International Journal of Multimedia and Ubiquitous Engineering
Vol.10, No.7 (2015), pp.183-190
http://dx.doi.org/10.14257/ijmue.2015.10.7.19

Time Series Analysis Based on Improved Kalman Filter Model

Yang Zhizhong and Xi Bao

Harbin Institute of Technology, School of Management, Harbin, 150001, China
yscls2013@163.com

Abstract

In common time series analysis methods, the prediction accuracy of the low-order model is poor, and the high-order model is difficult to calculate. Therefore, in this paper, we improve the construction process of the Kalman filtering model, and apply it into time series analysis. The concrete implementation for the improved method is to construct the low-order model with the ARMA method and intercept sufficient delay states, to deduce the state equation and measurement equation of the Kalman filtering model. As the experimental results show that,the improved Kalman filtering model can not only simplify the derivation of the state equation and measurement equation, but also achieve ideal prediction accuracy, the largest prediction error of the experimental data is -0.15%.

Keywords: *Time series, Kalman filter, data prediction, data analysis*

1. Introduction

Time series data is the data gathered on different time nodes, with no clear meaning when appearing alone, but they can reflect the time-varying status or degree of some certain things when coming as series. The characterization range of time series data is large, such as astronomical phenomena data, industrial technical data, economic fluctuation data, which can all described with time series data [1].

The analysis of time series data can make analysis or evaluation on some certain things or phenomena, and on the other hand, it can make prediction on the future development of something. There are many methods for time series data analysis, such as regression analysis, correlation analysis, wavelet analysis, genetic algorithm analysis, *etc*.[2-3].When constructing the time series data model, we generally need not consider its background, in other words, ignore whether it is technical data or economic data. This is because that, time series has already provided enough information for constructing the model with its internal timing characteristics [4]. However, there are also several problems in present time series analysis methods, for the poor accuracy in the low-order models of this kind of data analysis, and in the high-order models, it is difficult to estimate the model parameters because of the high complexity in calculation [5].

Kalman filtering technology is a common data processing method, which can estimate the status of the dynamic system in a series of data within measurement noise when the measurement variance is known [6]. Kalman filtering can also correct the estimated parameters dynamically, with high analytical precision [7]. Certainly, it is difficult to construct the Kalman state equation and measurement equation, which is similar as it difficult to construct high-order models in the time series analysis method.

Aiming at the problems in time series analysis method and the Kalman filtering in time series data processing, we improve the traditional Kalman filtering method in this paper, as constructing the prediction equation of the low-order model with the time series method firstly, then deducing the state equation and measurement

ISSN: 1975-0080 IJMUE

equation of Kalman filtering, to simplify the construction of the Kalman model, which can also ensure the accuracy in the time series analysis.

2. Construction of the Improved Kalman Filtering Model

When Kalman filtering is applied into the time series data analysis, the largest difficulty is to deduce the state equation and measurement equation. Therefore, in this paper, we improve the general construction of the Kalman filtering model, which is to construct a low-order model fitting the variation trend of the time series data by using the time series analysis method firstly, and then deduce the state equation and measurement equation of the Kalman filtering model according to the prediction equation of the low-order model, to achieve reducing the complexity in the construction process of the Kalman filtering model.

2.1 Construction Time Series Low-Order Model with the ARMA Method

In all kinds of the time series data analysis methods, the ARMA (Auto Regressive Moving Average) method is more common used, which combines the AR (Auto Regressive) method and the MA (Moving Average) method together, the constructed analysis model fits the characteristic of the time series data more accurately.

If Y_t indicates the time series, ε_t indicates an independent and identically distributed random variables sequence, then an AR model is as shown in Formula (1).

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon_t \tag{1}$$

Where, $Var(\varepsilon_t) = \sigma_\varepsilon^2 > 0$ and $E(\varepsilon_t) = 0$.

Formula (1) also shows that Y_t is an AR model subjecting to p -order. Inspecting the stationary condition of the AR model needs to inspect its lag operator $\varphi(B) = 1 - \beta_1 B - \beta_2 B^2 - \cdots \beta_p B^p$ .If the roots of the lag operator $\varphi(B)$ are all out of the unit circle, the roots of $\varphi(B) = 0$ are more than 1.

When describing the time series Y_t with the MA model, it can be shown as follows:

$$Y_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \cdots + \alpha_q \varepsilon_{t-q} \tag{2}$$

Here, Y_t is called the MA model subjecting to q -order.the MA model is stationary in any condition.

We can construct the ARMA model of the time series Y_t by combining AR and MR together,as is shown in Formula (3).

