

A Short-Term PV Power Forecasting Method Using a Hybrid Kmeans-GRA-SVR Model under Ideal Weather Condition

Jiemin Lin, Haiming Li*

School of Computer Science and Technology, Shanghai University of Electric Power, Shanghai, China

Email: *zjxulhm@163.com

How to cite this paper: Lin, J.M. and Li, H.M. (2020) A Short-Term PV Power Forecasting Method Using a Hybrid Kmeans-GRA-SVR Model under Ideal Weather Condition. *Journal of Computer and Communications*, 8, 102-119.
<https://doi.org/10.4236/jcc.2020.811008>

Received: October 16, 2020

Accepted: November 23, 2020

Published: November 26, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

With the continuous increase of solar penetration rate, it has brought challenges to the smooth operation of the power grid. Therefore, to make photovoltaic power generation not affect the smooth operation of the grid, accurate photovoltaic power prediction is required. And short-term forecasting is essential for the deployment of daily power generation plans. In this paper, A short-term photovoltaic power generation forecast method based on K-means++, grey relational analysis (GRA) and support vector regression (SVR) (Hybrid Kmeans-GRA-SVR, HKGSVR) was proposed. The historical power data was clustered through the multi-index K-means++ algorithm. And the similar days and the nearest neighbor similar day of the prediction day were selected by the GRA algorithm. Then, similar days and nearest neighbor similar days were used to train SVR to obtain an accurate photovoltaic power prediction model. Under ideal weather, the average values of MAE, RMSE, and R^2 were 0.8101 kW, 0.9608 kW, and 99.66%, respectively. The average computation time was 1.7487 s, which was significantly better than the SVR model. Thus, the demonstrated numerical results verify the effectiveness of the proposed model for short-term PV power prediction.

Keywords

Grey Relational Analysis, K-Means++, Nearest Neighbor Similar Day, Photovoltaic Power, Support Vector Regression

1. Introduction

In recent years, although breakthroughs have been made in the exploitation of shale gas and deep-sea combustible ice has also progressed, the fact that fossil energy reserves have limits has not changed. The development and utilization of

renewable energy technologies is still of great significance [1], and photovoltaic power generation is currently one of the most promising renewable power generation technologies [2]. Because photovoltaic energy is easily available, pollution-free and inexhaustible, it has become the best substitute for industrial and residential power generation [3]. According to the 2020 report of the International Renewable Energy Agency, in the past 8 years, the global PV power generation cost has fallen by more than 70%, and the global installed capacity has reached 578.553 GW [4].

However, the chaotic nature of the weather system makes the production of photovoltaic power generation highly random, volatile and intermittent, which greatly increases the difficulty of large-scale application of PV power generation [5] [6]. In April 2018, California was forced to cut its power generation because the electricity from solar and wind power plants could not be consumed and had to be abandoned.

In order to improve the power system's ability to consume photovoltaic energy, many solutions have been proposed, including energy storage optimization [7], demand response strategy [8] [9], power flow optimization [10], stand-alone microgrid [11], and PV power forecasting [12]. Considering economic and feasibility, photovoltaic power generation forecasting is one of the most promising solutions. Meanwhile, it is also an important technical support for the refined management of photovoltaic power generation [13] [14]. The voltage and frequency imbalance caused by the grid connection of photovoltaic power stations can be eliminated through accurate photovoltaic power prediction [15].

The current photovoltaic power generation forecasting technologies have three main directions: physical methods, statistical-time series methods and ensemble methods [13]. The physical methods include Sky Imagery, Satellite-Imaging, and Numerical Weather Prediction (NWP) models. The statistical methods consist of Regression, Autoregressive, Markov Chain, Support Vector Machine (SVM), and Artificial Neural Network (ANN) models. The ensemble method is an optimized combination of the aforementioned methods.

Satellite Images and Sky Images make predictions by tracking and predicting the trajectory of the cloud, but are limited by image resolution and processing algorithms [2] [16] [17] [18] [19]. The cost and practicality of this method need to be optimized. In [20] [21] [22] [23], NWP is proposed to predict solar irradiance. However, there are many restrictions on obtaining NWP information from domestic meteorological departments, which limit its application. In [24] [25] [26] [27], the time series method is applied to short-term solar forecasting, but as the time scale and output dimension increase, the forecast accuracy decreases. The prediction of solar energy is realized by artificial neural network in [28] [29] [30] [31]. The disadvantage of ANN is that the complexity of the system is greatly increased due to the nature of the multi-layer network architecture. In [32] [33], the support vector machine (SVM) is used for short-term photovoltaic output forecasting. SVM is a simple, powerful, non-linear and low

computational cost model. Compared with artificial neural networks, SVM can overcome the problem of local minimums and requires fewer samples (that is, it can learn without relying heavily on prior knowledge).

The ensemble method solves the limitations of a single model by mixing together different models with unique functions, thereby improving the prediction performance [34]. For the prediction of photovoltaic power generation, the ensemble method that mixes various effective methods is more effective and accurate [35]. For example, the hybrid GRA-GA-SVM model [36], which performed better than the SVM model.

Under ideal weather condition, a hybrid Kmeans-GRA-SVR model is proposed in this paper. The main contributions of this paper include:

- 1) A novel short-term PV power forecasting method that utilizes SVR, clustering and similarity algorithms was proposed.
- 2) In order to increase the operation speed and reduce the operation cost, a multi-index clustering algorithm is used to cluster historical power data to obtain ideal weather and non-ideal weather.
- 3) In order to improve the prediction accuracy of photovoltaic power generation under ideal weather conditions, clustering algorithm and GRA algorithm are used to obtain similar days in the same cluster as the forecasting day.
- 4) Through the GRA algorithm to obtain the nearest neighbor similar days to solve the problem of decreased prediction accuracy caused by the large time interval between the similar days and forecasting day.

The remainder of this paper is organized as follows. Section 2 describes the hybrid Kmeans-GRA-SVR model. Section 3 illustrates clustering and model evaluation metrics. Section 4 introduces the experiments and result analysis. Finally, conclusions are given in Section 5.

2. Hybrid Kmeans-GRA-SVR Model

2.1. K-Means++ Clustering Algorithm

K-means++ clustering algorithm is an improved version of K-means algorithm. This algorithm separates the K initial cluster centers more from each other. In this work, which is selected as the classifier due to its higher efficiency and improved robustness compared with others (e.g., standard K-means, K-medoids, Gaussian mixture models, etc.) [37]. The running process of K-means++ is as follows [38]:

Step 1: Randomly select a sample as the first cluster center c_1 ;

Step 2: Calculate the probability of each sample being selected as the next cluster center:

$$\frac{D(x)^2}{\sum_{x \in X} D(x)^2} \quad (1)$$

where, $D(x)$ represents the distance between the sample and the nearest cluster center.

