AUT Journal of Electrical Engineering

AUT J. Elec. Eng., 49(2)(2017)187-194
DOI: 10.22060/eej.2017.12487.5077

Combination of Transformed-means Clustering and Neural Networks for Short-Term Solar Radiation Forecasting

M. Ghayekhloo[1*] and M. B. Menhaj[2]

[1] Young Researchers and Elite Club, Qazvin Branch, Islamic Azad University, Qazvin, Iran
[2] Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran

ABSTRACT: In order to provide an efficient conversion and utilization of solar power, solar radiation data should be measured continuously and accurately over the long-term period. However, the measurement of solar radiation is not available to all countries in the world due to some technical and fiscal limitations. Hence, several studies were proposed in the literature to find mathematical and physical models to estimate and forecast the amount of solar radiation such as stochastic prediction models based on time series methods. This paper proposes a hybridization framework, considering clustering, pre-processing, and training steps for short-term solar radiation forecasting. The proposed method is a combination of a novel data clustering method, time-series analysis, and multilayer perceptron neural network (MLPNN). The proposed Transformed-Means clustering method is based on inverse data transformation and K-means algorithm that presents more accurate clustering results when compared to the K-Means algorithm; its improved version and also other popular clustering algorithms. The performance of the proposed Transformed-Means is evaluated using several types of datasets and compared with different variants of K-means algorithm. The proposed method clusters the input solar radiation time-series data into an appropriate number of sub-datasets which are then preprocessed by the time-series analysis. The preprocessed time-series data provide the input for the training stage where MLPNN is used to forecast the solar radiation. Solar time-series data with different solar radiation characteristics are also used to determine the accuracy and the processing speed of the developed forecasting method with the proposed Transformed-Means and other clustering techniques.

Review History:

Received: 2 February 2017
Accepted: 7 March 2017
Available Online: 5 April 2017

Keywords:
Data Mining
Time Series Analysis
Forecasting
Solar
K-Means

1- Introduction

Solar radiation is the most important parameter in solar conversion, renewable energy, and especially photovoltaic (PV) systems [1, 2]. Different input data and forecasting models depend on the forecast horizon. Statistical models are appropriate for very short-term time scales ranging from 5 min up to 6 h [3]. Moreover, in [4] is presented an overview of different approaches to forecast solar irradiance.

The references [5-12] present the forecasting of photovoltaic (PV) power generation. These techniques can be classified into two types as follows:

First type is related to direct methods forecast according to the historical data, and associated weather information forecast PV power output. Support Vector Regression (SVR), ANN [10], and hybrid ANN methods [12] are the techniques used in the direct method forecasting.

Second type is related to indirect forecasting methods, where, historical solar irradiance and weather data are used, and, then, they are converted to the PV power output.

Some techniques adopted wavelet analysis [6], fuzzy logic method [5], artificial neural network [7], and hybrid Artificial Neural Network (ANN)-based methods [9]. In addition, ANN-based methods [11] are used for the classification and forecasting problems.

In [13], a 1-day-ahead hourly forecasting method based on the combination of a Self-Organizing Map (SOM) [12], a Learning Vector Quantization (LVQ) network [13], SVR [14] method is developed. In [15] it is shown that ANN techniques have a better accuracy than other techniques such as the fuzzy approach and the nonlinear and linear ones. In addition, the applications of ANNs in renewable energies has been presented in [16]. Similarly, Adaptive Neuro-Fuzzy Inference systems (ANFIS) and ANN models of Ground-Coupled Heat Pump (GCHP) systems have been reviewed in [17].

In addition, Multi- Layer Perceptron (MLP) is used in [18] to forecast the daily horizon (d+1) of global irradiation. Their proposed model is compared with AR, ARMA, k-Nearest Neighbors (k- NN) and Markov Chains approaches. Clustering techniques have a great role in unsupervised pattern recognition [19]. This technique classifies the groups of data separately in the field of renewable energy forecasting, providing a better understanding of collected information.

Furthermore, it improves the accuracy of the final forecast results. A new time series clustering technique for demand forecasting and renewable energy prediction is proposed in [20]. And also [21] developed a short-term PV generation forecasting model by employing SOM algorithms and wavelet neural networks. In [22], a method based on LVQ and the clustering is used to forecast solar irradiation.

In reference [23], a combined Nonlinear Auto Regressive (NAR) and K-means clustering are introduced to forecast hourly solar irradiance.

