An Accurate Medium-Term Load Forecasting based on Hybrid Technique

Z.M. Yasin[1], N.F.A. Aziz[2], N.A. Salim[3], N.A. Wahab[4], N.A. Rahmat[5]
[1,3,4]Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam, Malaysia
[2,5]Department of Electrical Power Engineering, Universiti Tenaga Nasional, Malaysia

Article Info	ABSTRACT
	An accurate medium term load forecasting is significant for power generation scheduling, economic and reliable operation in power system. Most of classical approach for medium term load forecasting only consider total daily load demand. This approach may not provide accurate results since the load demand is fluctuated in a day. In this paper, a hybrid Ant-Lion Optimizer Least-square Support Vector Machine (ALO-LSSVM) is proposed to forecast 24-hour load demand for the next year. Ant-Lion Optimizer (ALO) is utilized to optimize the RBF Kernel parameters in Least-Square Support Vector Machine (LS-SVM). The objective of the optimization is to minimize the Mean Absolute Percentage Error (MAPE). The performance of ALO-LSSVM technique was compared with those obtained from LS-SVM technique through a 10-fold cross-validation procedure. The historical hourly load data are analyzed and appropriate features are selected for the model. There are 24 inputs and 24 outputs vectors for this model which represents 24-hour load demand for whole year. The results revealed that the high accuracy of prediction could be achieved using ALO-LSSVM.

Corresponding Author:

Z.M. Yasin
Faculty of Electrical Engineering,
Universiti Teknologi MARA, Shah Alam, Malaysia
E-mail: zuhailamy74@gmail.com

1.   INTRODUCTION

Electric load forecasting is important in planning, operation and regulation of electric power systems An accurate load forecasting will lead to substantial savings in operating and maintenance costs, increased the stability, reliability and security of the system [1]. Underestimate the load demand may cause the insufficient of power supply to the consumers. Furthermore, it may result the reduction of power quality in the system. On the other hand, overestimation may lead the provider to make unnecessary investment and does not meet the optimum economic power dispatch. Electric load forecast can be divided into 3 types. The first type is long-term load forecasting. The forecast whitin 5 to 20 years is classified as long term load forecasting [2]. This type of forecasting also has non-linear correlation with other factors. The second type is medium-term load forecasting. Medium-term load forecasting (MTLF) can be considered as forecast for monthly up to several year [3]. Operators can rely on MTLF in making decisions for unit commitment, system security analysis, dispatching schedule and load flow analysis. Therefore, improving MTLF accuracy is crucial for increasing the efficiency of systems and reducing the costs [4].

The third type is short-term load forecasting (STLF). STLF mainly covers the period of one week, and refers to the assessment of load per hour during the day [5]. This type of prediction is more specific in time as it considers hourly prediction. The more specific the more accurate the prediction can be. The short-term load forecasting is needed for control and scheduling of power system, power system maintenance, power system operation and contingency analysis [6-7].

Conventional methods such as linear regression methods [8], time-series modelling [9] and general exponential method [10] have been utilized for load forecasting. These methods were able to predict the linear load series and unable to predict non-linear character of load [11]. In line with the rapid development of artificial intelligence algorithms, especially algorithms with strong self-learning such as simulated annealing algorithm, artificial neural network, BP neural network and particle swarm fuzzy inference have been widely used in load predictions. However, all these methods have their own advantages and disadvantages. Recently, support vector machine (SVM), which is suitable for solving practical problems such as load forecasting [12]. An improved version of SVM, Least-Square Support Vector Machine (LS-SVM) applies equality constraints instead of inequality constraints to simplify the complex calculation and improve the training process. In this paper, the hybridized of LS-SVM and Ant-Lion Optimizer (ALO) is presented for medium-term load forecasting.

