Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 86 (2016) 273 – 276

Procedia Computer Science

2016 International Electrical Engineering Congress, iEECON2016, 2-4 March 2016, Chiang Mai, Thailand

A hybrid ARIMA and Neural Networks model for PM-10 pollution estimation: The case of Chiang Mai city moat area

Rati Wongsathan[a]*, Isaravuth Seedadan[b]

[a] *Electrical Engineering Department, North-Chiangmai University HangDong Chiang Mai Thailand 50230*
[b] *Electrical Engineering Department, North-Chiangmai University HangDong Chiang Mai Thailand 50230*

Abstract

Most of time series model are usually investigated and implemented by ARIMA and Neural Networks (NNs) model. However, ARIMA model may not be adequate for complex patterned problem while NNs model can well reveal the correlation of nonlinear pattern. Since, over-fitting due to a learning process is the main advantage of NNs as well as local trapped of parameters due to the large structure of the networks. To improve the forecast performance of both ARIMA and NNs for high accuracy, hybrid ARIMA and NNs model is alternate selected and employed to examine the Chiangmai city moat's PM-10 time series data. The experimental results demonstrated that the hybrid model outperformed best over NNs and ARIMA respectively.

Peer-review under responsibility of the Organizing Committee of iEECON2016

Keywords: ARIMA model; Neural Networks; Hybrid ARIMA-NNs; PM-10.

1. Introduction

People in Northern Thailand and nearby country such as Lao and Burma has annually faced and suffered from severe pollution related to particulate matter up to 10 micrometer or PM-10. Especially during the high season on February to April, the PM-10 level is exceed above mean PM-10 standard of 120 $\mu g/m^3$, specify by Thai government for daily PM-10 threshold monitoring [1]. The statistical of daily PM-10 during 2012-2014 is monitored in Chiangmai city moat area, which exhibit PM-10 level up to nearly 300 $\mu g/m^3$ and was illustrated in fig. 2. Wood burning is still the major evident on such time as was also found in Indonesia. Other factors normally come from increasing of private transport using in the city and industry also was found in Beijing, China.

* Corresponding author. Tel.: +66(81)2893400; fax: +66(53)8199998.
E-mail address: rati1003@gmail.com

Peer-review under responsibility of the Organizing Committee of iEECON2016
doi:10.1016/j.procs.2016.05.057

The measurement of PM-10 usually announces in the daily morning to warn people. However, like a weather forecast, the PM-10 forecast model should be implemented to predict PM-10 in advance. Traditional, most of the forecast model frequently used the historical values of PM-10 to estimate the current PM-10 value e.g. ARIMA (linear model) [2]-[3] and Neural Networks (NNs) (nonlinear model) [4]–[5]. Our previous work [4] has already done to predict PM-10 in the Chiangmai city moat area by using various NNs model. The accuracy resulted well, however the complexity of the model is the main problem which makes the model lack of efficiency learning and easy to saturate. The reliable of the forecast model not depends only on an accuracy result but suitable model structure. Further, forecast of PM-10 is not the easy task due to various statistically factors affected this value.

In this work, to solve the problem mentioned above, PM-10 forecast model is basically designed by ARIMA model to capture a suitable number of the historical values which are referred to the input of the model. Using this input number for a guideline of input node number of NNs, the task remains only tune the number of hidden node to yield the minimum mean square error (MSE). To improve the predictive performance of both ARIMA and ANN for high accuracy result, the theoretical and empirical findings have suggested an effective way by combining different models. The combining strategy of ARIMA and NNs is done to generate the hybrid model as hARIMA-NNs. The performance of the propose hybrid forecast model will be also assessed relative to an ARIMA and the NNs model.

2. Methodology and Method

2.1. An ARIMA Model

An ARIMA model typically consists of three parts i.e. auto regression AR(order p), moving average MA(order q) and differencing in order to strip off the integration of the series (order d) and then form ARIMA(p, d, q):

$$\Delta^d Y_t = \delta + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}\,. \tag{1}$$

Where Δ=(1-B), B refers to the backward shift operator for B(Y_t)=Y_{t-1}, Y_t is the observation data at time t, δ is the constant, φ_1, φ_2, …, φ_p are the autoregressive parameter, ε_t is the randomly error at time t and ~$N(0, \sigma^2)$, and θ_1, θ_2, …, θ_q are the moving average parameters.

A practical approach to building ARIMA model includes three iterative steps i.e. identification, parameter estimation, and diagnostic checking. By this approach, PM-10 data is nonstationary using preliminary investigarion by ACF (auto correlation function) and PACF (partial ACF) and unit root test by augmented Dickey-Fuller (ADF) test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. In the identification step, the differencing and power of data transformation is used to make the time series data stationary. The ACF and PACF of ΔPM identified the order q and p to 5 and 5 respectively, which yields ARIMA(5,1,5) model. After tentative models are identified, the set of ϕ and θ parameters are numerically estimated such that an overall measure of error is minimized and expressed in (2) at a 99% confidence interval level of statistics test. Diagnostic checking by several statistics assumption of the residuals such as Box-Pierce Chi-Square test verified that (2) is sufficient since no correlation of the residuals.

$$\begin{aligned}\Delta PM_t = {} & (0.405)\Delta PM_{t-1} + (0.171)\Delta PM_{t-2} - (0.064)\Delta PM_{t-3} - (0.808)\Delta PM_{t-4} + (0.326)\Delta PM_{t-5} \\ & - (0.600)\varepsilon_{t-1} - (0.268)\varepsilon_{t-2} + (0.034)\varepsilon_{t-3} + (0.864)\varepsilon_{t-4} - (0.583)\varepsilon_{t-5}\end{aligned} \tag{2}$$

2.2 A Neural Networks Model (NNs)

An NNs is regarded as multivariate, nonlinear and nonparametric method which can well reveal the correlation of nonlinear time series in delay state space. In this work, a feed-forward neural network type of Multi-layer perceptron (MLP) was selected to use as the forecast model. It typically consists of a three-layer i.e an input layer, a hidden layer and an output layer, and is shown in dash line box of fig. 1. The output of NNs is refered as predicting PM-10 at current time t and is weighted summation of each hidden layer neuron's output which can be expressed as

$$PM_t^{approx} = f\left(\mathbf{W}^{(2)} \times g\left(\mathbf{W}^{(1)} \times \mathbf{PM} + \mathbf{b}^{(1)}\right) + b^{(2)}\right), \tag{3}$$

where the number of input node corresponds with parameter p in ARIMA(p,d,q), D is the number of hidden node in the hidden layer, $\mathbf{W}^{(1)}$ and $\mathbf{b}^{(1)}$ is weight matrix and bias vector between input and hidden layer respectively, $\mathbf{W}^{(2)}$ and $b^{(2)}$ is the weight vector and bias value between hidden and output layer respectively, and **PM** is the input matrix