Load forecasting with support vector regression: influence of data normalization on grid search algorithm

Thanh Ngoc Tran[1], Binh Minh Lam[1], Anh Tuan Nguyen[1], Quang Binh Le[2]
[1]Faculty of Electrical Engineering Technology, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam
[2]Department of Engineering, Ho Chi Minh City Power Corporation, Ho Chi Minh City, Vietnam

Article Info

Article history:

Received Aug 3, 2021
Revised Jan 27, 2022
Accepted Feb 22, 2022

Keywords:

Data normalization
Grid Search
Load forecasting
Support vector regression

ABSTRACT

In recent years, support vector regression (SVR) models have been widely applied in short-term electricity load forecasting. A critical challenge when applying the SVR model is to determine the model for optimal hyperparameters, which can be solved using several optimization methods as the grid search algorithm. Another challenge that affects the response time and the precision of the SVR model is the normalization process of input data. In this paper, the grid search algorithm will be suggested based on data normalization methods including Z-score, min-max, max, decimal, sigmoidal, softmax, and then utilized to evaluate both the response time and precision. To verify the proposed methods, the actual electricity load demand data of two cities, including Queensland of Australia and Ho Chi Minh City of Vietnam, were utilized in this study.

Corresponding Author:

Tran Thanh Ngoc
Faculty of Electrical Engineering Technology, Industrial University of Ho Chi Minh City
12 Nguyen Van Bao, Ward 4, Go Vap District, Ho Chi Minh City, Vietnam
Email: tranthanhngoc@iuh.edu.vn

1. INTRODUCTION

Forecasting the electricity load plays a major role in an electricity system, which is composed of production planning, operational planning, and planning for future development plans [1]–[5]. There are a variety of solutions to predict the electricity load, including multiple regression, exponential smoothing, autoregressive integrated moving average (ARIMA), and artificial neural networks (ANNs), [6]–[12]. In the past decades, support vector regression (SVR) has emerged as a promising solution to electricity load forecasting [13]–[20]. Typically, the prediction precision of the SVR model relies on its hyperparameters, including ε (error tolerance), C (penalty parameter), Kernel functions, and their Kernel parameters. Therefore, it is crucial to find optimal values of the SVR hyperparameters. Several optimization methods such as grid search, random search, and genetic algorithm, have been studied for this challenge, of which the Grid Search algorithm is widely applied in many works [21]-[35].

Another factor that affects the running time and the precision of the SVR algorithm is the characteristics of the input data. Data normalization, therefore, was adopted in many studies for SVR models [13], [16], [36], [37]. However, data normalization was not of their consideration with the use of the grid search algorithm [28]–[35]. This might lead to the missing of the best results of hyperparameter tuning in the grid search algorithm because the data had not been normalized. Besides, the running time of the model might extensively increase without the use of data normalization.

Addressing these problems, this study suggests different data normalization techniques along with the grid search method for SVR hyperparameter tuning. At first, the input data are partitioned into two distinct sets, including training and testing datasets. The training dataset is then used for the training process, in which the grid search method is employed to obtain different sets of optimal hyperparameters corresponding to different data normalization methods. On the other hand, the prediction errors of the optimal SVR models are evaluated in the testing process with the use of the testing dataset. Finally, the electric load data of Ho Chi Minh City, Vietnam, and Queensland, Australia are used to verify the results.

This paper is structured: In sections 2 and 3, we present an introduction of the SVR model, SVR hyperparameters, the grid search algorithm, the mathematical models of data normalization techniques, and the SVR grid search algorithm based on data normalization. The experimental results and their evaluations are shown in section 4. Lastly, section 5 discusses the conclusions.

2. RESEARCH METHOD

2.1. SVR model

Considering a sample dataset as given: $\{x_i, y_i\}, \forall i = 1, \dots, N$, with N the length of the samples, $x_i \in R^n$ the input vector, and $y_i \in R$ the corresponding output vector. The crucial principle of SVR is the non-linear mapping of the input vector x into a feature space of higher dimensions by using a feature function $\varphi(): R^n \to R^h$. The SVR function, which defines the correlation between the input and the target, is acquired using (1) [13], [16]–[18]:

$$f(x) = \varpi^T \phi(x) + b \quad (1)$$

In (1), ω denotes the weight coefficient and b denotes the bias of the function. They can be determined by minimizing the regularized risk function R, given by (2):

$$R = \frac{1}{2}\|\varpi\|^2 + C\sum_{i=1}^{N} L_\varepsilon(y_i - f(x_i), x_i) \quad (2)$$

In (2), the first component, $\|w\|^2$, is known as the regulation term, the second component represents the empirical error between the actual and the predicted values, C is the penalty coefficient to regularize the relationship of these above-mentioned quantities, L_ε is the insensitive loss function that is defined by (3), and the error tolerance ε determines the constraints of $f(x)$ as presented by Figure 1.

