A bridge damage monitoring and assessment strategy based on Gaussian mixture modelling

Heng Li[1, 2], Yitong Wu[2] *, Shaohua Xu[2], Mingde Zheng[2], Wentao Zhang[3], Feiyu Zheng[2]

1. GuangXi BeiBu Gulf Investment Group Co., Ltd., Nanning, Guangxi, P.R. China.
2. Guangxi Beitou IT Innovation Technology Investment Group Co., Ltd., Nanning, Guangxi, P.R. China.
3. Guilin University of Electronic Technology, Guilin, Guangxi, P.R. China.

lih@bgigc.com, wuyt@bgigc.com, xush@bgigc.com, zhengmd@bgigc.com, glietzwt@163.com, zfy@ bgigc.com.

*Correspondingauthor: Yitong Wu, E-mail: wuyt@bgigc.com

Abstract: For the bridges equipped with health monitoring systems, it is essential to precisely assess the health status of the bridge from the data with noise and outliers. An improved Gaussian mixture model clustering algorithm is used to process the obtained bridge strain sensor monitoring data and generate the characteristic data of clustering in order to improve the accuracy of data analysis of bridge health monitoring system. The cluster and feature data are obtained by Expectation Maximization process, and the isolated clusters are filtered out by the threshold parameters of weight and the Euclidean distance between clusters center. Based on the feature data, a scoring strategy is established to assess the bridge health and sensor status. The proposed strategy is used to analyze the monitoring data of a bridge in western China, and the analysis results show the availability of this strategy. Compared with the directly collected data, the processed bridge health scoring curve variation is smooth, which can reduce the influence of noise and abnormal data. The sensor status scoring curve can track changes in collected data and respond to different transformation scenarios. This shows that the proposed strategy can evaluate the state of this bridge and the sensors. provided a reference for bridge maintenance decision during the operational period.

Keywords: Continuous steel structure bridge; Gaussian mixture model; structural health monitoring; bridge health assessment.

1. Introduction

After more than 20 years of development, the Bridge Health Monitoring (BHM) system in China has effectively used technologies such as structural damage monitoring and digital twin to greatly improve bridge safety [1-2]. The safety assessment mechanism established in the early warning of damage continues to be a focus of current research. Various approaches, including autoregressive models, wavelet energy entropy, Bayesian models, deep learning, etc., have been employed to formulate theories for monitoring damage and to develop assessment models [3-10].

At present, there is a lack of a readily transferable and standardized model for real-time diagnostic evaluation. Additionally, few studies have focused on models that capture the correlation between sensor data. Therefore, this study proposes an early warning strategy based on an improved Gaussian mixture model clustering algorithm, which is designed for monitoring the health status of bridge. First, the acquired bridge sensor monitoring data is processed to generate clustering feature data. Then, a sensor and bridge health scoring strategy will be developed based on this feature data. The effectiveness of the strategy is validated by analyzing monitoring data from an operational bridge.

2. Algorithm principle and analysis strategy

2.1 Mathematical Theory of Gaussian Mixture Model

The data returned can be conceptualized as a sequence of data streams of dimension m, representing the number of strain sensors, when a strain sensor monitors a bridge. The Gaussian Mixture Model (GMM), originally introduced by C. Stauffer et al.[11], is adept at handling data streams in the form of images with pixel data types, statistical data distributions, and multi-point clustering. In the GMM algorithm, the final segmentation clustering result can be obtained with the center point being the mean of each Gaussian distribution, as long as the number of required clustering results is known, The probability density function of a multivariate Gaussian distribution for an m-dimensional feature vector $\boldsymbol{x}$ is given by:

$$N(\boldsymbol{x}) = \frac{1}{(2\pi)^{m/2}|\boldsymbol{\Sigma}|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \tag{1}$$

Here, π is the weight of the cluster, $\boldsymbol{\mu}$ is the cluster center, $\boldsymbol{\Sigma}$ is the covariance of the cluster and $|\boldsymbol{\Sigma}|$ is the determinant of the covariance matrix.

Assuming there are total K components, each component i represents a multivariate Gaussian distribution. The mixture model constitutes a weighted sum of multiple Gaussian models:

$$P(\boldsymbol{x}) = \sum_{i=1}^{K} \pi_i \cdot N_i(\boldsymbol{x}) \tag{2}$$

Here, π_i is the weight of the i-th component, and $N_i(\boldsymbol{x})$ is the probability density function of the i-th component's multivariate Gaussian distribution.

