gradients descent algorithm using Keras SGD class. As a result, a total of 3674 trainable parameters for single layer model, 6224 trainable parameters for two layers model and 8774 trainable parameters for three layers model is required for this architecture.

2.1. Inputs selection

For input selection, the historical load data, day of week, week of month and month of year will be used for electrical load prediction [26]. Week in month is referred to the week sequence in a month which is whether it is the first week, second week and so on. Month in year is referred to the month sequence in a year whether it is first month, second month and so on. The historical data for previous day \((d - 1)\) will be used to forecast at forecast day \((d + 1)\) which is the next day of present time \((d)\) shown in Figure 2. A total of 24 input neurons will be allocated to load data of hour 1 to hour 24 of previous day.
To implement the input data into the ANN, bit encoding method is used in this paper. 7 input neurons were used for the day of week data. For instance, if it is Sunday, the first neurons of day of week neurons group will give 1 as input and 5 neurons will be used for the week of month. The month of year will used 12 neurons input following the similar method. It should be noted that the day of week, week of month and month of year inputs should be 24 hours load input day date. For instance, if the forecast day is 5 September 2018 and the input day is 3 September 2018, the day of week is Monday and will be encoded as 0100000, it is second week of the month so it will be encoded as 01000 and ninth month of the year and encode as 000000001000. Then the 24-hours load input data will be taken from 3 September 2018 to predict the load on 5 September 2018.

2.2. Activation function and data preprocessing

To determine the output of the ANN, a non-linear activation function will be used in the hidden layer while a linear activation will be used in the output layer. This includes exponential, Tanh, sigmoid and softsign activation function. For output layer, a rectified linear unit (ReLU) will be used for all ANN models. For data preprocessing, all input data will be normalized into range of 0 to 1 before feeding the data into ANN except for the bit encoded data since its value is either 0 or 1. For normalization, min-max feature scaling is used. The description of min-max feature scaling is shown in (1).

$$X' = \frac{X-X_{\text{min}}}{X_{\text{max}}-X_{\text{min}}}$$

As shown in (1), $X$ is the exact value of our data, $X'$ is the normalized value, $X_{\text{max}}$ represents the maximum value of dataset and $X_{\text{min}}$ represents the minimum value of the dataset. The maximum load value used is 1048 MWh and the minimum load value used is 291 MWh. Both of these values are the highest and lowest load from year 2016 until 2017. In addition, any load value from year 2018 that exceed this minimum and maximum range will be removed. Finally, the chosen parameter is measure based on the value of mean absolute percentage error (MAPE). Daily best and worst absolute percentage error were recorded for performance analysis. The testing MAPE will be plotted against the training MAPE for overfitting analysis in this paper. The description of overfitting analysis is shown in Figure 3.

For overfitting analysis, a secondary training loop has been set up based to Figure 3. At every 50 training epochs, the training MAPE will be inspected either it has been falling under certain range given in
the conditions shown in Figure 3. If it falls into the range, the current model will be stored. The original attempt was conducted to train the models to a very low training MAPE, as low as 3%. However, overfitting occurred much sooner and more often than expected. Therefore, this secondary training loop is the attempt to capture a model just before it over-trained occurs to the model.

Figure 3. Secondary training loop flowchart

3. RESULTS AND DISCUSSION

This section will discuss the results and analysis that have been conducted to forecast the electrical load from 2016 to 2018. First, the best testing MAPE will be chosen based on a single hidden layer model. Then, the chosen best testing MAPE will be used for further analysis by adding several hidden layers to investigate the best ANN configuration for electrical load forecasting.

3.1. Best testing MAPE

In this paper, the Test MAPE is calculated using the mean of every hour to find the absolute percentage error. To calculate the MAPE, 334-day load data of every hour in each day is used. Then, the 334 data will be multiplied by 24 hour and finally result in 8,016 pieces of hourly data. The testing MAPE is the summation of each individual day hourly absolute percentage error divided by the total number of data which is 8,016. Table 1 shows the results of the Test MAPE and its respective Train MAPE.