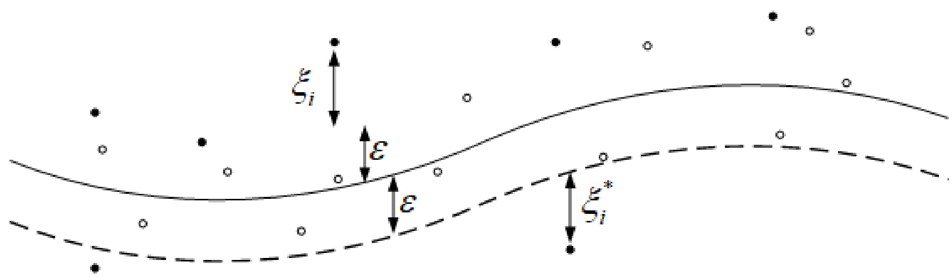


Figure 1. Illustration of ε, ξi, ξi* of the SVR model

$$L_\varepsilon(y_i - f(x_i), x) = \begin{cases} 0, & |y - f(x)| \le \varepsilon \\ |y - f(x)| - \varepsilon, & otherwise \end{cases} \quad (3)$$

The two slack variables ξ_i, ξ_i^* are introduced to indicate how much deviation the data points can be from the margin ε, so-called ε-tube. From Figure 1, ξ_i, ξ_i^* can be calculated as (4):

$$\begin{aligned} &|\text{y-f}(x)| - \varepsilon = \xi, \; over\ the\ tube \\ &|\text{y-f}(x)| - \varepsilon = \xi^*, \; under\ the\ tube \end{aligned} \quad (4)$$

By combining (4) with (3) and (2), the regularized risk function R can be re-written as (5) and follows the constraints (6):

$$R = \frac{1}{2}\|\varpi\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \quad (5)$$

$$y_i - (\varpi^T \phi(x) + b) \leq \varepsilon + \xi_i$$
$$(\varpi^T \phi(x) + b) - y_i \leq \varepsilon + \xi_i \quad (6)$$

The function $f(x)$ is determined by applying the Lagrange function as given in (7), where α_i, α_i^* represent the Lagrange multipliers and $K(x_i, x)$ denotes the Kernel function, that is defined as a dot (.) product of $\varphi(x_i)^T$and $\varphi(x)$:

$$f(x) = \varpi^T \phi(x) + b = \sum_{i=1}^{N} (\alpha_i^* - \alpha_i) K(x_i, x) + b \quad (7)$$

Some conventional Kernel functions widely used in SVR can be mathematically expressed in the formulas (8)-(10):

Linear: $K(x, y) = x^T y$ (8)

RBF: $K(x, y) = e^{-\gamma \|x-y\|^2}$ (9)

Sigmoid: $K(x, y) = tanh(\gamma\ x^T y + r)$ (10)

with x, y the inputs, r≥0 the intercept constant, and γ>0 the main parameter of the Kernel function.

2.2. SVR hyperparameters

A machine learning model can be composed of two different types of parameters. The first one consists of model parameters learned during the model training, and the second one of hyperparameters which can be randomly set before starting training instead. Based on the SVR model in section 2.1, the hyperparameters that control the model performance include the following parameters [21], [30]–[32]:

- The ε parameter indicating the constraints of $f(x)$;
- The C parameter implying the relationship between the regulation term and the empirical error;
- The Kernel function: linear, RBF, Sigmoid;
- The Kernel γ parameter.

Therefore, it is critical to find optimal values of these hyperparameters to enhance the prediction performance of the SVR model. Several optimization techniques can be applied for this purpose such as grid search, random search, and genetic algorithms. Of these methods, the grid search algorithm is selected in this study because of its simplicity and effectiveness.

2.3. Grid search method

The grid search is a searching process through a grid of subsets that were pre-specified by the combinations of different values of the hyperparameters. Optimal hyperparameters are those corresponding to the model that produces the smallest error [28]–[35]. Figure 2 shows an example of the grid search algorithm with two hyperparameters ε and γ. The ε hyperparameter is configured with three values {ε1, ε2, ε3}. Similarly, {γ1, γ2, γ3} are the configuration values of the γ hyperparameter. A combination of these two hyperparameters hence consists of 9 pairs. As a result, the grid search algorithm does the searching of the best model based on these 9 pairs.

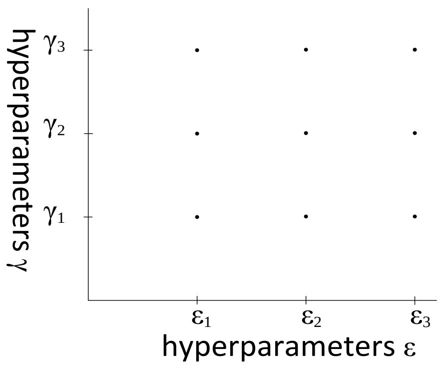


Figure 2. Illustration of grid search algorithm

In general, the performance of SVR models can be estimated using a variety of metrics or error measures that evaluate the error between the actual and predicted values. Some popular error measures, to name a few, include mean absolute error (MAE) and root mean square error (RMSE). Their formulas are shown in (11), given as [29], [35]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|^2}, MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (11)$$

2.4. Cross-validation procedure

Machine learning in general and SVR in particular can suffer from overfitting referring to a model that performs very well in the training process but poorly with new datasets. In this regard, one of the techniques called as k-fold cross-validation can be used to limit overfitting in the grid search algorithm [31]. This method allows a given dataset to be partitioned into k subsets (folds), of which (k-1) folds are used for training and the remaining fold is used for testing to validate the resulting model. As a result, the model is trained and tested in k times. The results of k times of cross-validation are then averaged to give an estimate of the model performance. Figure 3 shows an example of the k-fold cross-validation method with k=5. The given dataset was split into 5 folds. In the first cross-validation, the model took 4 folds from folds 2 to 5 for training, and the remaining fold (fold 1) was retained for validation. This process was then repeated from the second to the fifth cross-validations with the validation fold from fold 2 to fold 5. The averaged results of 5 times of cross validation helped improve the reliability of the model performance.