Then use the roulette method to select the next cluster center;

Step 3: Repeat step 2 until K cluster centers are selected;

Step 4: For each sample x_j in the datasets, calculate its distance to K cluster centers, and then put it into the class corresponding to the smallest distance cluster center;

Step 5: For each cluster, recalculate its cluster center c_i

$$c_i = \frac{1}{C_i} \sum_{x \in c_i} x \tag{2}$$

Step 6: Repeat steps 4 and 5 until the position of the cluster center does not change.

In this part, K-means++ clustering method is used to directly cluster the historical power data of each season. The reason for selecting historical power data for clustering is that the aging of the equipment itself and its own indicators are different under different weather conditions. It is difficult for us to finely measure these changes. The characteristics of historical power data will integrate these changes into it. After clustering the historical power data, the centroid value of each cluster is calculated by the minimum, average and maximum of global horizontal irradiance (GHI), diffuse horizontal irradiance (DHI), relative humidity (RH) and temperature (T). The Euclidean distance between the 12 meteorological factor characteristic values of the forecast day and each cluster centroid is calculated to determine the cluster to which the forecast day belongs.

2.2. Grey Relational Analysis Algorithm

Grey relational analysis refers to the quantitative description and comparison method of the development and change of a system. The basic idea is to judge the correlation degree by comparing the geometrical similarity between the reference data column and several data columns. It reflects the degree of correlation between the curves. Generally, the more consistent the change tendency of the reference sequence and the comparison sequence, the higher the degree of correlation between the two variables. The flow of the GRA algorithm is as follows [39]:

Step 1: Determine the reference sequence y that reflects the characteristics of the system behavior and the comparison sequence x_i that affects the system behavior:

$$y = \{y(k) | k = 1, 2, \dots, n\} \tag{3}$$

$$x_i = \{x_i(k) | k = 1, 2, \dots, n\}, i = 1, 2, \dots, m \tag{4}$$

where, n and m represent the dimension of the eigenvalues and the number of comparison sequence, respectively.

Step 2: Non-dimensionalization of variables:

$$d_j^*(k) = \frac{D_j(k) - D_{av}(k)}{D_{\max}(k) - D_{\min}(k)}, k = 1, 2, \dots, n; i = 0, 1, 2, \dots, m; j = 1, 2, \dots, m + 1 \tag{5}$$

where, $D_j(k)$ contains reference sequence and comparison sequence, $D_{av}(k)$,

$D_{\min}(k)$ and $D_{\max}(k)$ are the average, minimum and maximum values of each column, j represents sum of the number of reference sequence and comparison sequence.

Non-dimensionalization is used to solve the problem that the columns cannot be compared due to the different dimensions.

Step 3: Calculate correlation coefficient $\xi_i(k)$:

$$\xi_i(k) = \frac{\min_i \min_k |y(k) - x_i(k)| + \rho \max_i \max_k |y(k) - x_i(k)|}{|y(k) - x_i(k)| + \rho \max_i \max_k |y(k) - x_i(k)|} \quad (6)$$

where, ρ is called the resolution coefficient, here, ρ is 0.5.

Step 4: Calculate correlation degree.

Calculate the average value of the correlation coefficient at each moment (that is, each point in the curve) r_i :

$$r_i = \frac{1}{n} \sum_{k=1}^n \xi_i(k), k = 1, 2, \dots, n \quad (7)$$

Step 5: Sort correlation degree.

After determining the cluster to which the prediction day belongs, the correlation between the prediction day and each sample in the cluster is calculated by GRA based on 12 meteorological factor eigenvalues. And the date with the correlation degree greater than the threshold (an appropriate correlation value that takes into account the similarity and the number of samples) is regarded as the similar days. For the ideal weather, the sample with the highest correlation in the 7 days before the forecast date is set as the nearest neighbor similar day.

2.3. Support Vector Regression

SVM obtains the ability to linearly analyze the nonlinear characteristics of the sample by mapping low-dimensional data to high-dimensional space. Based on the structural risk minimization theory, SVM constructs the optimal classification surface in the feature space, thereby overcoming the local optimal problem and requiring fewer training samples. When the data type is complex, SVR can be used, which was first developed by Vapnik *et al.* [40]. The definition of SVR is as follows:

$$f(x) = \omega^T \phi(x) + b \quad (8)$$

where ω is a vector of weight coefficients, $\phi(x)$ is the nonlinear mapping function (mapping x to a high-dimensional feature space), and b denotes a bias constant. In addition, b and ω can be obtained by the following formula:

$$\text{minimize: } \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \xi_i + \xi_i^* \quad (9)$$

subject to:

$$\begin{cases} y_i - \langle \omega, \phi(x_i) \rangle - b \leq \varepsilon + \xi_i \\ \langle \omega, \phi(x_i) \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i \geq 0, \xi_i^* \geq 0 \end{cases} \quad (10)$$

where ξ_i and ξ_i^* are slack variables, and C denotes the penalty variable, ε is the insensitive loss function.

In this paper, the radial basis function (RBF) kernel is applied to construct the SVR model. The RBF kernel is presented as:

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \quad (11)$$

where γ is the kernel parameter.

2.4. Day-Ahead PV Power Ensemble Forecasting Model

2.4.1. Hybrid Kmeans-GRAs-SVR Model

The working process of the hybrid Kmeans-GRAs-SVR model is shown in **Figure 1**.

