Improvement of Bayesian Dynamic Linear Model for Predicting Missing Data of Bridges

Ge Chenghan¹, Wang Tao²

¹ Automation, Beijing Institute of Technology, Beijing, Beijing, 100081, China

² Automation, Beijing Institute of Technology, Beijing, Beijing, 100081, China

Abstract. The missing data in bridge operation will lead to the decline of the reliability of data analysis results. In this paper, the Bayesian dynamic linear model is improved by changing the parameter matrix of hidden state variables, and the model is optimized under the condition that the predefined variables are unchanged. The frequency of a strain measuring point of the bridge is taken as the observed value, and the collected frequency value of one month is used as the training set (the collection time interval is 30 minutes) to predict the data of the next week. By comparing the predicted result with the observed value, it is found that the absolute error is less than 14.05Hz and the relative error is less than 1.82% when the training frequency value varies from 756 Hz to 773.4 Hz.

1 Introduction

The status of bridge engineering symbolizes the development level of infrastructure in a country. With the development of economy, there are more and more highway bridges, railway bridges, pedestrian bridges and other special bridges (such as pipelines and cables), so it is imperative to inspect and maintain these infrastructures. Therefore, a series of sensors are deployed to monitor the condition of the structure in real time to alleviate the problems caused by the aging of the bridge structure. However, the actual operation of bridges will be affected by many factors, for example, many bridges are far from urban areas, and solar energy can only be used for power supply in the field, which leads to the fact that the batteries of equipment cannot be replenished with electricity from solar panels after the electricity is exhausted in continuous rainy weather, resulting in power failure of equipment and missing data. In the case of missing data, it is unfavourable to analyse the health status of bridges, and there is a great potential safety hazard.

The mainstream forecasting methods in China are polynomial regression analysis method [1], autoregressive moving average model [2], grey model [3] and so on. However, the effect of these prediction methods in practical bridge application is not satisfactory. For example, the polynomial regression analysis method will cause a sharp increase in the amount of calculation under the condition of multivariate dimension; the dimension of autoregressive moving average model in practical application does not exceed two dimensions, which is suitable for the problem of less sample data; the

grey model shows good results when the data has strong tendency, and only stays at the level of observation. However, the influencing factors of the bridge environment are complex and changeable, and the above methods can not solve the multi-dimensional variable situation.

The Bayesian dynamic linear model (BDLM) was first proposed by British statisticians M. West and J. Harrison [4] in 1976. In view of the particularity of the actual background of bridge structure, Solhjell first combined Bayesian dynamic linear model with structural health monitoring (SHM). K. Y. Koo [5] and others explored the influence of bridge temperature, traffic load and wind speed on bridge strain frequency, and determined that the influence of wind speed on frequency was ignored by principal component analysis. Goulet [6] defined the hidden state variables that affect the bridge strain frequency, and realized the prediction of the bridge frequency value by separating the observation into various general components according to the specific bridge model. However, the data selected by Goulet are all stable collected values without drastic fluctuations, which is lack of representativeness. Moreover, in order to improve the accuracy of the prediction value, Goulet used expectation maximization (EM) algorithm to train model parameters, leading to the result that the predefined of state variables are changed and the significance of multi-step prediction is lost.

In this paper, the time period with large and frequent temperature changes is selected to ensure the representativeness of the training data. On the premise that the data of the bridge meets the conditions of the model, the parameter matrix is changed. The parameters of components are set as constant 1 except the

*e-mail: 1414227979@qq.com

parameters of periodic components, which not only ensures the predefined of state variables unchanged, but also avoids the problem of long calculation period and falling into local optimal solution caused by the expectation maximization algorithm. After comparing the multi-step prediction results with the observed values, the error is in a reasonable range.

