Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 48-30 (2015) 137–142

Application of Graphical Modelling to Selecting Input Variables for Solar Radiation Forecasting

Hiroyuki Mori*

* Department of Network Design, Meiji University
Tokyo, Japan (e-mail: hmori@meiji.ac.jp).

Abstract: This paper proposes an efficient method for selecting input variables of solar radiation forecasting. The proposed method makes use of Graphical Modeling that clarifies the dependence of output variables on input ones. It is useful to estimate a causal relationship between variables though the direct graph. Recently, photovoltaic systems are widely spread in power systems from a standpoint of reducing CO_2 emissions. However, it is often pointed out that they offer unstable generation conditions due to meteorological conditions. To overcome the problem, an accurate model is required for the forecasting method. In this paper, a Graphical Modeling method is used to determine the appropriate input variables of ANN in solar radiation forecasting. It is important to calculate realistic correlation between input and output variables by excluding superficial correlation that brings about erroneous results. The proposed method is successfully applied to real data.

Keywords: Forecasts, Neural network models, Graphical modeling, Partial correlation, Radial basis function network

1. INTRODUCTION

In recent years, global warming is one of main concerns in the world so that each country aims at reducing the emissions of CO_2. Specifically, the introduction of renewable energy into power systems have been widely spread. In renewable energy, PV (photovoltaic) systems are the mainstream in Japan. However, it is difficult to introduce them into power systems due to the drawback that the weather conditions affect the output of the generation output. Power system operators are faced with the problem of uncertainties. The use of efficient forecasting method helps them to deal with the uncertainties. This paper, handles solar radiation forecasting that is related to PV generation output. So far, a lot of works have been done on solar radiation forecasting as follows:

- Multiple regression model and ARMA (Autoregressive Moving Average) model (Goh and Tan, 1977; Moreno-Munoz, *et al.*, 2008)
- ANN (Artificial Neural Network) such as MLP (Multi-Layer Perceptron) (Ghanbarzadeh, *et al.*, 2009) and RBFN (Radial Basis Function Network) (Mellit, *et al.,* 2009)

The ARMA model is one of statistical techniques that are based on the stationary time-series process. It has a drawback not to follow the sudden changes for luck of adaptive function that comes from the fixed parameters. On the other hands, the ANN model is more flexible than the ARMA model due to the good performance of nonlinear approximation. MLP is one of ANNs that are often used in the engineering fields. The use of the sigmoid function is useful for evaluating the weights between the neurons. RBFN is based on the weighted sum of Gaussian functions. The conventional studies have shown that RBFN is better than MLP (Mori and Iwashita, 2005).

This paper focuses on GRBFN (Generalized Radial Basis Function Network) that is one of the extension models of RBFN in a way that the centers and the widths of the radial basis function are adjusted by the learning process although RBFN does not consider data-driven learning process (Poggio and Girosi, 1989; Mori and Takahashi, 2012). This paper proposes Graphical Modelling (Whittaker, 1990; Borgelt and R. Kruse, 2001) to improve the model accuracy of GRBFN by selecting appropriate input variables. Graphical Modelling makes use of the partial correlation coefficient through information theory so that realistic correlation is obtained. The conventional studies applied Graphical Modelling the following areas in selecting input variables of the model:

- Short-term load forecasting (Mori and Kurata, 2007)
- Carbon price forecasting (Mori and Jiang, 2008)

The conventional methods often used the correlation between input and output variables (Yona, *et al.*, 2007). However, it is pointed out that the correlation coefficient brings about erroneous results due to the existence of the pseudo correlation. It includes the influence of other inputs variables on the correlation between a certain input and output variables. Thus, it is necessary to exclude the influence in the correlation coefficients. The proposed method is successfully applied to real data of solar radiation.

