Photovoltaic Output Prediction Using Auto-regression with Support Vector Machine

Anna Ogawa
Dept. of Open and Environmental Systems
Keio University
3-14-1 Hiyoshi, Kohoku-ku,
Yokohama, 223-8522, Japan
spice-up_your-life@z8.keio.jp

Akiko Takeda
Dept. of Mathematical Informatics
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo,
113-8656, Japan
takeda@mist.i.u-tokyo.ac.jp

Toru Namerikawa
Dept. of System Design Engineering
Keio University
3-14-1 Hiyoshi, Kohoku-ku, Yokohama, 223-8522, Japan
namerikawa@sd.keio.ac.jp

Abstract

This paper proposes a new method for short term photovoltaic (PV) power output forecasting. The method, which we call *auto-regression-with-exogenous-input using support vector regression* (ARX-SVR), uses auto-regression with exogenous input (ARX) and support vector machine (SVM). AR is a classical time series analysis method, while SVM is famous for its high prediction performance on non-time series datasets. By estimating parameters of ARX by SVM, we expect that *ARX-SVR* predicts more accurately PV power output because ARX catches a trend of the PV generation and SVM improves performance by using a historical dataset appropriately. Numerical experiments show that ARX-SVR outperforms not only AR and SVM but also existing models based on clustering.

1 Introduction

PV generation systems have become popular with the growing demand for renewable energy in terms of sustainability in energy industry. For energy management systems (EMS), many researchers have tried to achieve better prediction of renewable energy production, which is directly linked with the energy operating efficiency of the whole system. However, PV power output is difficult to forecast because it is influenced by weather and may change sharply within a short period of time.

Existing PV power output prediction methods can be classified roughly into three types: auto-regression (AR), neural network (NN) and support vector machine (SVM). AR is a classical approach of time series analysis that is used in, for example, financial engineering and signal processing. It has many variations with regard to error terms (ARMA), exogenous inputs (ARX), and so on. Such variants are used for predicting PV output in e.g., [10, 11, 13]. NN and SVM are popular supervised learning methods in machine learning. The NN model represents neural biological systems and models complex relationships between inputs and outputs by adjusting weights of the connections between elements. For predicting PV output by NN, existing works, e.g., [1, 5] used various meteorological factors as inputs. SVM has become popular in recent years for its high forecasting accuracy and small computational complexity. SVM has been used in PV power output forecasting models in [4, 7, 8, 12].

In this paper, we propose a novel PV output prediction method which constructs a time series model of PV power output based on AR and estimates its parameters by using SVM. As far as we have investigated, there is no existing method which combines AR and SVM for predicting PV power output. We expected that the combination of AR and SVM could well predict an 11-hour trend (from 8:00 to 18:00) of PV power output because AR catches a trend of the PV generation and SVM improves prediction performance by using a historical dataset appropriately. The combination of AR and SVM makes it possible to use a single prediction model for PV output prediction, though some existing methods [1, 5, 8, 13] need classifying the dataset into a few classes based on the weather type by clustering techniques and construct models for each to enhance their models' performance. Indeed, numerical experiments show that our model could predict PV power output better than such previous methods.

Section 2 describes the existing methods. Section 3 describes our model, ARX-SVR. Finally, Section 4 compares it with existing models in terms of prediction accuracy.

2 Existing Methods for PV Power Output Prediction

2.1 Auto-regression with Exogenous Input (ARX)

The auto-regression with exogenous input $ARX(p, q)$ is one of the most famous time series methods. This method is used for the output variable having high correlation with its own previous values. Using p dimensional historical input $P_{t-p}, \cdots, P_{t-1}$ (i.e., historical PV generation data) and $q+1$ dimensional exogenous input $T_{t-q}, \cdots, T_t$ (i.e., some meteorological factors), we can describe a relation between inputs and output as follows:

$$P_t = \textstyle\sum_{i=1}^{p} \varphi_i P_{t-i} + \sum_{j=0}^{q} \eta_j T_{t-j} + \varepsilon_t, \ \ t = 1, \ldots, l, \tag{1}$$

where $\boldsymbol{\varphi}$ and $\boldsymbol{\eta}$ are parameters, and the last term ε_t is white noise. The parameters are usually determined by using the least-squares method. Huang et al. [11] used ARMA (another kind of AR considering a time series of error terms $\boldsymbol{\varepsilon}$ instead of exogenous inputs $\boldsymbol{T}$) to predict the output P_t.

