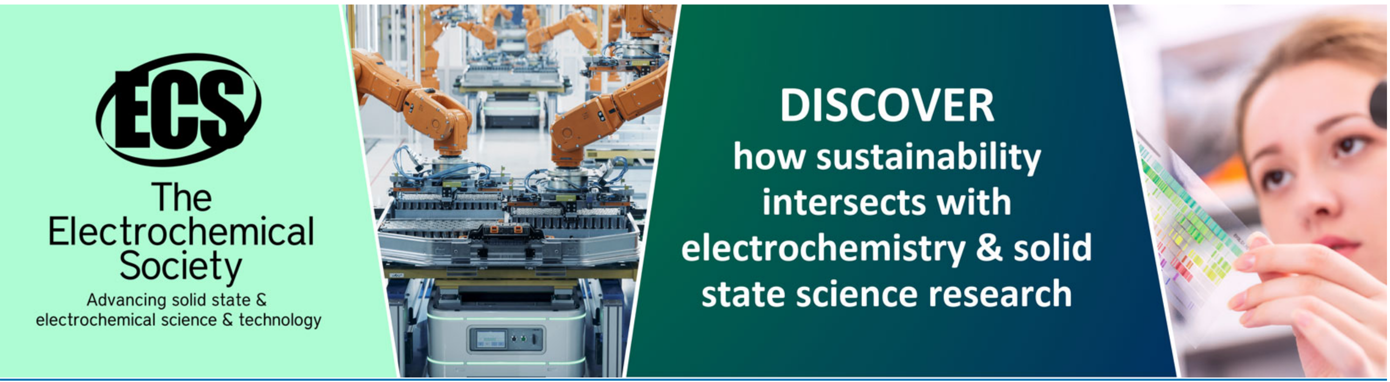
IOP Conference Series: Earth and Environmental Science

PAPER • **OPEN ACCESS**

SAGA-FCM-LSSVM model-based short-term power forecasting of photovoltaic power plants

To cite this article: Z N Peng *et al* 2018 *IOP Conf. Ser.: Earth Environ. Sci.* **188** 012079



IOP Conf. Series: Earth and Environmental Science **188** (2018) 012079 doi:10.1088/1755-1315/188/1/012079

SAGA-FCM-LSSVM model-based short-term power forecasting of photovoltaic power plants

Z N Peng[1,2], P J Lin[1,2,3], Y F Lai[1,2], Z C Chen[1,2], L J Wu[1,2] and S Y Cheng[1,2]

[1] College of Physics and Information Engineering, and Institute of Micro-Nano Devices and Solar Cells, Fuzhou University, 350116 Fuzhou, China
[2] Jiangsu Collaborative Innovation Center of Photovoltaic Science and Engineering, 213164 Changzhou, China

E-mail: linpeijie@fzu.edu.cn

Abstract. With the rapid development of social economy, vigorously developing solar energy has become a powerful means to solve energy and environmental problems. However, the instability of weather condition makes the output of PV power have strong randomness, fluctuations and intermittence. Thus accurate photovoltaic (PV) power forecast eliminates the negative impacts of the grid connection of PV power generation systems, which are very meaningful for effectively integrating the PV power systems into the grid. The paper presents a Simulate Anneal and Genetic Algorithm (SAGA), fuzzy c-means clustering (FCM) and least square support vector machine (LSSVM) (SAGA-FCM-LSSVM) model-based power short-term forecasting of PV power plants approach. The experimental effect of the proposed prediction method is verified by employing large datasets from the Desert Knowledge Australia Solar Center (DKASC) website. In this work, the FCM clustering algorithm is adopted to cluster the historical power datasets, and the LSSVM technique maps the multivariate meteorological factors and power data nonlinear relationship. The SAGA method is applied to improve the initial clustering centers of the FCM clustering algorithm to obtain a higher prediction performance. The prediction result of the method in this paper is contrasted with back propagation neural network (BPNN) and LSSVM models, and reveals excellent effect in improving the accuracy of prediction.

