Evaluation on Forecasting Algorithms of Time Series

Rongling Lang[1] Yongle Lv[3]
School of Electronic and Information Engineering
Beijing University of Aeronautics and Astronautics
Beijing, CHINA
ronglinglang@163.com

Qunfa Cui[2]
School of Mathematical Science
Anyang Normal University
Anyang, Henan, CHINA

***Abstract*—Time series forecasting exists in a lot of fields, and many forecasting algorithms were proposed at present. How to evaluate forecasting algorithm is requested in algorithm analysis. The techniques of evaluation of forecasting are studied here, and a novel measure, relative degree of association, is also proposed to evaluate forecasting algorithm. Some experimental results using these measures give performances of these measures in evaluating forecasting algorithms.**

Keywords-evaluation; forecasting; time series; metrics

I. INTRODUCTION

Time series exists in various fields extensively, such as industry, economy, finance, science observing and social science, etc. Therefore how to extract the potentially knowledge from these time series is an important problem. Some studies mainly focus on data mining of time series, such as trend analysis, similarity search, pattern mining and forecasting [1]. And also study on forecasting for time series is very important.

Forecasting would bring error inevitably. There are 3 kinds of errors brought by forecasting. One is induced by the indeterminacy of the problem, and it can not be eliminated. Another is induced by the half-baked messages, and the error can not be eliminated unless more information is available. The final one is induced by the forecasting algorithm itself and it is can be improved by choosing more suitable algorithms. What is the measure for choosing forecasting algorithm, namely how to evaluate the forecasting algorithm? There is not a universal method currently [2].

To solve the problem, evaluation techniques of forecasting algorithm are studied here. In Section II, several measures are dicussed, such as absolute error, relative error and error distribution. The degree of association such as correlative coefficient and degree of trend association are also discussed in this section. A novel comprehensive measure is proposed in Section III. And some experimental results are illustrated to be effective in section IV.

II. METRICS FOR ERROR OF TIME SERIES

Let $X=\{x_1,x_2,\cdots,x_N\}$ be a time series to be forecasted, and $\hat{X}=\{\hat{x}_i\},i=1,2,\cdots N$ is corresponding forecasted value of X. Therefore, $\varepsilon=\{\hat{x}_i-x_i\},i=1,2,\cdots N$ is the error sequence of the two series., As there are a lot of useful information for evaluation of the relationship between x_i and $\hat{x}_i$ in the error sequence ε, it is commonly used to evaluate forecasting algorithms.

There are many metrics to measure forecasting error, such as mean squared error (MSE), root mean square error (RMSE) and mean absolute error (MAE). They are defined as

$$MSE=\frac{1}{\mathrm{N}}\sum_{i=1}^{N}(\hat{x}_i-x_i)^2$$

$$RMSE=\sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{N}(\hat{x}_i-x_i)^2}$$

$$MAE=\frac{1}{\mathrm{N}}\sum_{i=1}^{N}|\hat{x}_i-x_i|$$

Although they are all capable of evaluation of the fluctuation of error series, and can be calculated easily, the result may be unfair for every point because of lackness of reference information. For example, different results must be got when x_i is 1 and 1000 respectively, even if $\varepsilon_i=\hat{x}_i-x_i=0.1$. Besides, they are sensitive to large-error, and MSE will magnify the error which is larger than 1. Therefore forecasting algorithm using above 3 criteria would be very sensitive to noise.

Relative error could improve error measurement of time series with lackness of reference information. $|x_i|$, $\|X\|$ and $\|\varepsilon\|$ are common estimated reference. Mean squared relative error (MSRE) and root mean square relative error (RMSRE) are most widely used relateive metrics. They are defined as

$$\mathrm{MSRE}=\frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left(\frac{\hat{x}_i-x_i}{x_i}\right)^2$$

$$RMSRE=\sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left(\frac{\hat{x}_i-x_i}{x_i}\right)^2}\ .$$

978-1-4244-4639-1/09/$25.00 ©2009 IEEE

As relative error is also sensitive to large-error, error distribution [3] is proposed for estimation of performance evaluation. The probability distribution function of error is appreciatively defined as

$$\mathrm{F}(\varepsilon < x \mid \hat{X}) = \frac{\delta}{\mathrm{N}}$$

δ is the number of points of ε that are larger than x and probability density function is defined as

$$f(\varepsilon \mid \hat{X}) = \frac{\zeta}{\mathrm{N}\Delta x}$$

ζ is the number of points of ε that are in the range of $[x, x+\Delta x]$.