From the table, ANN model with Tanh hidden layer activation function has produce the best test MAPE with 8.9% error while with sigmoid has the worst MAPE with 15.76% error. Therefore, Tanh function has been selected to investigate the effect of the number of hidden layers to the prediction performance for load forecasting which will be presented in the later part of this section.

Table 1. Comparison of different activation function for SLTF

| Single hidden layer model | Test MAPE, % | Train MAPE, % |
|---------------------------|-------------|---------------|
| Tanh                      | 8.9         | 7.01          |
| Sigmoid                   | 15.76       | 6.19          |
| Softsign                  | 9.61        | 8.11          |
| Exponential               | 12.06       | 8.12          |

3.2. Overfitting analysis

In ANN, overfitting occurs when a model tries to forecast data trend in the data that is too noisy to be predicted. This situation is happened due to the overly complex model that has many parameters. If the model is overfitted, the forecast output would become inaccurate due to the trend does not reflect the current data pattern. Figure 4 shows the comparison between the test MAPE and the train MAPE where Figure 4 (a) is for Tanh, Figure 4 (b) is for sigmoid, Figure 4 (c) is for softsign and Figure 4 (d) is for exponential.

Tanh activation function have the most consistent curve which indicate its test and train MAPE consistency. For Sigmoid activation function, the test and train MAPE relationship is very inconsistent as the graph have shown a zig-zag pattern. Softsign is slightly worse than Tanh when it comes to test and train MAPE curve but it placed at the second place in this situation. The relationship of exponential function starts off linear and as the train MAPE go further down after the point of overfitting, the test MAPE seems to be regaining its linear relationship with train MAPE.
3.3. Best and worst forecast sample

In choosing the best ANN parameter for electrical load forecast, two different prediction perspective has been chosen; best and worst to measure the effectiveness of the activation function throughout the forecast. The average day sample is taken from each individual day average absolute percentage. For example, to obtain these sample, calculation is performed on across all available 334 days hourly load to obtain its individual day average absolute percentage error, so there will be 334 pieces of daily average absolute percentage error. From there, the day which has the highest and lowest value of daily average absolute percentage error will be taken out and its hourly load forecast will be plot against the actual load. Table 2 shows the average absolute percentage error for the best and the worst day in electrical load prediction.

| Model    | Best Individual Day Average Absolute Percentage Error (%) | Worst Individual Day Average Absolute Percentage Error (%) |
|----------|----------------------------------------------------------|----------------------------------------------------------|
| Tanh     | 1.38                                                     | 48                                                       |
| Sigmoid  | 4.18                                                     | 68.23                                                    |
| Softsign | 2.32                                                     | 47.94                                                    |
| Exponential | 3.65                                               | 52.02                                                    |

By referring to the table, the lowest average absolute percentage Error for the best individual day is found to be from Tanh and then followed by Softsign, exponential and sigmoid. However, the lowest average absolute percentage error for the worst individual day is found to be from Softsign and then followed by Tanh, exponential and sigmoid. Although Tanh falls into second place, the percentage difference is low as it is only about 0.06%. Based on this analysis, Tanh activation function has been selected as the ANN model in investigating the effect of hidden layer number to the forecasting performance which will be explained in the next section.

3.4. Effect of ANN hidden layers to the forecasting performance

In this analysis, different number of hidden layers has been selected to choose the best ANN configuration. Figure 5 shows the comparison between the test MAPE and the train MAPE where Figure 5 (a) and Figure 5 (b) is for two and three hidden layers respectively. The two hidden layers shows a similar performance with single hidden layer shown in Figure 4 however, the three hidden layer model start overfitting around the value of 7.12% in train MAPE. Therefore, the three hidden layer model has shown a worst forecast stability performance compared to the two hidden layer model. Table 3 shows the performance of the test MAPE and train MAPE for two and three hidden layers ANN models respectively. Based on the
table, it can be shown that the two hidden layers model has shown superiorities compared to three hidden layers model with lower MAPE in test and train MAPE. Therefore, the two hidden layer model is recommended to be used in short term load forecasting.

