Engineering and Applied Science Research July – September 2019;46(3):267-275 Technical

EASR

Engineering and Applied Science Research

https://www.tci-thaijo.org/index.php/easr/index

Published by the Faculty of Engineering, Khon Kaen University, Thailand

Bayesian optimization in a support vector regression model for short-term electricity load forecasting

Su Wutyi Hnin and Chawalit Jeenanunta*

School of Management Technology, Sirindhorn International Institute of Technology, Thammasat University, Pathum Thani 12120, Thailand

Received 17 January 2019
Revised 30 April 2019
Accepted 3 May 2019

Abstract

Purpose: Support vector regression (SVR) has long been known as a great tool in forecasting. SVR combined with evolutionary algorithms have been used in many remarkable applications. Bayesian optimization (BO) has the ability to find good points in a search space without many function evaluations. This paper presents short-term load forecasting (STLF) in Thailand using SVR with Bayesian optimization (BO). The purpose of this paper is to improve forecasting accuracy by optimizing the hyperparameters of SVR.

Design/methodology/approach: The Electricity Generating Authority of Thailand (EGAT) provides 30 minute load data for Bangkok and the metropolitan region. The data from August 2015 to July 2017 is used for training and testing. Mean absolute percentage error (MAPE) and tracking signal (TS) are used to measure the performance of the proposed model. The hyperparameters of SVR are optimized using three algorithms, the genetic algorithm (GA), particle swarm optimization (PSO), and Bayesian optimization (BO). Findings: By comparing the MAPE results, the SVR-BO outperforms the other two algorithms.

Keywords: Short-term electricity load forecasting, Bayesian optimization, Support vector regression, Data preprocessing

1. Introduction

Short-term load forecasting (STLF) is an essential part of managing operations such as short-term maintenance, unit commitment, and power flow dispatch optimization [1]. STLF focuses on a time period of one minute to a week. It is used for daily load planning, load flow management and capacity scheduling for electricity generation. STLF uses daily and weekly cycles. There are many factors affecting load demand such as seasonality, the day of the week, the month of year and temperature. Complex and non-linear effects are difficult to forecast using traditional techniques. Moreover, the load pattern diverges from normal on holidays, the days close to holidays and weekends.

The foundation of a support vector machine (SVM) was developed by Vapnik (1995). SVM was developed to solve classification problems. Then, it was extended to the domain of regression problems and named Support Vector Regression (SVR) [2-4]. SVR can model non-linear relations and give great performance in forecasting [5]. It can effectively generalize and is robust in higher dimensions, which means there is no over-fitting. The objective of SVR is to prevent settling on a local optimum and reach the global optimum. It is characterized by hyperparameters that control the behavior of the function. These hyperparameters highly affect forecasting accuracy. Therefore, optimizing the hyperparameters in SVR is a major issue in training the model. Many researchers have suggested setting the hyperparameters of SVR using optimizing algorithms [6].

Bayesian optimization (BO) was first studied by Kusher in 1964 and then Mockus in 1978. Methodologically, it touches on several important machine learning areas: active learning, contextual bandits, and Bayesian nonparametric approaches. BO has received serious attention in machine learning since 2007. It is an excellent tool for finding good machine learning hyperparameters.

2. Literature review

There are various forecasting techniques that have been applied to STLF because of its significant economic and environmental implications [7]. It can be classified into two methods, traditional and artificial intelligence-based methods. The most well-known traditional techniques are exponential smoothing [8], regression [9], autoregression (AR) [10], autoregressive moving average (ARMA) [11-12], and autoregressive integrated moving average (ARIMA) [13]. The artificial intelligence based methods are fuzzy systems [14-16], neural networks (NN) [17] and support vector regression (SVR) [18-21]. A computational intelligence-based model is effective in solving non-linear and complex equations [22].

*Corresponding author. Tel: +66 2501 3505
Email address: chawalit@siit.tu.ac.th
doi: 10.14456/easr.2019.30

By introducing the structural risk minimizing principle, SVR gets better generalization capability. Due to its ability to solve non-linear problems, SVR became an efficient model among machine learning algorithms [23]. SVR has good MAPE and RMSE performance among machine learning algorithms [24]. In [25], the strength of SVR is seen by comparing traditional methods (AR) with larger training data sets. The SVR parameters significantly affect the accuracy of the model, as has been shown by many researchers. Therefore, choosing the best parameters is an important issue in SVR modelling. There are many algorithms that can be used to select the best SVR parameters. The most widely used algorithms are the genetic algorithm (GA) [26-27], chaotic genetic algorithms [28-29], particle swarm optimization (PSO) [6, 30], simulated annealing (SA) [31], ant colony optimization (ACO) [32], and Bayesian optimization (BO) [33].

Hybridizing Bayesian optimization with support vector regression (SVR) to enhance the forecasting accuracy is an active area in forecasting. This hybrid algorithm combines the advantages of SVR and Bayesian inference [34]. SVR is highly suitable for load forecasting as it can map non-linear relationships. BO has received attention in recent years because it is compatible with machine learning algorithms such as deep learning, neural networks and support vector machines [35]. It deals with the problem of local minima and computational complexity in training [33].

This paper progresses as follows. An overview of SVR and BO are given in Section 2. Section 3 covers the combination of SVR-BO and Section 4 gives a discussion of the data. The results of forecasting performance are discussed in Section 5. Then, the conclusions and suggested future research are presented in the last session.

3. Methodology

3.1 Support vector regression

The approach of support vector regression (SVR) is based on linear regression in a feature space [3]. Figure 1 demonstrates the process of SVR considering its error value. According to the figure, the regression relies on support vectors and the errors are ignored that are smaller than *ε*.