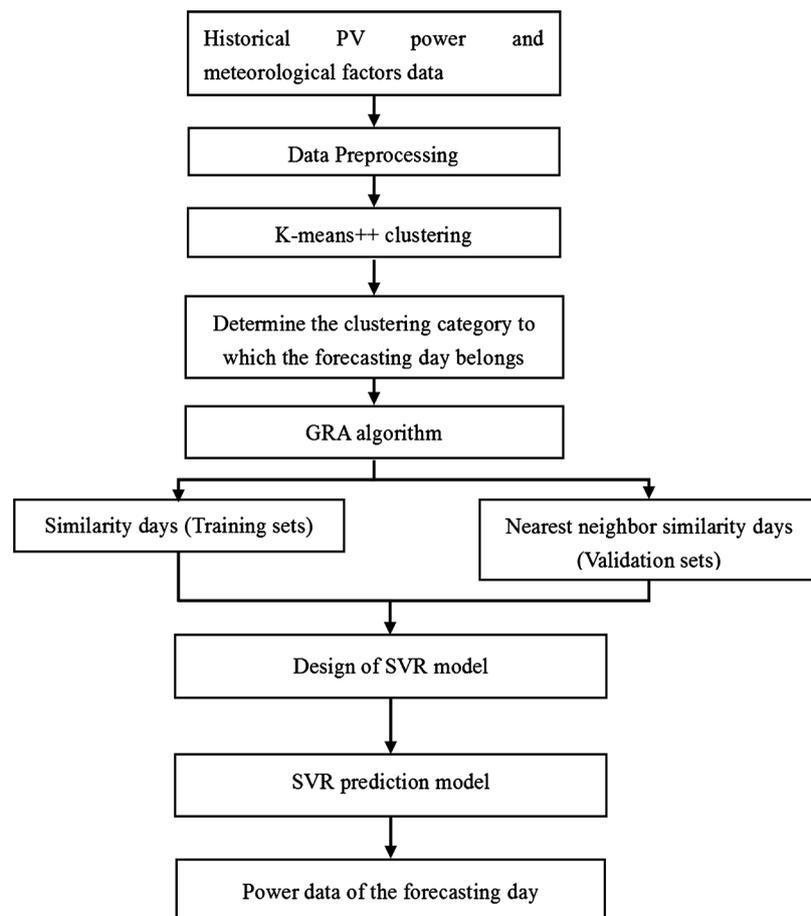


Figure 1. The working process of the Hybrid Kmeans-GRAs-SVR model.

Step 1: The historical power and meteorological factors data are used as training and test data, and the missing and abnormal data in the data set are processed.

Step 2: The historical photovoltaic power data of the four seasons are clustered separately through the K-means++ algorithm, and the minimum, average and maximum values of GHI, DHI, RT, and T are regarded as the central value of each cluster.

Step 3: According to the 12 meteorological factor eigenvalues, the Euclidean distance between the forecast day and each cluster center is calculated to determine the cluster category to which the forecast day belongs.

Step 4: The correlation between the prediction day and each sample in the cluster is calculated by GRA to obtain similar days and nearest neighbor similar days. We normalize the data of similar days and nearest neighbor similar days as training set and validation set.

Step 5: After determining the C and γ of SVR through grid search and cross-validation, the SVR is trained to obtain a prediction model and predict the output power on the prediction day.

2.4.2. Forecast Structure under Ideal Weather Condition

The structure of the prediction model under ideal weather conditions is shown in **Figure 2**. The power of the nearest neighbor similarity day and predicted day's global horizontal irradiance, diffuse horizontal irradiance, relative humidity and temperature minimum, average and maximum value as input. Because the annual photovoltaic power generation has obvious seasonal distribution characteristics, we train the HKGSVR forecasting model separately for the forecast days of each season.

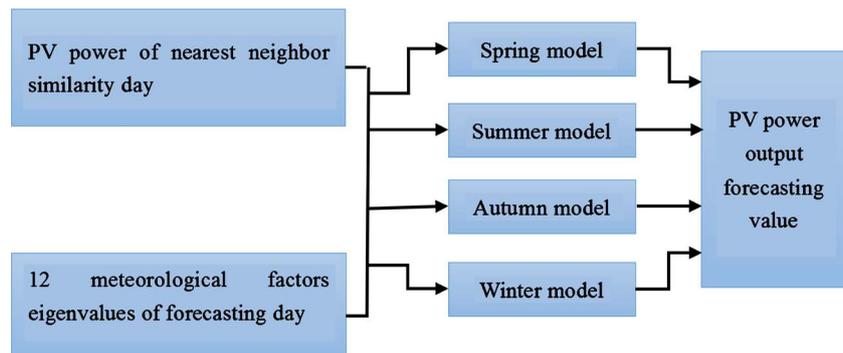


Figure 2. Forecast structure under ideal weather condition.

3. Evaluation Metrics

3.1. Clustering Evaluation Metrics

If the ground truth labels are not known, evaluation must be performed using the model itself. The Silhouette Coefficient is an example of such an evaluation, the score is higher when clusters are dense and well separated. Silhouette Coefficient $S(i)$ is defined as follows:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{12}$$

where, $a(i)$ is the mean distance between a sample and all other points in the same cluster, $b(i)$ is the mean distance between a sample and all other points in the next nearest cluster. Average the Silhouette Coefficient of all points, which is the total Silhouette Coefficient of the clustering result.

Davies-Bouldin index is defined as follows:

$$DBI = \frac{1}{n} \sum_{i=1}^n \max_{j \neq i} \left(\frac{\overline{S}_i + \overline{S}_j}{\|\omega_i - \omega_j\|_2} \right) \tag{13}$$

where, \overline{S}_i is the average distance from the points in the cluster to the cluster centroid, $\|\omega_i - \omega_j\|_2$ is the distance between the centroid of cluster i and j . The Davies-Bouldin index is lower if the model clusters have better separation.

Sum of squared errors (SSE) is also an effective metric. That is, the sum of squared errors of the distance between the centroid of each cluster and the points in the cluster. SSE is defined as follows:

$$SSE = \sum_{i=1}^K \sum dist(x, c_i)^2 \tag{14}$$

3.2. Metrics of Photovoltaic Power Forecasting Techniques

In order to evaluate the performance of the proposed method HKGSVR for photovoltaic power generation forecasting, the root mean square error (RMSE), average absolute error (MAE) and coefficient of determination (R^2) indicators were calculated. They are defined as follows.

1) The RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N P_{fi} - P_{ai}} \tag{15}$$

where, P_{ai} and P_{fi} are the actual and predicted value at i hour. N refers to the number of hours a sample contains.