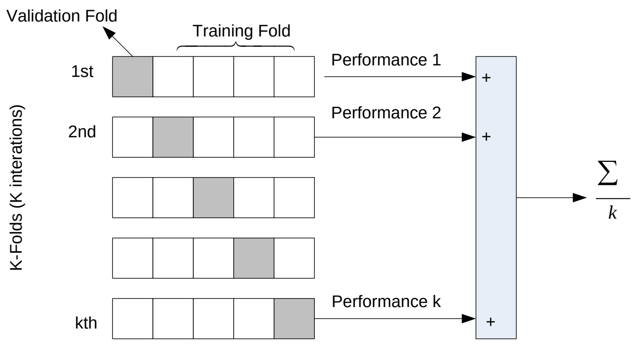


Figure 3. Cross validation procedure

2.5. Data normalization

Several studies have shown that the prediction performance of SVR is strongly affected by the size and the fluctuation of the input data. As a result, data normalization is required for both training and testing processes of the model. Different data normalization techniques have been investigated for SVR models in previous works [36], [37], of which zero-mean, min-max, max, decimal, sigmoid, and softmax are selected in this study. The mathematical equations of these techniques are shown in Table 1, where x_{mean}, x_{std}, x_{min}, and x_{max} are mean, standard deviation, min, and max values of x, respectively, and j is the smallest integer such that $\max(|x'| \leq) 1$.

Table 1. Equations of normalization techniques

Oder	Normalization	Equations
1	Zero-Mean	$x' = \frac{x - x_{mean}}{x_{std}}$
2	Min-Max	$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$
3	Max	$x' = \frac{x}{x_{max}}$
4	Decimal	$x' = \frac{x}{10^j}$
5	Sigmoid	$x' = \frac{1}{1 + e^{-a}}, \forall a = \frac{x - x_{min}}{x_{std}}$
6	Softmax	$x' = \frac{1 - e^{-a}}{1 + e^{-a}}, \forall a = \frac{x - x_{min}}{x_{std}}$

3. GRID SEARCH ALGORITHM BASED ON DATA NORMALIZATION FOR SVR MODEL

Based on section 2, the precision of the SVR model depends on its hyperparameters and the grid search algorithm that is combined with the cross-validation procedure provides an effective way to get these optimal hyperparameters. At the same time, data normalization also affects the response time and the performance of the SVR model. Therefore, the authors in this study suggest combining the grid search algorithm with different techniques of data normalization to evaluate the response time as well as the precision of the SVR model. The proposed method is shown in Figure 4. The algorithm was trained and tested following the steps:

- Step 1: The original sample data were processed to provide two pairs of input-target, named as (X_{train}, Y_{train}) and (X_{test}, Y_{test}) for training and testing datasets, respectively.
- Step 2: The training and testing datasets were normalized using each normalization technique as previously mentioned in Section 2.5.
- Step 3: The grid search method was applied to obtain the SVR optimal hyperparameters CFG_{opt}. Generally, CFG is a total combination of different sets of SVR hyperparameters. In particular, CFG={cfg_i}, i=1: N, with N the number of the combinations and cfg_i={ε_i, C_i, $kernel_i$, γ_i}.
- The cross-validation technique was also implemented in this step to enhance the performance of the grid search algorithm.
- Step 4: The SVR model with optimal hyperparameters was chosen to produce the predicted value $Y'_{predict}$.
- Step 5: $Y'_{predict}$ was the normalized value. Therefore, an inverse normalization process was required to obtain the original $Y_{predict}$.
- Step 6: Using (11), the prediction error of the SVR model was calculated based on the difference between $Y_{predict}$ and Y_{test}.

In the above procedure, step 3 was referred to as the training process, and steps 4, 5, and 6 implied the testing process. The whole process was applied with each data normalization technique that was mentioned in section 2.5. The corresponding results were recorded for evaluating the performance of these techniques.