2.2 Support Vector Regression (SVR)

Support vector regressions (i.e., ε-SVR and ν-SVR in [3]) are popular supervised learning methods due to their high forecasting accuracy and small computational complexity. Suppose non-time series data samples $\{(\boldsymbol{T}_1, P_1), \cdots, (\boldsymbol{T}_l, P_l)\}$, where $\boldsymbol{T}_t \in R^n$ is an input (i.e., some meteorological factors) and $P_t \in R^1$ is a target output (i.e., PV generation data). The standard form of ν-SVR is:

$$\begin{array}{ll} \min\limits_{\boldsymbol{w}, b, \xi_t, \varepsilon} & \dfrac{1}{2}\|\boldsymbol{w}\|^2 + C(\nu\varepsilon + \dfrac{1}{l}\displaystyle\sum_{t=1}^{l} \xi_t) \\ \text{s.t.} & |(\langle \boldsymbol{w} \cdot \boldsymbol{\phi}(\boldsymbol{T_t}) \rangle + b) - P_t| \leq \varepsilon + \xi_t, \ \xi_t \geq 0, \ t = 1, 2, \cdots, l, \end{array} \tag{2}$$

where $\boldsymbol{\phi}(\cdot)$ is a nonlinear transformation to a higher dimensional space; C is a parameter for measuring the trade-off in complexity and loss; ν is a parameter which shows the ratio of support vectors among all samples; ε is a variable that expresses the permissible error; $\boldsymbol{\xi}$ is a slack variable. From the optimal solution $(\boldsymbol{w}^*, b^*)$, we can predict P from the prediction formula $\langle \boldsymbol{w}^* \cdot \boldsymbol{\phi}(\boldsymbol{T}) \rangle + b^*$ for the input $\boldsymbol{T}$. We can express not only a linear regression formula but also a nonlinear one using the nonlinear transformation $\boldsymbol{\phi}(\cdot)$. When we take the dual of (2), the terms $\langle \boldsymbol{\phi}(\boldsymbol{T_i}) \cdot \boldsymbol{\phi}(\boldsymbol{T_j}) \rangle$, $\forall i, \forall j$, appear in the objective function. By replacing the inner products $\langle \boldsymbol{\phi}(\boldsymbol{T_i}) \cdot \boldsymbol{\phi}(\boldsymbol{T_j}) \rangle$ with the kernel function $K(\boldsymbol{T_i}, \boldsymbol{T_j})$ (this is called the *kernel trick*), we can easily obtain a nonlinear prediction formula. Some of the popular kernel functions (i.e., RBF kernel and polynomial kernel) are shown in [2]. J. Shi et al. [8] used ε-SVR, which is equivalent to ν-SVR with corresponding ε and ν, for data samples whose input is temperature and whose target output is PV power output after classifying historical dataset into a few classes based on the weather type by a clustering technique.

3 Proposed Method: Auto-regression with Exogenous Input Using Support Vector Regression (ARX-SVR)

We estimate the parameters, $\boldsymbol{\varphi}$ and $\boldsymbol{\eta}$, of the time series model ARX by using SVR. Zhang et al. [9] combined another kind of AR (seasonal AR, SAR, where seasonal term is considered instead of the exogenous input term in (1)) and SVR for predicting avian influenza epidemics. In this model, the target output was expressed with SAR and the parameters were set by SVR.