2) The MAE is expressed as:

$$MAE = \frac{1}{N} \sum_{i=1}^N |P_{fi} - P_{ai}| \tag{16}$$

3) The R^2 is given as:

$$R^2 = \frac{\left(N \sum_{i=1}^N P_{fi} P_{ai} - \sum_{i=1}^N P_{fi} \sum_{i=1}^N P_{ai} \right)^2}{\left(N \sum_{i=1}^N P_{fi}^2 - \left(\sum_{i=1}^N P_{fi} \right)^2 \right) \left(N \sum_{i=1}^N P_{ai}^2 - \left(\sum_{i=1}^N P_{ai} \right)^2 \right)} \tag{17}$$

4. Experimental Analysis

4.1. Data

In this paper, the general datasets on the DKASC (Desert Knowledge Australia Solar Center) website are used for related experiments. The photovoltaic array is composed of 22 polycrystalline silicon photovoltaic panels with a rated power of 265 W, whose total rated power is 5.83 kW. The photovoltaic array is located at the Desert Knowledge Precinct in Alice Springs (a town in the Northern Territory that enjoys one of the country’s highest solar resources in an arid desert environment). The geographic location, physical object and configuration infor-

mation of the photovoltaic array are shown in **Figure 3**, **Figure 4**, and **Table 1**. Meteorology (global horizontal irradiance, diffuse horizontal irradiance, relative humidity and temperature) and historical power data of PV arrays from March 1, 2018 to February 29, 2020 were used in the experiment. The experiment uses

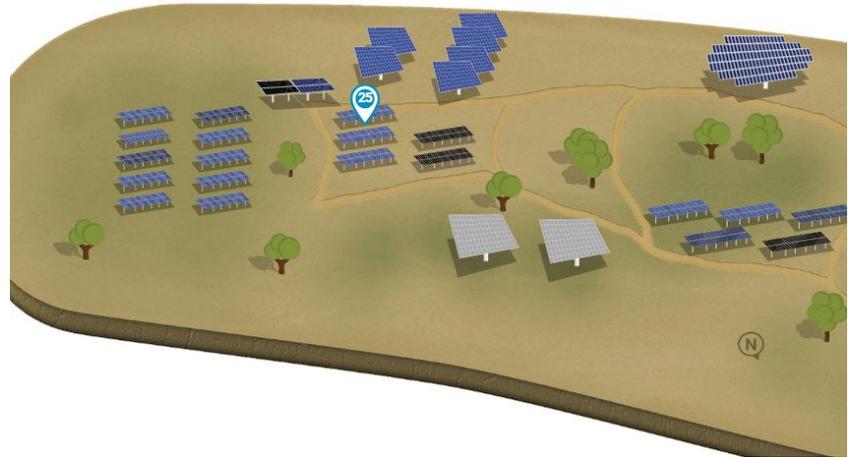


Figure 3. The geographic location of photovoltaic array (25 is the label of the array).



Figure 4. The physical object of photovoltaic array.

Table 1. The configuration information of photovoltaic array.

Item	Information
Array Rating	5.83 kW
Panel Rating	265 W
Number of Panels	22
Panel Type	HSL 60S
Array Area	36.74
Inverter Size/Type	SMA SMC 6000 A
Array Tilt/Azimuth	Tilt = 20, Azimuth = 0 (Solar North)
Installation Completed	Sat, 2 Jul 2016

data with an interval of 1 hour from 7:00 to 18:00 every day.

4.2. Number of Clusters

In order to obtain the appropriate number of clusters for each season, SSE, DBI and Silhouette Coefficient (S) are used for evaluation. Taking summer as an example, the experimental results are shown in [Figure 5](#), [Figure 6](#), [Figure 7](#), and [Figure 8](#). It can be seen from [Figure 5](#) that SSE decreases as the number of clusters increases, and when it reaches 2, the downward trend begins to slow down. Observe [Figure 6](#), DBI achieves the best value in 3 clusters. [Figure 7](#) shows that the maximum value of S appears in 2 clusters.

After comprehensively considering each evaluation index and cluster observation results, the number of clusters in summer clustering is set to 2, and the result is shown in [Figure 8](#). In [Figure 8](#), the blue cluster is defined as ideal weather cluster, and the red cluster is non-ideal weather cluster. The evaluation of clustering results in each season is shown in [Table 2](#). In order to prevent local

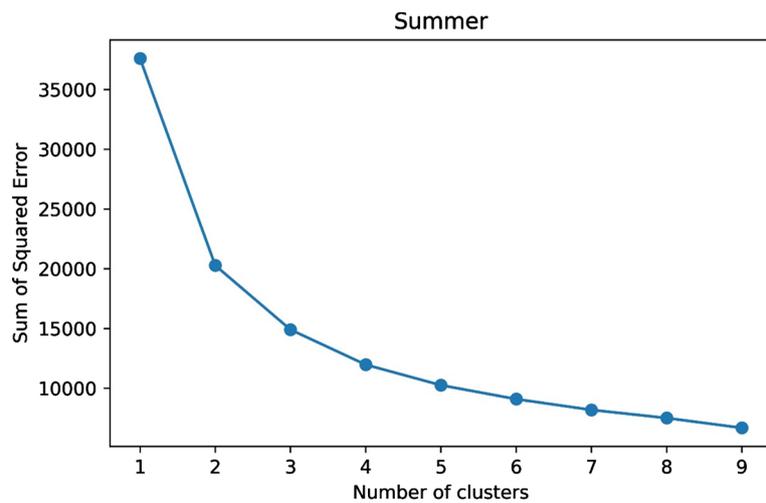


Figure 5. The relationship between the number of clusters and SSE in summer.

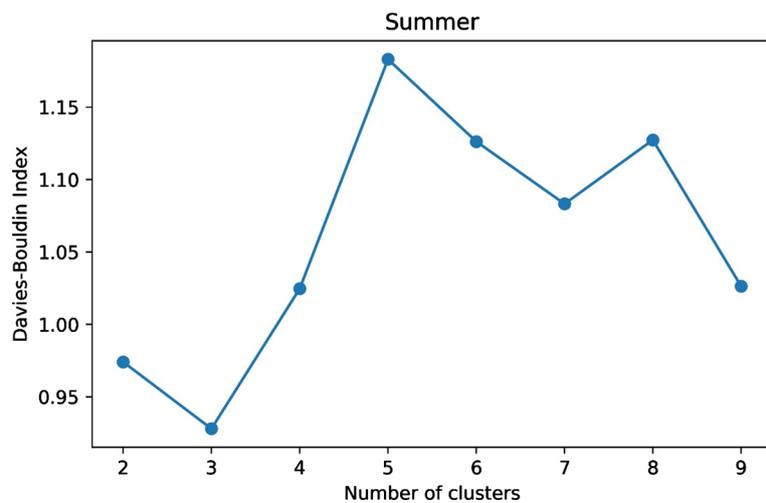


Figure 6. The relationship between the number of clusters and DBI in summer.