2 Programme

2.1 BDLM

In BDLM, the observation equation and state equation are defined as:

$$Y_t = F_t X_t + v_t, v_t \sim N[0, V_t] \quad (1)$$

$$X_t = G_t X_{t-1} + w_t, w_t \sim N[0, W_t] \quad (2)$$

$$X_t = [x_t^{B,LL} x_t^{B,AR} x_t^{T,LL} x_t^{T1,S1} x_t^{T1,S2} x_t^{T2,S1} x_t^{T2,S2} x_t^{T,AR} x_t^{L,LL} x_t^{L,AR} x_t^P]'$$

The state transition matrix is defined as:

$$G_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \cos\omega^{T1} & \sin\omega^{T1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\sin\omega^{T1} & \cos\omega^{T1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cos\omega^{T2} & \sin\omega^{T2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\sin\omega^{T2} & \cos\omega^{T2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The observation matrix is defined as:

$$F_t = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The model error variance matrix is defined as:

$$W_t = \text{diag}(0, (\sigma^{B,AR})^2, 0, 0, 0, 0, 0, (\sigma^{T,AR})^2, 0, (\sigma^{L,AR})^2, \infty)$$

The measurement error variance matrix is defined as:

$$V_t = \begin{bmatrix} (\sigma_B)^2 & 0 & 0 & 0 \\ 0 & (\sigma_T)^2 & 0 & 0 \\ 0 & 0 & (\sigma_L)^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

2.3 Probability recurrence

The specific steps for prediction and inference of observation variables and state variables are as follows:

Posterior distribution of state variables at time (t-1):

$$(X_{t-1}|D_{t-1}) \sim N[m_{t-1}, C_{t-1}] \quad (3)$$

Prior distribution at time t:

2.2 Global matrix

The observation variables can be directly obtained by the acquisition equipment. There are four observation variables in this paper, namely: frequency (B), temperature (T), traffic load (L) and traffic mode (P). The definition of hidden state variables is determined according to the characteristic of observation variables. There are three types: baseline component (LL), autoregressive component (AR) and cyclic component (S).

Observation variables are defined as:

$$Y_t = [y_t^B y_t^T y_t^L y_t^P]'$$

State variables are defined as:

$$(X_t|D_{t-1}) \sim N[a_t, R_t] \quad (4)$$

As in equation (4), $a_t = G_t m_{t-1}$, $R_t = G_t C_{t-1} G_t' + W_t$.

One-step prediction distribution at time t:

$$(Y_t|D_{t-1}) \sim N[f_t, Q_t] \quad (5)$$

As in equation (5), $f_t = E(Y_t|D_{t-1}) = F_t a_t$, $Q_t = \text{var}(Y_t|D_{t-1}) = F_t R_t F_t' + V_t$.

Posterior distribution at time t:

$$(X_t|D_t) \sim N[m_t, C_t] \quad (6)$$

As in equation (6), $m_t = E(X_t|D_t) = a_t + A_t e_t$, $C_t = \text{var}(X_t|D_t) = R_t - A_t A_t' Q_t^{-1}$, $A_t = R_t F_t Q_t^{-1}$, $e_t = Y_t - f_t$.

K-step state prediction distribution:

$$(X_{t+k}|D_t) \sim N[a_t(k), R_t(k)] \quad (7)$$

As in equation (7), $a_t(k) = G_{t+k}a_t(k-1)$, $R_t(k) = G_{t+k}R_t(k-1)G_{t+k}' + W_{t+k}$.

K-step observation prediction distribution:

$$(Y_{t+k}|D_{t-1}) \sim N[f_t(k), Q_t(k)] \quad (8)$$

As in equation (8), $f_t(k) = E(Y_{t+k}|D_{t-1}) = F_{t+k}a_t(k)$, $Q_t(k) = \text{var}(Y_{t+k}|D_{t-1}) = F_{t+k}R_t(k)F_{t+k}' + V_{t+k}$.

3 Application environment

3.1 Analysis of bridge under real situation

In this paper, the Shenzhen Bay Highway Bridge is taken as an engineering example, and the data of the M10 static strain measurement point of the main beam section at the

entrance of Hong Kong side of the bridge is taken as the analysis object. At present, the collected data since the installation of sensors in January 2020 have been obtained, as shown in figure 1. It can be seen from figure 2 that the daily variation of the traffic flow of the bridge is relatively stable, so it can be considered that the traffic load and the traffic mode defined in the model change consistently every day. According to the change of temperature data in figure 3, the range of temperature in half a year is 23.1°C , which can be approximately regarded as the range of temperature in one year. Among them, the temperature in April is located in a period that the temperature varies greatly and frequently, and the range of temperature reaches 12°C in only one month. In April, there are four cases where the range of temperature exceeds 10°C in two days. Therefore, taking the data of April as the training set can ensure the representativeness of data selection.