2. RESEARCH METHOD

The data is obtained from PJM website. PJM is a regional transmission organization (RTO) that coordinates electrical transmission systems in all or parts of Illinois, Delaware, Indiana, Kentucky, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, West Virginia and the District of Columbia. In order to verify the effectiveness of the proposed algorithm, historical load data from Duke is selected. The hourly data for whole days in 2010 and 2011 are used as an input and output for training data and testing data respectively. The hourly data in the first day of January is set to be input while the second day of January will be the output. The hourly data in the 1st day until 364th day will assigned as input while the hourly data in the 2nd day until 365th day will be the output. The data can be downloaded from [13].

2.1. Least-Square Support Vector Machine (LS-SVM)

The approach of LS-SVM is a reformulation of the principles of SVM, which applies equality instead of inequality constraints [14]. The optimization problem in LS-SVM is formulated as:

$$\min_{w,b,e} \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{N} e_i^2 \tag{1}$$

Where w is an unknown coefficient vector, γ is a regularization constant and e is assumed to be a white noise process. Equation (1) is subject to:

$$y_i = w^T \varphi(x_i) + b + e_i, i = 1, \dots, N, \tag{2}$$

Where xi is mapped into a high dimensional feature space with mapping φ. The problem can be solved using Lagrange multipliers and the solution is presented in form:

$$y(x) = \sum_{i=1}^{N} \propto_i K(x, x_i) + b \tag{3}$$

Where K(x,xi) represents kernel, defined as the dot product between the φ(x)T and φ(x). In this paper, Radial Basis function (RBF) is used.

$$K(x, x_i) = e^{\frac{-\lfloor x - x_i \rfloor 2}{\sigma^2}} \tag{4}$$

In LS-SVM with RBF kernel function, the selection of parameters between gamma and sigma is essential. These parameters need to be tuning to minimize training error and improved the prediction performance. This paper proposed 10-folds cross validation technique for the parameters selection. Mean absolute percentage error (MAPE) is used to quantify the performance of the prediction. The lower value of MAPE indicate that the prediction is good. The formula of MAPE is shown in Equation 5:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \times 100\% \tag{5}$$

Where A_t is the actual value and F_t is the forecast value.

Besides MAPE, the evaluation of the estimation is determined by the correlation of determination, R2 as shown in Eq. (6).

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(F_t - A_t)^2}{\sum_{t=1}^{n}(A_t - A_{t,avg})^2} \tag{6}$$

Where $A_{t,avg}$ is the average value of actual LS-SVM.

2.2. Ant Lion Optimizer

The ALO algorithm imitated from the interaction between ant-lions and ants in the trap [15]. This algorithm aspired from 5 important step in the true nature of the ant-lions hunting behavior. The ant-lion build the trap by digging the sand. After that the ant is randomly walk until trapping in the ant-lion's pits. This will give the ant-lion the chance to catch the ant but usually the prey will run away. This will lead to fourth step which the ant-lion will throw the sand making the ant sliding toward ant-lion. At the final step, ant-lion catch the prey and rebuild the pit. Random walks of ants are represented as Equation (7).

$$X_i^t = \frac{(X_i^t - a_i) \times (d_i - c_i^t)}{(d_i^t - a_i)} + c_i \tag{7}$$

Where a_i is the minimum of random walk of i-th variable, d_i is the maximum of random walk in i-th variable, c_i^t is the minimum of i-th variable at t-th iteration, and d_i^t indicates the maximum of i-th variable at t-th iteration. The new equations are formulated based on the random walks of prey that are affected by ant-lion's traps.

$$c_i^t = antlion_j^t + c^t \tag{8}$$

$$d_i^t = antlion_j^t + d^t \tag{9}$$

Where c^t is the minimum of all variables at t-th iteration, d^t indicates the vector including the maximum of all variables at t-th iteration, c_i^t is the minimum of all variables for i-th ant, d_i^t is the maximum of all variables for i-th ant, and $antlion_j^t$ shows the position of the selected j-th antlion at t-th iteration.