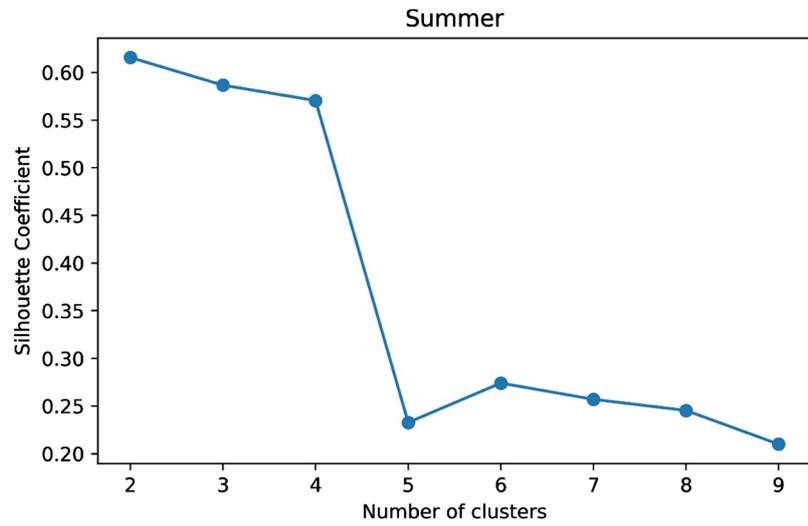


Figure 7. The relationship between the number of clusters and Silhouette Coefficient in summer.

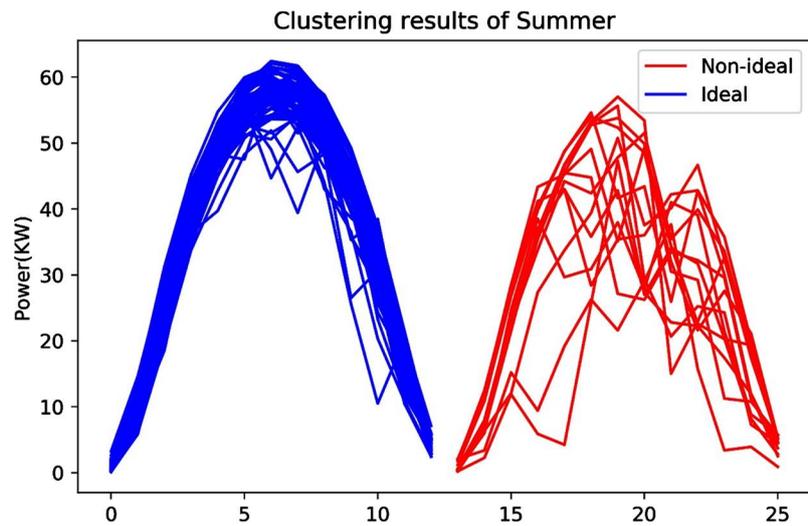


Figure 8. The clustering results of summer.

Table 2. Cluster evaluation metrics for each season.

Metrics	Spring	Summer	Autumn	Winter
2 cluster SSE	44,578.4063	20,278.6656	26,813.2643	4218.8007
3 cluster SSE	32,045.1222	14,902.9881	15,383.4952	2942.0887
4 cluster SSE	24,464.0061	11,971.8954	11,657.9236	2472.9994
2 cluster DBI	0.6723	0.9740	0.6410	0.7048
3 cluster DBI	0.7846	0.9440	0.7252	1.0029
4 cluster DBI	1.0187	1.0247	0.7773	0.8355
2 cluster S	0.6383	0.6156	0.7119	0.4861
3 cluster S	0.5893	0.5866	0.6390	0.5074
4 cluster S	0.5231	0.5705	0.4073	0.5092

optima or other abnormal situations, 100 rounds of experiments were carried out. Finally, the number of clusters in spring is 3, the number of clusters in summer is 2, the number of clusters in autumn is 3, and the number of clusters in winter is 3.

4.3. Correlation Degree Threshold for Similar Days

The selection of the threshold for similar days is not only related to the season, but also to the weather conditions on the forecast day. For ideal weather, it is observed that when the correlation degree is greater than 0.85, the power curve of similar days in each season is in a relatively ideal state (as shown in **Figure 9**, the power curve of August 19, 2018 (winter) with a predicted day correlation degree of 0.85). In order to increase the speed of training the model and reduce the computational cost, a higher correlation threshold is used in the experiment.

Under ideal weather conditions, the forecast days of each season, the correlation threshold of similar days, and the selection of nearest neighbor similar days are shown in **Table 3**. The thresholds of similar days in each season are 0.90, 0.88, 0.89 and 0.86.

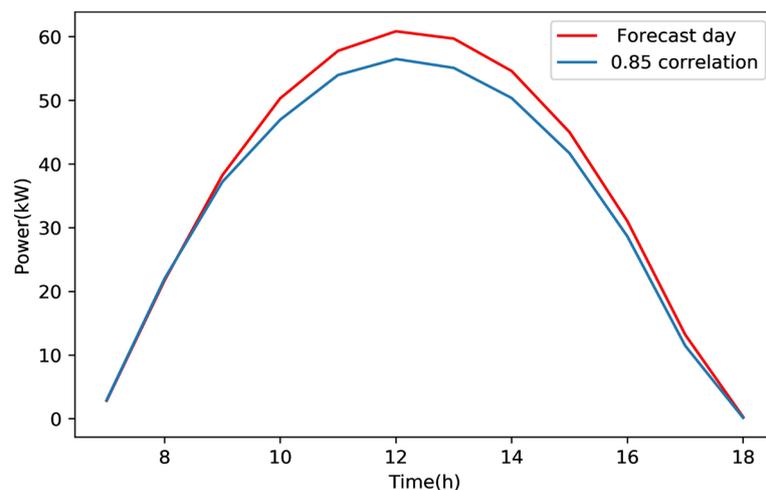


Figure 9. Forecast day and 0.85 correlation day curves.

Table 3. Forecasting day, similarity day's correlation threshold and nearest neighbor similarity day and correlation under ideal weather condition.

Item	Spring	Summer	Autumn	Winter
Forecasting day	September 15, 2019	February 09, 2020	April 18, 2019	August 17, 2019
Similar days threshold	0.90	0.88	0.92	0.86
Nearest neighbor similarity day	September 14, 2019	February 08, 2020	April 15, 2019	August 16, 2019
Correlation degree	0.9295	0.9452	0.9913	0.9579

4.4. Design of SVR Model

This part is mainly to explore the optimal C and γ of SVR, which are usually re-

lated to the characteristics of power generation in different seasons. Grid search and cross-validation are used to find the optimal number of C and γ for SVR. The optimal SVR structure for each season under ideal weather is shown in **Table 4**. Observing the following table shows that the models training time is 2.0342 s, 1.9506 s, 2.3272 s and 0.6826 s respectively.