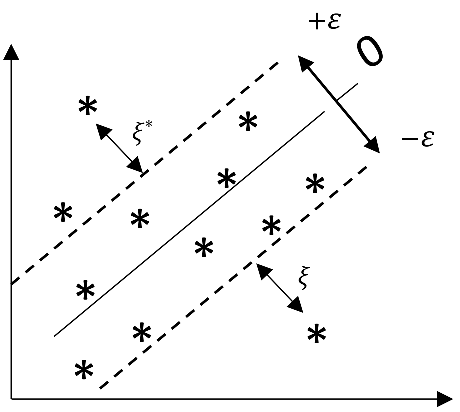

Figure 1 Simple example of SVR

$L = \{(a_i, b_i)\}_{i=1}^{D}$, where a_i is the input data and b_i is the target, the problem can be defined by the following equation:

$$f(x) = g\emptyset(x) + e \quad (1)$$

where, g represents the coefficients of regression and e is a bias term. It can be solved by the following structural risk minimization (SRM) function [2]:

$$SRM = \frac{1}{2}\|g\|^2 + C(\sum_{i=1}^{l} \xi + \xi^*) \quad (2)$$

Subjected to the constraints:

$$b_i - ga_i - e \leq \varepsilon + \xi$$
$$ga_i + e - b_i \leq \varepsilon + \xi^*$$
$$\xi \geq 0, \xi^* \geq 0$$

where, the width of the loss function is represented by ε, and this parameter that can be tuned to handle model complexity. The training data set is denoted as *C*. ξ_i^* and ξ_i are slack variables which must have non-negative values to ensure suitable constraints.

Three types of kernel functions are used in SVR. They are a linear kernel function (LKF), radial based function (RBF), and polynomial function (PLF). Among them, RBF is most applied since it only depends on the parameter γ to tune using the dataset. The RBF kernel function is $R(a, a_i)$:

$$R(a, a_i) = exp\gamma\|a - a_i\|^2 \quad (3)$$

where, a and a_i are the inputs for the specific dimensions and the diameter of the RBF represented by γ. SVR model performance significantly relies on hyperparameters (C_i, ε_i, γ_i). In this paper, Bayesian optimization (BO) is proposed to select the proper hyperparameters.

3.2 Bayesian optimization

The goal of the Bayesian optimization (BO) is to minimize or maximize an objective function *f(x)* for *x* in a bounded area. BO uses stochastic and deterministic functions. The results can be changed to different values since it is calculated at the same point. BO internally carries out a Gaussian process model (GPM) and trains the model by applying objective function evaluations [36]. One useful feature of BO is the application of an acquisition function (AF). It is an algorithm to evaluate the next point. BO is well-suited for optimizing parameters of other functions. It is developed for an objective function and is slow in evaluating values. BO does not need initial starting values to find a global solution unlike other algorithms [37]. It uses a GPM of *f*(*x*), updating the GPM at each new objective function. An acquisition function is formed in GPM to evaluate the next point *x*.

3.2.1 Gaussian Process Model (GPM)

The GPM that is managed by Bayesian Optimization (BO) is widely used in machine learning. GPM is a prior distribution which is effective in its handling of functions. The covariance and mean functions are considered in a Gaussian process [38].

The data set, $a = \{a_1, \ldots, a_n\}$ and the function $f = \{f(a_1), \ldots, f(a_n)\}$ are brought up by GPM.

$$f \sim GPM(P, L) \quad (4)$$

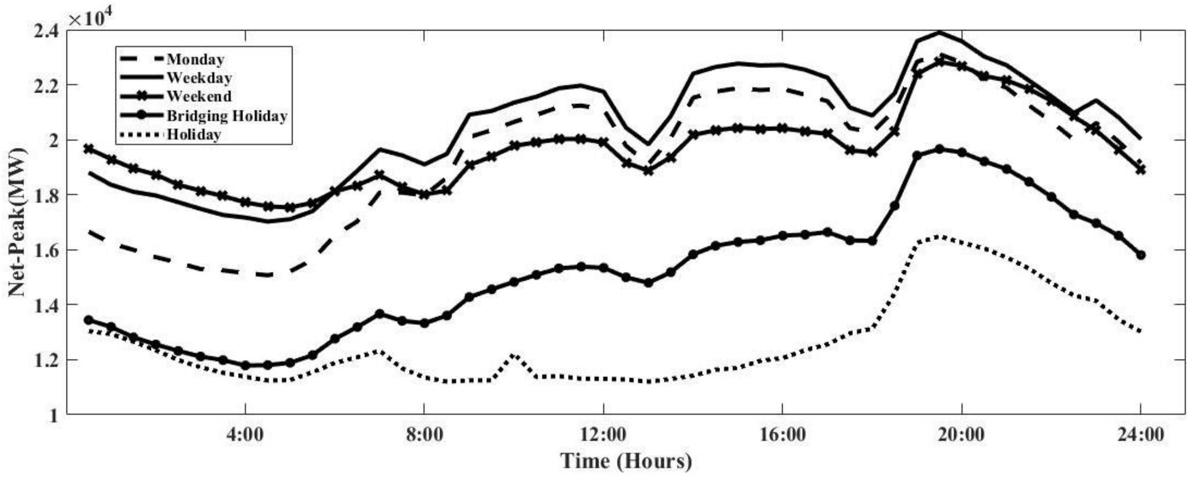

Figure 2 Various load demands for five significant groups in January 2016

where:

$p: A \rightarrow R$ is the mean function that manages $p(x) = E[f(a)]$.