The main contribution of this paper is given in the following. A hybrid method to forecast the solar radiation is proposed, which includes clustering, preprocessing and training stages. A novel clustering method based on the inverse transformation

The corresponding author; Email: m.ghayekhlou@aut.ac.ir

of data is developed to provide more accurate clustering results. The proposed Transformed-Means clusters the input solar radiation data into an appropriate number of subsets which are then preprocessed by the time-series analysis. The preprocessed data provide the input to the training stage where Multilayer Perceptron Neural Networks (MLPNN) are used to forecast the solar radiation.

The rest of the paper is organized as follows. Section II provides a brief description of the K-means algorithm. It also explains the proposed clustering and the hybrid solar irradiance forecasting methods. Section III demonstrates a case study where the clustering errors are calculated for different algorithms. The performance of the developed forecasting method with different clustering algorithms is also evaluated in this section. Finally, section IV concludes the paper.

2- Methodology

A. Proposed clustering method

Transformed-Means is proposed in this section. The proposed clustering algorithm uses a combination of a new technique to select the initial cluster centroids and a new approach for the reverse transformation of the data in order for remedying the shortcomings of the existing K-means algorithms and providing a better performance. The steps of the Transformed-Means algorithm are described in the following.

(i) Selecting initial centroids

Let $X=[x_1,...,x_n]$ be a set of n data. The selection of K initial centroids is as follows.

1. Set a unique dataset $X`=[(x`_1,x_r),...,(x`_m,r_m)]$ where r_i is the repetition number for each non-repetitive data vector x_i ($1 \le i \le m \le n$).
2. Calculate the variance of the unique dataset X' by (1) and sort the data vectors in the dataset X' in ascending order based on the Euclidean distance between each data vector and the variance of these data.

$$\mathrm{var}(X`) = E[(x`-\mu)^2] = E(X`^2) - (E(X`))^2, \mu = E[X`]$$
$$\mathrm{var}(X`) = \sigma^2 \equiv \frac{1}{N}\sum_{i=1}^{N}(x`_i - \overline{x`})^2 \tag{1}$$
$$\text{where}: \overline{x`} = \frac{1}{N}\sum_{i=1}^{N} x`_i$$

The Euclidean distance between each data vector $x`_i$ in the new dataset X' and $\sigma_X^2 = [\sigma_{x1}^2,...,\sigma_{xd}^2]$, i.e. the variance of the input dataset X' in the d-dimensional space R^d, is given by:

$$d(x'_i, \sigma_X^2) = \sqrt{(\sigma_{X1}^2 - x'_{i1})^2 + \cdots + (\sigma_{Xd}^2 - x'_{id})^2} \tag{2}$$

$$\begin{aligned} X'_1 &= [(x'_1, r_1),...,(x'_P, r_p)],\\ X'_2 &= [(x'_{P+1}, r_{P+1}),...,(x'_{2P}, r_{2P})],\\ X'_3 &= [(x'_{2P+1}, r_{2P+1}),...,(x'_{3P}, r_{3P})],\\ &\ldots\\ X'_K &= [(x'_{(K-1)\times(P)+1}, r_{(K-1)\times(P)+1}),...,(x'_{KP}, r_{KP})].\\ X' &= \bigcup_{k=1}^{K} X'_k \end{aligned} \tag{3}$$

3. Divide the dataset X`, consisting of m data, into K sub-datasets, with at most $P=\lceil m/k \rceil$ data, according to (3) such that the data elements of X` are distributed among the sub-datasets $X`_1$ to $X`_k$.
4. Now, we have K sub-datasets, of which one is used to determine only one of the K initial centroids. (4) is used to consider a weight attribute $w(x`_i)$ for each data entry $x`_i$ with the repetition number r_i in each of K subdatasets $\{X`_1, X`_2,..., X`_k\}$.

$$w(x'_i)_m = \frac{1}{\frac{1}{P}\sum_{j=1}^{P} dist(x'_i, x'_j)}(r_i)_m, (1 \le m \le K) \tag{4}$$

where $w(x`_i)_m$ is the weight attribute for $x`_i$ in the m-th sub-dataset.

5. In each of the K sub-datasets, the data entry with the highest weight attribute is selected as the initial centroid.

As a result, in each of the K sub-datasets, a data point located at the densest concentration of the related sub-dataset is chosen as the initial centroid.

