Enhancing the Performance of Building Load Forecasting Using Hybrid of GLSSVM – ABC Model

Mohammad Azhar Mat Daut[1, 2], Ahmad Sukri Ahmad[1, 2], Mohammad Yusri Hassan[1, 2, a], Hayati Abdullah[1, 3], Md Pauzi Abdullah[1,2] and Faridah Husin[1,2]

[1]Centre of Electrical Energy Systems (CEES), Institute of Future Energy, Universiti Teknologi Malaysia, 81310 Skudai, Johor Darul Ta'zim.
[2]Faculty of Electrical Engineering, Universiti Teknologi Malaysia (UTM), 81310 Johor Bahru, Johor, Malaysia
[3]Faculty of Mechanical Engineering, Universiti Teknologi Malaysia (UTM), 81310 Johor Bahru, Johor, Malaysia

Abstract. In conducting load forecasting, the accuracy of forecasting is an important aspect in planning and managing electricity. Thus, a new hybrid model is presented in this paper, which combines the Group Method of Data Handling, Least Square Support Vector Machine and Artificial Bee Colony (GLSSVM- ABC) for building load forecasting. Its performance accuracy has been compared with other methods by using the Mean Absolute Percentage Error (MAPE) and Root Means Square Error (RMSE). It was found that the proposed method has resulted in better performance accuracy in terms of both MAPE and RMSE. The MAPE analysis showed an increase in performance accuracy of more than 7 percent when compared to other methods. The RMSE analysis showed an increase in performance accuracy of more than 5 percent when compared to other methods. The results in this study showed that the proposed method is proven to be effective and has great potential for accurate building load forecasting.

1 Introduction

The development in electrical load forecasting is becoming interesting and load forecasting methods are constantly being improved. However, forecasting the electrical load is very difficult as electrical loads are frequently affected by several factors such as irregular behaviours, social, time and other non-linear factors [1].

Various methods have been used in modelling and analysing the data for forecasting purposes. Most of the improvements are for the purpose of increasing load forecasting accuracy. The accuracy of load forecasting can affect both users and suppliers. In the past few years, researchers have developed numerous forecasting model to increase the accuracy of load forecasting [1].

In general, methods for forecasting electrical load include using engineering methods, statistical methods and artificial intelligence methods. Among these methods, artificial intelligence method is the most frequently applied to conduct analysis [2]. There are also other models used in forecasting analysis such as decomposition and econometric models [3].

Artificial Neural Network (ANN) and Support Vector Machine (SVM) are also used widely in forecasting [2]. In the last few years, ANN is the most preferred method [4] due to its ability to deal with non-linear factors, and the accuracy of continuous function mapping can be achieved by a three layer neural network [1]. However, ANN requires a lot of training sample data and the selected initial weights can get the local optimal easily [5].

SVM is widely used in the research area and the industries due to its efficiency in solving non-linear problems even with small quantities of training data [2]. Bing Dong applied SVM method to forecast building energy consumption, and found that it performed better than other related models using neural network and genetic programming [6]. Hou and Lian proposed an application of SVM for cooling load forecasting. They concluded that the SVM could provide a promising alternative for cooling load forecasting [7].

Least Square SVM (LSSVM) is an improvement of SVM, proven better than ANN [8]. In the standard SVM model, the solution is addressed by quadratic programming. However, in the LSSVM model, the solution is addressed by a set of linear equations [9]. This improvement reduces the LSSVM complexity and requires less time. The important parts that play an important role in the LSSVM regression system are the regularization parameter and kernel parameter. Nevertheless, both parameters need to be selected properly by establishing a proper methodology to select the parameters [10].

Another sub-model of ANN is known as the Group Method of Data Handling (GMDH). This model has been used effectively with uncertainty. Linear and non-linear systems consist of a broad range of fields such as engineering, science and medicine [11]. The main idea of

[a] Corresponding author: yusrihutm@gmail.com

GMDH is to build an analytical function in a feed forward network. It is based on a quadratic node transfer function, and the coefficient of the model is obtained through the regression technique. Hongya et al [12] applied the GMDH to forecast electric load demand in Australia; the results showed that the GMDH performed better than the ARIMA model in the experiment. Tsado et. al. [13] also used GMDH for energy consumption forecasting, and the results showed the efficiency of GMDH in forecasting over the regression method.

Artificial Bee Colony (ABC) recently is seen as a competitor to the other existing optimization algorithm [8]. This method can conduct both global and local searches in each iteration, which is its main advantage from other method [14]. Furthermore, the parameters in ABC to control are less and require only simple mathematical equations [15]. These benefits can be applied further based on the optimization problems.

