SVMr-Based model for power prediction of a 20kWp grid-connected photovoltaic system

Moufida Bouzerdoum
Department of Electronics, Renewable Energy Laboratory (LER), University of Jijel,
Ouled Aïssa Zone, POB 98, Jijel, Algeria
m83.bouzerdoum@yahoo.com

Adel Mellit
Department of Electronics, Renewable Energy Laboratory (LER), University of Jijel,
Ouled Aïssa Zone, POB 98, Jijel, Algeria
a.mellit@yahoo.co.uk

***Abstract*— In this paper, a regression support vector machines (SVMr) for short-term power prediction of a grid-connected photovoltaic (GCPV) plant is described. The designed model is applied for a 20 kWp GCPV system at Trieste, Italy. An experimental database of output power is used to develop and verity the effectiveness of the model. Results show good ability of the model for short-term power prediction and have a practical value in practical applications.**

Keywords- photovoltaic systems; prediction; power output; support vector for regression(SVMr)

I. INTRODUCTION

The use of renewable energy is an essential step in modern life. It seeks to ensure for future generations a clean energy in a clean environment. Solar energy is the more powerful energy source because it is called to the more shared energy source which is the solar radiation. The prediction of solar energy is of great importance for many applications such as generation of electricity and providing portable clean water. Accurate prediction can greatly improve the performance of these systems.

The support vector machines are statistics learning tools introduced by Vapnik [1] in 1995. SVM are usually used in classification problems. The approach allows to define complex surfaces in spaces of large dimensions, with very concise representations. If traditional methods of learning based on the minimization of the training error (empirical risk), the main advantage of SVM is the possibility of determining an error (the risk structure) valid for validation. Based on the principle of SVM, the support Vector Regression (SVR) can treat problems of regression (linear or non linear). Recently several studies have been devoted to the use of SVR for function approximation and time series prediction [2, 3, 4].

In the rest of work, we will introduce some basic theoretical aspects of SVM for regression problems. Then, we will present and discuss the results of the application method on the prediction of power output of a 20kWp GCPV plant. We finish our study with a conclusion by making a summary of the main results obtained in this study

II. SUPPORT VECTOR MACHINE FOR REGRESSION (SVMR)

The basic idea of SVR prediction is described as follows:
Suppose we are given training data$\{(x_i, y_i)\}(i = 1,2, \ldots, n)$, where each $x_i \in R^n$ is the input vector with n dimension, $y_i \in R$ is the associated desired output value of x_i. The SVR algorithm is as follows [1, 5]: Given a positive real number ε, we find a function f such that:

– $f(x_i)$ returns a value that does not deviate of y_i more than ε ;
– f is "simple" (flat) as possible.

To find such a function, the SVR algorithm determines f as a linear function of the form:

$$f(x) = \langle w, \varphi(x)\rangle + b \quad (1)$$

Where $\varphi(x)$ is called the feature that is nonlinearly mapping from the input space x. w is the vector of the parameters (or weights) and b is a constant to be determined. To ensure the flatness of the function f, the standard weight $\|w\|$ is minimized. So the problem is to minimize this standard by ensuring that errors are less than ε and can be written

$$\min \frac{1}{2}\|w\|^2$$

$$\text{s.t } |y_i - \langle w, \varphi(x_i)\rangle - b| \leq \varepsilon, i = 1, \ldots, N \quad (2)$$

This description of the problem considers a linear function f that approximates all the examples with accuracy ε exists. In practice, this is not always the case. In the presence of excessive noise or outliers, it is more important to allow some errors. In this case, the concept of soft margin is used. It is to introduce slack variables ξ_i, ξ_i^* to make feasible the constraints of the optimization problem which becomes

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \quad (3)$$

ISBN : 978-81-925233-8-5

under the constraints

$$y_i - \langle w, \varphi(x_i)\rangle - b \le \varepsilon + \xi_i$$

$$\langle w, \varphi(x_i)\rangle + b - y_i \le \varepsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \ge 0, i = 1, \dots, N \qquad (4)$$

Where ξ_i and ξ_i^* respectively denote the positive and negative errors. The constant C > 0 is a hyper parameter be possible to adjust the compromise between the amount authorized error and the flatness of the function *f*. This formulation of the problem is to use an error function $|\xi|_\varepsilon$ called ε-insensitive of the form

$$|y - f(x)|_\varepsilon = \begin{cases} |y - f(x)| - \varepsilon, \text{for} |y - f(x)| > \varepsilon \\ 0 \qquad\qquad\qquad\qquad , \text{otherwise} \end{cases} \qquad (5)$$