Figure 5. Test MAPE vs train MAPE (a) two hidden layers and (b) three hidden layers

Table 3. Comparison of hidden layer number in ANN for SLTF application

| Multiple hidden layer model (Tanh) | Test MAPE, % | Training MAPE, % |
|-----------------------------------|-------------|-----------------|
| Two hidden layers                 | 8.9         | 7.01            |
| Three hidden layers               | 15.76       | 6.19            |

Figure 6 shows the best day forecast sample where Figure 6 (a) is for two hidden layer and Figure 6 (b) is for three hidden layers. Figure 7 shows the worst day forecast sample in a day where Figure 7 (a) is for two hidden layer and Figure 7 (b) is for three hidden layers. Table 4 shows the average absolute percentage error for the best and worst forecast day for two and three hidden layers. From the table, it can be shown that the two hidden layers model shows better performance compared to the three hidden layers. Therefore, it is highly recommended to use two hidden layers Tanh activation function in predicting the short-term electrical load.

Figure 6. Best day forecast sample in (a) two hidden layers and (b) three hidden layers

Figure 7. Worst day forecast sample in (a) two hidden layers and (b) three hidden layers

Table 4. Average absolute percentage error for the best and worst forecast day for different hidden layer

| Model                  | Best individual day average absolute percentage error (%) | Worst individual day average absolute percentage error (%) |
|------------------------|---------------------------------------------------------|----------------------------------------------------------|
| Two hidden layers (Tanh) | 1.29                                                   | 47.9                                                     |
| Three hidden layers (Tanh) | 2.49                                                   | 51.24                                                    |
4. CONCLUSION

The objective of this paper is to investigate the best ANN configuration for short term load forecasting. To find the most optimized ANN configuration, the performance of different ANN activation function is conducted where single hidden layer of Tanh, sigmoid, softsign and exponential function is chosen and the performance was measured and compared. Based on the comparison, Tanh activation function shows the best performance with 1.38% best individual day average absolute percentage error and 48% worst individual day average absolute percentage error. Due to its best performance in load forecasting using single hidden layer, Tanh activation function has been chosen to investigate the performance of different number of hidden layers in load forecasting. Based on the result, the ANN with two hidden layers has shown a better performance compared with ANN with three hidden layers. Therefore, the higher number of hidden layers does not indicate better forecasting performance. For future works, it is suggested that the number of inputs can be increased to include weather data to improve the load forecasting performance for different climate conditions.

ACKNOWLEDGEMENTS

The authors are grateful to the Universiti Teknologi Malaysia and the Ministry of Higher Education (MOHE), for their partial financial support through their research funds, Vote No. R.J130000.2651.17J53.

REFERENCES

[1] B. Michel, J. A. D. Massignan, T. M. O. Santos, J. B. A. London Jr., and C. D. Maciel, “Multiple Households Very Short-Term Load Forecasting using Bayesian Networks,” Electric Power Systems Research, vol. 189, pp. 106733, 2020, doi: 10.1016/j.epsr.2020.106733.

[2] S. Fallah, R. C. Deo, M. Shojafar, M. Conti, and S. Shamsirband, “Computational Intelligence Approaches for Energy Load Forecasting in Smart Energy Management Grids: State of the Art, Future Challenges and Research Directions,” Energies, vol. 11, pp. 596, 2018, doi: 10.3390/en1103596.

[3] D. Grzegorz, “Pattern-based Local Linear Regression Models for Short-Term Load Forecasting,” Electric Power Systems Research, vol. 130, pp. 139-147, 2016, doi: 10.1016/j.epsr.2015.09.001.