This paper proposes the combination of three algorithms; GMDH, LSSVM and ABC to improve the performance of forecasting. GLSSVM had been used to train the actual data with the other input, while the ABC was used to carry out the global and local searches for the best forecast. The forecast was then validated with the actual data as well as comparison with the other methods.

The remaining section of this paper is organized as follows; Section 2 introduces the fundamental of GLSSVM-ABC, while Section 3 explains the methodology of this study, and Section 4 discusses the result. The last section, Section 5 concludes this study.

2 The Fundamental of GLSSVM-ABC

This section introduces the fundamental of Group Method Data Handling with Least Square Support Vector Machine (GLSSVM) and Artificial Bee Colony (ABC0 in terms of theory and concepts.

2.1 GMDH and LSSVM (GLSSVM)

GLSSVM is the combination of GMDH and LSSVM models. This GMDH was introduced by Ivakhnenko [1] in early 1970s, while LSSVM was initiated by Vapnik [2]. The objective of the combination of these two models is to enhance their capability. The basic procedure of this hybrid model is carried out as follows:

Step 1: The training and testing data are separated from the normalized data.

Step 2: A combination of two input variable (x_i, x_j) is generated in each layer. The number of input variable is identified by using Equation (1). Regression of polynomial for this layer is created by establishing the quadratic expression which approximates the output in Equation (2).

$$^{M}C_2 = \frac{M!}{(M-2)!2!} \tag{1}$$

where M is the number of observations in the training set.

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{N} \alpha_i \varphi(x_i)$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i = 0$$

$$\frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \lambda e_i$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow w^T \varphi(x_i) + b + e_i - y_{i=0} \tag{2}$$

i = 1,2,…,N

Step 3: In the next layer, the new input is determined. The smallest root means square (RMSE) represents the output and variable for training data set, which are then combined as the input variable $\{x_1, x_2, ..., x_M, x'\}$ with $M = M + 1$. This new variable will be used as the input for the LSSVM model.

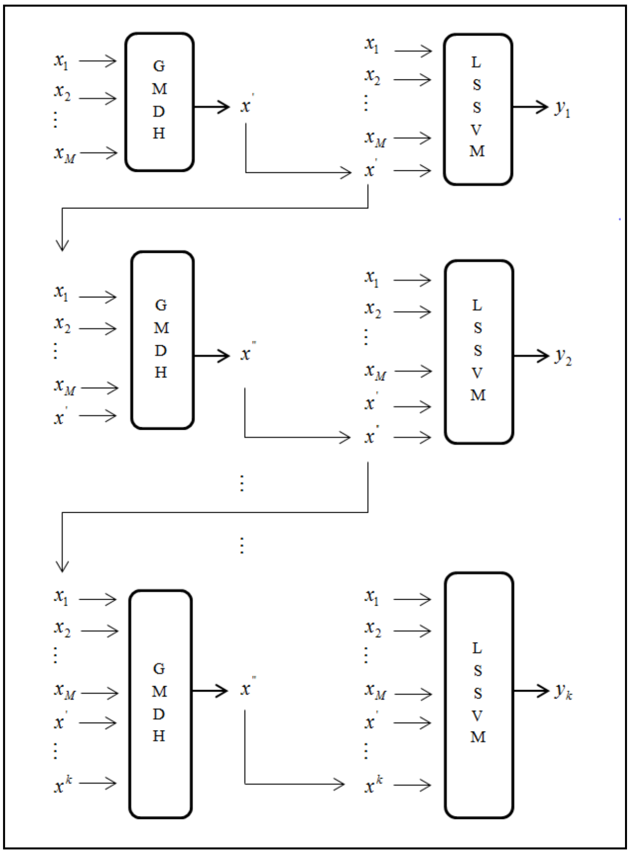

Figure 1. Structure of GLSSVM

Step 4: Steps 2 to 4 are repeated based on the number of iterations by using GLSSVM algorithm. The minimum value of RMSE from the GLSSVM algorithm is selected as the output model. The flow of GLSSVM process is illustrated in Fig. 1.

2.2 Artificial Bee Colony (ABC)

Karaboga and Basturk [16] developed the Artificial Bee Colony (ABC) algorithm inspired from the behaviour of honey bees searching for food. The main focus on this algorithm is to find the best amount of nectar (fitness) by finding the best position of food sources (solutions).