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \cdots + \alpha_q \varepsilon_{t-q} \tag{3}$$

Here, Y_t is called the ARMA model subjecting to (p, q)-order,which can be in shorthand as follows.

$$\varphi(B) Y_t = \theta(B) \varepsilon_t \tag{4}$$

Here, $\varphi()$, $\theta()$ indicates evaluation coefficient of the ARMA model.

2.2 Construction of Kalman Model Based on the ARMA Low-Order Model

For a common discrete system, the basic expression of its Kalman filtering model is as follows.

$$\begin{aligned} Y(t+1) &= \Re(t+1,t) + \eta(t+1,t) m(t) \\ S(t+1) &= T(t+1) Y(t+1) + n(t+1) \end{aligned} \tag{5}$$

Here, $Y(t)$ indicates the state vector, $S(t)$ indicates the observation vector, $m(t)$ indicates the system noise, $n(t)$ indicates the measurement noise, $\Re(t+1,t)$ indicates the state transition matrix, $\eta(t+1,t)$ indicates the excitation transfer matrix, $T(t+1)$ indicates the predicted output transfer matrix. The top equation in Formula (5) is the state equation, and the bottom one is the measurement equation.

When Kalman filtering is applied into time series data prediction, we can deduce the final prediction equation according to the mathematical induction and the orthogonality theorem *etc.*, as is shown in Formula (6-9).

$$\hat{Y}(t+1,t+1) = \Re(t+1,t)\hat{Y}(t,t) + K(t+1)[S(t+1) - T(t+1)\Re(t+1,t)\hat{Y}(t,t)]) \quad (6)$$

Here, $\hat{Y}(t+1,t+1)$ indicates the status evaluation to the data at the (t+1) moment, $K(t+1)$ indicates the Kalman gain matrix at the (t+1) moment.

$$K(t+1) = \frac{P(t+1,t)T^T(t+1)}{T(t+1)P(t+1,t)T^T(t+1) + C_m(t+1)} \quad (7)$$

Here, $P(t+1,t)$ indicates the single-step prediction error covariance matrix from t to (t+1) moment, $C_m(t)$ is the covariance matrix of $m(t)$.

$$P(t+1,t) = \Re(t+1,t)P(t,t)\Re^T(t+1,t) + \eta(t+1,t)C_n(t)\eta^T(t+1,t) \quad (8)$$

Here, $C_n(t)$ is the covariance matrix of $n(t)$.

$$P(t+1,t) = [I - K(t+1)T(t+1)]P(t+1,t) \quad (9)$$

It can be seen in Formula (6), the key performance of the Kalman filtering prediction is that, when predicting the status of the latest moment, it can correct the status evaluation of the previous moment. However, to achieve the Kalman filtering prediction, we have to deduce the right state equation and measurement equation.

In this paper, we firstly construct the low-order model of the time series with the ARMA method, then deduce the state equation and measurement equation of the Kalman filtering model with this low-order model. Discretely process the ARMA model, we can get:

$$\begin{aligned} Y_{t+1} &= Y(t+1) \\ Y_t &= Y(t) \\ Y_{t-1} &= Y(t-1) \\ Y_{t-2} &= Y(t-2) \\ &\dots \end{aligned} \quad (10)$$

In this way, we achieve the correspondence of the ARMA model and Kalman model in data types. Here, we retain 4 delay states for the ARMA model, to obtain the intercepted ARMA model, as Formula (11) shows.

$$Y(t+1) = \beta_1 Y(t) + \beta_2 Y(t-1) + \beta_3 Y(t-2) + \beta_4 Y(t-3) + \varepsilon_{t+1} \quad (11)$$

Therefore, we can construct the state equation and measurement equation of the Kalman filtering model, as are shown in Formula (12) and (13).

$$\begin{bmatrix} Y(t+1) \\ Y(t) \\ Y(t-1) \\ Y(t-2) \end{bmatrix} = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & \beta_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} Y(t) \\ Y(t-1) \\ Y(t-2) \\ Y(t-3) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} m(t+1) \quad (12)$$

$$S(t+1) = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} Y(t+1) \\ Y(t) \\ Y(t-1) \\ Y(t-2) \end{bmatrix} + n(t+1) \qquad (13)$$

3. Experimental Results and Analysis

In order to testify the performance of the improved Kalman filtering model mentioned in this paper, we select a time series with 24 data, as Table 1 shows.

Table 1 Time Series Data Selected in this Paper

Number	Data	Number	Data	Number	Data
1	4.609162	**9**	4.721825	**17**	4.812179
2	4.616138	**10**	4.748701	**18**	4.830121
3	4.622120	**11**	4.791390	**19**	4.855016
4	4.639072	**12**	4.817230	**20**	4.888809
5	4.658002	**13**	4.825219	**21**	4.921625
6	4.675946	**14**	4.828216	**22**	4.936525
7	4.688917	**15**	4.827215	**23**	4.941513
8	4.703879	**16**	4.814185	**24**	4.949502

Firstly, we construct the ARMA model of this group of time series data according to Formula (3), we pick the first eight data in the construction process. The Y_t curve and autocorrelation function curve formed by the first eight data are as shown in Figure 1 and 2.