1. Introduction

In recent years, the photovoltaic (PV) market has witnessed a tremendous growth due to the significant cost reduction of the PV modules on the market and the proposed supporting policies [1]. However, the integration of solar energy into an electrical network hampers the grid management due to the randomness, fluctuations and intermittence of weather condition [2]. Therefore, accurately forecast the PV power generation in advance is of great significance to the efficient utilization of PV energy and the grid management [3].

Up to now, a variety of PV power generation prediction methods have been applied, which are categorized as physical methods and statistical methods according to the forecasting models [4]. The physical methods are based on solar radiation, air temperature, cloud amount and other weather factors to predict PV generation [5]. However, the uncertainty of meteorological factors challenges to the accurate prediction of PV power [6]. Considering the changeable weather effectively based on historical power datasets, the statistical methods are widely accepted [7]. Khan *et al* develops a back propagation neural network (BPNN) model based on temperature, wind speed, humidity, solar radiation and air quality index to predict PV power production in haze weather [8]. Aziz et al. proposes a cuckoo search (CS)-least square support vector machine (LSSVM)-based approach for short-term power prediction with solar irradiation and ambient temperature as input attributes [9]. The method based on BPNN has a good approximation ability, but its convergence is slow and is likely to get into a local minimum extremely [10]. Furthermore, the prediction accuracy of LS-SVM approach is not high enough.

Owing to the shortcomings of BPNN and LS-SVM methods, the paper employs a SAGA-FCM-LSSVM model-based approach, which is valuable for the PV generation prediction with fast and exact forecasting performance. In particular, the proposed forecasting model can also be combined with the reconfiguration algorithm to determine the optimal array configuration under arbitrary operating conditions throughout the lifetime of the PV array, thus reducing the power derating caused by thermal related aging effects, which can optimize the life performances of the PV components [11,12].

2. Proposed architecture for the proposed PV power forecasting model

2.1. Proposed SAGA-FCM-LSSVM algorithm

2.1.1. SAGA-FCM algorithm. The FCM algorithm is a popular clustering method, which adopts the concept of geometric approximation. Nevertheless, FCM method is essentially a kind of improved local search approach.

In this paper, GA and SA algorithms are combined to cluster analysis. It overcomes the premature phenomenon of traditional GA because SA and GA can complement each other. Meanwhile, the genetic coding method and fitness function are designed according to the specific situation of the clustering problem, which makes the algorithm more effective and converges to the global optimal solution more quickly.

And the four statistical indicators of the PV output power which are *standard deviation σ*, *coefficient of variation cv*, *skewness Sk* and *kurtosis kur* after normalization are used as the eigenvalues of the cluster.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_i - \overline{P}\right)^2} \tag{1}$$

$$cv = \frac{\sigma}{\overline{P}} \tag{2}$$

$$Sk = \frac{N \cdot \sum_{i=1}^{N}\left(P_i - \overline{P}\right)^3}{\left(N-1\right)\left(N-2\right)\cdot\sigma} \tag{3}$$

$$kur = \frac{\sum_{i=1}^{N}\left(P_i - \overline{P}\right)^4}{\left(N-1\right)\cdot\sigma} \tag{4}$$

where P_i is the output power of a PV station at the time point. $\overline{P}$ is the output power average. N represents the number of sample points a day.

2.1.2. LS-SVM algorithm. LS-SVM is an improved algorithm for SVM, which is an effective approach for solving nonlinear issue, for instance, classification and function estimation. The principle of LSSVM is summarized as follows:

Let an n-dimensional vector sample be (x_1,y_1), (x_2,y_2), …, (x_n,y_n), the sample is mapped from space R^n to feature space $\varphi(x_i)$ by a nonlinear mapping $\varphi(\cdot)$, and the optimal decision function is constructed in the space:

$$f(x) = w^T \varphi(x) + b \tag{5}$$

where w^T is the weight coefficient vector of feature space, b is the bias.

The optimization issue of LS-SVM can be transformed into dual space as follows due to the principle of structural risk minimization:

$$L(w,b,e,\alpha) = \frac{1}{2} w^T w + c\sum_{i=1}^{l} e_i^2 - \sum_{i=1}^{l} \alpha_i \left[w^T \varphi(x_i) + b + e_i - y_i \right] \tag{6}$$

where α_i is lagrange multiplier, c is regularization parameter.