This algorithm can be divided into three phase which are employed bees (EB), onlooker bees (OB) and scout bees (SB). All phases have different task to solve. Generally, the EB phase will solve the position of the food sources. With the memory of the food source position, this bee will spread the information to the OB phase. Thus, the decision making needs to be done by OB to select the best of food source information given by EB. The last phase of this algorithm is SB. This SB is formed from of a few of employed bees, which leave the food sources and to find a new one. All the basic steps about this algorithm are enlightened in the following manner:

Step 1: Initializing the food source. In this stage, the solution will be generated randomly between the ranges of parameter by using Equation (3).

$$x_{ij} = x_j^{\min} + rand(0,1)\left(x_j^{\max} - x_j^{\min}\right) \quad (3)$$

where i represents i_{th}, the number of food source and j is the number of optimization variables associated with the i_{th} food source. Afterward, this process continues to evaluate the quality of the solution (fitness) by using Equation (4).

$$fitness_i = \begin{bmatrix} \dfrac{1}{(1+f_i)} & \text{if } f_i \geq 0 \\ 1+abs(f_i) & \text{if } f_i < 0 \end{bmatrix} \quad (4)$$

where the cost, f_i is cooperated with x_i.

Step 2: Assigning the food source to employee bees. Bees are employed to search for new food source in the memory by applying the Equation (5).

$$v_{ij} = x_{ij} + \varphi_{ij}(x_{ij} - x_{kj}) \quad (5)$$

where φ_i are consistently distributed number within a range of $[-1,1]$. A random optimization variable in the range of $[1,D]$ is represented as j ; where D is not a negative number. The randomly selected food source, k is different from i. Solution weight for higher probability is found by onlooker bees by using Equation (6).

$$p_i = \frac{fit_i}{\sum_i^{SN} fit_i} \quad (6)$$

where fit_i is the fitness solution and SN is the number of food source position.

Step 3: Onlooker bees will be in charge of the selection of the quality solution. Again, using Equation (5), the best solution based on the better solution weight is found, then evaluated by using Equation (6).

Step 4: Deciding the food source to be abandoned and assigning it to scout bees. After all the processes related to exploitation have been done and the number of food cannot be improved, the employed bees will become the scout bees and will process a random search by using Equation (7).

$$x_{id} = x_d^{\min} + rand(0,1)(x_d^{\max} - x_d^{\min}) \quad (7)$$

where $d = 1,2,...,n.$

Step 5: Memorizing the solution.

Step 6: Obtaining the output. If needed, steps 1 – 5 will be repeated.

3 Methodology

The description of data and evaluation of accuracy for building load forecasting are discussed in this section.

3.1 Description of Data

The set of data input which includes the dew point, dry bulb, humidity and pressure had been implemented in this study in order to evaluate the performance of the proposed method. From that, 70 % had been used as the training data while the remaining which is 439 sets of data was used for testing and represented as 'N'.

3.2 Evaluation of Accuracy

Three different types of evaluation had been used to evaluate the performance of accuracy, which include the Means Absolute Percentage Error (MAPE) and Root Means Square Error (RMSE). All these parameters are a major part in a forecasting study in order to differentiate the capability of the model. All their definitions are expressed as follows:

$$MAPE = \frac{1}{N}\left[\sum_{t=1}^{N}\left|\frac{F_t - A_t}{F_t}\right|\right] \times 100 \quad (8)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(A_t - F_t)^2} \quad (9)$$

where t=1,2,…,x

A_t = Actual Load

F_t = Forecast Load

N = Number of test data

4 Results and Analysis

The performance accuracy results of the proposed GLSSVM- ABC method has been compared with the results of the GMDH, LSSVM, GLSSVM and GABC methods. The performance accuracy results as well as the actual and forecasted loads from all the methods are shown in Table 1. The results achieved by the proposed method were compared using MAPE and RMSE. From the table, it can be seen that all the models have different performance accuracy.

In terms of MAPE, it can be seen that, the proposed method has resulted in an improvement of the performance accuracy when compared to GLSSVM, GABC, LSSVM and GMDH. The improvement in accuracy when compared to GLSSVM is approximately 7 percent. The improvement in accuracy when compared to GABC and LSSVM are 30 percent and 49 percent respectively. The improvement in accuracy when compared to GMDH is 60 percent which is also the highest improvement achieved by the GLSSVM-ABC.

In terms of RMSE, the proposed method again showed the lowest error compared to the other methods. It can be seen that the proposed method has improved performance accuracy when compared to GLSSVM by almost 5 percent, GABC by 31 percent, LSSVM by 49 percent and the highest improvement achieved is 55 percent when compared to GMDH.

From the analysis, it shows that the proposed method has performed better for both MAPE and RMSE compared to the other methods. This is because the incorporation of the ABC with the GLSSVM has resulted in a balanced exploration, where it does not get stuck in a local minimum. This is important to avoid the over-fitting problem between the actual and forecasted results.