Table 4. Parameters of HKGSVR model for each season under ideal weather conditions.

Structure	Spring	Summer	Autumn	Winter
Input form	(564, 2)	(408, 2)	(612, 2)	(156, 2)
C	100,000.00	100,000.0	100,000.00	1.00
γ	0.0001	0.0001	0.002	1.00
Time (s)	2.0342	1.9506	2.3272	0.6826

4.5. Forecasting Results and Discussion

Figure 10 shows the prediction results of the HKGSVR model in each season under ideal weather conditions. The feature combination is the power of the nearest neighbor similar day and the 12 meteorological factor eigenvalues of the forecast day (NP_W). Because the prediction accuracy is high under ideal weather condition, feature selection is considered from the difficulty of obtaining and the accuracy of the data. The daily weather eigenvalues of the forecast day are easier to obtain and more accurate than the hourly forecast value.

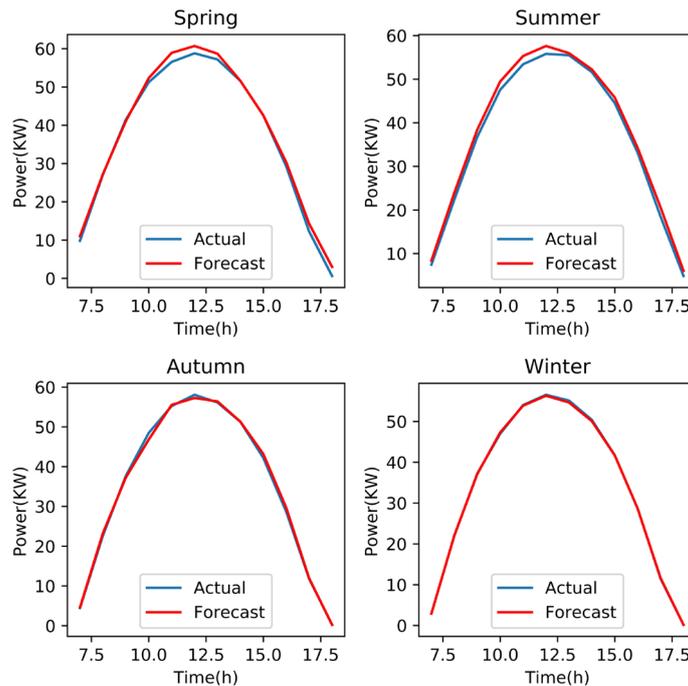


Figure 10. Forecast results of each season under ideal weather conditions.

It can be seen from **Figure 10** that the forecast performance of each season is superior under ideal weather conditions. From the evaluation metrics in **Table**

5, it can be seen that the average value of R^2 is 0.9966. The MAE are 1.4521, 1.4661, 0.7120, and 0.2132 kW respectively. The average value of RMSE is 0.9608 kW. The model has high prediction accuracy for ideal weather.

As shown in **Figure 11**, HKGSVR's forecast results in each season are significantly better than SVR. Through the comparison of **Table 5** and **Table 6**, it can be found that the training and optimization time of HKGSVR is much shorter than that of SVR within the same search range. Compared with SVR, the proposed model has an optimization of 74.99%, 73.86%, 69.18%, and 91.05% in training and optimization time for each season. The proposed model's MAE

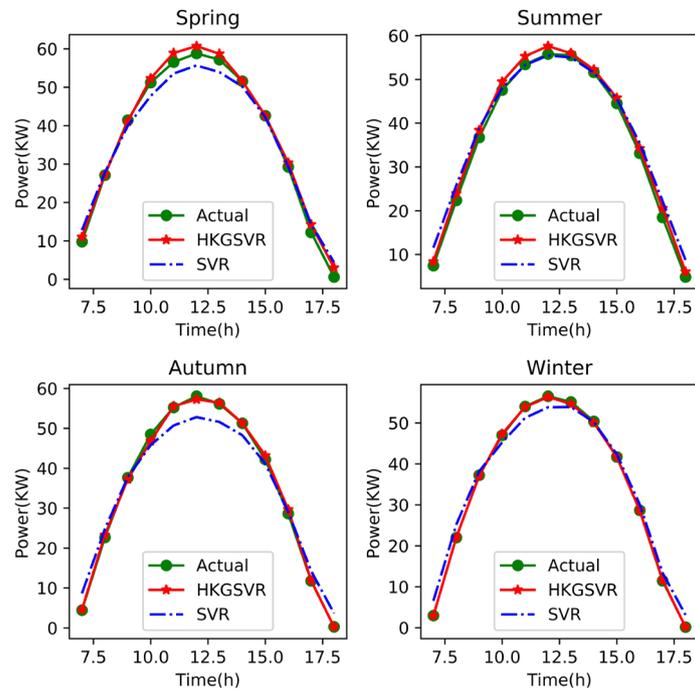


Figure 11. Comparison of SVR and HKGSVR prediction results.

Table 5. HKGSVR evaluation metrics under ideal weather conditions.

Metrics	Spring	Summer	Autumn	Winter	Average	Std
MAE (kW)	1.1763	1.3753	0.5269	0.1618	0.8101	0.5640
RMSE (kW)	1.4521	1.4661	0.7120	0.2132	0.9608	0.6103
R^2	0.9945	0.9932	0.9987	0.9999	0.9966	0.0032
Time (s)	2.0342	1.9506	2.3272	0.6826	1.7487	0.7288

Table 6. SVR evaluation metrics under ideal weather conditions.

Metrics	Spring	Summer	Autumn	Winter	Average	Std
MAE (kW)	2.1586	1.8429	2.8045	1.9542	2.1901	0.4300
RMSE (kW)	2.4877	2.3955	3.2315	2.2143	2.5822	0.4475
R^2	0.9837	0.9819	0.9737	0.9873	0.9817	0.0058
Time (s)	8.1325	7.4614	7.5504	7.6254	7.6924	0.3009

enhancement with respect to the SVR model is 45.51%, 25.37%, 81.21%, 91.72%, respectively. The presented model's RMSE improvement relative to the SVR model is 41.63%, 38.80%, 77.97%, 90.37%, respectively. The average R^2 of the proposed model is also better than the SVR model.