The mathematically modeling for the behavior of sliding ants toward ant-lion are formulated as Eq. (10) and Eq. (11). The formulations are based on the radius of ants' random walks that is decreased eventually.

$$c^t = \frac{c^t}{I} \tag{10}$$

$$d^t = \frac{d^t}{I} \tag{11}$$

At the final step, ant-lion catch the prey and rebuild the pit. The step is formulated as:

$$antlion_j^t = ant_i^t \text{ if } f(ant_i^t) > (antlion_j^t) \tag{12}$$

Elitism is a crucial characteristic of evolutionary algorithms that allows them to maintain the best solution(s) obtained at any level of optimization process. The best ant-lion obtained is saved in every iteration and considered as elite. The elite are the fittest ant-lion and should be able to affect the random walks of all the ants in iteration process. Thus, it is assumed that every ant randomly walks around a selected ant-lion by the roulette wheel and the elite simultaneously as follows:

$$ant_i^t = \frac{R_A^t - R_E^t}{2} \tag{13}$$

Where R_A^t is the random walk around the antlion selected by the roulette wheel at t-th iteration, R_E^t is the random walk around the elite at t-th iteration, and ant_i^t is the position of i-th ant_i^t t-th iteration.

Reference [15] proved that the proposed ALO algorithm shows high exploration and exploitation in solving mathematical functions. The proposed random walk mechanism and random selection of ant-lions stimulate exploration which facilitate the ALO algorithm to achieve global optima and solve local optima stagnation when solving complexity problems. Moreover, adaptive shrinking boundaries of ant-lions' traps and elitism emphasize exploitation as iteration increases, which leads to an accurate approximation of the global optimum. All these characteristics require the ALO algorithm to solve real optimization problems potentially and avoid local optima. Therefore, this paper presents the application of ALO for solving load forecasting problems.

2.3. Development of Hybrid LS-SVM

In this paper, a hybrid Ant-Lion Optimizer Least-square Support Vector Machine (ALO-LSSVM) is proposed to forecast 24-hour load demand. As mentioned earlier, in LS-SVM (with the RBF kernel), two parameters need to be tuning which are gamma (γ) and sigma ($\sigma2$). Sigma is the kernel function parameter (squared bandwidth) while gamma is the regularization parameter for determining the trade-off between the training error minimization and smoothness of the estimated function. If the value of sigma is too big, it will lead to under fitting phenomenon to sample data. On the contrary, if the value of sigma is too small, it will lead to over fitting phenomenon to sample data [16]. In ALO-LSSVM, ALO is used to enhance the performance of LS-SVM by optimizing the values of gamma and sigma. The objective of the optimization is to minimize the value of Mean Absolute Percentage Error (MAPE). The overall flowchart of ALO-LSSVM is shown in Figure 1.

Firstly, the ant-lion and ant matrices are initialized randomly. In every iteration, the position of each ant with respect to an ant-lion are updated. Then, the best fitness are selected by the roulette wheel operator and the elite. The boundary of position updating is defined as proportional to the current number of iteration. The updating position is then accomplished by two random walks around the selected ant-lion and elite. When all the ants randomly walk, they are evaluated by the fitness function. If any of the ants become fitter than any other ant-lions, their positions are considered as the new positions for the ant-lions in the next iteration. The best ant-lion is compared to the best ant-lion found during optimization (elite) and substituted if it is necessary. These steps are repeated until the termination criterion is met. The termination criterion is set based on the difference between maximum and minimum fitness which is less than 10-7. The maximum iteration is set to 300 and the number of the search agent is set to 20.
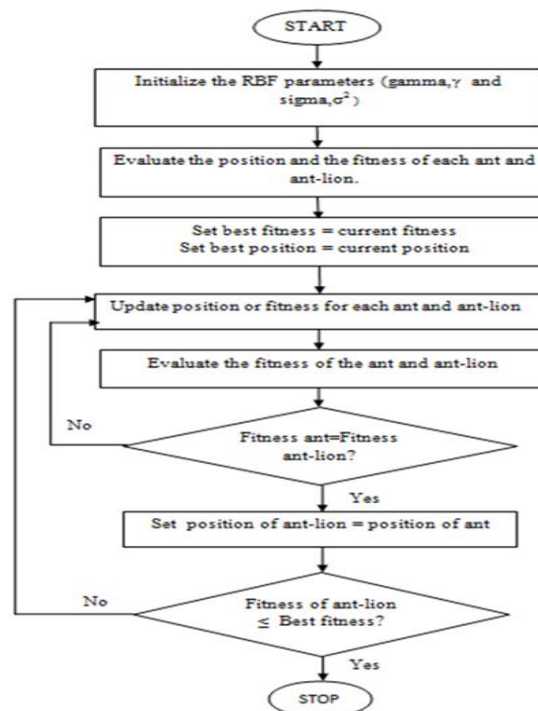