As for the PV output prediction, we can achieve good performance by using ARX-SVR without clustering the dataset. Our model uses time series data as inputs, and therefore, the clustering technique may destroy the trend of time series data.

3.1 Formulation

We predict the PV power output P_t using 3 hour historical PV output, $P_{t-3}, P_{t-2}, P_{t-1}$, because current PV generation tends to have a strongly high correlation with previous 3 hour historical data. We also use the temperature T_t, that is, the predicted temperature, as an exogenous input for predicting the PV generation P_t. Our model for predicting the PV power output at time t is *ARX-SVR*$(3,0)$ shown below with parameters $\boldsymbol{\varphi}$ and η:

$$\hat{P}_t = \textstyle\sum_{i=1}^{3} \varphi_i P_{t-i} + \eta T_t, \tag{3}$$

where the parameters $\boldsymbol{\varphi}$ and η are determined by ν-SVR as follows:

$$\begin{array}{ll} \min\limits_{\boldsymbol{\varphi},\eta,\xi_t,\varepsilon} & \dfrac{1}{2}\left\|\begin{pmatrix}\boldsymbol{\varphi}\\ \eta\end{pmatrix}\right\|^2 + C(\nu\varepsilon + \dfrac{1}{l}\displaystyle\sum_{t=1}^{l}\xi_t) \\ \text{s.t.} & \left|\sum_{i=1}^{3} \varphi_i P_{t-i} + \eta T_t - P_t\right| \leq \varepsilon + \xi_t,\ \xi_t \geq 0,\ t = 1, 2, \cdots, l. \end{array} \tag{4}$$

The term $\sum_{i=1}^{3} \varphi_i P_{t-i} + \eta T_t$ shows the predicted PV power output and it represents the term $\langle \boldsymbol{w} \cdot \boldsymbol{\phi}(\boldsymbol{x})\rangle + b$ in the SVR formulation (2). The term P_t shows the actual PV power output and it represents the output y_i in SVR formulation (2). Figure 1 illustrates the ARX-SVR process.

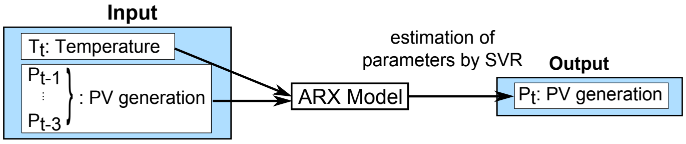

Figure 1: ARX-SVR process

4 Implementation

4.1 Data

To estimate $\boldsymbol{\varphi}$ and η in (3), we used PV generation data and the next day's temperature forecasts for the city of Yokohama, Japan, from the Japan Meteorological Agency [14]. The radiation angle and location of the PV system (on the Yagami Campus of Keio University) were fixed, and the PV installation capacity was $55[W]$. The time period was from 2012/02/01 to 2012/06/30 (no data for 04/14 and 04/27) with the interval of 1 hour. The daily radiation time was from 8:00 to 18:00.

4.2 Comparison with Existing Models

We used data that were obtained in the last 50 days before a certain point of time as the training data for building the models in numerical simulations and tested the performance of the models using the next day's data as test data. By using a rolling horizon (i.e., by shifting the point of time), we repeated the procedure of building models and testing the performance 96 samples. We calculated mean relative error (MRE), root mean square error (RMSE) (for details, see [8]) and the variance of RMSE (VAR) for such 96 test errors. In this paper, to estimate the parameters of ARX-SVR(3,0), we used the following basic parameter setting for (2), unless otherwise specified: $C = 1$, which is the best among $\{10^{-4}, 10^{-3}, \ldots, 1, 10^{1}\}$, $\nu = 0.6$, which is the best among $\{0.1, 0.2, \ldots, 0.9\}$, and in addition, $\gamma = 0.1$ for RBF kernel. Figure 2 shows RMSE of ARX-SVR(3,0) for linear kernel (dotted line) where $C = 0.01$ and RBF kernel (solid line) where γ is varied from 0.005 to 50. The figure implies that RBF kernel can achieve better prediction performance than linear kernel if a proper parameter γ is chosen.