(ii) Inverse transformation

The inverse data transformation approach was first used in [24] to solve problems associated with the K-means clustering algorithm, see the Animator [25]. However, the approach presented in [24] suffers from a number of shortcomings such as the finding of a suitable artificial data structure, the performing of the mapping, and the controlling of the inverse transformations. This algorithm cannot generally guarantee an optimal result. In this paper, in some cases, the data transformation leads to the deviation of data towards the incorrect cluster centroids. In addition, this algorithm is not appropriate for clustering one-dimensional data sets and does not provide optimal results, and with regard to the evaluation results, it is preferable to the K-means algorithm. For the inverse transformation of data, first, we generate an artificial data X_* as the input data with the same sizes, say n, and dimension, say d. Therefore data vectors are divided into distinct clusters, with K centroids, without any fluctuations. Then, we represent a one-to-one mapping from the input data onto the artificial data ($X \rightarrow X_*$). In the approach presented in [24], each data, in the artificial structure, is randomly placed in an initial centroid placement, where the K initial cluster centroids are distributed uniformly along a line. This random placement may break the clustering structure and deviate the data towards incorrect cluster centroids, consequently providing incorrect results. This may also affect the stability of the final algorithm results. To address these problems, after determining initial centroids, all the real data elements are placed in the vicinity of the initial centroid. In the next step, the inverse transformation of the artificial data the main data is done by a series of inverse transformation that gradually moves data elements to their real positions. During this process, K-means updates the clustering model after any change. By this procedure, data vectors move slowly to their original positions without breaking the clustering structure. The clustering algorithm suggested in this paper employs a different approach that minimizes the reported problems of K-means star. Here, the procedure of gradual movement of data is as follows. First, each initial centroid $initC_l$ ($1<l\le k$) must be placed into the position of the data vector d_i, which

has the minimum distance to the corresponding data. Then, using the inverse transformation they move back from artificial structure to their real positions.

Generally, for a dataset D=[d_1,...,d_n] of n data vectors, the steps of gradual inverse transformation of data into their real positions obey the following procedure:

1. sort the real data vectors D=[d_1,...,d_n] in ascending order based on the Euclidean length of the vectors and store them into a new dataset,
2. to construct the artificial data structure, D^{art}, as the initial position of the data, place each initial centroid $initC_l$ ($1<l\leq k$) in the position of the data vectors of the dataset D`, closer to that initial centroid ($d`_i \rightarrow initC_l$) compared to their distance from other K-1 initial centroids. This forms the artificial structure $D^{art}=[d_1^{art},..., d_n^{art}]$ where $d^{art} \in initC$. By doing this, in the artificial structure D^{art}, real data move into the location of initial cluster centroid that is closer to the actual position of related data,
3. sort the real data vectors D=[d_1,...,d_n] in descending order based on the Euclidean length of the vectors and store them to the new dataset,
4. determine the distance between initial artificial data (D^{art}) and sorted real data (D″), and cast them in the set Dist″=[$dist_1''$, $dist_2''$,..., $dist_n''$], where each element $dist_i''$ represents the distance of the i-th data vector (d_i^{art}) in the artificial dataset D^{art} from the position of the corresponding data (d_i'') in the dataset D″,
5. according to the number of iterations given by the user (Steps>1), divide each element of Dist″=[$dist_1''$,...,$dist_n''$] by the value of steps and update the new values of the data elements in Dist″. Thus, we have:

$$\text{Dist}'' = [(\text{dist}_1'' / \text{Steps}),...,(\text{dist}_n'' / \text{Steps})] \tag{7}$$

6. steps of the inverse transformation of data f
7. or each iteration is given by (8):

$$\text{D}^{\text{art}}_{\text{itrNum}} = \text{D}^{\text{art}}_{\text{itrNum-1}} + itrNum \cdot \text{Dist}'' \tag{8}$$
$$(1 < itrNum \leq Steps)$$

where D^{art} is the position of data in the artificial structure, itrNum is the corresponding iteration number and Dist″ is the distance of the data sorted in descending order from the data positions in the artificial structure. We should note that D_1^{art} is equal to the initial data positions in the artificial structure (initial artificial dataset). After any transformation in the artificial data, K-means runs the previous centroid and the transformed artificial dataset D^{art} as the input. In the first step, the value of initC as the initial centroids, $initC_0$, obtained in the first phase of the algorithm is fed to K-means algorithm as its input. After completing all steps, the real data positions are replaced by the artificial dataset (in ascending order), and the final centroid C is the output result for final centroids. Figure 1 shows the flowchart for Transformed-Means.

B. Proposed Hybrid Solar Forecasting Method

This section develops a hybrid solar forecasting method that combines the time-series analysis, the Transformed-Means, a cluster selection algorithm, and MLPNN. Figure 4 shows the proposed forecasting method.

The proposed hybrid forecasting method consists of three stages of clustering, pre-processing, and training. Figure 2 shows the proposed forecasting method.

