Solar Energy 144 (2017) 254–266

Contents lists available at ScienceDirect

Solar Energy

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

Comparisons among Bat algorithms with various objective functions on grouping photovoltaic power patterns

Amr A. Munshi [a,b,*], Yasser A.-R.I. Mohamed [a]

[a] Electrical and Computer Engineering Department, University of Alberta, Edmonton, Alberta, Canada
[b] Computer Engineering Department, Umm Al-Qura University, Makkah, Saudi Arabia

ARTICLE INFO

Article history:
Received 29 July 2016
Received in revised form 27 November 2016
Accepted 14 January 2017

Keywords:
Clustering
Bio-inspired algorithms
Photovoltaic systems
Power patterns

ABSTRACT

Photovoltaic power pattern clustering is fundamental in providing enhanced knowledge on the impacts of integrating photovoltaic systems into the electrical grid without extensive analysis and simulations. This paper investigates a set of clustering algorithms and validity indices to find the most efficient ones in grouping photovoltaic power patterns data. Furthermore, the introduction of the recently-developed bio-inspired optimization method, Bat, with various objective functions in clustering photovoltaic power patterns is presented. In order to evaluate the clustering results in a comprehensive manner, six internal validity indices are employed and a method to determine the optimum number of clusters is introduced. The clustering results on two datasets show that bio-inspired clustering algorithm Bat based on within-cluster-sum-of-squares to between-cluster variation (Bat WCBCR) as an objective function produces significantly high separated and well compact clusters.

1. Introduction

In recent years, generating power from photovoltaic (PV) systems has been of significant research interest. The motivations behind this are to overcome the financial and environmental issues of conventional power resources. For example, the unstable prices of fossil fuels, and the significant portion of environmental pollution and greenhouse gas emissions are two main concerns for industrialized countries. PV systems convert the abundant solar energy into electrical energy. The advances in PV technology such as the ability to generate electricity for a quarter of a decade with minimal regular maintenance and the continuous reduction in capital cost are encouraging factors to consider integrating PV systems into the electrical grid. The output power of PV systems is influenced by the irradiation level and the ambient temperature (Farivar et al., 2011). Also, fluctuations in the output power of PV systems could appear due to shadowing or power quality disturbance. For the purpose of studying the impacts of these output power fluctuations prior to the installation of PV systems in an efficient manner, extensive analysis and simulations using long historical data with sub-hourly time steps are required. However, dealing with such data is time consuming and computationally expensive. For this purpose developing solutions that can reduce the burden of extensive studies and simulations related to integrating PV systems into the electrical grid are of interest. For that, clustering techniques are employed to group PV power patterns (PVPPs) that have similar features. Hence, a representative power pattern from each group can be utilized in the simulations. Moreover, accurate clustering of PVPPs provides statistical information about the occurrence of particular patterns, and accordingly, assists in operating and planning for such systems.

Bio-inspired optimization methods mimicking swarm behavior have been recently developed. Although, each member of the swarm moves individually, the swarm as a whole collaborate together in order to achieve a common optimization objective. That is, each member stores its local best solution and updates it when a better solution is found. The behavior of swarm is adjusted according to the overall best member's solution (global best solution). Bio-inspired algorithms have been efficient in solving many optimization-related problems (Lee et al., 2011; Lee and Chong, 2011; Sun et al., 2014; Yang, 2013; Yang, 2011). However, applying such algorithms on power pattern problems has not gained much attention. Specifically, the application of Bat algorithms to PVPP clustering is not reported. For this purpose, Bat clustering algorithms with various objective functions are addressed in this paper to investigate their performance on clustering PVPPs.

In (Chicco et al., 2013) an Electrical Pattern Ant Colony Clustering (EPACC) algorithm has been applied to group load patterns of non-residential customers to assist the tariff formation process. The results of the EPACC outperformed the results of the

* Corresponding author at: Electrical and Computer Engineering Department, University of Alberta, Edmonton, Alberta, Canada.
E-mail address: aamunshi@ualberta.ca (A.A. Munshi).

http://dx.doi.org/10.1016/j.solener.2017.01.034
0038-092X/

K-means algorithm. In (Tang et al., 2012a; Tang et al., 2012b), several bio-inspired clustering algorithms have been applied on several real-world datasets from the University of California (UCI) machine learning repository (Blake and Merz, 1998). The objective function of these algorithms was based on the minimum squared error function. The results show that bio-inspired clustering algorithms achieve more accurate results than the classical K-means algorithm on most datasets. More precisely, the Cuckoo and Bat clustering algorithms were comparable and presented the best clustering results. The reason for choosing 'Bat' over 'Cuckoo' is due to the significant shorter time to converge to an optimized solution (Tang et al., 2012a). In addition, due to its efficiency, Bat algorithm has been applied in many areas such as, optimization, feature selection, data mining and others (Yang, 2013; Yang, 2011). PVPP clustering has been adopted in analyzing the output power fluctuation effects on interconnecting PV systems into the electrical grid (Omran et al., 2010) and for determining the optimal location and size of PV plants (Haghdadi et al., 2012; Zagouras et al., 2014). The previous research has been based on K-means (Omran et al., 2010; Haghdadi et al., 2012; Zagouras et al., 2014), Ward's Hierarchical (Omran et al., 2010), and Fuzzy C-means (FCM) Haghdadi et al., 2012. In (Mori and Takahashi, 2012) a Radial Basis Function Network model to predict short-time PV generation was proposed by clustering historical time series PV power data, then constructing a prediction model at each cluster so that the prediction is based on data similarity. Also, (Hosoda and Namerikawa, 2012) proposed a predicting model by clustering historical data and using the weather forecast. The results of Omran et al. (2010), Haghdadi et al. (2012), Mori and Takahashi (2012), Hosoda and Namerikawa (2012) showed that the results depend significantly on the accuracy of the clustered data. Thus, developing clustering algorithms that produce efficient partitioning of PV power data is in interest. In addition, the potential of PV in becoming a major power resource world-wide (European Photovoltaic Industry Association, 2013) motivates the investigation of applying various clustering techniques to investigate the most appropriate technique for clustering PVPPs.

