Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Forecasting of ozone concentration using frequency MA-OWA Model

CHING-HSUE CHENG[1]
Department of Information Management
National Yunlin University of Science & Technology
Douliou, Yunlin, Taiwan
chcheng@yuntech.edu.tw

SUE-FEN HUANG[2]
Department of Information Management
National Yunlin University of Science & Technology
Douliou, Yunlin, Taiwan
g9623801@yuntech.edu.tw

Abstract—**Air pollution can cause the human health, plants growth and daily mortality in numerous studies over the past decade. Therefore, forecasting and analysis of air quality are important topics of research today. The causes of poor air quality are global warming, greenhouse effects, acid rain, etc. Air pollution problems are related to the emissions of sulfur dioxides (SO_2), nitrogen dioxide (NO_2), suspended particulates (PM_{10}), ozone (O_3), carbon monoxide (CO), and unburned hydrocarbons (HC) and so on. O_3 adverse effects on human health has received intensively concern in recent years, It has been recognized as one of the principal pollutants that degrades air quality, thus we uses O_3 attribute to evaluate air quality.**

This study proposed a frequency moving average - order weight average (MA-OWA) model to forecast air quality by daily O_3 concentration. Due to O_3 data is belong to time series pattern, MA can simple calculation and OWA operator can aggregate multiple lag periods into single aggregated value by different situation parameters α. Frequency MA-OWA based time series model can efficiently and accurately forecast O_3. To demonstrate of the proposed, air quality monitoring the urban sites of Hsinchu (Taiwan), is selected for the numerical experiment from 2007 was utilized. From the results, the proposed methods outperform the listing methods in RMSE and MAPE.

Keywords—*Air pollution, Air Quality, ozone, moving average, order weight average.*

I. INTRODUCTION

Air pollution is one of the most serious problems in Taiwan. Automotive industry, the introduction of motorized vehicles, and the explosion of the population, are factors contributing toward the growing air pollution problem. Air pollution can have serious consequences for the health of human beings, and severely affects natural ecosystems.

The primary air pollutants found in most urban areas are carbon monoxide (CO), nitrogen oxides (NO_2), ozone (O_3), sulfur oxides (SO_2), hydrocarbons, and particulate matter (PM_{10}) and other toxins [1]. Focus of health studies and control efforts has increasingly turned to PM_{10} and O_3 as the most important air pollutant species of concern. In the past, the primary focus on the current understanding of the health is affected by PM_{10} and O_3 in the Taiwan.

In the early stages of development, major industry is small traditional manufacture firm in Taiwan, by which PM_{10} mainly produced. Last decade, the manufacturing industries have moved to oversea, thus PM_{10} concentration has been dropping gradually. Volatile organic compounds (VOCs) exhausted by vehicle combine with NO_x in sunlight recantation to form O_3 [2]. Taiwan environmental protection administration points out O_3 concentration gradually raised by vehicle of high density. Ozone is a serious pollutant can cause respiratory problems and other healthy conditions. Thus, monitoring and predicting air quality is an important topic.

This paper we adopt OWA operator to fusion lag periods with different weight for the attribute to match different situation. The proposed MA-OWA method can solve the time

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

series prediction problem, i.e., utilizes OWA based moving average to predict next period predictors, which can adjust each attribute weight by residual frequency for predict in different situation. The proposed method can provide an easily explanation and computation mechanism.

II. THE RELATED WORKS

This section briefly reviews the related literature, including two sections: literature reviews of time series model, OWA operator, and forecast accuracy.

A. Time Series Model

Predict standardized a lot of method, each method to different from relative accuracy predicted in a short time or a long time. Due to predict logic foundation not same, so analysis complexity different, therefore the predicted will be different in the future.

The time series is a group of observing value that the same variable according to the order happening. There are two kinds factors the time and the variable that corresponds to time, therefore can find out the variable to change the characteristic, trend and rule of development in the Time Series, thus the effective prediction to the future. The general Time Series contain four kinds: long trend variations, seasonal variations, cyclical variations, random variations. General quantitative analysis predicts the method such as simple arithmetic average, the weighting arithmetic average etc... The simple arithmetic average method is simple, the mean real data time in the past can next predict, suitable for using a short time predict. The weighting arithmetic average is that each real time offers different important degree to the past, this method can delete seasonal influence or regularity [3,4].