4.2.1 Comparison of PV Output Prediction Models

We compared our method to existing methods; ARMA(2,3) [11], Clustering + AR [13], Clustering + SVR [8], and ARX-SVR(3,0). The hyper parameters were as follows:

- ARMA(2,3) [11]: The PV output was predicted using the PV historical dataset. The input dataset was 2 hour historical PV output and 3 hour residual errors.
- Clustering + AR [13]: The dataset was classified into three groups based on the weather forecast: sunny day, cloudy day, rainy day. AR model [13] was used for each groups. In this paper, we do not apply a Kalman filter to the AR model. If this filtering technique is needed, we can also apply it to our proposed model.
- Clustering + SVR [8]: Similarly to the above, the dataset was classified into three groups based on the weather forecast. We followed [8] and used ε-SVR with the RBF kernel. C was 1 and ε was 10^{-4}. Each parameter was set to the best among $\{10^{-4}, 10^{-3}, \ldots, 1, 10^{1}\}$.

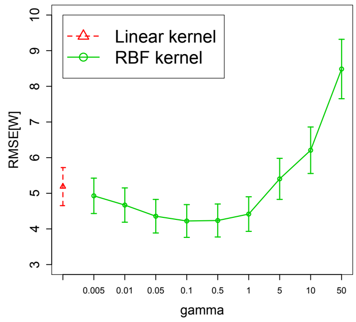

Figure 2: Comparison of kernel functions, RBF with varied γ vs Linear.

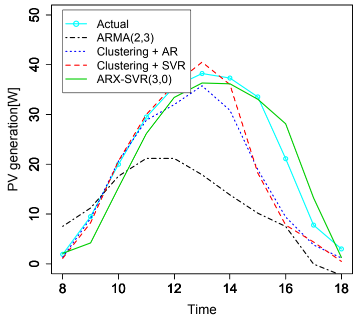

Figure 3: PV power output prediction is relatively easy in a sunny day.

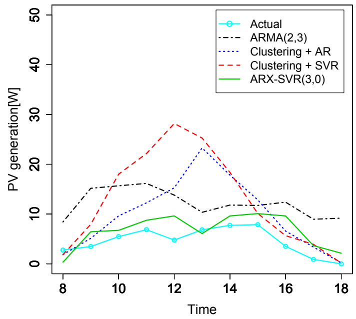

Figure 4: PV power output prediction is relatively difficult in a cloudy day.

Table 1 shows the averages of RMSEs, MREs and VARs among 96 samples for ARMA(2,3), Clustering + AR, Clustering + SVR, and our model: ARX-SVR(3,0). ARX-SVR(3,0) performed the best. In this way, our proposed method using SVR for a time series inputs achieved better performance than other methods using AR or SVR by themselves. Our model tends to predict the PV power output more accurately even on the days with ambiguous weather prediction (e.g., cloudy days or days that start overcast and end up with rainy weather) as Figures 3 and 4 imply.

Table 1: Comparison of four PV prediction models

	RMSE $[W]$	MRE $[\%]$	VAR $[W^2]$
ARMA(2,3)	8.302	12.788	43.088
Clustering + AR	5.819	7.550	30.090
Clustering + SVR	5.825	7.860	30.482
ARX-SVR(3,0)	**4.543**	**6.602**	**24.198**

4.2.2 Comparison of Parameter Estimation Methods

We compared the following parameter estimation methods for φ and η in (3) with the same input dataset that includes 3 hour historical PV output and predicted temperature; ARX(3,0) uses the least-squares method (LS), our model ARX-SVR(3,0) uses ν-SVR, and NN uses the backpropagation method. Table 2 shows the averages of RMSEs, MREs and VARs for PV power output prediction among 96 samples using the above mentioned three parameter estimation methods. This table shows that ARX-SVR(3,0) that uses ν-SVR as the parameter estimation method outperformed others and it could considerably reduce the volatility of RMSE.