Figure 1. Flowchart of ALO-LSSVM

3. RESULTS AND ANALYSIS

LS-SVM with 10-fold cross validation technique is used to find the value of gamma (γ) and sigma ($\sigma2$) in this paper. The accuracy of prediction is determined by calculating Mean Absolute Percentage Error (MAPE) and correlation of determination (R2). LS-SVM is simulated ten times to determine the best prediction performance. The best, average and worst results in term of MAPE value are tabulated in Table 1. The results revealed that the best value for gamma and sigma are 132.3344 and 44.3020 which produce MAPE of 4.3796%. The lower the MAPE is better, while the R2 should approach to 1 which indicates the good regression line.

Table 1. MAPE and R2 obtained from LS-SVM

	Cross validation technique		
	Best	Average	worst
MAPE (%)	4.3796	4.5453	4.7096
R2	0.8873	0.8756	0.8703

In order to optimize the value of RBF parameters, a new algorithm namely ALO-LSSVM is proposed as described in section 4. In ALO-LSSVM, the Kernel parameters were optimized using ALO. The performance of training data using ALO-LSSVM is illustrated in Figure 2. From the results obtained in Figure 2, the optimum value for gamma (γ) is 340.2442 while for sigma ($\sigma2$) is 321.1076. These values produce 4.356% of MAPE.
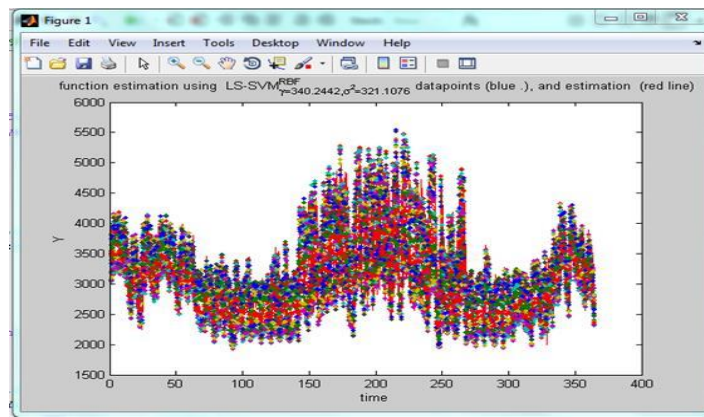

Figure 2. Training data produced by ALO-LSSVM

Table 2 shows the comparison of prediction performance between LS-SVM with cross validation technique and ALO-LSSVM in terms of MAPE and R2. From the results tabulated in Table 2, it can be seen that ALO-LSSVM produced better performance in terms of MAPE value and R2.