[4] W. Zeyu, X. Zhou, J. Tian, and T. Huang, “Hierarchical Parameter Optimization based Support Vector Regression for Power Load Forecasting,” Sustainable Cities and Society, vol. 71, pp. 102937, 2021, doi: 10.1016/j.scs.2021.102937.

[5] N. Hannah, S. New, A. Cohen, and B. Ramachandran, “Load Forecasting using Multiple Linear Regression with Different Calendars,” Distributed Energy Resources in Microgrids, pp. 405-417, 2019, doi: 10.1016/B978-0-12-817774-7.00016-8.

[6] F. Juan Rendon-Sanchez, and M. Lilian de Menezes, “Structural Combination of Seasonal Exponential Smoothing Forecasts Applied to Load Forecasting,” European Journal of Operational Research, vol. 275, pp. 916-924, 2019, doi: 10.1016/j.ejor.esr.2018.12.013.

[7] E. Meira de Oliveira, and F. Luiz Cyrino Oliveira, “Forecasting mid-long Term Electric Energy Consumption through bagging ARIMA and Exponential Smoothing Methods,” Energy, vol. 144, pp. 776-778, 2018, doi: doi.org/10.1016/j.energy.2017.12.049.

[8] N. Thomas and G. Markus, “A Generally Applicable, Simple and Adaptive Forecasting Method for the Short-Term Heat Load of Consumers,” Applied Energy, vol. 241, pp. 73-81, 2019, doi: 10.1016/j.apenergy.2019.03.012.

[9] C. Qing, W. Zhang, K. Zhu, D. Zhou, H. Dai, and Q. Wu, “A Novel Trilinear Deep Residual Network with Self-Adaptive Dropout Method for Short-Term Load Forecasting,” Expert Systems with Applications, vol. 182, pp. 115272, 2021, doi: 10.1016/j.eswa.2021.115272.

[10] F. Tingting and L. Risto, “Evaluation of a Multiple Linear Regression Model and SARIMA Model in Forecasting Heat Demand for District Heating System,” Applied Energy, vol. 179, pp. 544-552, 2016, doi: 10.1016/j.apenergy.2016.06.133.

[11] Y. Ailing, L. Weide, and Y. Xuan, “Short-Term Electricity Load Forecasting based on Feature Selection and Least Squares Support Vector Machine,” Knowledge-Based Systems, vol. 163, pp. 159-173, 2019, doi: 10.1016/j.knosys.2018.08.027.

[12] Aasim, S. N. Singh, and A. Mohapatra, “Data Driven Day-Ahead Electrical Load Forecasting through Repeated Wavelet Transform Assisted SVM Model,” Applied Soft Computing, vol. 111, pp. 107730, 2021, doi: 10.1016/j.asoc.2021.107730.

[13] M. Malekizadeh, H. Karami, M. Karimi, A. Moshari, and M. J. Sanjari, “Short-Term Load Forecast using Ensemble Neuro-Fuzzy Model,” Energy, vol. 196, pp. 117-127, 2020, doi: 10.1016/j.energy.2020.117127.

[14] A. Danladi, M. Yohanna, M. I. Pwu, and B. M. Garkida, “Long-Term Load Forecast Modelling using a Fuzzy Logic Approach,” Pacific Science Review A: Natural Science and Engineering, vol. 18, pp. 123-127, 2016, doi: 10.1016/j.psera.2016.09.011.

[15] L. Chen, “A Fuzzy Theory-based Machine Learning Method for Workdays and Weekends Short-Term Load Forecasting,” Energy and Buildings, vol. 245, pp. 111072, 2021, doi: 10.1016/j.enbuild.2021.111072.

[16] H. Azim, M. M. Nezhad, E. Pshihayian, D. A. Garcia, F. Keynia, and L. D. Santoli, “Short-Term Electricity Price and Load Forecasting in Isolated Power Grids based on Composite Neural Networks and Gravitational Search Optimization Algorithm,” Applied Energy, vol. 277, pp. 115503, 2020, doi: 10.1016/j.apenergy.2020.115503.