5. Conclusion

A hybrid day-ahead photovoltaic power generation prediction model (HKGSVR) based on K-means++, GRA and SVR was proposed. Both historical power data and weather data were used to train the model. Moreover, samples of similar days and nearest neighbor similar days were used to train the prediction model. The average values of MAE, RMSE, and R^2 were 0.8101 kW, 0.9608 kW, and 99.66%, respectively. The average computation time was 1.7487 s, which was significantly better than the SVR model. Thus, the demonstrated numerical results verify the effectiveness of the proposed model for short-term PV power prediction.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Tsai, Y.-C., Chan, Y.-K., Ko, F.-K. and Yang, J.-T. (2018) Integrated Operation of Renewable Energy Sources and Water Resources. *Energy Conversion and Management*, **160**, 439-454. <https://doi.org/10.1016/j.enconman.2018.01.062>
- [2] Wang, F., Zhen, Z., Liu, C., Mi, Z., Hodge, B.-M., Shafie-khah, M., *et al.* (2018) Image Phase Shift Invariance Based Cloud Motion Displacement Vector Calculation Method for Ultra-Short-Term Solar PV Power Forecasting. *Energy Conversion and Management*, **157**, 123-135. <https://doi.org/10.1016/j.enconman.2017.11.080>
- [3] Koo, C., Hong, T., Jeong, K., Ban, C. and Oh, J. (2017) Development of the Smart Photovoltaic System Blind and Its Impact on Net-Zero Energy Solar Buildings Using Technical-Economic-Political Analyses. *Energy*, **124**, 382-396. <https://doi.org/10.1016/j.energy.2017.02.088>
- [4] IRENA (2020) Renewable Energy Statistics 2020. The International Renewable Energy Agency, Abu Dhabi.
- [5] Jiang, H. and Dong, Y. (2017) Forecast of Hourly Global Horizontal Irradiance Based on Structured Kernel Support Vector Machine: A Case Study of Tibet Area in China. *Energy Conversion and Management*, **142**, 307-321. <https://doi.org/10.1016/j.enconman.2017.03.054>
- [6] Köhler, C., Steiner, A., Saint-Drenan, Y.-M., Ernst, D., Bergmann-Dick, A., Zirkelbach, M., *et al.* (2017) Critical Weather Situations for Renewable Energies—Part B: Low Stratus Risk for Solar Power. *Renew Energy*, **101**, 794-803. <https://doi.org/10.1016/j.renene.2016.09.002>
- [7] Wang, F., Zhou, L., Ren, H., Liu, X., Talari, S., Shafie-khah, M., *et al.* (2018) Multi-Objective Optimization Model of Source-Load-Storage Synergetic Dispatch for a Building Energy Management System Based on TOU Price Demand Response. *IEEE Transactions on Industry Applications*, **54**, 1017-1028.

- <https://doi.org/10.1109/TIA.2017.2781639>
- [8] Talari, S., Shafie-khah, M., Wang, F., Aghaei, J. and Catalao, J.P.S. (2017) Optimal Scheduling of Demand Response in Pre-Emptive Markets Based on Stochastic Bilevel Programming Method. *IEEE Transactions on Industrial Electronics*, 66, 1453-1464. <https://doi.org/10.1109/TIE.2017.2786288>
- [9] Chen, Q., Wang, F., Hodge, B.M., Zhang, J., Li, Z., Shafie-Khah, M., *et al.* (2017) Dynamic Price Vector Formation Model-Based Automatic Demand Response Strategy for PV-Assisted EV Charging Stations. *IEEE Transactions on Smart Grid*, 8, 2903-2915. <https://doi.org/10.1109/TSG.2017.2693121>
- [10] Biswas, P.P., Suganthan, P.N. and Amaratunga, G.A.J. (2017) Optimal Power Flow Solutions Incorporating Stochastic Wind and Solar Power. *Energy Conversion and Management*, 148, 1194-1207. <https://doi.org/10.1016/j.enconman.2017.06.071>
- [11] Wang, F., Zhou, L., Wang, B., Wang, Z., Shafie-khah, M. and Catalão, J. (2017) Modified Chaos Particle Swarm Optimization-Based Optimized Operation Model for Stand-Alone CCHP Microgrid. *Applied Sciences*, 7, 754. <https://doi.org/10.3390/app7080754>
- [12] Zhen, Z., Xuan, Z., Wang, F., Sun, R., Duić, N. and Jin, T. (2019) Image Phase Shift Invariance Based Multi-Transform-Fusion Method for Cloud Motion Displacement Calculation Using Sky Images. *Energy Conversion and Management*, 197, Article ID: 111853. <https://doi.org/10.1016/j.enconman.2019.111853>
- [13] Sobri, S., Koohi-Kamali, S. and Rahim, N.A. (2018) Solar Photovoltaic Generation Forecasting Methods: A Review. *Energy Conversion and Management*, 156, 459-497. <https://doi.org/10.1016/j.enconman.2017.11.019>
- [14] Wang, F., Li, K., Liu, C., Mi, Z., Shafie-Khah, M. and Catalao, J.P.S. (2018) Synchronous Pattern Matching Principle-Based Residential Demand Response Baseline Estimation: Mechanism Analysis and Approach Description. *IEEE Transactions on Smart Grid*, 9, 6972-6985. <https://doi.org/10.1109/TSG.2018.2824842>
- [15] Yu, H.J.J. (2017) Virtuous Cycle of Solar Photovoltaic Development in New Regions. *Renewable and Sustainable Energy Reviews*, 78, 1357-1366. <https://doi.org/10.1016/j.rser.2017.05.018>
- [16] Dev, S., Savoy, F.M., Lee, Y.H. and Winkler, S. (2016) Short-Term Prediction of Localized Cloud Motion Using Ground-Based Sky Imagers. 2016 *IEEE Region 10 Conference*, Singapore, 22-25 November 2016, 2563-2566. <https://doi.org/10.1109/TENCON.2016.7848499>
- [17] Engerer, N.A. and Morrissey, M. (2012) The Impact of Cloud Events on Simulated Photovoltaic Arrays. The University of Oklahoma, Norman. <https://doi.org/10.1016/j.ifacol.2017.08.1488>
- [18] Zaher, A., Thil, S. and Nou, J. (2017) Comparative Study of Algorithms for Cloud Motion Estimation Using Sky-Imaging Data. *IFAC-Papers OnLine*, 50, 5934-5939. <https://doi.org/10.1016/j.renene.2016.12.023>
- [19] Cheng, H.Y. (2017) Cloud Tracking Using Clusters of Feature Points for Accurate Solar Irradiance Nowcasting. *Renewable Energy*, 104, 281-289. <https://doi.org/10.1016/j.renene.2015.11.005>
- [20] Lima, F.J.L., Martins, F.R., Pereira, E.B., Lorenz, E. and Heinemann, D. (2016) Forecast for Surface Solar Irradiance at the Brazilian Northeastern Region Using NWP Model and Artificial Neural Networks. *Renewable Energy*, 87, 807-818. <https://doi.org/10.1016/j.solener.2013.05.005>
- [21] Perez, R., Lorenz, E., Pelland, S., Beauharnois, M., Van Knowe, G., Hemker, K., *et al.* (2013) Comparison of Numerical Weather Prediction Solar Irradiance Forecasts