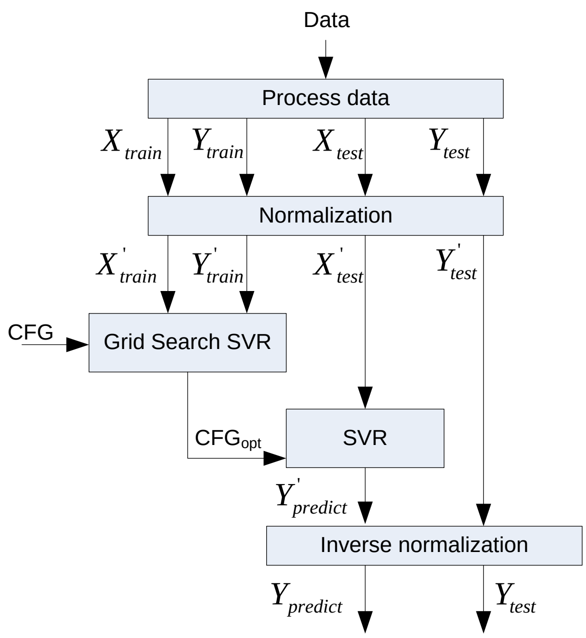


Figure 4. The SVR grid search algorithm based on data normalization

4. RESULTS AND DISCUSSION

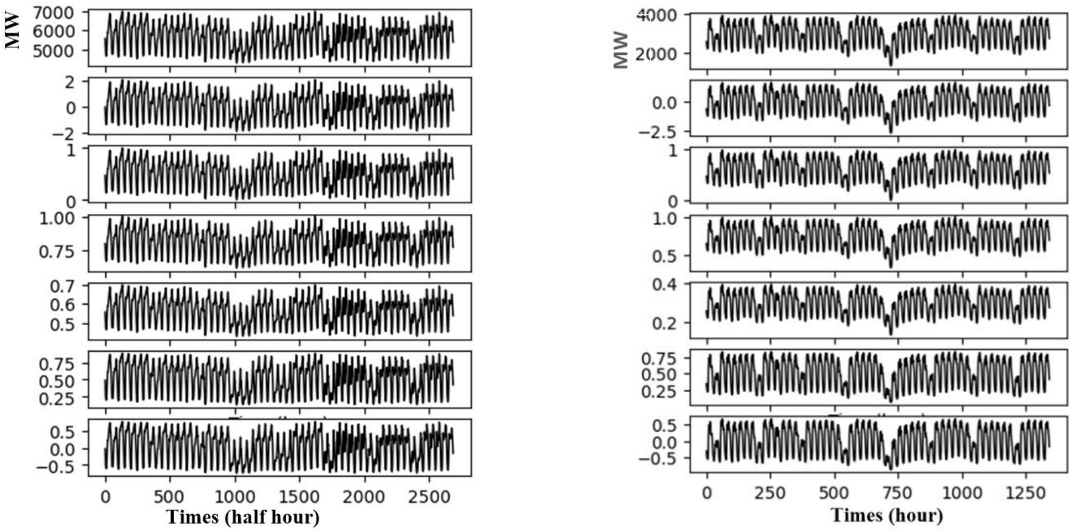
4.1. Data description

To verify the reliability of the suggested algorithm, half-hourly load demand data of Queensland (Australia) and hourly load demand data of Ho Chi Minh City (Vietnam) were used as inputs in the experiments. In addition, each dataset was divided into datasets #1 and #2 with different time lengths and different statistical characteristics, as shown in Table 2. These datasets #1 and #2 are independent, and the experiments will be performed on datasets #1 and #2, respectively.

Table 2. The characteristics of datasets #1 and #2

Description	Datasets #1				Datasets #2			
	Queensland		Hochiminh city		Queensland		Hochiminh city	
	X_{Train}	X_{Test}	X_{Train}	X_{Test}	X_{Train}	X_{Test}	X_{Train}	X_{Test}
Time	26/04/14-23/05/14	24/05/14-30/05/14	25/11/18-22/12/18	23/12/18-29/12/18	29/03/14-23/05/14	24/05/14-30/05/14	28/10/18-22/12/18	23/12/18-29/12/18
Size	(1344, 48)	(336, 48)	(672, 24)	(168, 24)	(2688, 48)	(336, 48)	(1344, 24)	(168, 24)
Min	4304.46	4404.48	1347.7	1873.9	4279.21	4404.48	1347.70	1873.90
Median	5532.44	5591.45	2917.94	2844.65	5589.60	5591.46	2952.69	2877.92
Max	6982.23	6824.76	3945.9	3695.2	6984.78	6824.76	3945.90	3760.3

Figure 5 shows the waveforms of Y_{train} for dataset #1 in Table 2 with different data normalization methods, including none (unnormalized data), zero-mean, min-max, max, decimal, sigmoid, and softmax, from top to bottom respectively. In particular, Figure 5(a) illustrates the waveforms corresponding to Queensland data, while Figure 5(b) shows those with respect to the data of Ho Chi Minh City. For dataset #1, the measurement error used was RMSE, along with k-fold=4. For dataset #2, the measurement error MAE with k-fold=2 was applied. Thus, there will be 02 data sets (#1 and #2) corresponding to 02 values of k-fold (4, 2) as well as 02 measurement error values (RMSE, MAE). This allows the data to be processed under different circumstances, thereby improving the reliability of the experiment results.



(a) (b)

Figure 5. Waveforms of Y_{train} of dataset #1 (a) Queensland and (b) Ho Chi Minh City

4.2. Hyperparameter tuning

As mentioned earlier, the SVR hyperparameters contain the tube size ε, the regularized constant C, the Kernel functions L, the Kernel functions parameters γ. Their tuning values are shown in Table 3 for both datasets #1 and #2. Combining all these values gives 176 cases corresponding to 176 possible SVR models.

Table 3. Tuning hyerparameter values

Items	Values
Tube size ε	1e-4, 1e-3, 1e-2, 1e-1
Regularized constant C	0.1, 1, 10, 100
Kernel functions K	linear, RBF, sigmoid
Kernel function parameter γ	1e-4, 1e-3, 1e-2, 1e-1, 1
Number of combination CFG	176

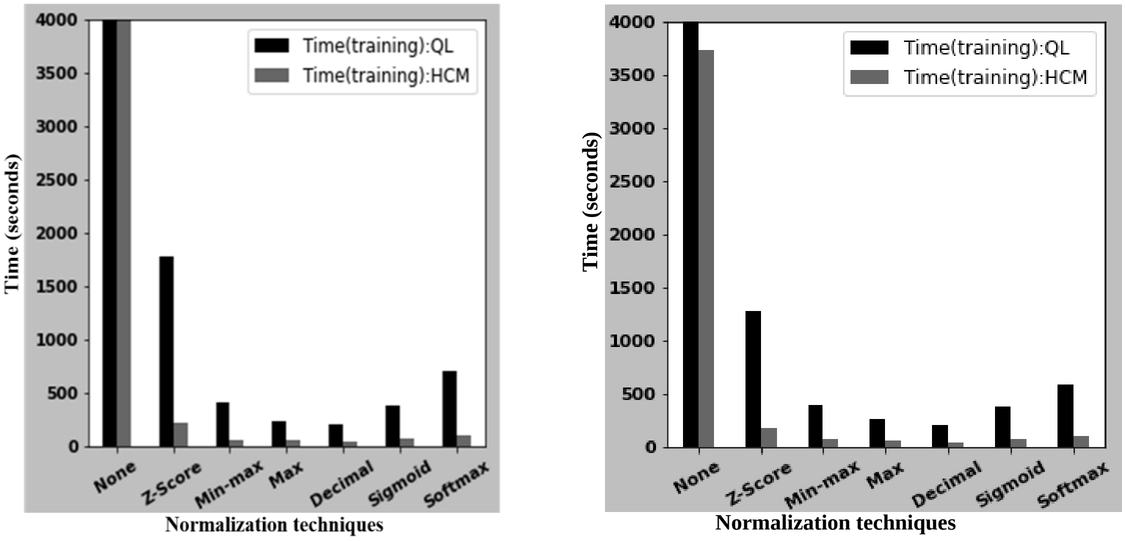
4.3. Experimental results

Table 4 shows the running time in seconds of the training process for Queensland (QL) and Ho Chi Minh City (HCM) corresponding to datasets #1 and #2. The running time of the training process that is shown in Table 4 and illustrated in Figure 6. Particularly, Figure 6(a) corresponds to dataset #1, and Figure 6(b) corresponds to dataset #2.