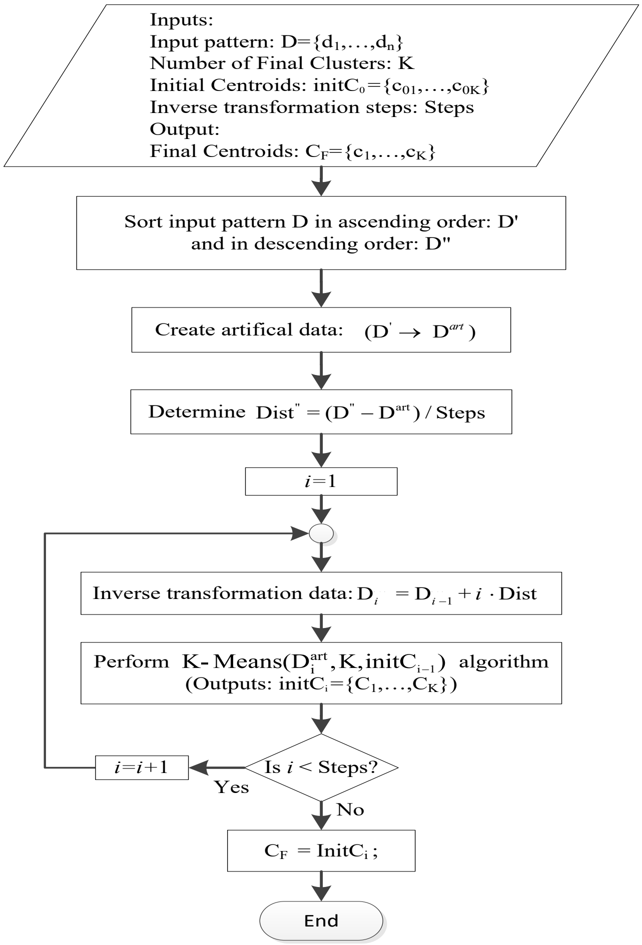

Fig. 1. Flowchart of Transformed-Means Algorithm

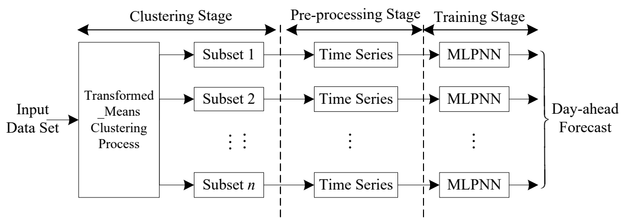

Fig. 2. Proposed Forecasting Method

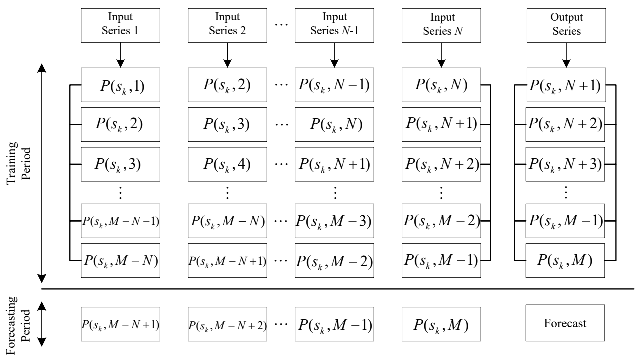

Fig. 3. Structure of the Input and Output Time Series for the MLPNN

The proposed Transformed-Means clusters the input data set into an appropriate number of subsets, say n. The method, below, is used to determine the number of clusters [26]. The pre-processing stage uses the time series analysis to provide the most appropriate inputs for MLPNN.

Figure 3 shows the structure of the time series for training and forecasting periods. $P(s_k,N)$ and M are the N^{th} data and the total number of data in the price subset k, respectively.

The average distance determines the subset used to provide the forecast for a specific hour. To this end, the distance between that hour's measurement and the average of the data within each subset is calculated for all clusters. The subset with the minimum distance is used in the pre-processing stage to provide the inputs for the MLPNN and forecast the price for a specific hour. The distance for subset k is calculated by:

$$\left|Average\left(P\right)-P_{meas.}\right|=\left|\left(\frac{1}{M}\sum_{m=1}^{M}P\left(s_k,m\right)\right)-P_{meas.}\right| \quad (9)$$

where $P_{meas.}$ is the price measurement.