The performance can be judged by checking the closeness between relevant error distribution and that of reference[3]. The correlation coefficient can be used for checking the closeness between two probability density function. Given $f(x)$, $f_R(x)$ be relevant error distributions of data and reference, so the correction coefficient is defined as

$$\rho = \frac{\int f(x) f_R(x) dx}{[\int f(x)^2 dx \int f_R(x)^2 dx]^{\frac{1}{2}}} \quad (1)$$

In equation, the reference must be gotten, but it is impossible in some cases. Therefore a new method using error distribution is proposed here.

For a better algorithm, more points of error sequence should located around zero. That means the value of $p = \mathrm{F}(\varepsilon <|x|| \hat{X})$ should be bigger. For example, there are two error sequences in figure 1, which are error sequences of forecasting algorithms. Let $|x| = 1.5$, $\sigma = 0.735$, and σ is standard deviation. p is 0.815 and 0.875 respectively in (a) and (b), so the algorithm with $p = 0.875$ is better.

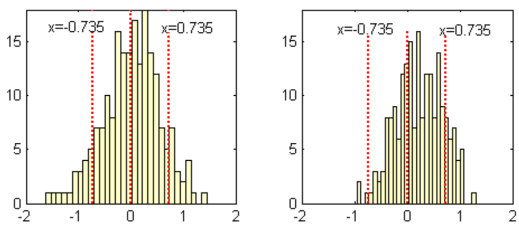

Figure 1. error distribution

There is shortcoming in despite of using any information of error alone. For example, in figure 2, RMSRE of these two algorithms are equal, and the value of $p = \mathrm{F}(\varepsilon <|x|| \hat{X})$ are also equal. However the result gotten from algorithm I gives better trend of X, and the effect can be improved by eliminating the unitary windage. The algorithm I should be better then algorithmII, but the result can not be got only by using information of error alone. Therefore the degree of association are needed to be considered, which can check the relationship between X and $\hat{X}$.

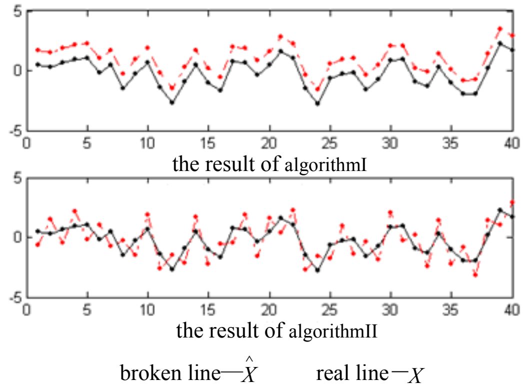

Figure 2. the result of forecasting

Correlation coefficient shows relationship between two random variables [4], so it can be chosen as a measure for evaluating forecasting algorithms. Let $a = \{a_1, a_2, \cdots, a_{n-k+1}\}$ and $b = \{b_1, b_2, \cdots, b_{n-k+1}\}$ be local trend sequences of two different time series [5], corrlation coefficient between above two sequnces give the local trend association between the corresponding time series..

Local trend association also have shortcoming when they are used alone. For example, when $\hat{X} = 100X$, the correlative coefficient $r(X, \hat{X}) = 1$, $\hat{X}$ is chosen as the forecasting value of X, but the forecasting is not proper. Therefore a comprehensive measure is proposed here which can integrate the information of error sequence and the relationship between time series.

III. Prepare Your Paper Before Styling

In time series forecasting, precise is always the most important. In common sense, the error sequence should be more concentrated around zero. In another word, forecasting sequence should have the same characteristics as original time series, namely the correlation between them should be bigger. This is because relativity expresses the feasibility of modifying the systemic error of forecasting algorithm. The systemic error can be modified by move parallelly or by other methods if the relationship between the series is bigger. Therefore the degree of relative sequence association is proposed which integrates the factors of error sequence and relativity.