Table 1. Performance results of load forecasting

Algorithm	Actual Data (MW)	Forecasted Data (MW)	MAPE	RMSE
GMDH	30.059	29.067	0.00752	0.0473
LSSVM	30.059	30.846	0.00581	0.0421
GABC	30.059	30.632	0.00426	0.0307
GLSSVM	30.059	29.638	0.00319	0.0225
GLSSVM - ABC	30.059	30.457	0.00298	0.0213

5 Conclusions

This paper has described a proposed method for GLSSVM which incorporates the ABC algorithm based on the short term load forecasting, covering aspects such as dry bulb, dew point, humidity, pressure and historical load as input. The simulation result has shown that the proposed method has outperformed the other methods. In terms of MAPE, the proposed method has improved the accuracy of forecasted load in the range of 7% to 60%. The RMSE analysis showed that the accuracy of the forecasted load has been improved in the range of 5% to 55%. From these results, it can be concluded that the GLSSVM-ABC gave better results than the other methods in terms of MAPE and RMSE. It is envisaged that the proposed method has great potential for accurate load forecasting and can be very useful for the purpose of energy management in buildings.

Acknowledgment

This work was supported by the Ministry of Education Malaysia and Universiti Teknologi Malaysia through the Research University Grant (GUP) vot 07H57.

References

1. J. Wang, L. Li, D. Niu, and Z. Tan, An annual load forecasting model based on support vector regression with differential evolution algorithm, *Applied Energy,* **vol. 94**, (2012).
2. H. X. Zhao and F. Magoulès, A review on the prediction of building energy consumption, *Renewable and Sustainable Energy Reviews,* **vol. 16**, pp. 3586 - 3592, (2012).
3. L. Suganthi and Anand A. Samuel, Energy Models for Demand Forecasting-A Review, *Renewable and Sustainable Energy Reviews,* **vol. 16**, pp. 1223-1420, (2012).
4. K. Metaxiotis, A. Kagiannas, D. Askounis, and J. Psarras, Artificial intelligence in short term electric load forecasting: a state-of-the-art survey for the researcher, *Energy Conversion and Management,* **vol. 44**, pp. 1525–1534, (2003).
5. E. Avci, Selecting of the optimal feature subset and kernel parameters in digital modulation classification by using hybrid genetic algorithm–support vector machines, *Expert Systems with Applications,* **vol. 36**, pp. 1391-1402, (2009).
6. B. Dong, C. Cao, and S. E. Lee, Applying support vector machines to predict building energy consumption in tropical region, *Energy and Buildings,* **vol. 37**, pp. 545-553, 5//(2005).
7. Zhijian Hou and Z. Lian, An Application of Support Vector Machines in Cooling Load Prediction, presented at the International Systems and Application 2009, Wuhan, (2009).
8. Z. Mustaffa, Y. Yusof, and S. S. Kamaruddin, Application of LSSVM by ABC in Energy Commodity Price Forecasting, in *Power Engineering and Optimization Conference (PEOCO), 2014 IEEE 8th International*, Langkawi, (2014), pp. 94 - 98.
9. X. Yang, Comparison of the LS-SVM Based Load Forecasting Models, in *IEEE*, ed. Harbin, Heilongjiang, China: IEEE, (2011), pp. 2942 - 2945.
10. O. Hegazy, O. S. Soliman, and M. A. Salam, LSSVM-ABC Algorithm for Stock Price prediction,

International Journal of Computer Trends and Technology (IJCTT), **vol. 7**, pp. 81-92, (2014).

11. R. Samsudin, P.Saad, and A. Shabri, River flow time series using least squares support vector machines, *Hydrology and Earth System Sciences,* vol. 15, pp. 1835-1852, (2011).
12. Hongya Xu, Yao Dong, Jie Wu, and Weigang Zhao, Application of GMDH to Short Term Load Forecasting, *Advanced in Intelligent System,* vol. 138, pp. 27-32, (2012).
13. T. Jacob, U. A. Usman, S. Bemdoo, and A. A. Susan, Short-term Electrical Energy Consumption Forecasting Using GMDH-type Neural Network, *Journal of Electrical and Electronic Engineering,* vol. 3, pp. 42-47, (2015).
14. W.-C. Hong, Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm, *Energy,* vol. 36, pp. 5568-5578, 9// (2011).
15. Z. Mustaffa, Y. Yusof, and S. S. Kamaruddin, Application of LSSVM by ABC in energy commodity price forecasting, in *Power Engineering and Optimization Conference (PEOCO), 2014 IEEE 8th International*, (2014), pp. 94-98.
16. D. Karaboga and B. Basturk, A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm, *Journal of Global Optimization,* vol. 39, pp. 459-471, 2007/11/01 (2007).