$K: A^2 \rightarrow R$ is the covariance of function that is determined as:

$$R(a, a_i) = [E(f(a) - p(a))(f(a_i) - p(a_i)\ ] \quad (5)$$

3.2.2 Acquisition Function (AF)

In the recent years, the use of an acquisition function (AF) in BO was innovative [39]. It is used to determine the next best point. It can balance the sampling points that explore the areas that could not be trained effectively and get a minimum fitness function. Expected Improvement (EI), Lower Confidence Bound (LCB), and Probability of Improvement (PI) are widely applied in AF. In this paper, EI is applied to select the next best point.

$$EI(a) = \ [E_q \max\left(0, \mu_q(a_{best}) - f(a)\right)] \quad (6)$$

where:

a_{best} = the point for the minimum posterior mean
$\mu_q(a_{best})$ = the minimum value of the posterior mean

4. Dataset

The data was provided by the Electricity Generating Authority of Thailand (EGAT). It is collected at half-hour intervals each day since it is easy to see the load pattern and avoids the problem of having too much data. It has 48 periods (or bins) per day. We use the peak load data from August 2015 to July 2017.

4.1 Data cleaning

Since there are different load patterns in daily electricity demand, the load demand needs to be filtered and cleaned [40]. In Figure 2, there are five significant load patterns: Mondays, weekdays, weekends, bridging holidays, and holidays. The days before and after the holidays are bridging holidays. Both holiday and bridging holiday loads are lower than on normal weekdays. Monday loads are the highest among the other days. Therefore, we filter the load data to find similar load patterns for each type of day.

There are four processes for data cleaning. Replacing the holidays by a weighted moving average is the first step. Then, the bridging holidays are also replaced with a weighted moving average. For the third step, a time-window filtering band $B_t(d)$ is created to detect outliers. Last, the outliers are detected using a time-window filtering band and replaced by a k window moving average (k=4).

$$B_t(d) = \frac{\left[\sum_{i=1}^{k} L_t(d - 7 \times i)\right]}{k} \pm N \times SD(V_t(d)) \quad (7)$$

$$V_t(d) = \left[L_t(d), L_t(d-7), L_t(d-14).....L_t(d-7 \times m)\right] \quad (8)$$

where:

$L_t(d)$ = Peak load at period t for day d,
$L_t(d\text{-}7)$ = Previous week the same day peak load at period t for day d,
m = Week number,
$V_t(d)$ = Time-window based filtering band of day d,
$SD(V_t(d))$ = Standard deviation of all periods in the time-window $V_t(d)$,
k = Period number,
N = Size of the filtering band,
t = Period (bin) number (1, 2,….48)

The width of the window filtering band depends on the size of N. Based on experience, the optimal size of N is set as 1.6 to detect outliers in this research. After cleaning and filtering the load, there are no big changes in the dataset. The load demand has a similar pattern. Figure 3 presents the load patterns for five different groups after cleaning.

4.2 Data segmentation for training and testing

First, the data is arranged into seven segments representing a different day of the week. If we forecast a Monday load, the training data set contains only Monday data. Therefore, there are seven separate dataset segments for training and testing. A walk-forward testing routine [41] is applied to training and testing sets. The model is trained with 52 dataset and then tested with one dataset. Then, the data window slides forward one window at a time and the same process is performed for another 51 datasets. Figure 4 shows

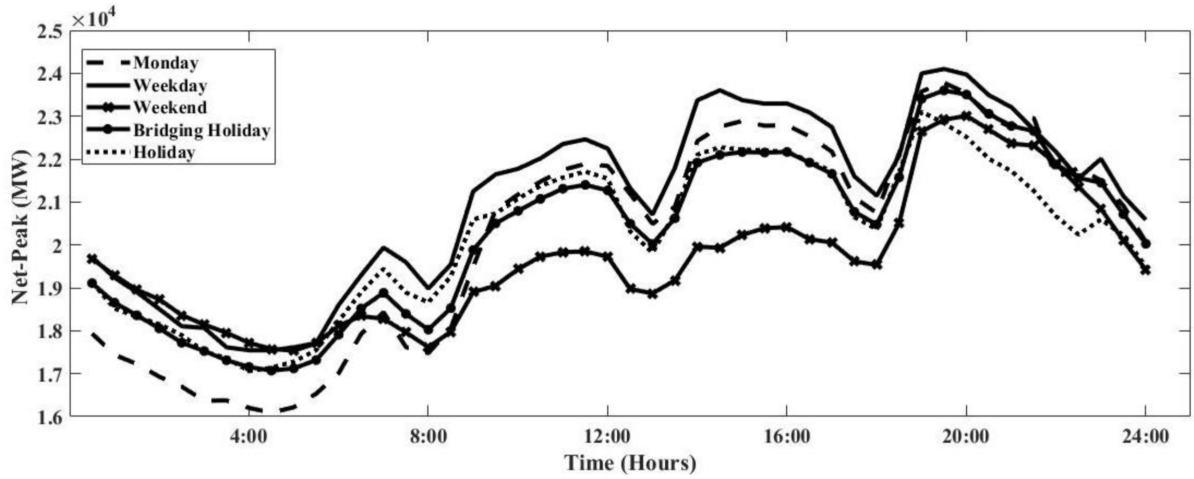

Figure 3 Various load demands after replacing outliers of five different groups in January 2016