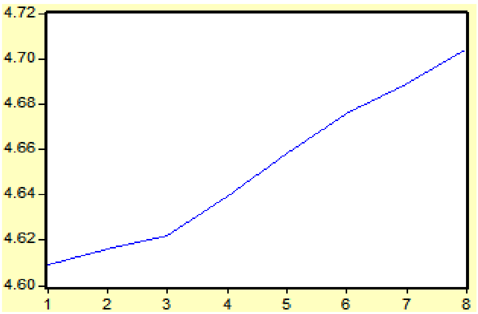

Figure 1. Curve of the First Eight Data

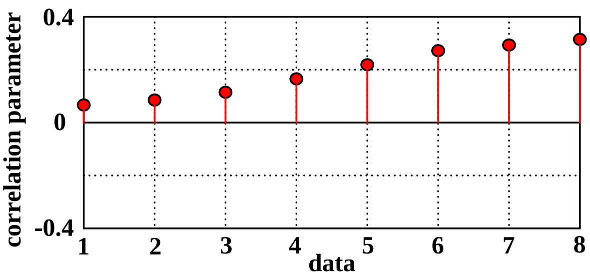

Figure 2. Autocorrelation Coefficient of the First Eight Data

Next, we calculate the first order differential form to the first eight data, and draw the curves again according to the 8 first order differential data, as is shown in Figure 3.The autocorrelation coefficient of the 8 first order differential data is as shown in Figure 4.

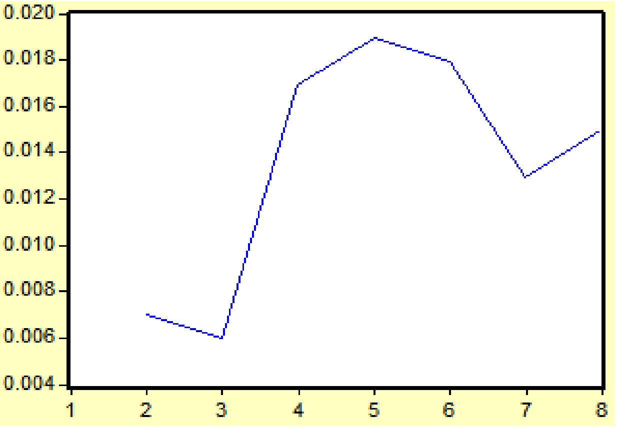

Figure 3. First Order Differential Curve of the First Eight Data

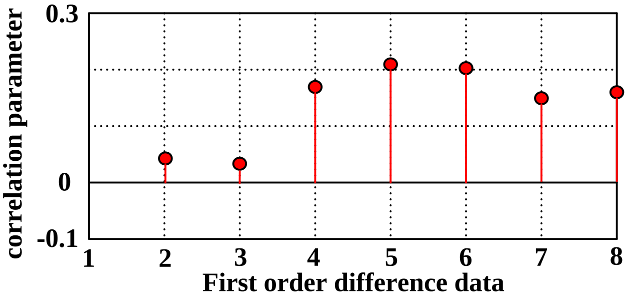

Figure 4. First Order Differential Autocorrelation Coefficient of the First Eight Data

Later, we retain 4 delay states according to the first eight data and the ARMA model, to get its low-order model as follows:

$$Y(t+1) = 0.3621Y(t) + 0.2273Y(t-1) + 0.0178Y(t-2) + 0.1844Y(t-3) + \varepsilon_{t+1}$$

Hence, we can construct the state equation and measurement equation of Kalman filtering model corresponded with this group of the time series data, as is shown in following.

$$\begin{bmatrix} Y(t+1) \\ Y(t) \\ Y(t-1) \\ Y(t-2) \end{bmatrix} = \begin{bmatrix} 0.3621 & 0.2273 & 0.0178 & 0.1844 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} Y(t) \\ Y(t-1) \\ Y(t-2) \\ Y(t-3) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} m(t+1)$$

$$S(t+1) = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} Y(t+1) \\ Y(t) \\ Y(t-1) \\ Y(t-2) \end{bmatrix} + n(t+1)$$

After obtaining the state equation and measurement equation of the Kalman filtering model corresponded with this group of the time series data, we predict all the 24 data, the original data curve and the prediction curve is as Figure 5 shows.