Table 5 introduces the optimal SVR hyperparameters that were obtained by the grid search algorithm during the training process for QL and HCM in cases of datasets #1 and #2, respectively. It should be noted that the optimal Kernel function in most cases of normalization techniques for all datasets #1 and datasets #2 was ‘rbf’. Only in the case of none (unnormalized) data, it was ‘linear’.

Table 4. The running time (seconds) of the training process

Normalization	Dataset #1		Dataset #2	
	QL	HCM	QL	HCM
None	25,958	6,129	15,230	3,725
Z-Score	1,771	214	1,272	164
Min-max	405	58	396	68
Max	232	56	253	52
Decimal	201	32	202	33
Sigmoid	376	67	370	63
Softmax	705	102	582	90



(a) (b)

Figure 6. Running time of the training process (a) dataset #1 and (b) dataset #2

Table 5. Optimal hyperparmeters of the training process

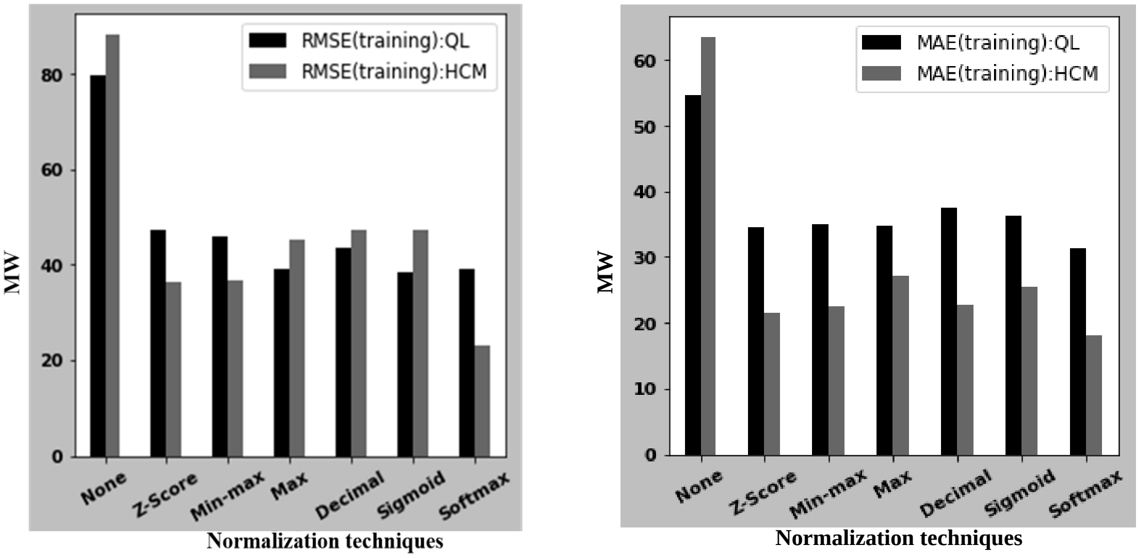
Normalization	Dataset #1						Dataset #2						Kernel function
	ε		C		γ		ε		γ		C		
	QL	HCM	QL	HCM	QL	HCM	QL	HCM	QL	HCM	QL	HCM	
None	1e-3	1e-2	1e-1	1e-1			1e-4	1e-4	1e-1	1e-1			linear
Z-Score	1e-4	1e-2	1e2	1e2	1e-3	1e-2	1e-4	1e-4	1e2	1e2	1e-3	1e-2	rbf
Min-max	1e-4	1e-2	1e2	1e2	1e-2	1e-1	1e-2	1e-4	1e2	1e2	1e-2	1e-1	rbf
Max	1e-3	1e-2	1e2	1e2	1e-1	1e-1	1e-3	1e-4	1e1	1e2	1e-1	1e-1	rbf
Decimal	1e-3	1e-3	1e2	1e1	1e-1	1	1e-4	1e-3	1e1	1e2	1e-1	1	rbf
Sigmoid	1e-2	1e-2	1e1	1e2	1e-1	1e-1	1e-4	1e-3	1e2	1e2	1e-2	1e-1	rbf
Softmax	1e-2	1e-2	1e1	1e2	1e-1	1e-1	1e-4	1e-3	1e2	1e2	1e-2	1e-1	rbf

Table 6 shows the error measures with respect to the optimal hyperparameters that were determined from the training process. These error measures are plotted in Figure 7. Specifically, Figure 7(a) shows the RMSE of the model in case of dataset #1, while Figure 7(b) shows the MAE in case of dataset #2.

The testing performance of the optimal SVR models that were obtained from the training process is introduced in Table 7. Moreover, Figure 8 shows the testing performance of the optimal SVR models error measures regarding Table 7. Figure 8(a) shows the RMSE between the testing and predicated values, while Figure 8(b) shows the MAE between them.