Most time series patterns can be described in terms of two basic classes of components trend and seasonality. There are two main goals of time series analysis: (a) identifying the nature of the phenomenon represented by the sequence of observations, and (b) predicting future values of the time series variable.

From the literatures, it indicates that the top tree quantitative predict techniques used are simple moving average, weighted moving average, and exponential smoothing. In this section, we introduce the basic definition as follows:

[Definition 1] Simple moving average (MA) forecasting model

$$F_{t+1} = \frac{\sum_{i=t-n+1}^{t} A_i}{n} \tag{1}$$

where F_{t+1} is forecast for Period t+1, n = number of periods used to calculate moving average, and A_i =actual demand in Period i.

[Definition 2] Weighted moving average (WMA) forecasting model

$$F_{t+1} = \sum_{i=t-n+1}^{t} w_i A_i \tag{2}$$

where F_{t+1} is forecast for Period t+1, n = number of periods used to calculate moving average, A_i = actual demand in Period I, and w_i =weight assigned to Period I (with $\Sigma w_i = 1$).

[Definition 3] Exponential smoothing forecasting model

$$F_{t+1} = F_{t+1} + \beta(A_t - F_t) \tag{3}$$

where F_{t+1} is forecast for Period t+1, F_t is forecast for Period t, A_i = actual demand in Period I, and β =a smoothing constant ($0 \leqslant \beta \leqslant 1$).

B. OWA Operator

The concept of OWA operators was first introduced by Yager in 1988. Many approaches have been proposed to calculate the weights based on OWA operators and apply this concept to many fields. In this section, we introduce the basic definition and some operations of OWA [5].

[Definition 4] An OWA operator of n dimension is a mapping $F: R^n \to R$, which has an associated weighting vector $w^*[w_1^* w_2^* ... w_n^*]^T$ and has the properties:

$$\sum_i w_i^* = 1, \forall w_i^* \in [0,1], i = 1,2,...,n$$
$$\text{and such that} \quad f(a_1,...,a_n) = \sum_{j=1}^{n} w_j^* b_j \tag{4}$$

where b_j is the jth largest element of the collection of the aggregated objects $a_1, a_2, \ldots, a_n$. Fuller and Majlender use the method of Lagrange multipliers to transfer equation (7) to a polynomial equation, which can determine the optimal weighting vector. By their method, the associated weighting vector is easily obtained by (8)-(9).

$$Orness(w^*) = \frac{1}{n-1}\sum_{i=1}^{n}(n-i)w_i^* \tag{5}$$

$$Disp(w^*) = -\sum_{i=1}^{n} w_i^* \ln w_i^* \tag{6}$$

$$Maximize \quad \sum_{i=1}^{n} w_i^* \ln w_i^*$$
$$Subject\ to\ \alpha = \frac{1}{n-1}\sum_{i=1}^{n}(n-i)w_i \tag{7}$$

$$\ln w_j^* = \frac{j-1}{n-1}\ln w_n^* + \frac{n-j}{n-1}\ln w_1^*$$
$$w_j^* = \sqrt[n-1]{w_1^{*n-j} w_n^{*j-1}} \tag{8}$$

$$w_1^*[(n-1)\alpha + 1 - nw_1^*]^n = [(n-1)\alpha]^{n-1}[((n-1)\alpha - n)w_1^* + 1] \tag{9}$$

$$if \quad w_1^* = w_2^* = ... = w_n^* = \frac{1}{n}$$
$$disp(w^*) = \ln n$$
$$w_n^* = \frac{((n-1)\alpha - n)w_1^* + 1}{(n-1)\alpha + 1 - nw_1^*} \tag{10}$$

Hence, the optimal value of w_1^* should satisfy equation (9). When w_1^* is computed, we can determine w_n^* from equation (10), and then the other weights are obtained from equation (8). In a special case, when $w_1^* = w_2^* = ... = w_n^* = \frac{1}{n} \Rightarrow disp(w^*) = \ln n$ which is the optimal solution for α = 0.5.

The parameter α can be treated as a magnifying lens for the optimistic decision makers to determine the most important attribute based on the sparest information (i.e. optimistic and α =0 or 1) situation. On the other hand, when α =0.5 (moderate situation), this method can get the attributes' weights (equal weights of attributes) for the pessimistic decision makers based on maximal information (maximal entropy).