The problem (4) is solved by minimizing the Lagrangian*L*given by:

$$L = -\frac{1}{2}\sum_{i,j=1}^{N}(\alpha_i - \alpha_i^*)\left(\alpha_j - \alpha_j^*\right)k\left(x_i, x_j\right) - \varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) \qquad (6)$$

whereα_i, α_i^*positives are Lagrange multipliers.
The weight of the model are determined by

$$w = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\,\varphi(x_i) \qquad (7)$$

and the model can be written as

$$f(x) = \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*)\,k(x_i, x) + b \qquad (8)$$

The bias parameter *b* can be calculated by the conditions of Karush-Kuhn-Tucker(KKT).
The function *k* is called kernel function. The most commonly used for SVM kernels are polynomial kernels, sigmoidal and radial basis function (RBF) defined as follows:

- Linear : $k(x, x') = \langle x, x'\rangle$;
- Polynomial : $k(x, x') = (\gamma\langle x, x'\rangle + c)^d$;
- Sigmoidal : $k(x, x') = tanh(\gamma\langle x, x'\rangle + c)$;
- RBF : $k(x, x') = exp\left(\frac{-\|x, x'\|^2}{2\sigma^2}\right)$

Her γ, *c* and *d* are kernel parameters

III. DATABASE DESCRIPTION

In The used power output data were collected from a 20 kWp GCPV plant installed on the roof top of the municipality of Trieste, Italy. We have used data from 29 January to 25 May 2009 to evaluate the performance of the prediction model (a sample is taken every 10 min).

A database of 3437 patterns, as shown in Figure1, has been divided into two parts: a set of 3079 (89,5%) patterns is used for training the SVM model, while another set of 358 (10,5%) patterns is used for testing and validating the proposed SVM model.

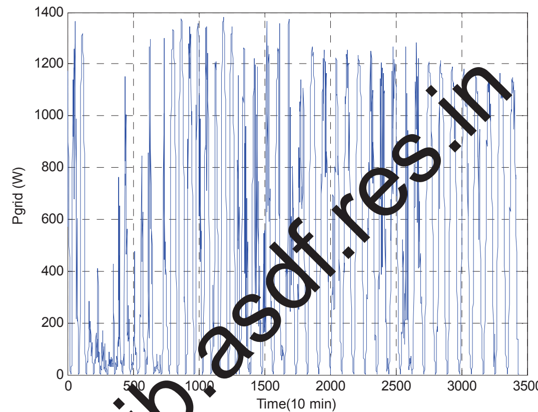

Figure 1. Power output data

IV. RESULTS AND DISCUSSIONS

Before applying the training algorithm, the data (input/output) should be normalized to [0, 1] using equation (9) [6, 7].

$$y_i^* = \frac{y_i - y_{min}}{y_{max} - y_{min}} \qquad (9)$$

Where y_i is the original data value, y_i^* is the corresponding normalized variable, y_{min} is the minimum values in $\{y_i\}$, y_{max} is the maximum value in $\{y_i\}$.

The procedure applied in the development of the SVMr model was as follows: first, data inputs are normalized as described before, and then training, test and validation sets were selected. After that, the parameters were chosen to create and train the models. Finally, the data were unormalized and the performance of the models was checked based on the error between the outputs values and the inputs ones.

The implementation of a SVM problem requires selecting the kernel function and the kernel parameter, the upper bound C and the ε-insensitive loss. The optimum parameters should be chosen.

Previous work has shown that Gaussian Radial Basis Function (RBF) gives better results in time series predictions than other parameters function [4, 8].The kernel SVM have a great affection the accuracy of the prediction. Different combinations of C, γ and ε were examined and the best combination of performance was selected. The values are determined by their tests as follows: C =0.0024, γ=0.021and ε=0.93.The optimum parameters were selected based on the lowest error in the validation step and makes SVM have high generalization.

Figure 2 shows the measured and the predicted power produced by GCPV plant using SVMr model for 4 days (from May 21 to May 24). As can be seen, the SVM model achieved high precision.

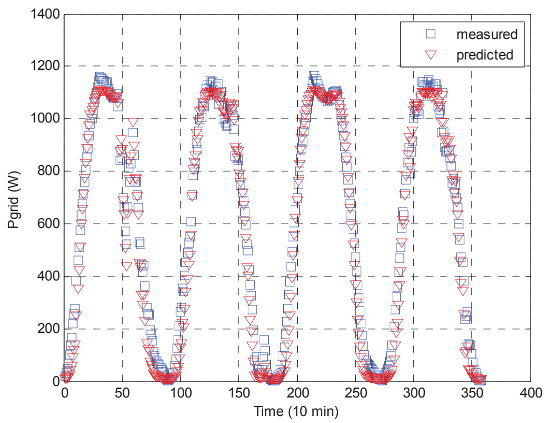

Figure 2. Measured versus forecasted power using SVMr

In order to evaluate the performance of the model which allows us to estimate and predict the power output, statistical tests, which are common for all types of prediction such as the NRMSE (Normal Root Mean Square Error), the NMBE (Normalized Mean Bias Error) and MPE (Mean Percentage Error) are calculated to compare the measured values to those calculated using the model [9, 10].