Let $X = \{x_i\}$ and $Y = \{y_i\}$, $(i = 1, 2, \cdots N)$ are two time series, and $\varepsilon = \{x_i - y_i\}$, $(i = 1, 2, \cdots N)$. The degree of relative sequence association (DRSA) is defined as

$$\mathrm{D}RSA(X, Y) = \frac{e^{|SA|} + e^{\rho_\varepsilon}}{e^{\sigma_\varepsilon^2} + e^{m_\varepsilon^2}} \quad (2)$$

where SA is the measure of association between X and Y, ρ_ε describes the concentrative degree of $\varepsilon = \{x_i - y_i\}$ around zero, $\sigma_\varepsilon{}^2$ and m_ε are variance and mean of ε.

There are lots of merits when DRSA is used as measure for evaluating forecasting algorithm, such as:

- $m_\varepsilon{}^2$ indicates agonic property of an algorithm. Therefore DRSA accords with the rule that it is a better algorithm if the algorithm is more agonic.
- $\sigma_\varepsilon{}^2$ indicates the dispersive degree of ε, and ρ_ε indicates the concentrative degree of ε being around zero. Therefore DRSA expresses the idea of people that they want to control the error of the forecasting algorithm.
- SA indicates the relationship between the two sequences. Therefore DRSA accords with the rule that the forecasting sequence should reflect the unitary trend of original sequence.

IV. Experiment Results

An typical time series, exhaust gas temperature margin (EGTM) is used in experiment, which is an important parameter of aeroengine [6]. The original data is showed in figure3 (a). there are some noises and odd points in the sequence, which affects forecasting result. some preprocessing should be performed before forecasting. The odd points are removed by low order polynomial moving fitting [7], and the noise are smoothed by wavelet method, and preprocessed result is showed in figure3 (b).

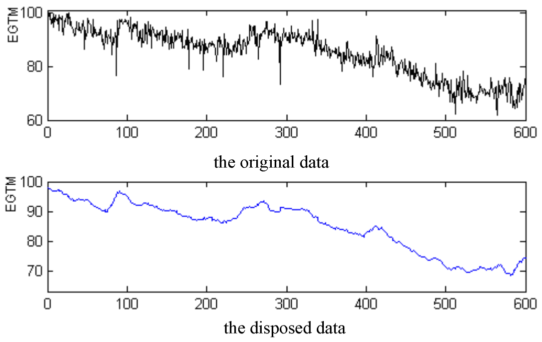

Figure 3. EGTM time series

The algorithms used in the experiment are autoregressive moving average (ARMA) [8] and R RBF Prediction Neural Network (BFPNN) [9].

TABLE I. THE RESULT OF ONE STEP FORECASTING

algorithm	MSE	MSRE (×10-5)	ρ	correlative coefficient
ARMA	0.0659	1.2062	0.870	0.9988
RBFPNN	0.0738	1.3459	0.845	0.9985
algorithm	**degree of trend association**	**DSRA[1]**	**DSRA[2]**	
ARMA	0.7615	2.467	2.189	
RBFPNN	0.7560	2.428	2.146	

TABLE II. THE RESULT OF FOUR STEP FORECASTING

algorithm	MSE	MSRE (×10-5)	ρ	correlative coefficient
ARMA	0.2955	5.3671	0.825	0.9940
RBFPNN	0.2458	4.4049	0.830	0.9960
algorithm	**degree of trend association**	**DSRA[1]**	**DSRA[2]**	
ARMA	0.4520	2.125	1.643	
RBFPNN	0.4910	2.205	1.731	

TABLE III. THE RESULT OF TEN STEP FORECASTING

algorithm	MSE	MSRE (×10-5)	ρ	correlative coefficient
ARMA	1.8507	32.903	0.815	0.9634
RBFPNN	1.4071	25.291	0.875	0.9752
algorithm	**degree of trend association**	**DSRA[1]**	**DSRA[2]**	
ARMA	0.4168	0.658	0.509	
RBFPNN	0.0589	1.108	0.759	