III. PROPOSED METHOD

Time series prediction is the use of a model to predict future data points before they are measured. Time series prediction has been successfully used in several application areas, such as university enrolment forecasting [6,7], financial forecasting [8,9], air quality forecasting [10,11] temperature forecasting [12,13] etc., and a number of techniques have been developed for modeling and prediction time series. Moreover, such as a number of techniques calculate processing excessively complex. Such as simple moving average (MA) method, each attribute weight is equal, but it does not consider changing trend in the several near periods, therefore predicted value produce larger error. In weighted moving average (WMA) method, each attribute weight is different, but it does not consider changing each time order to the influence that predicted value which more attention near term. In exponential moving average (EMA) method, each attribute give different α degree, it can reduces the predict error, but it is α base on random or human subjective assign which, thus adjust the frequency α will lead to predict accurate rate.

For overcoming these drawbacks mentioned above, we adopt OWA operator to fusion lag periods with different weight for the attribute to match different situation. The proposed MA-OWA method can solve the time series prediction problem, i.e., utilizes OWA based moving average to predict next period predictors, which can adjust each attribute weight by residual frequency for predict in different situation. The proposed method can provide an easily explanation and computation mechanism.

We proposed method consists of the following major steps:

Step1: Data preprocessing and attributes selection

Step2: Determine the number of periods

Step3: Calculate the OWA weights

Step4: Determine each attribute weight ranking

Step5: Predict ozone by MA-OWA aggregate operator

Step6: Evaluation

Next, the detailed algorithm of the proposed method is introduced, the proposed method which consists of 6 steps (as Figure 1) is presented as follows:

Step1: Data preprocessing and attributes selection

In predict, selecting correct attributes is important step, and attribute selection is to discover a subset of attributes that are relevant for the target data mining task [14].

Step2: Determine the number of periods

Given the period number n, rank the important degree of period t. $(i.e., t_n > t_{n-1} > ..t_1)$.

Step3: Calculate the OWA weights

We have known the number of attribute, and use the algorithm of OWA as Equation (4)~(10); and situation parameter α to calculate OWA weights $w_1^* \sim w_n^*$.

Step4: Determine each attribute weight ranking

First-stage:

To calculate each period residual value, that is $r_1 = n_2 - n_1$, $r_2 = n_3 - n_2$, …, $r_i = n_i - n_{i-1}$, where r_n is the n period subtraction the $n-1$ period produced residual value.

Two-stage:

Define the range $R=(\max r_i - \min r_i)$ and partition into u intervals as follow: $u_1 = [x_1, x_2) = [D_{\min}, x_1 + d)$, $u_2 = (x_2, x_3) = (x_2, x_2 + d)$,…, $u_i = (x_n, x_{n+1}] = (x_n, D_{\max}]$, where $d = R/u$ denotes the uniform cut value and $D_{\min} \le x_1 \le x_2 \le \ldots \le D_{\max}$, $D_{\min}$ denotes the minimum value in the range R, $D_{\max}$ denotes the maximum value in the range R.

Three-stage:

To assign each period residual value to the corresponding interval (u_i), to calculate total counts (C_i) number, if $C_1 > C_2 > \ldots > C_i$ then this interval (u_i) is most critical, where C_1 is the highest frequency of total counts r_i.

Four-stage

For computing the aggregated value, we multiply the values of attributes ordering by the corresponding OWA weights. According to OWA produced set $C_1 = w_1$, $C_2 = w_2$,…, $C_i = w_i$ where w_1 is the OWA most important weight.

Step5: Predict ozone by MA-OWA aggregate operator

From four-stage, the each period actual data (A_i) weight is obtained. Hereby, to generate a forecast $n+1$ value, it can be expressed as follows:

$$F_{n+1} = w_i^* * A \tag{11}$$

Step6: Evaluation

The ultimate goal of any forecasting endeavor is to get high accurate and unbiased forecast. We use MAPE and RMSE as evaluation criterion for forecast error to identify the difference between actual value and the forecast. MAPE and RMSE are define as follows [15,16]:

1. Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} e_t^{\,2}}{n}} \tag{12}$$

2. Mean absolute percentage error (MAPE)

$$MAPE = \frac{1}{n} \sum \left| \frac{e_t}{A_t} \right| (100\%) \tag{13}$$

where e_t is the forecast error of period t, $e_t = \mathrm{A_t} - \mathrm{F_t}$, A_t presents the actual demand of period t, and n denotes the number of periods of evaluation.