Figure 1. Frequency change in half a year.

Figure 2. Load change in half a year

Figure 3. Temperature change in half a year.

3.2 Normality verification

The data of M10 measuring point in April are verified to be normal. It can be seen from figure 4 that most of the frequency values are distributed in the interval $(\mu \pm \sigma)$, which approximately obeys the normal distribution with expectation of 761.3 and variance of 4.4.

Figure 4. Frequency distribution in April.

Figure 5. Correlation between frequency and temperature in April.

3.3 Correlation verification

Taking the frequency value of M10 measuring point in April as the ordinate and the temperature value as the abscissa, it can be seen from figure 5 that the frequency value and the temperature value are approximately linear.

4 Result

4.1 Separated hidden state variables

After bringing the data of M10 measuring point in April into the model, the hidden state variables can be separated from the observation variables. It can be seen from figure 6 that the strain frequency is only separated into two components, which is considered as the superposition of the baseline component and the autoregressive component of the frequency. It can be seen from figure 7 that the temperature is affected by daily and seasonal factors, and there are daily and seasonal components, so there are four components. It can be seen from figure 8 that the component type of the traffic load is the same as the strain frequency that does not include cyclic component.

Figure 6. The hidden state variables of frequency, in which (a) is the collected value curve of frequency, (b) is the curve of $x_t^{B,LL}$, and (c) is the curve of $x_t^{B,AR}$.

Figure 7. The hidden state variables of temperature, in which (a) is the collected value curve of temperature, (b) is the curve of $x_t^{T,LL}$, (c) is the curve of $x_t^{T1,S1}$, (d) is the curve of $x_t^{T2,S1}$, and (e) is the curve of $x_t^{T,AR}$.

Figure 8. The hidden state variables of traffic load, in which (a) is the collected value curve of traffic load, (b) is the curve of $x_t^{L,LL}$, and (c) is the curve of $x_t^{L,AR}$.

4.2 Treatment of missing data

This section introduces that the missing data in structure health monitoring is predicted by BDLM. In figure 9 and figure 10, the comparison between one-step prediction values and observation values is shown in the part without data missing, and multi-step prediction values

are given in the part with data missing, at this time, there are no observation values as the feedback values.

In order to explain the difference between prediction values and observation values more vividly, the relative error and absolute error are used as judgment indexes. In figure 11 and figure 12, the maximum relative error is 1.82%, and the maximum absolute error is 14.05Hz within 336 time steps.

Figure 9. Predicted value curve in April.

Figure 10. Predicted value curve of missing data.

Figure 11. Relative error curve.

Figure 12. Absolute error curve.

5 Conclusion

By applying Bayesian dynamic linear model to the special background of bridge health monitoring, the missing data can be predicted. At the level of observation variables, observation values can be used as feedback values to optimize one-step prediction values. At the level of state variables, there is no need to consider the parameter training of the model, but only explore the changes of the components themselves, and realize the multi-step prediction under the condition of ensuring the predefined variables unchanged. In addition, the prediction results can be kept within a reasonable error range, which proves the rationality of the selection of hidden state variables and the accuracy of prediction.

References

1. Zhou, Y., Zhu, Y. (2019) Research Status Analysis of Intelligent Health Monitoring of Bridges. *Guangzhou Architecture.*, 47(4): 3-7.
2. Wang, Y., Jin, G.Q. (2014) On-line structural analysis and state evaluation method of bridge health monitoring system. *Bridge Construction.*, 44(1): 25-30.
3. Fan, X.P. (2014) Reliability updating and Bayesian prediction of bridges based on proof loads and monitoring data. *Civil Engineering.*, 624(4): 39-63.
4. West, M., Harrison, J. (1998) Bayesian Forecasting and Dynamic Models. *Journal of the Operational Research Society* volume., 49: 179-180.
5. Koo, K.Y., Brownjohn, M.W. (2016) Structural health monitoring of the Tamar suspension bridge. *Struct Control and Health Monitor.*, 20(4): 609-625.
6. Goulet, J. (2017) Bayesian dynamic linear models for structural health monitoring. *Struct Control and Health Monitor.*, 24(12): 20-35.