Table 2. Comparison of MAPE and R2

Technique	Gamma (γ)	Sigma ($\sigma2$)	MAPE (%)	R2
LS-SVM	132.3344	44.30203	4.3806	0.8873
ALO-LSSVM	340.2442	321.1076	4.3560	0.8908

The performance of ALO-LSSVM for medium term load forecasting is measured through testing process as shown in Figure 3. The figure shows the comparison between predicted and actual data for one year testing data. From the results presented in Figure 3, it can be observed that the predicted and actual data are quite similar. For clear observation on the performance of ALO-LSSVM, graph of testing data for a month (January 2011) and week (first week of January 2011) are plotted in Figure 4 and Figure 5 respectively.
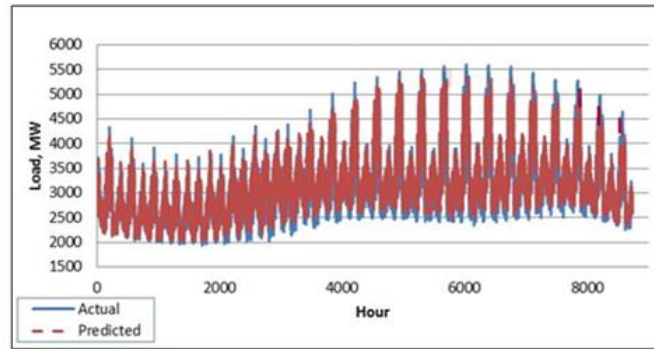
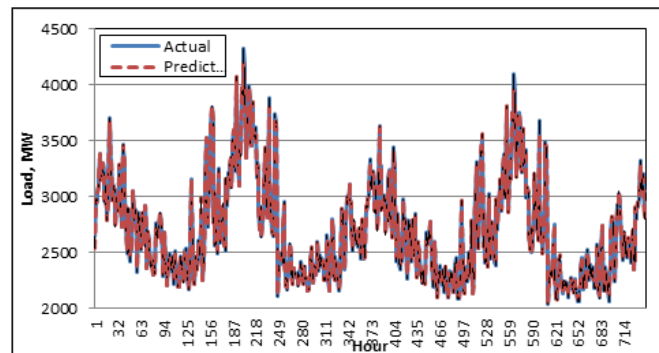Figure 3. Comparison between predicted and actual data in a year

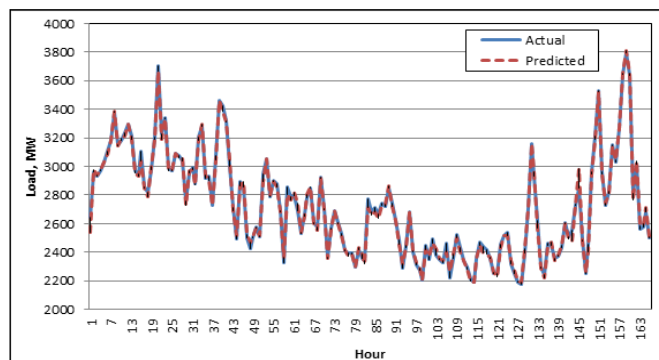Figure 4. Comparison between predicted and actual data in January

Figure 5. Comparison between predicted and actual data in the first week of January

It can be seen from Figure 3 that the electrical usage is highest in summer season while the lowest usage in February to March which are spring season. The highest electrical usage in summer is due to increasing the usage of air-conditioner and also increasing the human activities since it is a holiday. The major maintenance work is best to be done in February to March.

By referring to Figure 5, the first day is Saturday and continues until Friday. Starting from Friday's night, the electricity consumed is increase until Saturday. This is due to more activities in weekend and peoples start to have a great time after working. Based on the graph, the power provider should increase the generation in weekend compared to weekdays. This analysis will help the electricity provider to determine optimal unit commitment and plan the schedule. From all the scenarios have been discussed above, Medium-term Load Forecasting is essential to power Supply Company to determine the electricity consumption in specific time. From the forecasting, the company might not over generate thus will cut-off the operating cost. Besides, the electricity collapse or trip can be avoided.