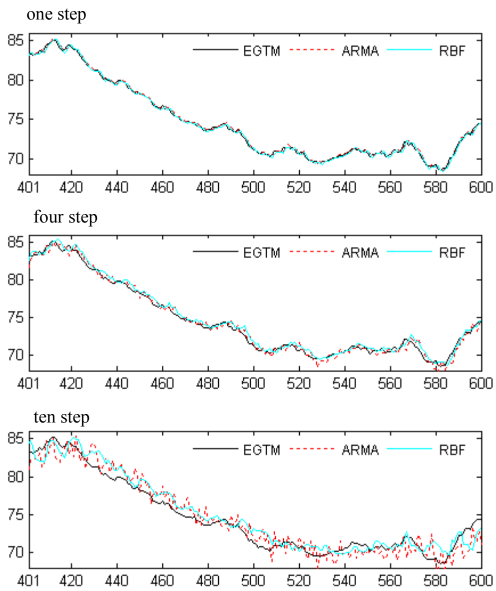

Figure 4. the forecasting result of EGTM

The forecasting results of ARMA and RBFPNN are showed in figure4. It shows that both algorithms can estimate the original time series and it is hard to do evaluation by naked eyes. Some evaluation metrics are illustrated in table 1, table 2 and table 3, where $\rho = F(\ \varepsilon < |x| |\hat{X})$, SA is correlative coefficient in DSRA[1], and SA is degree of trend association in DSRA[2], x is 1.5 times of standard deviation. It can be concluded that the performance of RBFPNN is better than that of ARMA in spite of the degree of trend association in step ten. It is because ARMA adapts to the linear and stationary situation, but it is not needed to limit the time series when RBFPNN is used.

V. CONCLUSIONS

Several metrics to evaluate the forecasting algorithm are discussed in this paper, such as absolute error, relative error, error distribution, correlative coefficient and degree of trend association .A new comprehensive measure is proposed here. Some experimental results using EGMT data are illustrated also. The conclusions can be got as follows:

- The measure of error adapts the situation in which the precision is more important, and there are not large-error.
- The degree of association adapts the situation in which the whole trend of the sequence is more important.
- The new comprehensive measure adapts the situation in which precision and whole trend both are needed to be considered.

The measures for evaluating the forecasting algorithm can not only be used to evaluate forecasting algorithms but also calculate the coefficient of combined prediction.

REFERENCES

[1] [1] KDnuggets: Polls: Time-Series Data Mining (Nov 2004). What types of time-series data mining you've done? URL:http://www.kdnuggets.com/polls/2004/time_series_data_mining.htm

[2] [2] CHEN You-hua, Entropy method and application to determine weights of combination forecasting [J], Journal of Anhui Univ. Natural Science Edition, 2003, 27(4): 1-6.

[3] [3] X. Rong Li, Zhanlue Zhao, Zhansheng Duan, Error Spectrum and Desirability Level for Estimation Performance Evaluation, ??????

[4] [4] Beilai Hu, Baogang Liu, Xihui Cao, Study on Time Series Forcasting by Neural Networks, Journal of Nankai Univ. Natural Science Edition, 1999, 32(3): 95-98.

[5] [5] Ildar Batyrshin, etc, Association Networks in Time Series Data Mining, Meeting of the North American, Fuzzy Information Processing Society, 2005, 754-759.

[6] [6] Yaoming Fu, Management of EGT Margin of Civil Turbofan in the Use and Maintenance [J], Aviation Maintenance & Engnering, 2005, 1: 44-45.

[7] [7] Xuguang Wu, system modeling and parameter estimating - theory and algorithm [M].China machine press, 2002.

[8] [8] E. P. B. George, M. J. Gwilym, C. R. Gregory, Time Series Analysis: Forecasting and Control [M], 3th Edition, Prentice-Hall, Inc., 1994.

[9] [9] Feng Liu,Chai Quek and Geok See Ng, Neural Network Model for Time Series Prediction by by Reinforcement Learning, IJCNN2005, IEEE: 809-814.