Table 6. The error measures of the training process

Normalization	RMSE (MW), dataset #1		MAE (MW), dataset #2	
	QL	HCM	QL	HCM
None	79.88	88.42	54.66	63.61
Z-Score	47.19	36.44	34.42	21.52
Min-max	46.06	36.79	34.95	22.42
Max	39.23	45.43	34.73	27.12
Decimal	43.72	47.33	37.49	22.73
Sigmoid	38.35	47.19	36.21	25.53
Softmax	39.24	23.00	31.30	18.04



(a) (b)

Figure 7. The training performance with respect to (a) dataset #1 and (b) dataset #2

Table 7. The error measures of the testing process

Normalization	RMSE (MW), dataset #1		MAE (MW), dataset #2	
	QL	HCM	QL	HCM
None	76.51	81.75	53.43	60.32
Z-Score	44.93	55.21	35.68	35.31
Min-max	44.53	54.69	34.95	34.70
Max	40.29	57.19	35.82	36.20
Decimal	42.82	59.94	37.38	33.17
Sigmoid	41.52	60.46	37.39	36.04
Softmax	41.14	53.51	32.86	32.87

4.4. Evaluation and discussion

The first metric to be evaluated is the running time of the grid search algorithm corresponding to different normalization techniques used for SVR. Table 4 and Figure 6 showed that applying data normalization techniques significantly reduced the running time of the program. Interestingly, executing the popular Z-Score technique took a longer duration than other methods, while performing the Max and Decimal methods seemly produced the shortest duration.

Analyzing Table 5, it was clearly observed that different data normalization techniques presented different optimal values of SVR hyperparameters, which had been achieved using the grid search algorithm in the training process. Besides, these optimal hyperparameters were different with datasets #1 and #2. Moreover, it was shown that the 'rbf' function was the optimal Kernel function for all methods of data normalization, except for the none case where the model used 'linear' Kernel function with unnormalized data. Tables 6, 7 and Figures 7, 8 clearly showed that the training and testing errors of the models using data normalization were much smaller than those of the model using unnormalized data (none case). At the same time, the data normalization demonstrated an obvious influence on the grid search algorithm. Specifically, each data normalization has different precision scores in the training and testing processes.

It is worth to note that the softmax produced the least errors in most cases for training and testing processes. Indeed, let us analyze Table 6 and Figure 7 for the training process in case #1, Ho Chi Minh City data. For the softmax normalization, the value of RMSE is 23 MW. This value is much smaller than that of other normalization types, especially, in comparison with common ones of Z-scores (36.44 MW) and min-max (36.79 MW). Similar results were also received for Queensland data. These results clearly indicated that selecting a suitable data normalization as softmax normalization in this study can give better precision score in the training and testing processes of the grid search algorithm for SVR model.

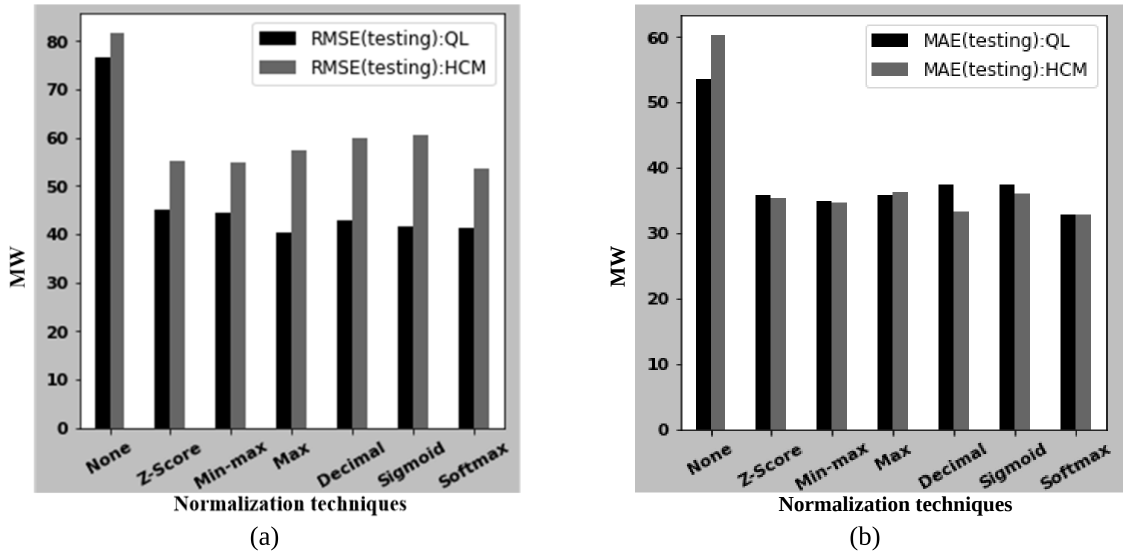


Figure 8. The testing performance (a) dataset #1 and (b) dataset #2

5. CONCLUSION

The study has successfully utilized an effective approach to analyze the effects of a variety of data normalization techniques on the grid search algorithm for determining SVR optimal hyperparameters in the case of electricity load forecasting. The running time and the error measures (RSME and MAE) were evaluated during training and testing processes. Both the daily electric loads of Queensland, Australia, and Ho Chi Minh City, Vietnam, were used to verify the suggested model. The total dataset was split into two subsets of training and testing datasets to enhance the reliability of the study. The results showed that using data normalization helped greatly reduce the running time and obtain much smaller errors in terms of MAE and RMSE. The results also indicated that conventional data normalization techniques such as Z-Score and Min-Max did not guarantee the shortest running time and the smallest errors. This conclusion demonstrates the feasibility of applying different data normalization methods with the Grid Search algorithm problem in SVR models.

ACKNOWLEDGEMENTS

This work is supported by Industrial University of Ho Chi Minh City under Grant No. 70/HĐ-ĐHCN.