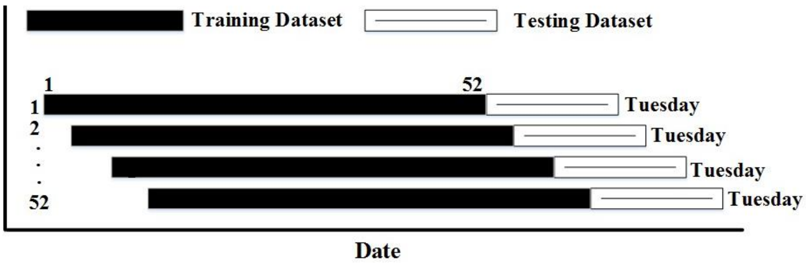

Figure 4 Walk-forward testing routine

Table 1 Sample arrangement of data training and testing for forecasting 2 August 2016

			Training			
No.	$L_t(d-1)$	$L_t(d-7)$	$T_t(d-1)$	$T_t(d)$	MoY	**Target** $F_t(d)$
1	09-08-15 (Mon)	3-08-15 (Tue)	09-08-15 (Mon)	10-08-15 (Tue)	8	10-08-15 (Tue)
	. . .	. . .	. . .	. . .	. . .	. . .
52	25-07-16 (Mon)	19-07-16 (Tue)	25-07-16 (Mon)	26-07-16 (Tue)	7	26-07-16 (Tue)
			Testing			
No.	$L_t(d-1)$	$L_t(d-7)$	$T_t(d-1)$	$T_t(d)$	MoY	$F_t(d)$
1	01-08-16 (Mon)	26-07-16 (Tue)	01-08-16 (Mon)	02-08-16 (Tue)	8	02-08-16 (Tue)

that the testing data window slides one dataset forward and the SVR is trained with a new training dataset using the same process.

4.3 Data arrangement for training and testing

The training data are from August 2015 to July 2016 and the performance is evaluated on the data from August 2016 to July 2017. There are five inputs into the model, yesterday's load (L_t (*d-1*)), previous week-same day load (L_t(*d-7*)), yesterday's temperature (T_t (*d-1*)), today's temperature T_t(*d*) and month of the year (*MOY*) to forecast the load F_t (*d*). Table 1 shows the data arrangement of training and testing data to forecast the load on 2 August 2017.

4.4 Forecasting performance

The performance of the model is calculated from the mean absolute percentage error (*MAPE*) and tracking signal (*TS*). There is a total of 48 observations or 48 forecasted values for each day, *d*. *MAPE* (*d*) and *TS* (*d*) are calculated using the following equations:

$$MAPE(d) = \left[\left(\frac{1}{t}\sum_{t=1}^{48}\left|\frac{L_t(d)-F_t(d)}{L_t(d)}\right|\right)\right]\times 100\,\% \tag{9}$$

$$AFE = \sum_{t=1}^{48}(L_t(d)-F_t(d)) \tag{10}$$

$$MAD = \sum_{t=1}^{48}\left|L_t(d)-F_t(d)\right| \tag{11}$$

$$TS = \frac{AFE}{MAD} \tag{12}$$

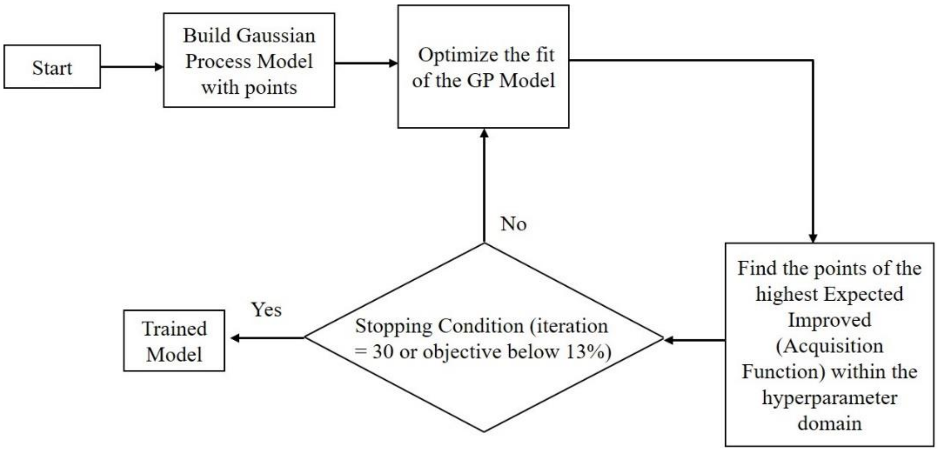

Figure 5 SVR-BO flow chart

Table 2 Upper bound and lower bound of Hyperparameters

SVR (Hyperparameters)	**Interval (Upper Bound – Lower Bound)**
C	1×10^{-3} - 1000
γ	1×10^{-2} - 300
ε	1×10^{-1} - 1000

Table 3 Hyperparameters of the three models

	SVR-GA			**SVR-PSO**			**SVR-BO**		
	C	γ	ε	**C**	γ	ε	**C**	γ	ε
Mon	10	1	0.3	156.787	125.52	1.654	10.431	15.369	13.066
Tue	4.890	120.784	6.778	455.678	23.668	2.989	934.410	117.530	2.909
Wed	67.472	699.765	0.0981	8.946	123.778	4.112	5.869	14.392	826.511
Thurs	876.898	78.669	1.655	187.995	6.889	0.456	279.132	93.674	193.254
Fri	98.767	0.884	120.898	845.923	0.033	14.766	0.046	0.612	475.752
Sat	232.344	998.657	7.893	6.789	0.562	78.339	983.293	282.723	5.353
Sun	0.098	125.773	9.347	723.766	1.200	129.551	63.638	20.428	84.273

where:

$F_t(d)$ = Forecast load at period t for day d,
$L_t(d)$ = Actual load at period t for day d,
t = 1, 2, 3, …, 48 periods,

