G Model
ASOC 3705 1–10

Applied Soft Computing xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

A novel soft computing framework for solar radiation forecasting

Mahmoud Ghofrani [a,*], Mohadeseh Ghayekhloo [b], Rasool Azimi [b]

[a] School of Science, Technology, Engineering and Mathematics (STEM), University of Washington, Bothell, USA
[b] Young Researchers and Elite Club, Qazvin Branch, Islamic Azad University, Qazvin, Iran

ARTICLE INFO

Article history:
Received 2 October 2015
Received in revised form 13 June 2016
Accepted 14 July 2016
Available online xxx

Keywords:
Clustering
Data preprocessing
Forecasting
Soft computing
Solar radiation
T.S.C K-means

ABSTRACT

Accurate forecasting of renewable-energy sources plays a key role in their integration into the grid. This paper proposes a novel soft computing framework using a modified clustering technique, an innovative hourly time-series classification method, a new cluster selection algorithm and a multilayer perceptron neural network (MLPNN) to increase the solar radiation forecasting accuracy. The proposed clustering method is an improved version of K-means algorithm that provides more reliable results than the K-means algorithm. The time series classification method is specifically designed for solar data to better characterize its irregularities and variations. Several different solar radiation datasets for different states of U.S. are used to evaluate the performance of the proposed forecasting model. The proposed forecasting method is also compared with the existing state-of-the-art techniques. The comparison results show the higher accuracy performance of the proposed model.

1. Introduction

The number of solar energy plants has increased spectacularly in recent years. However, the intermittent and unpredictable nature of renewable energy such as solar and wind leads to defects on reliability and stability of power grid systems while increasing its integration and operational costs [1]. This requires strategies to enable integrating renewable energy sources with the traditional resources of energy production [2]. The need for forecasting energy generated by these plants has become a decisive factor in competitive power markets. Solar irradiance forecasting requires further investigation to provide more accurate and efficient methods. Therefore, we propose a hybrid forecasting framework to enhance the accuracy of predicting solar photovoltaic (PV) generation to reduce the associated costs to enable a more economic and reliable operation of electric power systems. Analyzing solar irradiance time series using pattern recognition techniques can significantly improve the accuracy of the forecast. Clustering analysis is one of the most common unsupervised machine learning methods used in pattern recognition. The main goal of clustering is to generate compact groups of objects or data that share similar patterns within the same cluster, and isolate these groups from those which contain elements with different characteristics. In the field of renewable energy forecasting, this technique allows handling groups of data separately, which provides a better understanding of the collected information and improves the accuracy of the final forecast results. Several clustering methods have been used to identify patterns and provide a pattern-based prediction technique for renewable energy resources such as solar irradiance [3–19]. In [3] an optimization of interval type-2 fuzzy integrators was proposed for the Mackey-Glass time series forecasting. ANFIS (adaptive neuro-fuzzy inferences systems) models and genetic algorithms (GAs) were combined for this purpose. Type-2 fuzzy logic was also used in [4] to better manage uncertainties of real data for enhancing the learning process of the backpropagation method and improving the forecast accuracy. Two novel evolutionary neural network approaches including sparsely connected evolutionary ANN (SEANN) and time lag feature selection EANN (TEANN) were proposed in [5] for time series forecasting. The SEANN approach provides more flexible ANN structures for multi-step ahead forecasting. The TEANN approach adjusts the ANN parameters with a set of time lags which are fed into the forecasting model. A new time series forecasting model was proposed in [6] which is based on the belief rule (BR) inference methodology. The efficiency of the proposed method depends on both the structure and the belief degrees. Akaike's information criterion (AIC) was used to provide the appropriate delay step to determine the structure. A novel approach based on neuro-fuzzy modeling was proposed in [7] for time series prediction. The training patterns are extracted after selecting the proper relevant variables. A set of Takagi-Sugeno-Kang (TSK) fuzzy rules is constructed based on the extracted training patterns. The

* Corresponding author at: UWBB room 227, 18807 Beardslee Blvd, Bothell, WA 98011, USA.
E-mail address: mghofrani@uwb.edu (M. Ghofrani).

http://dx.doi.org/10.1016/j.asoc.2016.07.022
1568-4946/© 2016 Published by Elsevier B.V.