REFERENCES

[1] M. Q. Raza and A. Khosravi, “A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings,” *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 1352–1372, Oct. 2015, doi: 10.1016/j.rser.2015.04.065.

[2] G. Mohi Ud Din, A. U. Mauthe, and A. K. Marnerides, “Appliance-level short-term load forecasting using deep neural networks,” in *2018 International Conference on Computing, Networking and Communications (ICNC)*, Mar. 2018, pp. 53–57, doi: 10.1109/ICCNC.2018.8390366.

[3] J. Walther, D. Spanier, N. Panten, and E. Abele, “Very short-term load forecasting on factory level-A machine learning approach,” *Procedia CIRP*, vol. 80, pp. 705–710, 2019, doi: 10.1016/j.procir.2019.01.060.

[4] A. A. Aydarous, M. A. Elshahed, and M. A. Moustafa Hassan, “Short-term load forecasting approach based on different input methods of one variable: conceptual and validation study,” in *2018 Twentieth International Middle East Power Systems Conference (MEPCON)*, Dec. 2018, pp. 179–184, doi: 10.1109/MEPCON.2018.8635115.

[5] K. M. Powell, A. Sriprasad, W. J. Cole, and T. F. Edgar, "Heating, cooling, and electrical load forecasting for a large-scale district energy system," *Energy*, vol. 74, pp. 877–885, Sep. 2014, doi: 10.1016/j.energy.2014.07.064.
[6] Y. Hong, Y. Zhou, Q. Li, W. Xu, and X. Zheng, "A deep learning method for short-term residential load forecasting in smart grid," *IEEE Access*, vol. 8, pp. 55785–55797, 2020, doi: 10.1109/ACCESS.2020.2981817.
[7] Y. Liang, D. Niu, and W.-C. Hong, "Short term load forecasting based on feature extraction and improved general regression neural network model," *Energy*, vol. 166, pp. 653–663, Jan. 2019, doi: 10.1016/j.energy.2018.10.119.
[8] A. K. Singh, Ibraheem, S. Khatoon, M. Muazzam, and D. K. Chaturvedi, "Load forecasting techniques and methodologies: A review," in *2012 2nd International Conference on Power, Control and Embedded Systems*, Dec. 2012, pp. 1–10, doi: 10.1109/ICPCES.2012.6508132.
[9] E. E. Roussineau, P. Otto, and P. Gratzfeld, "Short-term forecasting in electric power systems using artificial neural networks," in *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, Oct. 2018, pp. 1–6, doi: 10.1109/ISGTEurope.2018.8571819.
[10] A. Kapoor and A. Sharma, "A comparison of short-term load forecasting techniques," in *2018 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*, May 2018, pp. 1189–1194, doi: 10.1109/ISGT-Asia.2018.8467788.
[11] T. Panapongpakorn and D. Banjerdpongchai, "Short-term load forecast for energy management systems using time series analysis and neural network method with average true range," in *2019 First International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP)*, Jan. 2019, pp. 86–89, doi: 10.1109/ICA-SYMP.2019.8646068.
[12] R. Sethi and J. Kleissl, "Comparison of short-term load forecasting techniques," in *2020 IEEE Conference on Technologies for Sustainability (SusTech)*, Apr. 2020, pp. 1–6, doi: 10.1109/SusTech47890.2020.9150490.
[13] L. Tian and A. Noore, "A novel approach for short-term load forecasting using support vector machines," *International Journal of Neural Systems*, vol. 14, no. 05, pp. 329–335, Oct. 2004, doi: 10.1142/S0129065704002078.
[14] W.-C. Hong, "Electric load forecasting by support vector model," *Applied Mathematical Modelling*, vol. 33, no. 5, pp. 2444–2454, May 2009, doi: 10.1016/j.apm.2008.07.010.
[15] S. De Cosmis, R. De Leone, E. Kropat, S. Meyer-Nieberg, and S. Pickl, "Electric load forecasting using support vector machines for robust regression," in *Proceedings of the Emerging M&S Applications in Industry and Academia/Modeling and Humanities Symposium*, 2013, pp. 1–8.
[16] B. E. Turkay and D. Demren, "Electrical load forecasting using support vector machines: A case study," *International Review of Electrical Engineering (IREE)*, vol. 6, no. 5, pp. I-49-I–53, 2011.
[17] X. Li, D. Gong, L. Li, and C. Sun, "Next day load forecasting using SVM," in *Advances in Neural Networks-ISNN 2005*, Springer Berlin Heidelberg, 2005, pp. 634–639.
[18] T. Thanh Ngoc, L. Van Dai, and C. Minh Thuyen, "Support vector regression based on grid search method of hyperparameters for load forecasting," *Acta Polytechnica Hungarica*, vol. 18, no. 2, pp. 143–158, 2021, doi: 10.12700/APH.18.2.2021.2.8.
[19] M. Matijaš, M. Vukićević, and S. Krajcar, "Supplier short term load forecasting using support vector regression and exogenous input," *Journal of Electrical Engineering*, vol. 62, no. 5, Jan. 2011, doi: 10.2478/v10187-011-0044-9.
[20] S. Qiang and Y. Pu, "Short-term power load forecasting based on support vector machine and particle swarm optimization," *Journal of Algorithms and Computational Technology*, vol. 13, Jan. 2019, doi: 10.1177/1748301818797061.
[21] N. Sano, K. Higashinaka, and T. Suzuki, "Efficient parameter selection for support vector regression using orthogonal array," in *2014 IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC)*, Oct. 2014, pp. 2256–2261, doi: 10.1109/SMC.2014.6974261.
[22] J.-Y. Hsia and C.-J. Lin, "Parameter selection for linear support vector regression," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 12, pp. 5639–5644, Dec. 2020, doi: 10.1109/TNNLS.2020.2967637.
[23] F. Jiang and W. Wu, "Hybrid genetic algorithm and support vector regression performance in CNY exchange rate prediction," *Proceedings of the 2016 International Conference on Engineering Science and Management*, 2016, doi: 10.2991/esm-16.2016.32.
[24] M. Jiang, S. Jiang, L. Zhu, Y. Wang, W. Huang, and H. Zhang, "Study on parameter optimization for support vector regression in solving the inverse ECG problem," *Computational and Mathematical Methods in Medicine*, vol. 2013, pp. 1–9, 2013, doi: 10.1155/2013/158056.
[25] S. Lessmann, R. Stahlbock, and S. F. Crone, "Optimizing hyperparameters of support vector machines by genetic algorithms," *Proceedings of the 2005 International Conference on Artificial Intelligence, ICAI 2005*, 2005.
[26] S. K. Lahiri and K. C. Ghanta, "The support vector regression with the parameter tuning assisted by a differential evolution technique: study of the critical velocity of a slurry flow in a pipeline," *Chemical Industry and Chemical Engineering Quarterly*, vol. 14, no. 3, pp. 191–203, 2008, doi: 10.2298/CICEQ0803191L.
[27] H. Wang and D. Xu, "Parameter selection method for support vector regression based on adaptive fusion of the mixed Kernel function," *Journal of Control Science and Engineering*, vol. 2017, pp. 1–12, 2017, doi: 10.1155/2017/3614790.
[28] X. Ma, Y. Zhang, and Y. Wang, "Performance evaluation of Kernel functions based on grid search for support vector regression," in *2015 IEEE 7th International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)*, Jul. 2015, pp. 283–288, doi: 10.1109/ICCIS.2015.7274635.
[29] H. Zhang, L. Chen, Y. Qu, G. Zhao, and Z. Guo, "Support vector regression based on grid-search method for short-term wind power forecasting," *Journal of Applied Mathematics*, vol. 2014, pp. 1–11, 2014, doi: 10.1155/2014/835791.
[30] M. Yang, W. Li, H. Zhang, and H. Wang, "Parameters optimization improvement of SVM on load forecasting," in *2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Aug. 2016, pp. 257–260, doi: 10.1109/IHMSC.2016.269.
[31] S. Lin, S. Zhang, J. Qiao, H. Liu, and G. Yu, "A parameter choosing method of SVR for time series prediction," in *2008 The 9th International Conference for Young Computer Scientists*, Nov. 2008, pp. 130–135, doi: 10.1109/ICYCS.2008.393.
[32] Qiujun Huang, Jingli Mao, and Yong Liu, "An improved grid search algorithm of SVR parameters optimization," in *2012 IEEE 14th International Conference on Communication Technology*, Nov. 2012, pp. 1022–1026, doi: 10.1109/ICCT.2012.6511415.
[33] A. Zakrani, A. Najm, and A. Marzak, "Support vector regression based on grid-search method for agile software effort prediction," in *2018 IEEE 5th International Congress on Information Science and Technology (CiSt)*, Oct. 2018, pp. 1–6, doi: 10.1109/CIST.2018.8596370.
[34] S. F. Crone, J. Guajardo, and R. Weber, "The impact of preprocessing on support vector regression and neural networks in time series prediction," *Proceedings of the 2006 International Conference on Data Mining, DMIN 2006*, 2006.
[35] M. Zahid *et al.*, "Electricity price and load forecasting using enhanced convolutional neural network and enhanced support vector regression in smart grids," *Electronics*, vol. 8, no. 2, Jan. 2019, doi: 10.3390/electronics8020122.
[36] W. li and Z. Liu, "A method of SVM with normalization in intrusion detection," *Procedia Environmental Sciences*, vol. 11, pp. 256–262, 2011, doi: 10.1016/j.proenv.2011.12.040.
[37] S. Ali and K. A. Smith-Miles, "Improved support vector machine generalization using normalized input space," in *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2006, pp. 362–371.