3- Case studies

This section evaluates the accuracy of proposed clustering method. The datasets used in the experiment are available online [27-29]. More information regarding this data collection is presented in Figure 6. Mean Absolute Error (MAE) is used to determine clustering errors by

$$MAPE=\frac{1}{N}\sum_{n=1}^{N}\left|\frac{S(n)-S_{Actual}(n)}{S_{Actual}(n)}\right| \quad (10)$$

where N is the number of data points in cluster k, and $X_i^{(k)}$ is a data point in cluster k. The MAE values are calculated for different clustering techniques, namely the proposed Transformed-Means, K-means*, K-means++[30] , K-means, SOM [13], and GTSOM [31]. These data are given in Table I. The calculated MAE values show that Transformed-Means improves the quality of the clustering in comparison with other clustering methods.

A. Evaluation of proposed hybrid solar forecasting method

This section evaluates the developed hybrid solar forecasting method with the proposed Transformed-Means and other clustering techniques. The proposed forecasting is tested on several different solar datasets to provide a comprehensive performance analysis. MLPNNs are trained using 80 percent of the solar data and the remaining data are used for the test. MLP with three layers provide the solar forecast for one hour-, twenty four hour-, and forty eight hour-ahead prediction. The hidden and output layers use tansig and purline functions. The number of inputs for each MLP is equal to the length of the lagging window N=15. The structure of the network consists of six hidden and one output neurons.

The datasets are from different stations with different solar radiation characteristics [20]. Mean Absolute Percentage Error (MAPE) is used as the accuracy performance indicator and is given by:

$$MAPE=\frac{1}{N}\sum_{n=1}^{N}\left|\frac{S(n)-S_{Actual}(n)}{S_{Actual}(n)}\right| \quad (11)$$

where N is the total number of hours. In (11), $\hat{S}(n)$ and $S_{Actual}(n)$ are the solar radiation forecast and the actual solar radiation for the Nth hour, respectively.

The performance of the proposed forecasting method with Transformed-Means is compared to the results of the same forecasting method with different clustering techniques. Table II provides MAPE values and the processing time of the proposed forecasting method using each clustering technique on June 27th, 2013. The accuracy performance of the forecasting method with the Transformed-Means algorithm is better than those with other clustering algorithms. The proposed forecasting method with Transformed-Means clustering has a faster processing time compared to the forecasting with K-means++ and K-means* and it competes with SOM and GTSOM methods. The forecasting processing time with Transformed-Means is higher than that of K-means algorithm. This is due to the number of steps that the proposed method executes the K-means algorithm sequentially for the inverse data transformation. Table III provides the performance indicators of the forecast with the proposed Transformed-Means for different forecast horizons (1-hour, 24-hour and 48-hour ahead) on January 2nd, 2013.

Figures 4 to 6 show the performance of the proposed forecasting method with Transformed-Means in different weather conditions of sunny, cloudy and rainy. A one-week interval is considered for each weather condition: 08/25/2013 - 08/31/2013 the sunny week, 12/01/2013 - 12/07/2013 the cloudy week, and 04/09/2013 - 04/15/2013 the rainy week. The figures demonstrate the efficiency of the proposed method to forecast the solar radiation with different characteristics and variations. It can also be inferred from MAPE values calculated for the sunny, cloudy and rainy weeks, that are 3.8560%, 6.1616%, and 7.7580%, respectively.

4- Conclusions

The hybrid forecasting method proposed in this paper consists of three stages of clustering, pre-processing, and training to predict solar radiations. Time-series analysis is used in the pre-processing stage to provide the most appropriate inputs for NNs. Then, MLPNN is used to train the NNs and forecast the solar radiations. The proposed Transformed-Means clustering provides the clusters that give a higher resolution in preparation for the input into an MLPNN. Training is done using the MLPNN which is then used to predict solar radiation for that particular hour. The performance of the proposed Transformed-Means is evaluated using several different datasets and compared with K-means, K-mean++, K-mean*, SOM, and GTSOM algorithm. Our results demonstrate the enhanced efficiency of the proposed clustering algorithm in comparison with that of other clustering techniques. The performance of the proposed forecasting method is evaluated using solar datasets with different characteristics and variations to determine the accuracy and the processing speed of the proposed forecasting method with Transformed-Means and other clustering techniques. The comparison demonstrates a significant improvement in the forecast accuracy.