IV. Numerical Analysis and Comparison

For verifying the proposed algorithm, we practically collect air quality datasets to illustrate the proposed method, and compare with the listing methods. The detailed demonstration is step by step introduced as follows:

Step1: Data preprocessing and attributes selection

We practically collect air quality data with 365 records of air quality inspection station dataset with O_3 attribute from January 1, 2007 to December 31, 2007, in Hsinchu city, Taiwan.

Step2: Determine the number of periods

We can select $3th$ lag period number.

Step3: Calculate the OWA weights

We can to calculate the OWA weights $w_1^* \sim w_3^*$, it is shown in Table 1. Given $n=3$, If situation parameter $\alpha = 0.7$ then $[w_1^*, w_2^*, w_3^*] = [0.5540,\ 0.2920,\ 0.1540]$.

Step4: Determine each attribute weight ranking

First-stage:

We can to calculate each period real data residual value. For example,

$A_1 = 21.3$, $A_2 = 20.5$, $A_3 = 18.4$, $A_4 = 25$

$r_1 = 20.5 - 21.3 = -0.8$, $r_2 = 18.4 - 20.5 = -2.1$

Two-stage:

According to residual value, we can obtain $R = 67.7$, and give $3th$ lag period number to partition range into 3 intervals. For example, $R = 67.7$, $u = 3$

$u_1 = [-31.7, -9.13)$, $u_2 = (-9.13, 13.43)$, $u_3 = (13.43, 36]$

Three-stage:

This step, assign each period residual value to the corresponding interval , we can obtain $C_1 = 47$, $C_2 = 273$, $C_3 = 14$, thus $C_2 > C_1 > C_3$.

Four-stage

According to Table 1., we can give situation parameter $\alpha = 0.7$ weight is $[w_1^*, w_2^*, w_3^*] = [0.5540,\ 0.2920,\ 0.1540]$ $C_1 \to w_2^*$, $C_2 \to w_1^*$, $C_3 \to w_3^*$.

Step5: Predict ozone by OWMA aggregate operator

Calculate forecasting value: for example, the aggregated value F_{71} by MA-OWA operator is:

$$F_{71} = 0.2920 * 22 + 0.1540 * 37.6 + 0.5540 * 15.1 = 20.5789$$

Step 6: The evaluated results are shown as Table 2. We can see that the forecast error of RMSE and MAPE are smaller than listing methods. That is, the proposed method outperforms the listing methods.

V. CONCLUSION

This paper has proposed a new time series based on MA-OWA method to predict air quality by daily maximum O_3 concentration. From the result (Table 2), we can see that the proposed approach is better than the existing methods in RMSE, MSE and MAPE. That is, the estimated accuracy rate of the MA-OWA is better than the listing methods.

From the air quality pollutant standards index, the air quality pollutes main is caused by O_3 in Taiwan. This phenomenon has already influenced the air quality of environment in Hsinchu City. Future research can test whether the air quality have influenced the nearby human health and the security under the high-tech operation of Hsinchu Science-based Industrial in Hsinchu city.

REFERENCES

[1] R. Afroz, M. N. Hassan and N. A. Ibrahim, Review of air pollution and health impacts in Malaysia, Environmental Research, 92, 2003, pp.71–77.

[2] S. I. V. Sousa, F. G. Martins, M. C. Pereira, and M. C. M. Alvim-Ferraz, Prediction of ozone concentrations in Oporto city with statistical approaches, Chemosphere, vol. 64, 2006, pp. 1141–1149.

[3] K. Huarng, and H. K. Yu, "The application of neural networks to forecast fuzzy time series", Physica A, 363, , 2006a pp.481-491.

[4] K. Huarng and H. K. Yu, Ratio-based lengths of intervals to improve fuzzy time series forecasting, IEEE Trans. on Systems, Man, and Cybernetics-Part B: Cybernetics, 36(2) , 2006b, pp.328-340.

[5] R. R. Yager, Including importance in OWA aggregations using fuzzy systems modeling, IEEE Trans. Fuzzy Syst., vol. 6, 1998, pp. 286–294.

[6] S.M. Chen, Forecasting enrollments based on fuzzy time series, Fuzzy Sets and Systems vol. 81, 1996, pp. 311-319.

[7] S.M. Chen and C.-C. Hsu, A New Method to Forecast Enrollments Using Fuzzy Time Series, International Journal of Applied Science and Engineering, vol. 2, 2004, pp. 234-244.