4.   CONCLUSION

This paper had presented a medium-term load forecasting by using ALO-LSSVM to predict the load demand for every hour in a year. It is become a responsibility to power industry in making precise prediction in order to keep a healthy power supply and competition between the companies in terms of economy. In power planning, it is important not to make overestimation in order to avoid over spent. The determination of tariff also takes the load forecasting as the input to analyze. The most important thing to take into account is the stabilization of the electrical distribution especially at the receiving ends. In order to avoid electrical collapse at a particular area, medium-term load forecasting is needed to give the precise prediction since load demand varies from according to time. The results showed that the accurate prediction based on hourly load demand could be achieved using ALO-LSSVM algorithm. In future, it is suggested that ALO-LSSVM is also utilized for long-term load forecasting to verify the robustness and the nonlinearity of this hybrid technique.

REFERENCES

[1] Almeshaiei E, Soltan H. "A methodology for Electric Power Load Forecasting". *Alexandria Engineering Journal*. 2011: 50(2); 137–144.
[2] Hong T, Fan S. "Probabilistic electric load forecasting: A tutorial review". *International Journal of Forecasting*. 2016: 32(3); 914–938.
[3] Xia C, Wang J, McMenemy K." Short, medium and long term load forecasting model and virtual load forecaster based on radial basis function neural networks". *International Journal of Electrical Power and Energy Systems*. 2010: 32(7); 743–750.
[4] Laouafi A, Mordjaoui M, Laouafi F. "An evaluation of conventional and computational intelligence methods for medium and long-term load forecasting in Algeria". *3rd International Conference on Control, Engineering and Information Technology. Algeria*. 2015. 1-6.
[5] Santika G.D, Mahmudy W.F, Naba A. "Rule optimization of fuzzy inference system sugeno using evolution strategy for electricity consumption forecasting". *International Journal of Electrical and Computer Engineering(IJECE)*. 2017: 7(4); 2241–2252.
[6] Zizi Zhang, Hong Li, Yang Zhao, Xiaobo Hu. "Short-term load forecasting based on the grid method and the time series fuzzy load forecasting method". *International Conference on Renewable Power Generation*. Beijing. 2015. 1-6 .
[7] Bunnoon P. "Electricity Peak Load Demand using De-noising Wavelet Transform integrated with Neural Network Methods". *International Journal of Electrical and Computer Engineering (IJECE)*: 2016: 6(1); 12-20.
[8] Qiu X, Zhang L, Ren Y, Suganthan P, Amaratunga G. "Ensemble deep learning for regression and time series forecasting". *IEEE Symposium Series on Computational Intelligence (CIEL 2014)*. 2014. Florida. 1-6.
[9] Chaouch M. "Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves". *IEEE Transactions on Smart Grid*. 2014: 5(1); 411–419.
[10] Taylor J.W. "Short-term load forecasting with exponentially weighted methods". *IEEE Transactions on Power Systems*. 2012: 27(1); 458–464.
[11] Zeng M, Xue S, Wang Z, Zhu X, Zhang G. "Short-term load forecasting of smart grid systems by combination of general regression neural network and least squares-support vector machine algorithm optimized by harmony search algorithm method". *Applied Mathematics and Information Sciences*. 2013: 7(1); 291–298.
[12] Zheng Y, Zhu L, Zou X. "Short-Term Load Forecasting Based on Gaussian Wavelet SVM". *Energy Procedia*. 2011: 12(1); 387–393.
[13] Historical load data retrieved from: http:///www.jpm.com.
[14] Nguyen K, Le T, Lai V, Nguyen D, Tran D, Ma W. "Least square Support Vector Machine for large-scale dataset". *Proceedings of the International Joint Conference on Neural Networks*. Ireland. 2015. 1-8.
[15] Mirjalili S. "The ant lion optimizer". *Advances in Engineering Software*. 2015: 83(1); 80–98.
[16] De-yue M, Wen-ying L, "Application of Least Squares Support Vector Machine (LSSVM) based on Time Series in Power System Monthly Load Forecasting", *Asia-Pacific Power and Energy Conference*, 2011. 1-4,