The equation is simplified by using the optimization condition. According to Mercer, the inner product operation of $\varphi(x_i)^T\varphi(x_i)$ can be replaced by the kernel function $k(x_i,x_j)$. Finally, the regression function of LS-SVM is obtained by solving it.

Table 1 indicates the input parameters for LS-SVM model. There are 8 parameters related to the actual multivariate meteorological factors and PV power datasets.

Table 1. Input variables of LS-SVM model.

Input	Variables
x_1	global horizontal radiation of 1 h before
x_2	air temperature of 1 h before
x_3	relative humidity of 1 h before
x_4	PV power of 1 h before
x_5	global horizontal radiation of 2 h before
x_6	air temperature of 2 h before
x_7	relative humidity of 2 h before
x_8	PV power of 2 h before

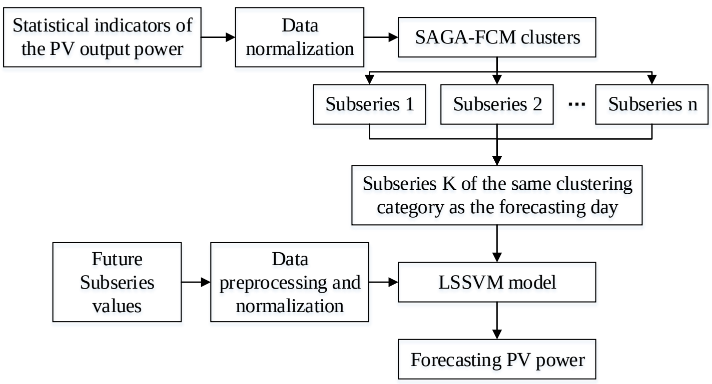


Figure 1. Process of the proposed PV power forecasting.

2.2. *Proposed SAGA-FCM-LSSVM process*

Figure 1 displays the PV generation forecasting process based upon the proposed SAGA-FCM-LSSVM model. First, the normalized statistical indicators of the PV output power are chosen as the inputs for the FCM clustering algorithm to categorize the historical power datasets into different subseries. The decomposed original datasets of these subsets are applied to train LS-SVM model. The future multivariate meteorological factors and power values are adopted to predict the future PV power employing LS-SVM. The SAGA method is applied to improve the initial clustering centers of the FCM clustering algorithm to improve performance. The proposed approach in the paper is evaluated by adopting evaluation metrics and by contrasting with other prediction methods.