[17] C. Gopal, M. Pipattanasomporn, V. Garg, and S. Rahman, “Robust Short-Term Electrical Load Forecasting Framework for Commercial Buildings using Deep Recurrent Neural Networks,” Applied Energy, vol. 278, pp. 115410, 2020, doi: 10.1016/j.apenergy.2020.115410.

[18] D. Hanane, F. Bagheri, H. Le, and L. Desaissat, “Neural Network Model for Short-Term and Very-Short-Term Load Forecasting in District Buildings,” Energy and Buildings, vol. 203, pp. 109408, 2019, doi: 10.1016/j.enbuild.2019.109408.

[19] S. H. Chan Zeke, H. W. Ngan, A. B. Rad, A. K. David, and N. Kasabov, “Short-Term ANN Load Forecasting from Limited Data using Generalization Learning Strategies,” Neurocomputing, vol. 70, pp. 409-419, 2006, doi: 10.1016/j.neucom.2005.12.131.
BIOGRAPHIES OF AUTHORS

Oon Yi Her received his B.Eng in Electrical Engineering from Universiti Teknologi Malaysia (UTM) in 2020. He has been currently graduated and actively work in industry. He can be contacted at email: yoon2@live.utm.my.

Mohd Saiful Azimi Mahmud received his B.Eng in Electrical Engineering, majored in Control and Mechatronics from Universiti Teknologi Malaysia (UTM) in 2015 and PhD in Electrical Engineering from Universiti Teknologi Malaysia (UTM) in 2019. Currently, he is a Senior Lecturer at Universiti Teknologi Malaysia (UTM) in the Division of Control and Mechatronics, School of Electrical Engineering. His research interest is related in the field of multi-objective optimization, Artificial Intelligence, robotics, control system and image processing applications. He can be contacted at email: azimi@utm.my.

Mohamad Shukri Zainal Abidin received his B.Eng in Electrical Engineering from Universiti Teknologi Malaysia (UTM) in 1998, Msc in Electrical Engineering from Universiti Teknologi Malaysia (UTM) in 2001 and PhD in Agriculture Engineering from Tokyo University of Technology, Japan in 2014. Currently, he is an Associate Professor at Universiti Teknologi Malaysia (UTM) in the Division of Control and Mechatronics, School of Electrical Engineering and his current research interest involve the adaptive control strategies in fibrous capillary irrigation system and agricultural robotics. He can be contacted at email: shukri@utm.my.

Razman Ayop received the bachelor’s degree in electrical engineering, the master’s degree in electrical engineering with specialization in power system, and the PhD degree in electrical engineering from Universiti Teknologi Malaysia (UTM), Johor, Malaysia, in 2013, 2015, and 2018, respectively. He is a Senior Lecturer with UTM and a member of Power Electronics and Drives Research Group, School of Electrical Engineering. His research interests include renewable energy and power electronics. This includes the photovoltaic (PV) emulator, maximum power point tracking (MPPT) converter, standalone energy management and sizing, PV modelling, vehicle to grid (V2G) system, particle swarm optimization, and fuzzy logic controller. He can be contacted at email: razman.ayop@utm.my.

Salinda Buyamin received her B. Eng in Electrical Engineering from University of Toledo, USA in 1998, Msc in Automation and Control (Distinctio) from University of Newcastle, United Kingdom in 2003 and PhD in Control of Electrical Drives from University of Newcastle, United Kingdom in 2007. Currently, she is an Associate Professor at Universiti Teknologi Malaysia (UTM) in Control and Mechatronics Department, and her current research interest involve the Modelling and Simulation of Dynamic Systems, Control and Development of Electric Drives System, Intelligent Control, Sensorless Control and Smart Agriculture System. She can be contacted at email: salinda@utm.my.