4.5 SVR-BO Model

Many optimization algorithms are employed to determine the hyperparameters of SVR. Bayesian optimization is applied to tune the SVR hyperparameters in this study. The most important part before modelling the hybrid function is encoding. The hyperparameters that have to be optimized in SVR are C, ε, and γ. The upper bound and lower bound of the hyperparameters are shown in Table 2. The particle is encoded as $A_i = (C_i, \varepsilon_i, \gamma_i)$. Figure 5 shows an outline of the SVR-BO algorithm. The model updates the Gaussian process and a new *a* value from the acquisition function is used to get the best hyperparameter. The model stops after 30 iterations or the objective value is under 13% of the cross-validation error rate. If the stopping conditions are not satisfied, the Gaussian process is updated using the acquisition function. Finally, the model is ready to forecast electricity demand.

5. The forecasting results and discussion

5.1 Hyperparameter settings

Since the data arrangement is divided into seven groups, one for each day, the suitable hyperparameter set also has seven different values groups. There are three optimization algorithms to be compared using the forecasting accuracy in this study. The first one is the genetic algorithm, which is a herustic method based on the process of natural selection. The fittest values are chosen for reproduction to produce the offspring of the next generation. The population size of the GA is 50 and the same number of chromosomes are used for the next generation from the parents of the previous generation. In the next generation, 50% result are from crossover, 45% from mutation and 5% become elite.

Secondly, particle swarm optimization is chosen to optimize the hyperparameters of SVR since it has had good performance in forecasting. The number of swarm size is 50 and the weight of global and local factors, c_1 and c_2, is 2. We set the tolerance value to correlate with the testing data to force improvement. The most suitable tolerance value for this experiment is 1×10^{-4}.

The last algorithm is the proposed model, Bayesian optimization. It tends to seek very good points in the search space with relatively few function evaluations. The hyperparameters of three models for each day are in Table 3.

5.2 The forecasting performance of the proposed model (SVR-BO)

For the SVR-BO model, to clearly see the daily load forecasting performance, the forecast results of each day

Table 4 Yearly and monthly MAPE of each day for SVR-BO

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
August (2016)	4.23	5.02	2.66	1.64	2.14	3.33	4.90
September (2016)	3.13	2.81	1.99	3.61	2.36	2.47	3.46
October (2016)	3.70	4.09	2.84	2.38	1.31	2.06	4.68
November (2016)	2.70	3.99	1.24	2.91	1.03	2.09	4.39
December (2016)	9.80	4.39	5.89	4.78	2.50	4.20	5.23
January (2017)	10.02	1.40	1.66	4.23	2.99	1.47	3.39
February (2017)	1.52	5.40	4.12	7.03	3.17	8.26	4.52
March (2017)	3.18	5.66	3.72	3.93	1.62	1.98	3.81
April (2017)	7.51	5.12	6.49	0.95	2.45	3.25	7.11
May (2017)	8.73	4.44	3.141	2.89	5.41	3.49	6.93
June (2017)	2.69	4.99	3.84	4.65	2.00	2.86	4.76
July (2017)	1.83	1.92	3.84	3.34	1.62	2.04	2.95
Yearly Average	4.98	4.10	3.50	3.53	2.38	3.12	4.68