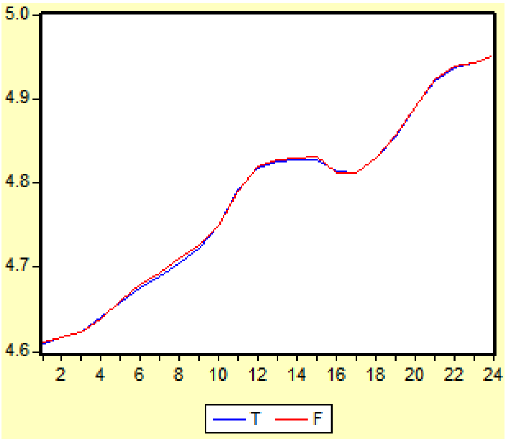

Figure 5. Truth-Value Curve and Prediction Curve of Time Series Data in this Paper

In Figure 5,"T"indicates the truth-value curve of the time series data, and"F"indicates the prediction curve of the time series data. It can be seen in Figure 5 that, the Kalman filtering model constructed by the improved method in this paper, achieves predicting the time series data ideally, the two curves are essentially coincident with each other, without big differences. In order to quantitatively analyze the difference between the prediction data and the truth-value,we analyze the error of the prediction curve. The error of the prediction curve is as shown in Figure 6.

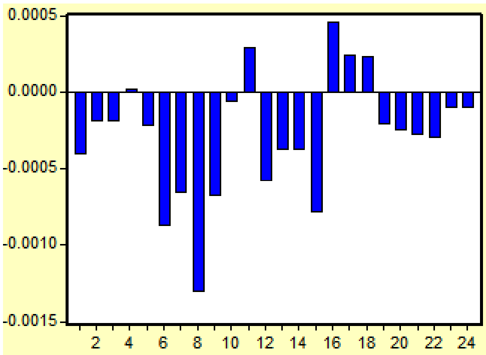

Figure 6. Error of the Prediction Curve

It can be seen in figure 6 that， the largest error between the prediction data and the truth-value is only -0.15%, which achieves ideal prediction effect.

4. Conclusion

Aiming at the shortage existing in common analysis methods of the time series data, we apply the Kalman filtering method into the time series data processing. Firstly, construct low-order model of time series data according to the ARMA time series analysis method. Then intercept sufficient delay states on the basis of this low-order model, and deduce the state equation and measurement equation of the Kalman filtering model according to the estimated parameters of these states, to achieve the complete construction of the Kalman filtering model. In the verification experiment, we selected a time series with 24 data, and constructed the ARMA low-order model with the first eight data, and then intercepted 4 delay states, to deduce the Kalman filtering model of these data. As the experimental results show that, in the Kalman prediction model constructed by the improved processing, the largest error of the prediction accuracy to the time series data is only -0.15%.

References

[1] Tak C. F. "A review on time series data mining," Engineering Applications of Artificial Intelligence, vol. 24 no. 1, **(2011)**, pp. 164-181.
[2] Xiaoyue W., Abdullah M., Hui D., Goce T., Peter S. and Eamonn K., "Experimental comparison of representation methods and distance measures for time series data," Data Mining and Knowledge Discovery, vol. 26 no. 2, **(2013)**, pp. 275-309.
[3] Frans V. D. B., Konrad J. W., Simeon M. and Terence L. V. Z., "HiTempo: a platform for time-series analysis of remote-sensing satellite data in a high-performance computing environment," International Journal of Remote Sensing, vol. 33 no. 15, **(2012)**, pp. 4720-4740.
[4] Iain M. and Gil M., "Estimating selection coefficients in spatially structured populations from time series data of allele frequencies," Population and Evolutionary Genetics, vol. 193 no. 3, **(2013)**, pp. 973-984.
[5] Peter M. A., C. Jeganathan, Jadu D. and Clement A., "Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology," Remote Sensing of Environment, vol. 123, **(2012)**, pp. 400-417.
[6] P. L. Houtekamer, Xingxiu D. and Herschel L., "Higher resolution in an operational ensemble Kalman filter," Monthly Weather Review, vol. 142, **(2014)**, pp.1143-1162.
[7] G. Triantafyllou, I. Hoteit, X. Luo, K. Tsiaras and G. Petihakis, "Assessing a robust ensemble-based Kalman filter for efficient ecosystem data assimilation of the Cretan Sea," Journal of Marine Systems, vol. 125, **(2013)**, pp. 90-100.
[8] Jungmin K., Kyunghoon J., Jaeyong K., Hajun S. and Sungshin K., "Positioning accuracy improvement of laser navigation using unscented Kalman filter," Intelligent Autonomous Systems, vol. 193 no. 12, **(2013)**, pp. 807-816.