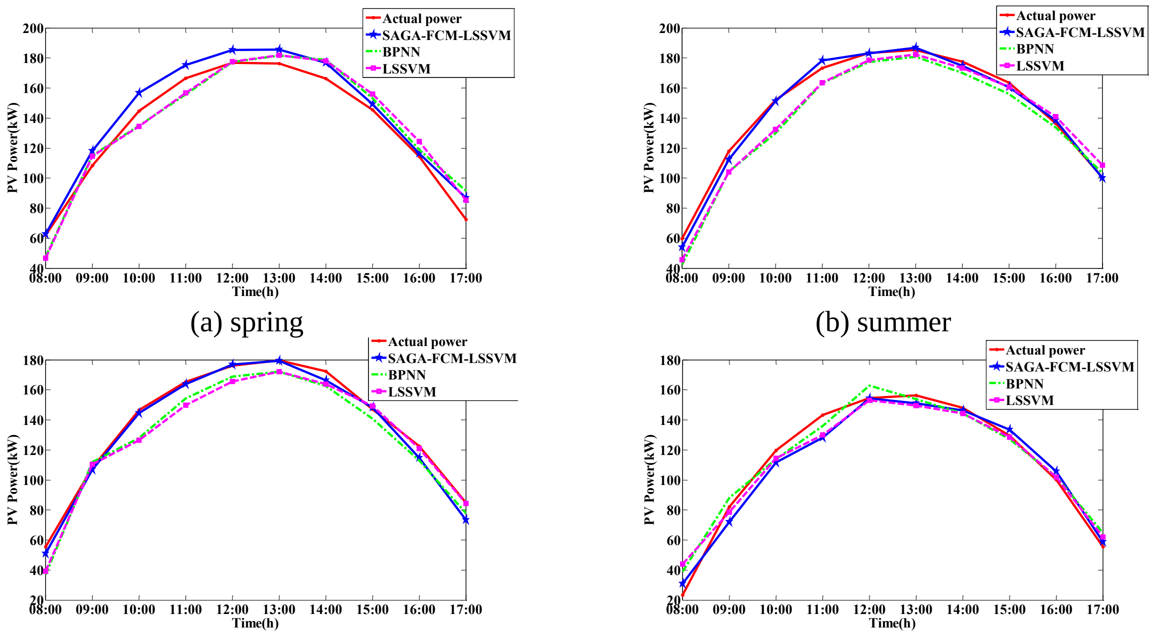
3. Experiments and result analysis

To verify the effect of the proposed prediction method based on SAGA-FCM-LSSVM model, BPNN method and LS-SVM method are applied in this paper. The multivariate meteorological factors (global horizontal radiation, air temperature and relative humidity) and historical power datasets from April 1, 2016 to March 30, 2018 are applied for the forecasting. The selected four forecasting days are September 14, 2017 (spring in Australia), February 26, 2018 (summer in Australia), March 30, 2018 (autumn in Australia) and July 29, 2017 (winter in Australia), which represent four seasons respectively.

The number of clustering categories and the Silhouette of the forecasting day are displayed in table 2. As can be obtained from the table 2, the Silhouette values of the selected four forecasting days are 0.4517, 0.4002, 0.3962 and 0.4428 when the number of clustering categories is 3.

Table 2. The number of clustering categories and the Silhouette of the forecasting day.

	Spring	Summer	Autumn	Winter
Categories	3	3	3	3
Silhouette	0.4517	0.4002	0.3962	0.4428

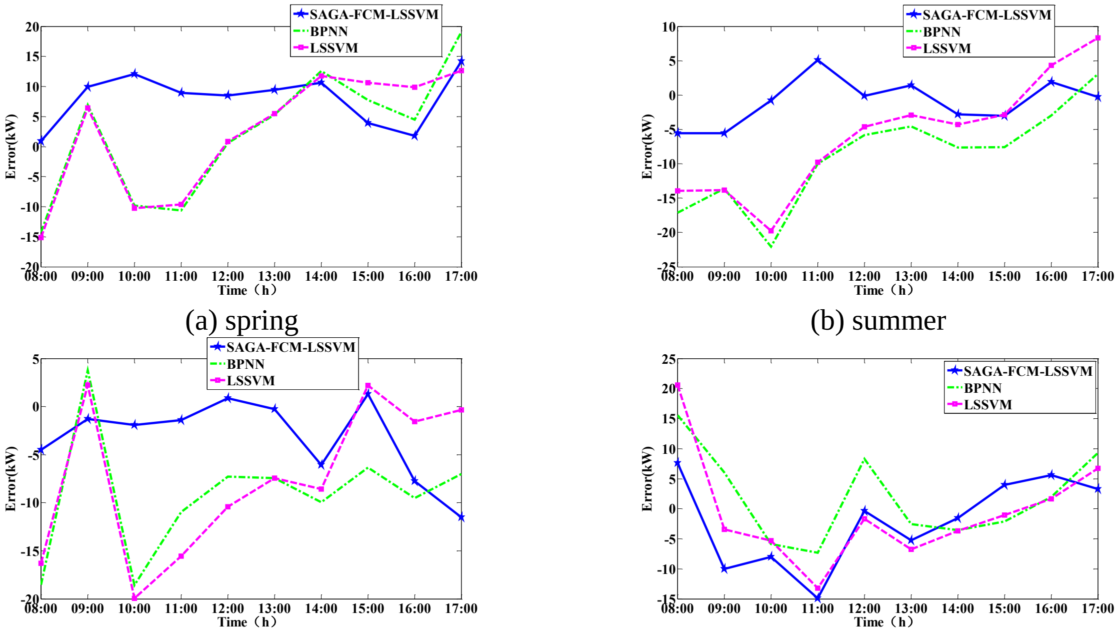


(a) spring (b) summer

(c) fall (d) winter

Figure 2. Prediction curves in different seasons.