Table 1. MAE Measures for Different Clustering Techniques

Dataset	Transformed-means	K-means*	K-means++	K-means	SOM	GTSOM
	MAE	MAE	MAE	MAE	MAE	MAE
IRIS	0.276	0.282	0.371	0.282	0.282	0.284
Magic	0.057	0.059	0.059	0.059	0.06	0.059
Bridge	0.0302	0.0305	0.0312	0.0314	0.0323	0.033
Thyroid	0.164	0.171	0.171	0.176	0.171	0.17
Shuttle	0.0134	0.0178	0.0177	0.0177	0.0206	0.0281
Pendigit	0.0874	0.092	0.0938	0.0925	0.0922	0.0932
Yeast	0.0823	0.0871	0.0886	0.0882	0.0874	0.0902
M.Libras	0.0801	0.0825	0.0839	0.0821	0.0823	0.0828
Spambase	0.0079	0.0138	0.0101	0.0101	0.0096	0.0092
Ames.solar	0.0587	0.0602	0.0674	0.0736	0.0681	0.206
Average	0.0857	0.0896	0.0994	0.0913	0.0906	0.1056

Table 2. MAPE Measures and the Processing Time for the Proposed Forecasting Method with Different Clustering Algorithms

Dataset	Proposed Forecasting with											
	K-means		K-means++		K-means*		SOM		GTSOM		Transformed-Means	
	MAPE (%)	TIME (S)	MAPE (%)	TIME (S)	MAPE (%)	TIME (S)	MAPE (%)	TIME (S)	MAPE (%)	TIME (S)	MAPE (%)	TIME (S)
Calmar	9.1504	3.4183	14.0135	208.8255	13.5207	185.1354	4.8961	10.9337	9.7382	11.6271	4.3884	18.505
Ames	6.5181	3.284	5.4013	179.6701	12.7408	70.8081	5.1879	10.7655	7.6717	11.6987	5.683	17.3066
Castana	9.3901	6.5178	15.3609	167.8548	19.5436	85.6386	6.17	11.476	15.9441	17.1637	3.9738	18.7752
Cedar Rapids	6.0561	3.4172	15.0728	178.3335	20.7103	78.1649	6.5241	9.2266	18.1845	12.7802	6.0908	16.8523
Chariton	7.273	4.6392	8.5269	161.3657	7.3104	56.1865	4.8451	10.4359	9.7728	12.8222	4.7236	16.7573
Crawfordsville	11.066	4.0127	7.7033	137.9742	7.0627	74.1403	5.3383	12.9632	7.5057	12.2886	5.6961	19.9653
Gilbert	9.7371	4.2676	11.6963	137.5263	8.8621	75.741	4.4	13.067	4.0131	12.2541	3.9322	15.6152
Lewis	7.6904	5.5625	13.5705	141.4503	7.1955	56.3463	5.8014	9.5478	5.9993	11.578	5.0457	13.9992
Muscatine	15.336	3.5019	10.810	167.46	8.8842	58.154	8.2512	10.8511	12.958	12.735	3.9061	15.3278
Nashua	7.103	5.0785	7.1265	163.67	4.6086	57.221	6.8671	10.898	10.077	13.151	4.0297	16.4659
Sutherland	17.3624	4.5716	27.501	106.5261	7.9467	73.826	7.3993	10.655	12.1469	11.7123	5.2679	15.2902
Average	9.6985	4.2065	12.434	159.149	10.762	79.214	5.971	10.983	10.364	12.710	4.7034	16.8054

Table 3. Performance Indicators for Forecast with Transformed-Means and Different Horizons

Dataset	Forecast Horizon					
	1 hour		24 hour		48 hour	
	MAPE (%)	TIME (S)	MAPE (%)	TIME (S)	MAPE (%)	TIME (S)
Calmar	3.1491	7.8608	3.9026	9.019	7.081	8.809
Ames	4.9079	9.7218	7.2025	7.372	8.2645	7.5425
Castana	4.889	9.526	4.4832	8.1952	5.5724	9.7542
Cedar Rapids	4.6938	9.0656	4.5938	8.0892	5.8605	8.8066
Chariton	2.354	8.8705	3.6352	7.9829	4.9314	8.8969
Crawfordsville	3.4381	9.7724	5.5753	10.1221	5.0063	11.6164
Gilbert	3.9741	8.2319	3.786	8.2241	3.379	8.5127
Lewis	2.6104	7.8469	4.5813	7.9123	4.7763	9.9176
Muscatine	3.9652	8.9005	4.1495	7.9697	6.2414	8.6616
Nashua	4.3548	8.4111	4.1003	9.6412	4.7808	8.4708
Sutherland	3.7997	7.8414	4.0242	8.7442	6.3851	7.9678

5- Nomenclature

PV: Power system
SVR: Support vector regression
HDD: Heating degree days.
LVQ: Learning vector quantization
SOM: Self-organizing map
ANFIS: Adaptive neuro-fuzzy inference systems
GCHP: Ground-coupled heat pump
MLP: Multi-layer perceptron
KNN: K-Nearest Neighbors
NAR: Nonlinear autoregressive