- in the US, Canada and Europe. *Solar Energy*, **94**, 305-326.
<https://doi.org/10.1016/j.solener.2013.05.005>
- [22] Verzijlbergh, R.A., Heijnen, P.W., de Roode, S.R., Los, A. and Jonker, H.J.J. (2015) Improved Model Output Statistics of Numerical Weather Prediction Based Irradiance Forecasts for Solar Power Applications. *Solar Energy*, **118**, 634-645.
<https://doi.org/10.1016/j.solener.2015.06.005>
- [23] Mathiesen, P., Collier, C. and Kleissl, J. (2013) A High-Resolution, Cloud-Assimilating Numerical Weather Prediction Model for Solar Irradiance Forecasting. *Solar Energy*, **92**, 47-61. <https://doi.org/10.1016/j.solener.2013.02.018>
- [24] Prema, V. and Uma Rao, K. (2015) Development of Statistical Time Series Models for Solar Power Prediction. *Renewable Energy*, **83**, 100-109.
<https://doi.org/10.1016/j.renene.2015.03.038>
- [25] Hirata, Y. and Aihara, K. (2017) Improving Time Series Prediction of Solar Irradiance after Sunrise: Comparison among Three Methods for Time Series Prediction. *Solar Energy*, **149**, 294-301. <https://doi.org/10.1016/j.solener.2017.04.020>
- [26] Shireen, T., Shao, C., Wang, H., Li, J., Zhang, X. and Li, M. (2018) Iterative Multi-Task Learning for Time-Series Modeling of Solar Panel PV Outputs. *Applied Energy*, **212**, 654-662. <https://doi.org/10.1016/j.apenergy.2017.12.058>
- [27] Bigdeli, N., Salehi Borujeni, M. and Afshar, K. (2017) Time Series Analysis and Short-Term Forecasting of Solar Irradiation, a New Hybrid Approach. *Swarm and Evolutionary Computation*, **34**, 75-88. <https://doi.org/10.1016/j.swevo.2016.12.004>
- [28] Wang, F., Mi, Z., Su, S., Zhao, H. (2012) Short-Term Solar Irradiance Forecasting Model Based on Artificial Neural Network Using Statistical Feature Parameters. *Energies*, **5**, 1355-1370. <https://doi.org/10.3390/en5051355>
- [29] Xue, X. (2017) Prediction of Daily Diffuse Solar Radiation Using Artificial Neural Networks. *International Journal of Hydrogen Energy*, **42**, 28214-28221.
<https://doi.org/10.1016/j.ijhydene.2017.09.150>
- [30] Alzahrani, A., Shamsi, P., Dagli, C., Ferdowsi, M. (2017) Solar Irradiance Forecasting Using Deep Neural Networks. *Procedia Computer Science*, **114**, 304-313.
<https://doi.org/10.1016/j.procs.2017.09.045>
- [31] Bou-Rabee, M., Sulaiman, S.A., Saleh, M.S. and Marafi, S. (2017) Using Artificial Neural Networks to Estimate Solar Radiation in Kuwait. *Renewable and Sustainable Energy Reviews*, **72**, 434-438. <https://doi.org/10.1016/j.rser.2017.01.013>
- [32] Felice, M.D., Petitta, M. and Ruti, P.M. (2015) Short Term Predictability of photovoltaic Production over Italy. *Renewable Energy*, **80**, 197-204.
<https://doi.org/10.1016/j.renene.2015.02.010>
- [33] Zeng, J.W. and Qiao, W. (2013) Short-Term Solar Power Prediction Using a Support Vector Machine. *Renewable Energy*, **52**, 118-127.
<https://doi.org/10.1016/j.renene.2012.10.009>
- [34] Leva, S., Dolara, A., Grimaccia, F., Mussetta, M. and Ogliari, E. (2017) Analysis and Validation of 24 Hours Ahead Neural Network Forecasting of Photovoltaic Output Power. *Mathematics and Computers in Simulation*, **131**, 88-100.
<https://doi.org/10.1016/j.matcom.2015.05.010>
- [35] Du, P., Wang, J., Yang, W. and Niu, T. (2018) Multi-Step Ahead Forecasting in Electrical Power System Using a Hybrid Forecasting System. *Renewable Energy*, **122**, 533-550. <https://doi.org/10.1016/j.renene.2018.01.113>
- [36] Wang, J., Ran, R., Song, Z. and Sun, J. (2017) Short-Term Photovoltaic Power Generation Forecasting Based on Environmental Factors and GA-SVM. *Journal of Elec-*

-
- tical Engineering & Technology*, **12**, 64-71.
<https://doi.org/10.5370/JEET.2017.12.1.064>
- [37] Luo, X., Zhu, X. and Lim, E.G. (2019) A Parametric Bootstrap Algorithm for Cluster Number Determination of Load Pattern Categorization. *Energy*, **180**, 50-60.
<https://doi.org/10.1016/j.energy.2019.04.089>
- [38] Arthur, D. and Vassilvitskii, S. (2007) K-Means++: The Advantages of Careful Seeding *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete algorithms, Society for Industrial and Applied Mathematics*, Philadelphia, January 2007, 1027-1035.
- [39] Hu, Y.-C. (2007) Grey Relational Analysis and Radial Basis Function Network for Determining Costs in Learning Sequences. *Applied Mathematics & Computation*, **184**, 291-299. <https://doi.org/10.1016/j.amc.2006.05.162>
- [40] Vapnik, V., Golowich, S.E. and Smola, A.J. (1997) Support Vector Method for Function Approximation, Regression Estimation and Signal. In: Mozer, M.C., Jordan, M. and Petsche, T., Eds., *Advances in Neural Information Processing Systems* 9, MIT Press, Cambridge, 281-287.