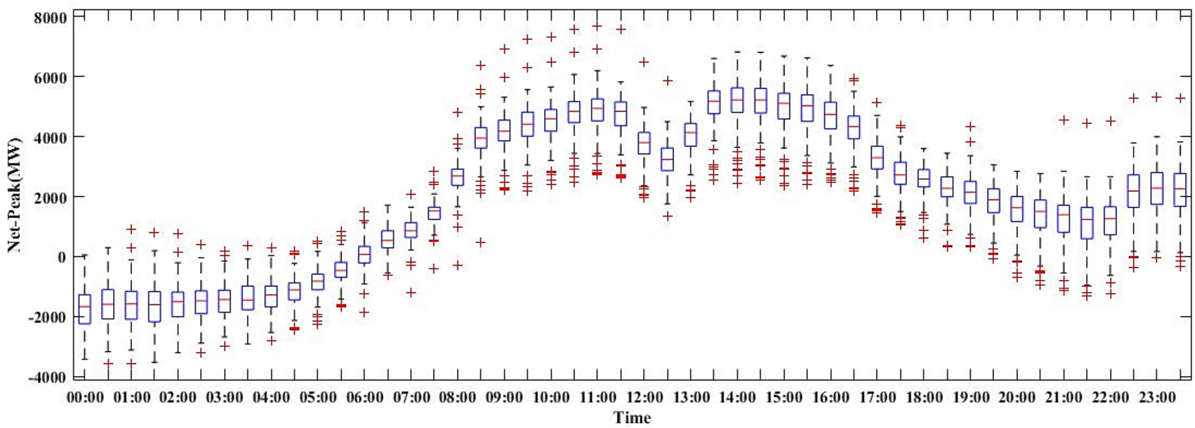

Figure 6 The load difference between Monday and Sunday

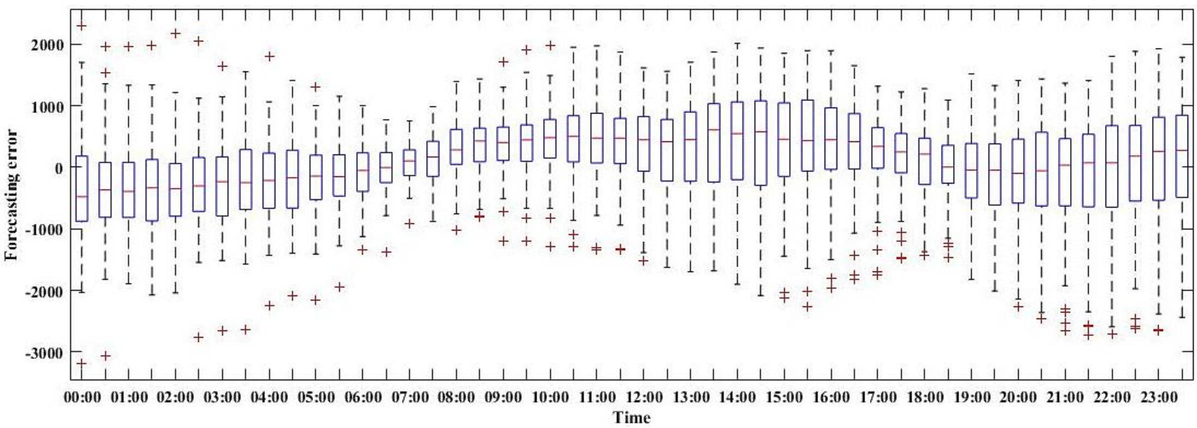

Figure 7 Forecasting error of Monday for SVR-BO

from August 2016 to July 2017 are shown in Table 4. The MAPE of the Friday group is better than the other groups as its training input is Thursday, which has a similar load pattern. Therefore, the group of Wednesday, Thursday, and Saturday MAPE are related with similar values in the training method. Sunday, Monday and Tuesday groups give the highest MAPE. The MAPE of Monday group is especially high, 4.98. This group has a high MAPE because the Saturday load data includes the training input dataset for Sunday. Although Saturday and Sunday are weekend holidays, the load demand of Saturday is higher than Sunday as some businesses are open on Saturday.

The load demand of Monday is significantly lower than Tuesday, mostly in December, January, and February. It causes the Tuesday group to have the highest MAPE. Moreover, the group with the worst MAPE is Monday. It is clear that the load difference between Monday and Sunday is large, as is graphically shown in Figure 6. There are six data summaries, outliers, maximum, upper quartile, median, lower quartile, and minimum. The outliers are more than 3/2 times the upper and lower quartiles. The load demand in the morning for Monday is lower than for Sunday. During the day and night times, except for some periods (20:00 to 23:30), the load demand for Monday is much larger than the load demand of Sunday. Small variations occur between 5:00 and 7:30. The largest variations affect the forecasting performance. Figure 7 gives details of the Monday forecasting errors, which are periods with the highest error.

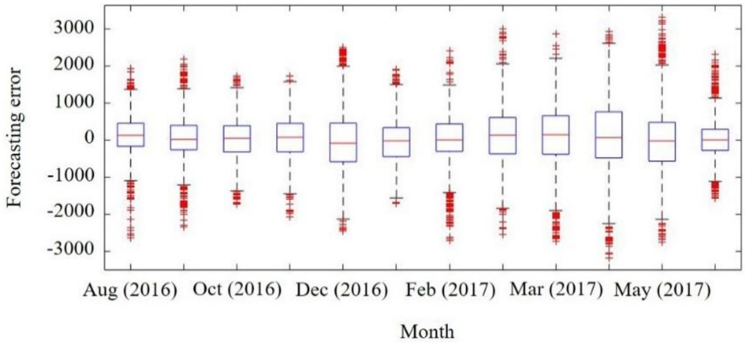

Figure 8 Boxplot for monthly forecasting error variation from August 2016 to July 2017 for SVR-BO

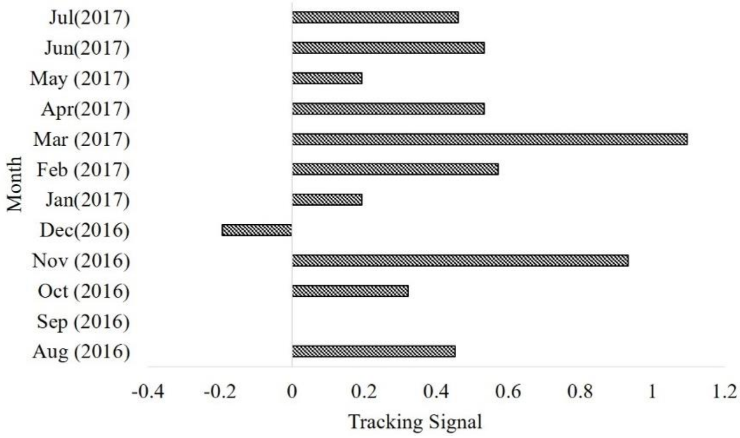

Figure 9 Monthly tracking signal from August 2016 to July 2017 for SVR-BO

Comparing Figures 6 and 7, it can be seen that larger load differences result in higher forecasting errors.

The box plot in Figure 8 helps to investigate the distribution of errors for each month. There are big variations in December and April since there are many holidays. Although holiday and bridging holiday load data are cleaned, the electricity usage in December is still less than the other months.

Moreover, the tracking signal (TS) provides insight which months are subject to over-forecasting and under-forecasting. If the TS value is less than zero, over-forecasting is evidenced. Positive values of TS indicate under-forecasting. Almost all the months are under-forecast except December which has a tracking signal of -0.23, as can be seen in Figure 9. December is the hardest month to forecast since it has low electricity use. So, it is always over forecast. The TS value in March and November are the largest among the under-forecasted months. The forecast performance of March, November, and December needs improvement, as can be seen in Figure 9.