$$MPE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{P_{i,meas}-P_{i,pred}}{P_{i,meas}}\right) * 100 \quad (10)$$

$$MBE = \frac{1}{N}\sum_{i=1}^{N}\left(P_{i,meas} - P_{i,pred}\right) \quad (11)$$

$$NMBE = \frac{MBE}{\frac{1}{N}\sum_{i=1}^{N} P_{i,pred}} * 100 \quad (12)$$

$$RMSE = \left[\frac{1}{N}\sum_{i=1}^{N}\left(P_{i,meas} - P_{i,pred}\right)^2\right]^{1/2} \quad (13)$$

$$NMRSE = \frac{RMSE}{\frac{1}{N}\sum_{i=1}^{N} P_{i,pred}} * 100 \quad (14)$$

Where $P_{i,meas}$ and $P_{i,pred}$, are respectively the measured and predicted values of the power output at time i and N is the number of measured values (predicted).

Table I shows the values of the errors of the forecasts obtained with the SVM model developed to predict the power output.

TABLE I. STATISTICAL TEST FOR POWER OUTPUT

RMSE KW	NRMSE %	MBE	NMBE %	MPE %	R
59.954	9.851	-28.94	-4.755	-1.219	0.9946

From table I, we can see that the SVMr model performs well in PV power output prediction. The average prediction precision for 10 min ahead is 59.954 kW in RMSE and -28.939 in MBE with a coefficient of correlation of 99.46%. So the SVM model has successfully implemented with high precision.

V. CONCLUSION

In this work, a support vector machines for regression problem (SVMr) has been successfully developed for prediction (10min ahead) the power output produced by a 20 kWp GCPV system. Various statistical tests such as RMSE, MBE, and the correlation coefficient, were calculated to validate our model. On forecasts of power output, we can say with 59.954 kW in RMSE, -28.939 in MBE and a correlation coefficient of 99.46% our model is quite satisfactory and valid to forecast power output of the photovoltaic system. In order to improve the results, future research will focus on the use of hybrid models based on SVM and times series models.

ACKNOWLEDGMENT

Authors would like to thank Dr. A. Massi Pavan (Trieste Univ., Italy) for providing a database. This work was partially supported by the TWAS under grant Ref. 09-108 RG/REN/AF/AC_C: UNESCO FR: 3240231224.

REFERENCES

[1] V.Vapnik, The nature of statistical learning theory, New York: Springer-Verlag, NY, USA, 1995.
[2] J. Shena, Y.R. Syaub, and E.S. Lee, “Support vector fuzzy adaptive network in regression analysis”, Computers and Mathematics with Applications, vol. 54, no.11-12, pp 1353-1366, December 2007.
[3] B.R. Chang and H.F. Tsai, “Forecast approach using neural network adaptation to support vector regression grey model and generalized auto-regressive conditional heteroscedasticity”, Expert Systems with Applications, vol. 34, no. 2, pp.925-934, February 2008.
[4] Z. Xin and C.Tian-Lun, “Nonlinear time series forecast using radial basis function neural networks”, Commun. Theor. Phys.,vol. 40, no. 2, pp. 165-168, August 2003.
[5] A. J. Smola and B. Schölkopf, “A tutorial on support vector regression”, Statistics and Computing, vol.14, no. 3, pp.199-222, August 2004.
[6] C. Chen, S. Duan, T. Cai, B. Liu, “Online 24-h solar power forecasting based on weather type classification using artificial neural network”, Solar Energy, vol.85, no.11, pp. 2856-2870, November 2011.
[7] S. Jhi, W. Lee, Y. Liu, Y.Yang, and P.Wang, “Forecasting power output of photovoltaic systems based on weather classification and support vector machines”, IEEE transactions on industry applications, vol. 48, no. 3, pp. 1064-1069, May/June 2012.
[8] P. Yee, S. Haykin, “A dynamic regularized radial basis function network for nonlinear, nonstationary time series prediction”, IEEE Transactions on signal processing, vol. 47, no.9, September 1999.

[9] S. A. Kalogirou, "Artificial neural networks in renewable energy systems applications: a review", Renewable and Sustainable Energy Reviews, vol.5, no.4, pp. 373-401, June 2001.

[10] A. Mellit, S. A. Kalogirou, "Artificial intelligence techniques for photovoltaic applications: A review", Progressin Energy and Combustion Science, vol.1, no.1, pp. 52-76, November 2008.