related rule parameters are then refined by the learning method. The refined fuzzy rules are then used for time series forecasting. A hybrid of linear autoregressive integrated moving average (ARIMA) and nonlinear artificial neural network (ANN) models was used in [8] for one-step and multi-step ahead time series forecasting. A hybrid model of self-organizing maps (SOM), support vector regression (SVR) and particle swarm optimization (PSO) was proposed in [9] to forecast hourly global solar radiation. SOM algorithm was applied in the first step to divide the entire input space into several disjointed regions or clusters. SVR was then used to model each cluster for detecting characteristic correlation between the predicted and the actual values. In order to deal with the volatility of SVR with different parameters, PSO algorithm was used to improve the forecasting performance of the SVR models. However, the SOM algorithms may converge to non-optimal clustering results depending on the initialization and learning rate considered for the algorithm. In addition, neighborhood violations occur if the output space topology does not match with the data shape. A new model based on fuzzy set theory was developed in [10] for forecasting hourly global solar irradiation. A Fuzzy C-Means clustering was used for establishing the membership functions from the input variable's attributes. Performance and limitations of the model were tested using data from ten stations with latitudes between 40° and 50° north. It was concluded that the accuracy of the model is adequate for forecasting the performance of photovoltaic systems in this band of latitudes. A solar radiation forecast technique based on the fuzzy logic and neural network was proposed in [11]. The clustering process was improved by using fuzzy logic and the number of sky classes was reduced via the clustering characteristics of the uniform process with more intuitive interpretation. In [12] a new time series clustering technique was proposed to reduce the computational complexity of smart grid optimization problems in applications such as demand forecasting and renewable energy prediction. A simple spectral clustering was applied to the household's electricity load and supply data to capture their behavioral patterns. A hybrid forecasting technique was presented in [13] to predict the hourly power output for a photovoltaic (PV) system. The proposed hybrid method combined the generalized radial basis function network (GRBFN), deterministic annealing (DA) clustering technique, weight decay (WD) method and an improved version of the particle swarm optimization (PSO) algorithm for the prediction. SOM algorithms and wavelet neural networks were employed in [14] to develop a short-term PV generation forecasting model. The SOM was used for clustering the weather data to account for seasonal variations, and the wavelet neural network was utilized to build prediction models for each cluster sample. However, limitations of the SOM mentioned above still held true for this study.

A method based on clustering and learning vector quantity (LVQ) was presented in [15] to forecast solar flares. First, the K-means algorithm was used to cluster the dataset with a skewed data distribution. This resulted in more balanced training datasets for LVQ to build a training model that can classify data into class labels indicating whether the solar flare occurs in the next 24 h or not. However, the accuracy of the LVQ method is highly dependent on the initialization of the model and the learning parameters used as well as the class distribution of the training data. A combined K-means clustering and nonlinear auto regressive (NAR) neural network model was introduced in [16] to improve forecast accuracy of the hourly solar irradiance. A forecasting technique based on the K-means clustering algorithm was presented in [17] for wind and PV energy production. However, the initial cluster centroids randomly selected in the first phase of the algorithm may provide incorrect results for K-means clustering. A combination of an enhanced version of K-means algorithm and L-method approach was used in [18] to develop a new forecasting technique that provides maps to determine suitable locations for the deployment of solar power plants. An efficient combination of the affinity propagation (AP) and the K-means clustering algorithms was employed to address the sensitivity of the K-means to the initial cluster centroids. However, the inherent weakness of the AP method for not detecting the non-spherical clusters makes it more effective for the synthetic dataset (which are generally spherical shape) rather than the real dataset. K-means algorithm was combined with cumulative probability distribution functions to forecast the hourly solar irradiance using a daily clearness index [20]. The performance of the K-means based clustering is highly dependent on the initial cluster centroids. The random selection of the initial cluster centroids for the K-means algorithm may result in non-optimal solutions [20]. Different variants of k-means algorithm have been proposed to address this limitation [21–24]. A global K-means algorithm was developed to provide optimal solution for clustering problems [21]. K-means++ initialization algorithm was presented in [22] to obtain an initial set of centroids that is near-optimal. A combination of a new initialization technique, K-means algorithm and a data transformation approach was proposed to solve the K-means clustering problems [23]. A novel data clustering algorithm was introduced in [24] that combines granular theory and Fuzzy C-Means (FCM) clustering algorithm for more reflective data structuring.

In this paper, we propose a hybrid forecasting method that combines a time-series analysis, a novel clustering technique for solar time series data, a new cluster selection algorithm and MLPNNs to predict solar radiations. The paper includes two significances: improving clustering algorithm as an important method in the field of data mining and enhancing solar forecasting as a planning strategy for renewable energy integration. The hybrid forecasting method provides a soft computing framework with the following contributions:

1 An improved version of K-means algorithm, named time series clustering (T.S.C) K-Means, is proposed to provide fixed, definitive clustering results for different runs of the algorithm as opposed to different answers obtained for different runs of the original K-Means.
2 A modified initialization is proposed to select initial centroids which are closer to the optimum centroids' locations to solve the problem of empty cluster generation that frequently occurs due to the random initialization of K-means based clustering methods.
3 A new classification approach is developed to classify hourly solar data based on clustering results of the average daily data to better characterize irregularities and variations of solar radiation. This leads to deep learning of neural networks to provide more accurate forecasts as compared to direct clustering of hourly data. In addition, using the average daily data rather than the hourly data in the classification process significantly decreases the associated computational time.
4 After splitting the training data into multiple sub-trains by the proposed classification approach, a new method based on Principal Component Coefficients (PCCs) and correlation analysis is used to select the most appropriate sub-train as the input to the neural network. This significantly accelerates the forecast process as the most appropriate portion of the data rather than the whole data is used for the NN training. The proposed technique is particularly important for very short-term forecasting where the forecast horizon can be as short as a few seconds ahead.

The proposed clustering algorithm addresses the limitations of the K-means based clustering and provides a faster processing than the existing state-of-the-art clustering methods. It also offers more accurate clustering results than the original K-means algorithm and its variants on real time-series data such as solar data. The envisioned widespread application of the developed clustering in

Please cite this article in press as: M. Ghofrani, et al., A novel soft computing framework for solar radiation forecasting, Appl. Soft Comput. J. (2016), http://dx.doi.org/10.1016/j.asoc.2016.07.022