5.3 Comparison of the forecasting performance of the proposed model with the other two algorithms

To clearly see the forecast performance of each month using the three optimization algorithms, the forecast results are shown in Table 5. The SVR-BO model gives better performance except for October (2016) and December (2016), where the MAPE of SVR-BO is larger than SVR-PSO. Since the influence of long holidays and low load demand, the MAPE has larger values in December, February, April and May.

Table 5 Yearly and monthly average MAPE of the three optimizing algorithms

	SVR-BO	**SVR-PSO**	**SVR-GA**
August (2016)	3.21	3.50	4.15
September (2016)	2.84	3.07	3.92
October (2016)	2.96	2.83	4.18
November (2016)	2.62	2.93	3.82
December (2016)	4.38	4.74	3.90
January (2017)	3.12	3.74	4.26
February (2017)	4.35	4.39	4.06
March (2017)	3.39	3.84	3.60
April (2017)	4.51	4.57	4.35
May (2017)	4.74	4.76	4.33
June (2017)	3.66	4.03	3.67
July (2017)	2.52	3.16	4.06
Yearly Average	3.53	3.80	4.02

The yearly average MAPE is used to select the best optimization algorithm to train SVR for short-term load forecasting. According to the yearly average MAPE, SVR-BO has the smallest MAPE, 3.53. The other two optimizing algorithms, SVR-PSO and SVR-GA, have MAPE values of 3.8 and 4.02, respectively. SVR-BO improves forecast accuracy more than others.

Alternatively, the electricity load has several load patterns and there are other ways to investigate the forecast performance of the three optimization algorithms. Thailand has five different load patterns with five separate groups for representing the outcomes of forecasting in Figure 10. This figure summarizes the yearly average MAPE of the different categories, Mondays, weekdays, weekends,

Figure 10 Yearly average MAPE of different categories for three optimization algorithms

holidays, and bridging holidays using three optimization algorithms. Comparing the three algorithms, SVR-BO gives the minimum yearly average MAPE in all categories. The maximum yearly average MAPE is seen with SVR-GA except for the weekdays group. Since the data is cleaned and filtered for holidays and bridging holidays, the performance of all algorithms of these two groups is good, especially for SVR-BO on the holiday group. The Monday group yields the largest MAPE for all algorithms according to the training method. Since the load on Sunday is significantly lower than Monday, it affects the performance of Monday regardless of the training method.

6. Conclusions

This research proposes short-term load forecasting using support vector regression with Bayesian optimization. Historical data from 2015 to 2017, which are collected from the Electricity Generating Authority of Thailand (EGAT), is used for this research.

We compare our proposed model (SVR-BO) with two optimization algorithms, SVR-GA, and SVR-PSO, by testing a one-year dataset (August 2016 – July 2017). The motivation of this research is to optimize the hyperparameters to get good forecasting performance. PSO and GA are easily trapped by local minima. The superiority of BO is that it moves beyond the local minima as the pattern is not stable at one objective value. We observe that the forecast of SVR-BO is more accurate than the other algorithms, especially on holidays and bridging holidays. SVR-BO is very effective in selecting suitable hyperparameters.

From our findings, the MAPE of December, April and May are still high even though we optimized the hyperparameters of SVR. The electricity demand in December is lower than the other months, so the demand is over-forecast and the forecasting error variation is higher. Therefore, the December results could be improved by training using only the December load. Moreover, the forecast results on Monday can be improved by using a different data cleaning method. Accordingly, future research may investigate the training inputs and modify the filtering techniques.

7. Acknowledgements

This research is partially supported by the Logistics and Supply Chain Systems Engineering Research Unit (LogEn). Sirindhorn International Institute of Technology (SIIT), Thammasat University (TU), Thailand. Data used in this research is provided by EGAT. Therefore, we acknowledge their support for completing this research successfully.

8. References

[1] Fan GF, Peng LL, Zhao X, Hong WC. Applications of hybrid EMD with PSO and GA for an SVR-based load forecasting model. Energies. 2017;10(11):1-22.

[2] Smola AJ, Schölkopf B. A tutorial on support vector regression. Stat Comput. 2004;14(3):199-222.

[3] Vapnik VN, Vapnik V. Statistical learning theory. New York: Wiley; 1998.

[4] Vapnik V. The nature of statistical learning theory. New York: Springer; 2013.

[5] Li YC, Fang TJ, Yu EK. Study of support vector machines for short-term load forecasting. Proceedings of the CSEE. 2003;23(6):55-9.

[6] Lin SW, Ying KC, Chen SC, Lee ZJ. Particle swarm optimization for parameter determination and feature selection of support vector machines. Expert Syst Appl. 2008;35(4):1817-24.

[7] Fan GF, Peng LL, Hong WC. Short term load forecasting based on phase space reconstruction algorithm and bi-square kernel regression model. Appl Energ . 2018;224:13-33.

[8] Gardner ES. Exponential smoothing: The state of the art. J Forecast. 1985;4(1):1-28.

[9] Papalexopoulos AD, Hesterberg TC. A regression-based approach to short-term system load forecasting. IEEE Trans Power Syst. 1990;5(4):1535-47.

[10] Akaike H. Fitting autoregressive models for prediction. Ann Inst Stat Math. 1969;21(1):243-7.

[11] Huang SJ, Shih KR. Short-term load forecasting via ARMA model identification including non-Gaussian process considerations. IEEE Trans Power Syst. 2003;18(2):673-9.

[12] Taylor JW. Triple seasonal methods for short-term electricity demand forecasting. Eur J Oper Res. 2010;204(1):139-52.

[13] Conejo AJ, Plazas MA, Espinola R, Molina AB. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. IEEE Trans Power Syst. 2005;20(2):1035-42.