The forecasting result cures by the SAGA-FCM-LSSVM forecasting model and the other forecasting models (BPNN and LS-SVM) are shown in figures 2(a)-2(d), respectively for spring, summer, fall and winter. From the results, it is clear that the SAGA-FCM-LSSVM model-based prediction method is more exact and reliable. Furthermore, figures 3(a)-3(d) show the corresponding absolute errors curves on the selected four days. It can be obtained that the error of the SAGA-FCM-LSSVM model in each season is little and the prediction error distribution is limited in [−15,15] kW.



(a) spring (b) summer

(c) fall (d) winter

Figure 3. Absolute error curves in different seasons.

The root mean square error (RMSE) [13], the mean absolute percentage error (MAPE) and the coefficient of determination (R^2) metrics, shown in equations (7)–(10), are chose to evaluate the prediction performance.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_{f,i} - P_{m,i}\right)^2} \tag{7}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left(\left|\frac{P_{f,i} - P_{m,i}}{\overline{P_{m,i}}}\right|\right)\times 100 \tag{8}$$

$$\overline{P_{m,i}} = \frac{1}{N}\sum_{i=1}^{N} P_{m,i} \tag{9}$$

$$R^2 = \frac{(N\sum_{i=1}^{N} P_{f,i}P_{m,i} - \sum_{i=1}^{N} P_{f,i}\sum_{i=1}^{N} P_{m,i})^2}{(N\sum_{i=1}^{N} P_{f,i}{}^2 - (\sum_{i=1}^{N} P_{f,i})^2)(N\sum_{i=1}^{N} P_{m,i}{}^2 - (\sum_{i=1}^{N} P_{m,i})^2)} \tag{10}$$

where $P_{f,i}$ is the ith prediction value of the power, $P_{m,i}$ is the ith actual value of the power, N is the sample point numbers in the PV power generation period.

Table 3 indicates the prediction accuracy evaluation by adopting the RMSE, MAPE and R^2, considering the proposed SAGA-FCM-LSSVM model-based prediction technology and the other comparative prediction approaches (BPNN and LS-SVM). It can be concluded that the proposed model results in better prediction accuracy: the RMSE, MAPE and R^2 have 6.1970 kW, 4.006% and 0.9890 average values. The average RMSE enhancement of the proposed model with respect to the previous comparative models is 41.37% and 37.29%. And the average MAPE enhancement is 40.85% and 35.36%, respectively.

Therefore, compared with BPNN and LS-SVM approaches, the average amplitudes of prediction error of the proposed method are smaller. Synthetically, the proposed SAGA-FCM-LSSVM model is a novel and effective short-term PV power generation prediction model.

Table 3. Comparison of the prediction error by adopting various prediction methods.

		BPNN	LS-SVM	SAGA-FCM-LSSVM
RMSE (kW)	Spring	11.0058	10.0473	9.0523
	Summer	11.7520	10.0885	3.3372
	Fall	11.1385	10.7465	5.0945
	Winter	8.3804	8.6486	7.4041
	Average	10.5692	9.8827	**6.1970**
MAPE (%)	Spring	7.1923	6.9588	6.0262
	Summer	6.6781	5.8421	1.8333
	Fall	7.3188	6.2311	2.7140
	Winter	5.9008	5.7562	5.4505
	Average	6.7725	6.1971	**4.006**
R^2	Spring	0.9341	0.9459	0.9903
	Summer	0.9665	0.9602	0.9957
	Fall	0.9770	0.9637	0.9938
	Winter	0.9779	0.9828	0.9761
	Average	0.9639	0.9632	**0.9890**

