Neurocomputing 211 (2016) 72–83

Contents lists available at ScienceDirect

Neurocomputing

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

Photovoltaic forecast based on hybrid PCA–LSSVM using dimensionality reducted data

M. Malvoni, M.G. De Giorgi*, P.M. Congedo

Dipartimento di Ingegneria dell'Innovazione, Università del Salento, via per Monteroni, 73100 LECCE, Italy

ARTICLE INFO

Article history:
Received 31 July 2015
Received in revised form
31 December 2015
Accepted 9 January 2016
Available online 11 June 2016

Keywords:
Photovoltaic forecast
Least squares support vector machines
Principal component analysis
Dimensionality reduction
Quadratic Renyi entropy

ABSTRACT

The power forecasting plays a significant role in the electrical systems. Furthermore the high-dimensional data reduction without losing essential information represents an important advantage in the forecasting models. Low computational costs and short execution time together with high predicted performance are the main goals to be reached in the development of a prediction method. In this paper a hybrid method based on an active selection of the support vectors, using the quadratic Renyi entropy criteria in combination with the principal component analysis (PCA), is shown to dimensionally reduce the training data in the forecasting models. The reduced data have been used to implement the Least Squares Support Vector Machines (LS-SVM) in order to predict the photovoltaic (PV) power in the day-ahead time horizon. The model has been validated using historical data of a PV system in the Mediterranean climate. Additionally the weather variations have been taken into account to evaluate the outcome of the sunny and cloudy condition in the PV forecasting models. The proposed technique gives fulfill results. A training data size same as 30% original dimension allows to improve the forecasting accuracy and reduces the computational time of 70% respect to an implementation without dimensionality reduction data.

1. Introduction

Nowadays the implementation of models for the prediction of the output of a renewable power plant is a popular research topic [1,2]. The generation systems prediction, as well the load forecasting, is essential into the electricity network. The predictions are the basis for an efficient scheduling and dispatch of the energy in the electric grid. Accurate forecasting methods reduce the operating costs and enhance reliability associated with the integration of renewable systems into the existing electricity grid.

In the literature intelligent forecasting methods as artificial neural networks (ANNs) have been widely used in developing wind and solar forecast models [3–6]. However, ANNs require a complex training process. In the recent years Kernel based estimation techniques, such as Support Vector Machines (SVMs) and Least Squares Support Vector Machines (LS-SVM), have been also applied as powerful nonlinear regression methods suitable for renewable power forecasts [7–9]. SVMs are more resistant to the over-fitting problem with respect to ANNs and permit to achieve high generalization performance in solving forecasting problems of various time series. LS-SVM is simpler and computationally less expensive, even if presents the same advantages of the ANNs and SVMs models [10].

In [8] ANN and LS-SVM were compared. Authors underlined that LS-SVM based models outperform ANNs in the prediction of the photovoltaic power.

Usually the PV power time series show important seasonal patterns (yearly, weekly, intra-daily patterns) that need to be taken into account in the modeling strategy. Furthermore the model is trained on the times series data, but when the available dataset is large, the learning model becomes more expensive in terms of time and computational resources.

The PV power is affected by weather and topographic factors, such as temperature, irradiance, humidity, which lead to large PV power variations. Furthermore the measured data could contain random fluctuations given by measurement errors and random factors. The meteorological fluctuations influence the prediction results with different impact on the forecasting accuracy if the weather is sunny or cloudy [8]. Previous works [8,11,12] applied the wavelet transform to reduce the noise contained in the data to be used in the forecasts. The weather data are well correlated, this permits the use of historical time series of these data to implement forecasting models. Nevertheless the redundant information can lead difficult modeling if the historical data, used as the model input, are a highly correlated. The principal component analysis (PCA) is one of traditional technique, extensively applied, to eliminate redundant information and improve the accuracy of the forecast of renewable power [13–15]. The PCA permits to extract essential

* Corresponding author.
E-mail address: mariagrazia.degiorgi@unisalento.it (M.G. De Giorgi).

http://dx.doi.org/10.1016/j.neucom.2016.01.104
0925-2312/

features and to reduce high-dimensional data into low-dimensional ones, which serve as inputs for the forecasting methods, as neural network or support vector machine, with a reduction of the CPU time. Some researchers used dimensionally reduction techniques in the support vectors for short term load forecasting [16] or wind prediction [17,18]. Despite its advantages, this technique has been rarely applied in the field of photovoltaic power prediction.

The main goals are to build a forecasting model that takes into account meteorological conditions together with the seasonal patterns of the PV power and to investigate the model accuracy when a dimensionally reduction is applied to the training dataset. The historical time series which consist of hourly values of the PV power, the module temperature, the ambient temperature and the plane of array irradiance, recorded for a period of about 1 year, are used to learn the model. In order to take into account the weather variation, as sunny and cloudy conditions, additional parameters are introduced in the learning set. This leads to an increase of the data set size. Furthermore, an efficient method to dimensionally reduce the learning data set is proposed based on the active selection of the recorded hourly samples in accordance to the quadratic Renyi entropy criteria and the implementation of the PCA to eliminate redundant information. The reduction technique is tested to predict the PV output power in a one-day head frame of a system located in a Mediterranean climate implementing the LS-SVM. A detailed analysis is carried out to evaluate the implemented models performance. The executive time and the computational complexity of the hybrid method are investigated.