[14] Mbarek MB, Feki R. Using fuzzy logic to renewable energy forecasting: a case study of France. Int J Energ Tech Pol. 2016;12(4):357-76.

[15] Li W, Yang Y, Yang Z, Zhang C. Fuzzy system identification based on support vector regression and genetic algorithm. Int J Model Ident Contr. 2011;12(1-2):50-5.
[16] Draidi A, Labed D. A Neuro-fuzzy approach for predicting load peak profile. Int J Electr Comput Eng. 2015;5(6): 1304-10.
[17] Jeenanunta C, Abeyrathna KD. Combine particle swarm optimization with artificial neural networks for short-term load forecasting. Int Sci J Eng Tech. 2017;1(1):25-30.
[18] Basak D, Pal S, Patranabis DC. Support vector regression. Neural Inform Process-Lett Rev. 2007;11(10):203-24.
[19] Drucker H, Burges CJ, Kaufman L, Smola AJ, Vapnik V. Support vector regression machines. In: Mozer MC, Jordan MI, Petsche T, editors. Advances in neural information processing systems. USA: MIT Press; 1997. p. 155-61.
[20] Chawalit J, Su WH. A Support vector regression model(SVR) for short term electricity forecasting in Thailand. 2016 International Conference on Applied Statistics; 2016 Jul 13-15; Phuket, Thailand.
[21] Su WH, Chawalit J. Short-term electricity load forecasting in Thailand: an analysis on different input variables. IOP Conf Ser: Earth Environ Sci. 2018;192(2018):1-8.
[22] Erişen E, Iyigun C, Tanrısever F. Short-term electricity load forecasting with special days: an analysis on parametric and non-parametric methods. Ann Oper Res. 2017:1-34.
[23] Barman M, Choudhury ND, Sutradhar S. A regional hybrid GOA-SVM model based on similar day approach for short-term load forecasting in Assam, India. Energy. 2018;145:710-20.
[24] Sousa JC, Jorge HM, Neves LP. Short-term load forecasting based on support vector regression and load profiling. Int J Energ Res. 2014;38(3):350-62.
[25] Mohandes M. Support vector machines for short-term electrical load forecasting. Int J Energ Res. 2002;26(4):335-45.
[26] Hong WC. Hybrid evolutionary algorithms in a SVR-based electric load forecasting model. Int J Electr Power Energ Syst. 2009;31(7-8):409-17.
[27] Hong WC, Dong Y, Chen LY, Lai CY. Taiwanese 3G mobile phone demand forecasting by SVR with hybrid evolutionary algorithms. Expert Syst Appl. 2010;37(6):4452-62.
[28] Hong WC, Dong Y, Zhang WY, Chen LY, Panigrahi EP. Cyclic electric load forecasting by seasonal SVR with chaotic genetic algorithm. Int J Electr Power Energ Syst. 2013;44(1):604-14.
[29] Ju FY, Hong WC. Application of seasonal SVR with chaotic gravitational search algorithm in electricity forecasting. Appl Math Model. 2013;37(23):9643-51.
[30] Abeyrathna KD, Jeenanunta CJ. Hybrid particle swarm optimization with genetic algorithm to train artificial neural networks for short-term load forecasting. Int J Swarm Intell Res. 2019;10(1):1-14.
[31] Sadati N, Zamani M, Mahdavian HRF. Hybrid particle swarm-based-simulated annealing optimization techniques. IECON 2006 - 32nd Annual Conference on IEEE Industrial Electronics; 2006 Nov 6-10; Paris, France. USA: IEEE; 2006. p. 644-8.
[32] Niu D, Wang Y, Wu DD. Power load forecasting using support vector machine and ant colony optimization. Expert Syst Appl. 2010;37(3):2531-9.
[33] Gopi G, Dauwels J, Asif MT, Ashwin S, Mitrovic N, Rasheed U, et al. Bayesian support vector regression for traffic speed prediction with error bars. 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013); 2013 Oct 6-9; The Hague, Netherlands. USA: IEEE; 2013. p. 136-41.
[34] Chu W, Keerthi SS, Ong CJ. Bayesian support vector regression using a unified loss function. IEEE Trans Neural Network. 2004;15(1):29-44.
[35] Klein A, Falkner S, Bartels S, Hennig P, Hutter F. Fast bayesian optimization of machine learning hyperparameters on large datasets. arXiv:1605.07079. 2016:1-9.
[36] Snoek J, Larochelle H, Adams RP. Practical bayesian optimization of machine learning algorithms. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, editors. Proceeding Advances in neural information processing systems; 2012 Dec 3-6; Nevada, USA. USA: NIPS. p. 2951-9.
[37] Martinez-Cantin R. Bayesopt: a bayesian optimization library for nonlinear optimization, experimental design and bandits. J Mach Learn Res. 2014;15(1):3735-9.
[38] Gao JB, Gunn SR, Harris CJ, Brown M. A probabilistic framework for SVM regression and error bar estimation. Mach Learn. 2002;46(1-3):71-89.
[39] Law T, Shawe-Taylor J. Practical bayesian support vector regression for financial time series prediction and market condition change detection. Quant Finance. 2017;17(9):1-14.
[40] Jeenanunta C, Abeyrathna KD, Dilhani MS, Hnin S W, Phyo PP. Time series outlier detection for short-term electricity load demand forecasting. Int Sci J Eng Tech. 2018;2(1):37-50.
[41] Cao LJ, Tay FEH. Support vector machine with adaptive parameters in financial time series forecasting. IEEE Trans Neural Network. 2003;14(6):1506-18.