Table 2: Comparison of three different parameter estimation methods

	RMSE $[W]$	MRE $[\%]$	VAR $[W^2]$
LS	9.164	14.399	59.238
ν-SVR	**4.543**	**6.602**	**24.198**
NN	9.548	9.278	2303.595

5 Conclusion

We proposed a new model for one-hour ahead PV power output prediction using ARX with SVM.

1. The ARX-SVR model is a mixture of classical time series analysis and popular machine learning approach; ARX expresses the trend of the current PV power output, whereas SVR improves the forecasting accuracy of the model.
2. Many existing methods classify the dataset based on the weather forecast before building the model. On the other hand, our model does not need to classify the dataset beforehand and outperforms some existing works.

The ARX-SVR model is expandable. Our model might predict the PV power output more accurately by considering another meteorological factors (e.g., adding seasonal term or another exogenous input using weather forecast).

Acknowledgment

This research was supported by the Japan Science and Technology Agency (JST, CREST).

References

[1] A. Yona, T. Senjyu, A. Y. Saber, T. Funabashi, H. Sekine and C. H. Kim, "Application of Neural Network to 24-hour-Ahead Generating Power Forecasting for PV System", Power and Energy Society General Meeting Conversion and Delivery of Electrical Energy in the 21st Century, IEEE, pp.281-297, 2008.

[2] B. Schölkopf and A. J. Smola, "Learning with Kernels", MIT Press, Cambridge, MA, 2002

[3] B. Schölkopf, A. Smola, R. Williamson and P. Bartlett, "New Support Vector Algorithms", Neural Computation, Vol.12, No.5, pp.1207-1245, 2000

[4] B. Wolff, E. Lorenz and O. Kramer, "Statistical Learning for Short-Term Photovoltaic Power Predictions", ECML PKDD, 2013

[5] C. Chen, S. Duan, T. Cai and B. Lin, "Online 24-h solar power forecasting based on weather type classification using artificial neural network", Solar Energy 85, 2856-2870, 2011

[6] F. P. Cruz, J. Weston, D. J. L. Hermann and B. Schölkopf, "Extension of the ν-SVM Range for Classification", IOS Press , pp.179-196, 2003

[7] J. Fonseca, T. Oozeki, T. Takashima, G. Koshimizu, Y. Uchida and K. Ogimoto, "Photovoltaic Power Production Forecast with Support Vector Regression: A Study on the Forecast Horizon", 37th IEEE Photovoltaic Specialists Conference, pp.2579-2583, 2011.

[8] J. Shi, W.J. Lee, Y. Liu, Y. Yang and P. Wang, "Forecasting Power Output of Photovoltaic Systems Based on Weather Classification and Support Vector Machines", IEEE Transaction on Industry Applications, Vol.48, No.3, pp.1064-1069, 2012.

[9] J. Zhang, J. Lu and G. Zhang, "A Seasonal Auto-regressive Model Based Support Vector Regression Prediction Method For H5n1 Avian Influenza Animal Events", International Journal of Computational Intelligence and Applications Vol. 10, No. 2 (2011) 199-230

[10] P. Bacher, H. Madsen and H. A. Nielsen, "Online short-term solar power forecasting", Solar Energy 83, 1772-1783, 2009

[11] R. Huang, T. Huang and R. Gadh, "Solar Generation Prediction using the ARMA Model in a Laboratory-level Micro-grid"

[12] R. Xu, H. Chen and X. Sun, "Short-term Photovoltaic Power Forecasting with Weighted Support Vector Machine", IEEE International Conference on Automation and Logistics, pp.248-253, 2012

[13] Y. Hosoda and T. Namerikawa, "Short-term Photovoltaic Prediction by using $H\infty$ Filtering and Clustering", SICE Annual Conference 2012, pp.119-124, 2012

[14] Japan Meteorological Agency. URL `http://www.jma.go.jp/jma/indexe.html`.