BIOGRAPHIES OF AUTHORS

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

Minh Binh Lam received the B.E. degree in Electrical - Electronics Engineering from BachKhoa University, Vietnam, in 2002, and the M.S. degree in Telecommunication and Information Systems from University of Essex, U.K, in 2004. She is currently a lecturer of the Faculty of Electrical Engineering Technology of the Industrial University of Ho Chi Minh City, Vietnam. She is also pursuing a Ph.D. degree at Department of Electronic Engineering, Pukyong National University, Busan, Korea. Currently she is a Lecturer with the Faculty of Electrical Engineering Technology, Industry University of Ho Chi Minh City, Vietnam. Her current research interests include high-frequency radio frequency identification, far field communication, and low-power sensor networking. She can be contacted at email: lambinhminh@iuh.edu.vn.

Nguyen Anh Tuan received the B.Sc degrees from the Industrial University of Ho Chi Minh City, Vietnam in 2012 and M.Sc degrees from the HCMC University Of Technology and Education, Vietnam in 2015, respectively. Currently, he is a Lecturer with the Faculty of Electrical Engineering Technology, Industry University of Ho Chi Minh City, Vietnam. His main research interests include HVDC transmission systems, active filters, power quality and load forecasting, Scada, PLC, Automatic system. He can be contacted at email: nguyenanhtuan@iuh.edu.vn.

Le Quang Binh achieved the Bachelor and Master in Power System at the Ho Chi Minh University of Technology, Vietnam in 1997 and 1999, respectively. Currently, he is a Director of Engineering Department, Ho Chi Minh City Power Corporation, Viet Nam. His main research interests include Power system protection, high-voltage transmission, grid automation, SCADA and digital substation. He can be contacted at email: binhlq@hcmpc.com.vn.