This research will allow to improve the forecasting models accuracy of the PV power production in view of the computational resources needed to implement them.

This paper is organized as follows: the Section 2 describes the theory regarding the LS-SVM, the quadratic Renyi entropy, the PCA decomposition and the proposed procedure. In the Section 3 the dataset and problem statement are presented. The metric to evaluate the performance of model are illustrated in Section 4. The results of the method to dimensionality reduce are discussed in Section 5. Finally Section 6 states the conclusions of the paper.

2. Theory and methods

This section explains the theory that is the base of the proposed model in the present work and how it is applied to forecast PV power output in the proposed method.

2.1. Least square support vector machine

The Least-squares support vector machines allows to build a nonlinear representation of the original inputs using positive-definite kernel functions, based on a primal-dual formulation.

Given a training set of N data points, $\mathcal{D}_N=\{x_k, y_k\}_{k=1}^{N}$ where $x_k\epsilon\mathbb{R}^d$ is the k-th input data and $y_k\epsilon\mathbb{R}$ is the k-th output data, a regression model can be constructed using $\varphi:\mathbb{R}^d\rightarrow\mathbb{R}^s$ unknown function map of the input space:

$$y_k=w\varphi(x_k)+b \quad k=1\ldots N \tag{1}$$

where $w\epsilon\mathbb{R}^s$ is the weight vector and $b\epsilon\mathbb{R}$ is the bias term. The regression equation is transformed into an optimization problem with constraint that means to minimize a cost function:

$$\underset{w,e}{\text{mix}}\,\mathcal{J}(w, e)=\frac{1}{2}w^Tw+\frac{\gamma}{2}\sum_{k}^{N} e_k^2 \quad k=1\ldots N \tag{2}$$

where e_k is an artificial variable and γ is the regularization factor. In order to solve this optimization problem, the Lagrange function is defined as:

$$L(w, b, e, \alpha)=\mathcal{J}(w, e)-\sum_{k=1}^{N} \alpha_k\left\{y_k\left[w\varphi(x_k)+b\right]-1+e_k\right\} \quad k=1\ldots N \tag{3}$$

where $\alpha_k\epsilon\mathbb{R}$ are the Lagrange multipliers. The optimal conditions are:

$$\begin{cases} \dfrac{\partial L}{\partial w}=0 \rightarrow \mathrm{w}=\sum_{j=1}^{N} \alpha_j\varphi(x_j) \\ \dfrac{\partial L}{\partial b}=0\rightarrow \sum_{k=1}^{N} \alpha_k=0 \\ \dfrac{\partial L}{\partial e_j}=0 \rightarrow \alpha_j=e_j\gamma j=1\ldots N \\ \dfrac{\partial L}{\partial \alpha_j}=0 \rightarrow y_j=w\varphi(x_j)+b+e_j \end{cases} \tag{4}$$

Applying the Mercer's theorem [19]:

$$\varphi^T(x_k)\varphi(x_j)=\mathrm{K}(x_k, x_j) \quad k, j=1\ldots N \tag{5}$$

The Radial Basis Functi on (RBF) is introduced as kernel function K and defined as:

$$\mathrm{K}(x_k, x_j)=\exp\left(-\frac{\|x_k-x_j\|_2^2}{\sigma^2}\right) \tag{6}$$

where σ is a tuning parameter. So, the solution in matrix notation is:

$$\begin{bmatrix} \Omega+\frac{1}{\gamma}I & 1 \\ I^T & 0 \end{bmatrix}\begin{bmatrix} \alpha \\ b \end{bmatrix}=\begin{bmatrix} y \\ 0 \end{bmatrix} \tag{7}$$

where $\Omega_{kj}=\mathrm{K}(x_k, x_j)$ is the kernel matrix and $\alpha=[\alpha_1,\ldots\ldots\alpha_N]^T$ are the variables in dual space. So, the approximate model of y is expressed in following form:

$$y(x)=\sum_{k=1}^{N} \alpha_k\mathrm{K}(x,x_k)+\mathrm{b} \tag{8}$$

So calculating α and b, it is possible obtain an estimation of y in dual representation.

2.2. Approximation of the feature map for LS-SVM dual representation

The solution of Eq. (8) permits to determinate α and b, it is need to solve a system of size NxN that is unfeasible when the number N is more large. Furthermore it is advantageous to have a finite-dimensional approximation of the feature map $\hat{\varphi}:\mathbb{R}^d\rightarrow\mathbb{R}^M$, $M\ll N$. So, an expression for the approximation of the feature map, based on the Nyström approximation [20–22], is given as follows:

$$\hat{\varphi}=\frac{\mathrm{M}}{\sqrt{\lambda_{i,M}}}\sum_{m=1}^{M} u_{mi,M}\mathrm{K}(x,x_m) \tag{9}$$

where $\lambda_{i,M}$ is the i-th eigenvalue and the $u_{mi,M}$ is the m-th element of the i-th eigenvector for a MxM kernel matrix with $i=1,\ldots$M. Hence, an M-approximation of y can be obtained replacing the expression for the approximation of the feature map given by Eq. (9) in Eq. (1):

$$\hat{y}_m=w\hat{\varphi}(x_m)+b \quad m=1\ldots M \tag{10}$$

The Eq. (10) requests a subsample of size M of the original training set.