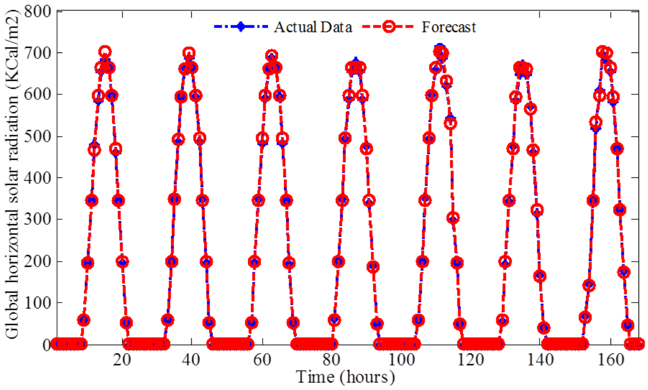

Fig. 4. Forecast Results for the Sunny Week (8/25/2013 - 8/31/2013)

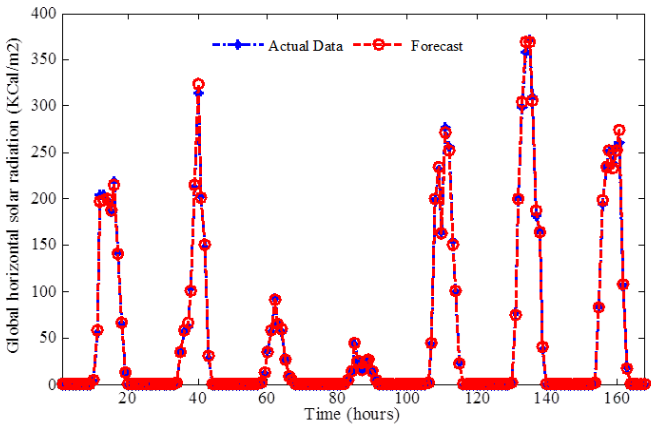

Fig. 5. Forecast Results for the Cloudy Week (12/01/2013 - 12/07/2013)

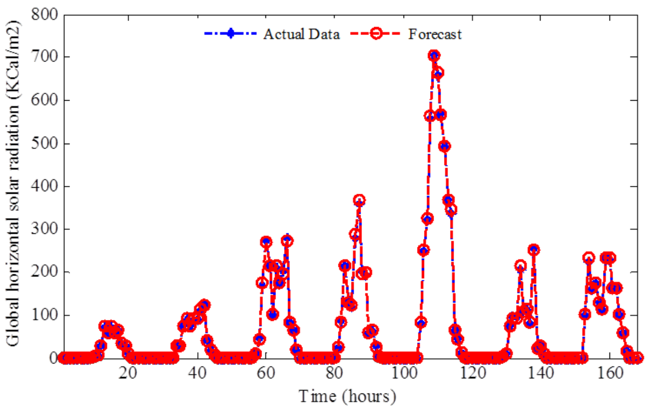

Fig. 6. Forecast Results for the Rainy Week (04/09/2013 - 04/15/2013)

References

[1] A. Mellit, M. Benghanem, A.H. Arab, A. Guessoum, Modelling of sizing the photovoltaic system parameters using artificial neural network, in: Proc. of IEEE, CCA, 2003, pp. 353-357.

[2] M. Nehrir, C. Wang, K. Strunz, H. Aki, R. Ramakumar, J. Bing, Z. Miao, Z. Salameh, A review of hybrid renewable/alternative energy systems for electric power generation: Configurations, control, and applications, IEEE Transactions on Sustainable Energy, 2(4) (2011) 392-403.

[3] E. Lorenz, D. Heinemann, Prediction of solar irradiance and photovoltaic power, (2012).

[4] A. Mellit, Artificial Intelligence technique for modelling and forecasting of solar radiation data: a review, International Journal of Artificial intelligence and soft computing, 1(1) (2008) 52-76.

[5] K. Tanaka, K. Uchida, K. Ogimi, T. Goya, A. Yona, T. Senjyu, T. Funabashi, C.-H. Kim, Optimal operation by controllable loads based on smart grid topology considering insolation forecasted error, IEEE transactions on smart grid, 2(3) (2011) 438-444.

[6] P. Zhang, Generation Scheduling for Supply and Demand Balancing in Power Systems with Renewable Power Generation, Kyushu University, 2013.

[7] A. Yona, T. Senjyu, T. Funabshi, H. Sekine, Application of neural network to 24-hours-ahead generating power forecasting for PV system, IEEJ Transactions on Power and Energy, 128 (2008) 33-39.