In our previous research on clustering PVPPs (Munshi and Mohamed, 2016), various clustering algorithms from different clustering categories were applied to investigate the most appropriate algorithm for clustering PVPP data. The clustering algorithms included: K-means from partitional clustering, Ward's Hierarchical clustering from agglomerative clustering, FCM from fuzzy clustering, Self-organizing Maps (SOM) from neural network based methods, and Ant Colony and Bat from bio-inspired optimization methods. It was observed that the Bat clustering algorithm with the minimum mean square error function (J) as an objective function (Bat J) outperformed the other clustering algorithms and produced sufficient clusters. For this purpose, it is in our interest to use the Bat clustering algorithm with various objective functions to enhance the clustering formation of PVPP data. Therefore, in this paper five different versions of Bat algorithms are proposed as an attempt to enhance the clustering process of PVPPs.

This paper presents the results of a detailed investigation of the performance of Bat clustering algorithms based on various objective functions to establish the grouping process of PVPPs. The main scope of the paper is to enhance the clustering formation of the previous best clustering algorithm, Bat J, for clustering PVPP data. Also, comparing the performance of K-means and the various versions of Bat clustering algorithms, then find the best combination of clustering algorithm and validity index that can present the optimum number of clusters for PVPP data. The K-means has been included due to its extensive utilization in many applications and in clustering PVPP (Omran et al., 2010; Haghdadi et al., 2012). The effectiveness of the clustering algorithms is compared by using properly defined metrics and indicators. The clustering algorithm that produces well defined clusters can hence, provide more accurate results in the PV power systems simulations.

The remainder of this paper is structured as follows. Section 2 illustrates the general layout of the methodology. In Section 3, the clustering techniques are introduced. Section 4 presents the data dimension reduction technique. The results of applying the clustering algorithms and data dimension reduction techniques on two real PVPP datasets are presented and discussed in details in Section 5. A practical implementation using the PVPP representatives is briefly presented in Section 6. Finally, the conclusions are drawn in Section 7.

2. Methodology

The investigation of the most appropriate clustering technique for grouping PVPPs is achieved by applying a pattern recognition methodology on historical PVPP time series data. This historical data consists of daily patterns of irradiance and ambient temperature at a certain site for a couple of past years with an appropriate time resolution. The time resolution should be able to capture the short-term fluctuations in the irradiance and ambient temperature. The evaluations of the clustering results are compared based on the compactness and separation of the produced clusters. The general layout of the method is presented in Fig. 1 and the basic steps are discussed in the following subsections.

2.1. Data preprocessing

The irradiance and ambient temperature time series data are divided into segments where each segment represents a daily pattern. This data may be incomplete (missing observation values), and noisy (containing errors, or outlier values which deviate from the expected). Incomplete data can occur for a number of reasons: (1) Observations of interest may not always be available, such as the ambient temperature at a certain time. (2) Relevant data may

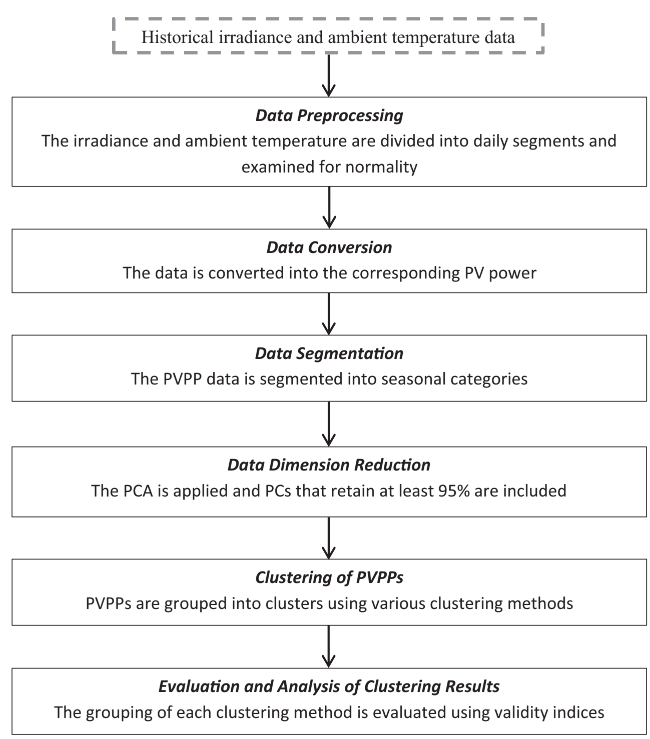

Fig. 1. General flow chart of the method.