[8] T. L. Chen, C. H. Cheng, and H. J. Teoh, Fuzzy time-series based on Fibonacci sequence for stock price forecasting, Physica A, vol. 380, 2007, pp. 377-390.

[9] T. L. Chen, C. H. Cheng, and H. J. Teoh, High-order fuzzy time-series based on multi-period adaptation model for forecasting stock markets, Physica A, vol. 387, 2008, pp. 876-888.

[10] D. Wang and W.-Z. Lu ┌, Forecasting of ozone level in time series using MLP model with a novel hybrid training algorithm, Atmospheric Environment, vol. 40, 2006, pp. 913–924.

[11] J. Zabkar, R. Zabkar, D. Vladuˇsiˇc, D. C. emas, D. Sˇuc, and I. Bratko, Q2 Prediction of ozone concentrations, Ecological Modelling, vol. 191, 2006, pp. 68–82.

[12] H. K. Yu, Weighted fuzzy time-series models for TAIEX forecasting, Physica A, vol. 349, 2005, pp. 609-624.

[13] L. W. Lee, L. H. Wang, and S. M. Chen, Temperature prediction and TAIFEX forecasting based on fuzzy logical relationships and genetic algorithms, Expert Systems with Applications, vol. 33, 2007, pp. 539–550.

[14] J. Han and M. Kamber, Data Mining: Concepts and Techniques, 2006.

[15] K. Huarng and T. H. K. Yu, "The application of neural networks to forecast fuzzy time series," Physica A, vol. 363, 2006, pp. 481-491.

[16] H. K. Yu, "Weighted fuzzy time-series models for TAIEX forecasting," Physica A, vol. 349, 2005, pp. 609-624.

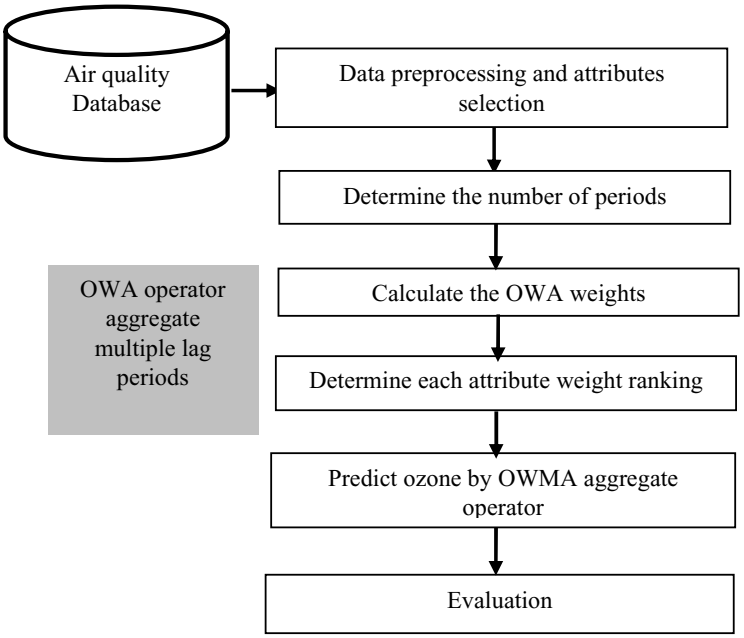

Figure 1. Research process

TABLE I. THE $w_1^* \sim w_3^*$ VALUES FOR DIFFERENT SITUATION

$n=3$	PARAMETER VALUES α					
	$\alpha=0.5$	$\alpha=0.6$	$\alpha=0.7$	$\alpha=0.8$	$\alpha=0.9$	$\alpha=1$
w_1^*	0.33333334	0.4383545	0.5539551	0.68185425	0.82629395	1
w_2^*	0.33333334	0.3232422	0.2919922	0.23583984	0.14697266	0
w_3^*	0.33333334	0.23839216	0.15399875	0.08189219	0.026305677	0

TABLE II. THE RESULTS OF THE MA, WMA AND ARMA FORECASTING MODELS FOR DATA

O_3	MA(3)	EMA $\alpha=0.2$	ARMA(3,3)	Proposed method MA-OWA(3) , $\alpha=0.7$
MAPE (%)	29.89	23.82	26.89	18.48
RMSE	8.3147	6.3567	7.4717	5.4737