4. Conclusions

In the study, a new SAGA-FCM-LSSVM model-based method is proposed for short-term PV power generation forecasting. The original hourly global horizontal radiation, air temperature, relative humidity and PV power are taken as the inputs of the prediction model. Owing to the utilization of historical hourly multivariate meteorological factors and power, the input variables become more accurate, which is of great significance to the real-time prediction of PV power. Furthermore, the FCM clustering algorithm is applied to cluster the historical power datasets, the LS-SVM technique is adopted to train the forecasting model. The SAGA method is used to optimize the initial clustering centers of the FCM clustering algorithm. The datasets on the DKASC website are employed to verify the proposed method, by contrasting with BPNN and LS-SVM methods. The experimental results show that the smaller prediction error can be obtained by applying the proposed model, and its prediction performance is superior to the comparative models.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61574038, 61601127, and 51508105), the Science Foundation of Fujian Education Department (Grant No. JAT160073), and the Science Foundation of Fujian Science & Technology Department (Grant Nos. 2018J01774, 2015J05124 and 2016H6012), the Fujian Provincial Economic and Information Technology Commission of China (Grant Nos. 83016006 and 830020).

The performance of the proposed forecasting approach was tested by using a large database from the Desert Knowledge Australia Solar Center (DKASC) website.

References

[1] Daliento S, Chouder A, Guerriero P, Pavan A M, Mellit A, Moein R and Tricoli P 2017

Monitoring, diagnosis, and power forecasting for photovoltaic fields: A review *Int. J. Photoenergy* **2017** 13
[2] Voyant C, Notton G, Kalogirou S, Nivet M L, Paoli C, Motte F and Fouilloy A 2017 Machine learning methods for solar radiation forecasting: A review *Renew. Energ.* **105** 569-82
[3] Kazem H A and Yousif J H 2017 Comparison of prediction methods of photovoltaic power system production using a measured dataset *Energ. Convers. Manage.* **148** 1070-81
[4] Sobri S, Koohi-Kamali S and Rahim N A 2018 Solar photovoltaic generation forecasting methods: A review *Energ. Convers. Manage.* **156** 459-97
[5] Wang H Z, Yi H Y, Peng J C, Wang G B, Liu Y T, Jiang H and Liu W X 2017 Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network *Energ. Convers. Manage.* **153** 409-22
[6] Wang F, Zhen Z, Liu C, Mi Z, Hodge B M, Shafie-Khah M and Catalão J P S 2018 Image phase shift invariance based cloud motion displacement vector calculation method for ultra-short-term solar PV power forecasting *Energ. Conver.s Manage.* **157** 123-35
[7] Das U K, Tey K S, Seyedmahmoudian M, Mekhilef S, Idris M Y I, Deventer W V, Horan B and Stojcevski A 2018 Forecasting of photovoltaic power generation and model optimization: A review *Renew. Sust. Energ. Rev.* **81** 912-28
[8] Khan I, Zhu H L, Yao J X, Khan D and Iqbal T 2017 Hybrid power forecasting model for photovoltaic plants based on neural network with Air Quality Index *Int. J. Photoenergy* **2017** 1-9
[9] Aziz M A A, Yasin Z M and Zakaria Z 2017 Prediction of photovoltaic system output using hybrid least square support vector machine *2017 7th IEEE International Conference on System Engineering and Technology (ICSET) Shah Alam Malaysia* 151-6
[10] Liu F, Li R R, Li Y, Yan R F and Saha T 2017 Takagi–Sugeno fuzzy model-based approach considering multiple weather factors for the photovoltaic power short-term forecasting *Iet. Renew. Power Gen.* **11** 1281-7
[11] Balato M, Costanzo L and Vitelli M 2016 Reconfiguration of PV modules: A tool to get the best compromise between maximization of the extracted power and minimization of localized heating phenomena *Sol. Energy* **138** 105-18
[12] Balato M, Costanzo L and Vitelli M 2015 Series–Parallel PV array re-configuration: Maximization of the extraction of energy and much more *Appl. Energ.* **159** 145-60
[13] Eseye A T, Zhang J H and Zheng D H 2018 Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information *Renew. Energ.* **118** 357-67