[8] S. Cao, W. Weng, J. Chen, W. Liu, G. Yu, J. Cao, Forecast of solar irradiance using chaos optimization neural networks, in: Power and Energy Engineering Conference, 2009. APPEEC 2009. Asia-Pacific, IEEE, 2009, pp. 1-4.

[9] G. Capizzi, C. Napoli, F. Bonanno, Innovative second-generation wavelets construction with recurrent neural networks for solar radiation forecasting, IEEE Transactions on neural networks and learning systems, 23(11) (2012) 1805-1815.

[10] J. Shi, W.-J. Lee, Y. Liu, Y. Yang, P. Wang, Forecasting power output of photovoltaic systems based on weather classification and support vector machines, IEEE Transactions on Industry Applications, 48(3) (2012) 1064-1069.

[11] T.-C. Yu, H.-T. Chang, The forecast of the electrical energy generated by photovoltaic systems using neural network method, in: Electric Information and Control Engineering (ICEICE), 2011 International Conference on, IEEE, 2011, pp. 2758-2761.

[12] S. Wang, N. Zhang, Y. Zhao, J. Zhan, Photovoltaic system power forecasting based on combined grey model and BP neural network, in: Electrical and Control Engineering (ICECE), 2011 International Conference on, IEEE, 2011, pp. 4623-4626.

[13] T. Kohonen, The self-organizing map, Proceedings of the IEEE, 78(9) (1990) 1464-1480.

[14] C. Cortes, V. Vapnik, Support-vector networks, Machine learning, 20(3) (1995) 273-297.

[15] A.K. Yadav, S. Chandel, Solar radiation prediction

using Artificial Neural Network techniques: A review, Renewable and Sustainable Energy Reviews, 33 (2014) 772-781.

[16] S.A. Kalogirou, Artificial neural networks in renewable energy systems applications: a review, Renewable and sustainable energy reviews, 5(4) (2001) 373-401.

[17] H. Esen, M. Inalli, A. Sengur, M. Esen, Artificial neural networks and adaptive neuro-fuzzy assessments for ground-coupled heat pump system, Energy and Buildings, 40(6) (2008) 1074-1083.

[18] C. Paoli, C. Voyant, M. Muselli, M.-L. Nivet, Forecasting of preprocessed daily solar radiation time series using neural networks, Solar Energy, 84(12) (2010) 2146-2160.

[19] M.S. Bobi, Use, operation and maintenance of renewable energy systems: Experiences and future approaches, Springer, 2014.

[20] N. Sengupta, S. Aloka, B. Narayanaswamy, H. Ismail, S. Mathew, Time series data mining for demand side decision support, in: Innovative Smart Grid Technologies-Asia (ISGT Asia), 2013 IEEE, IEEE, 2013, pp. 1-6.

[21] Y. Yang, L. Dong, Short-term PV generation system direct power prediction model on wavelet neural network and weather type clustering, in: Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2013 5th International Conference on, IEEE, 2013, pp. 207-211.

[22] R. Li, H. Wang, Y. Cui, X. Huang, Solar flare forecasting using learning vector quantity and unsupervised clustering techniques, SCIENCE CHINA Physics, Mechanics & Astronomy, 54(8) (2011) 1546-1552.

[23] K. Benmouiza, A. Cheknane, Forecasting hourly global solar radiation using hybrid k-means and nonlinear autoregressive neural network models, Energy Conversion and Management, 75 (2013) 561-569.

[24] M.I. Malinen, R. Mariescu-Istodor, P. Fränti, K-means∗: Clustering by gradual data transformation, Pattern Recognition, 47(10) (2014) 3376-3386.

[25] http://cs.uef.fi/sipu/clustering/animator/.

[26] D.J. Ketchen Jr, C.L. Shook, The application of cluster analysis in strategic management research: an analysis and critique, Strategic management journal, (1996) 441-458.

[27] http://cs.uef.fi/sipu/datasets.

[28] https://archive.ics.uci.edu/ml/datasets.

[29] http://mesonet.agron.iastate.edu

[30] D. Arthur, S. Vassilvitskii, k-means++: The advantages of careful seeding, in: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, 2007, pp. 1027-1035.

[31] J. Herbert, J. Yao, A game-theoretic approach to competitive learning in self-organizing maps, Advances in Natural Computation, (2005) 418-418.

Please cite this article using:

M. Ghayekhloo and M. B. Menhaj, Combination of Transformed-means Clustering and Neural Networks for Short-Term Solar Radiation Forecasting, *AUT J. Elec. Eng.,* 49(2)(2017)187-194.

DOI: 10.22060/eej.2017.12487.5077