10.1016/j.ifacol.2015.12.367

2. GENEALIZED RADIAL BASIS FUNCTION NETWORK

This section briefly describes GRBFN (Generalized Radial Basis Function Network)(Poggio and Girosi, 1989). It is based on RBFN that consists of a weighted sum of the radial basis functions. RBFN was extended to adjust the center and the width of the radial basis functions through the learning process. Now, let us explain RBFN that consists of three layers; input, hidden and output as shown in Fig. 1, where symbols $x_1, x_2, \cdots, x_n$ denotes the input variables, symbol y means the output variable and $w_1, w_2, \cdots, w_n$ implies the weights between the hidden and the output layers. Unlike MLP, RBFN has the feature the weights between the input and the hidden are set to be 1 and the nonlinear transform of the Gaussian function is made at the hidden layer. . The output of neuron i at the hidden layer may be written as

$$a(x)_i = \exp\left(-\frac{(x-\mu_i)^2}{\sigma_i^2}\right) \tag{1}$$

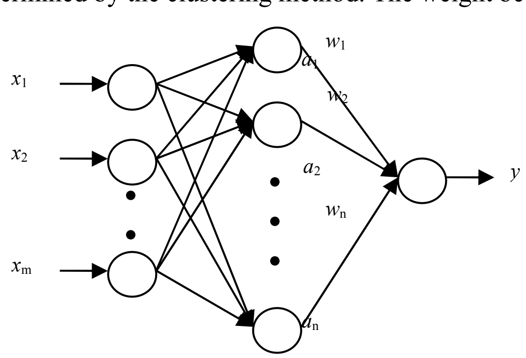
where,
a_i : output of the *i-th* unit at the hidden layer
x : input vector at the input layer
μ_i : center of the *i-th* unit at the hidden layer
σ_i : width of the *i-th* unit at the hidden layer

The output may be written as

$$y = \sum_{i=1}^{n} w_i a_i(x) \tag{2}$$

where,
y : output at the output layer
w_i : the *i-th* weight between the hidden and output layers
n : number of hidden units

The centers and widths of the Gaussian function are determined by the clustering method. The weight between the



Note) x_m: input data, w_n: weights of between input and hidden layers, a_n: output of hidden layer, y: output

Fig. 1. Structure of RBFN

hidden and the output layers may be evaluated by minimizing the following cost function:

$$f = \frac{1}{2}\sum_{j=1}^{l}\left(y_j - t_j\right)^2 \tag{3}$$

where,
y_j : output j of RBFN obtained by input pattern j
t_j : target j of learning data

GRBFN is different from RBFN is way that GRBFN has the learning process of the center and the width of the radial basis functions. The center and the width are determined by minimizing the cost function of (3) like the weights between the hidden and the output layers repeatedly. Thus, GRBFN outperforms the conventional RBFN due to the additional learning process.

3. GRAPHICAL MODELING

This section describes Graphical Modelling. It is one of the methods that give the structure of multiple variables through graph (Whittaker, 1990). It aims at classifying the relationship between specified two variables. In other words, it selects the appropriate covariance matrix that is more reasonable from a standpoint of statistics. Now, suppose the conventional method that gives the following correlation coefficients between variables:

$$r_{uv} = \frac{\sum_{i=1}^{N}(u_i - u_m)(v_i - v_m)}{\sqrt{\sum_{i=1}^{N}(u_i - u_m)^2}\sqrt{\sum_{i=1}^{N}(v_i - v_m)^2}} \tag{4}$$

where,
r_{uv} : correlation coefficient
u_i , v_i : sample data
u_m : average of u_i
v_m : average of v_i
N: number of data samples

However, if more than three kinds of samples exist, the correlation coefficients above brings about erroneous results due to the pseudo correlation that means overestimation or underestimation caused by other variables. In other words, it is not appropriate to use the correlation coefficient in selecting input variables in such a case. To overcome the drawback, Graphical Modelling focuses on the partial correlation coefficients in evaluating the relationship between two variables. Now, suppose Matrixes $R = (r_{mn})$ and $R^{-1} = (r^{mn})$ that denote the correlation coefficient matrix and its inverse, respectively. According to the principal of parsimony, the small partial correlation coefficients between variables should be ignored numerically. Namely, simple modelling is more preferable from a standpoint of statistics. Thus, it is necessary to remove them in evaluating the partial