To estimate the parameters of the GMM, it is necessary to calculate the probability that each sample belongs to each component. π_i, $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are determined by repeated iteration, so the Expectation Maximization (EM) process is introduced and the derivatives of π_i, $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are calculated respectively. That is, E-step and M-step, and repeat the calculation of E-step and M-step until convergence, as shown in Equation (3) to (6):

E-step: Based on the existing parameters, the likelihood of each data coming from a sub-model is calculated：

$$w_i^{(n)} = \frac{\pi_i \cdot P_i\left(\boldsymbol{x}^{(n)}\right)}{\sum_{k=1}^{K} \pi_k \cdot N_k\left(\boldsymbol{x}^{(n)}\right)} \tag{3}$$

Here, $w_i^{(n)}$ is the probability that the i-th component belongs to the n-th cluster sample, $P_i\left(\boldsymbol{x}^{(n)}\right)$ is the probability density function of the i-th Gaussian component.

M-step: Calculate the model parameters of the new iteration, calculate the derivative and set the derivative to 0 to obtain, as Equation (4), (5) and (6):

$$\pi_i = \frac{1}{N}\sum_{n=1}^{N} w_i^{(n)} \tag{4}$$

$$\boldsymbol{\mu}_i = \frac{\sum_{n=1}^{N} w_i^{(n)} \cdot \boldsymbol{x}^{(n)}}{\sum_{n=1}^{N} w_i^{(n)}} \tag{5}$$

$$\boldsymbol{\Sigma}_i = \frac{\sum_{n=1}^{N} w_i^{(n)} \cdot \left(\boldsymbol{x}^{(n)} - \boldsymbol{\mu}_i\right) \cdot \left(\boldsymbol{x}^{(n)} - \boldsymbol{\mu}_i\right)^T}{\sum_{n=1}^{N} w_i^{(n)}} \tag{6}$$

2.2 Algorithm Design

In the context of long-term analysis, the occurrence of false positives or offsets during the actual data collection process is inevitable. Individual sensors often generate short cycle abnormal points, possible reasons being continuous changes in the state of the bridge itself or the deviation of the measurement of the sensor itself. Therefore, the algorithm design should be based on the analysis mean values, mean square deviations, time series characteristics and isolated values screening. The pseudo-code of the algorithm is given as follows:

Table 1. Clustering pseudocode

	Notes
Input: the collected data X with upper limit r, the weight threshold w_acc, the distance coefficient d, the real-time data new_X	
Output: cluster center coordinates Mu, cluster covariance matrix Var	
def Value_GMM (new_X)：	
If_not_define:	
Mu, Var, W = GMM_Init();	Initializing the cluster center Mu, the covariance Var, the weight of each data W
Mu_point = Count_Statistics(W, w_acc);	The weight W and the weight threshold w_acc are used to distinguish each point belonging to a cluster, and the number Mu_point of each point in the cluster is counted. If Mu_point is 0, the isolated cluster will be removed
For x in range(len(Mu_point)):	
If(M_point(x) == 0):	
Remove_cluster();	
For y in range(len(M_point)):	Calculate the Euclidean distance between the real-time data new_x and Mu. If the distance exceeds the distance coefficient d, a new cluster is added
If Distence(new_X, Mu[y]) > o:	
New_cluster();	
Mu, Var, W = EM(list(X));	Repeat the E-M procedure
Return Mu,Var	

Here, the distance coefficient d and the weight threshold w_acc are introduced to improve the speed of convergence rate and clustering accuracy.

The health status of the sensor is revealed by apparent through the return value of the Gaussian distribution variance (Var), which indicates the dispersion and correlation in the corresponding dimension. Typically, a larger Var in the cluster implies a more scattered distribution, suggesting the presence of interference in the analysis of the sensor data. The correlation coefficient is used to represent the sensor health score if the sensors are correlated and the changes are synchronous:

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1m} \\ \vdots & \ddots & \vdots \\ \sigma_{m1} & \cdots & \sigma_{mm} \end{bmatrix} \tag{7}$$

$$mark_s = \frac{1}{p}\sum_{n=1}^{N}\sum_{i=1}^{m-1}\sum_{j=1}^{m} count_k \cdot \left|\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}\right| \tag{8}$$

Here, m is the actual coordinate dimension.

The health of bridge is reflected by the location of the cluster center. The confidence rating of the safe state is then calculated by the following formula:

$$mark_b = \frac{1}{p}\sum_{n=1}^{N} \frac{count_n}{\sum_{l=1}^{m} \mu_{nl}^2} \tag{9}$$

Here, $count_n$ is the number of points corresponding to cluster n and is the coordinate value of the cluster center in the l-th dimension. It is assumed that a higher score indicates a safer condition of the bridge.

3. Examples of data analysis

This study is based on the BHM system data of a continuous steel structure bridge in western China. The bridge is a three-span bridge, as shown in the Fig. 1. The data acquisition architecture is shown in the Fig. 2, and the sensor group is shown in the Fig. 1[10]. The data was cleaned and pre-processed at the edge device based on time domain, frequency domain, correlation, trend and mode analysis.

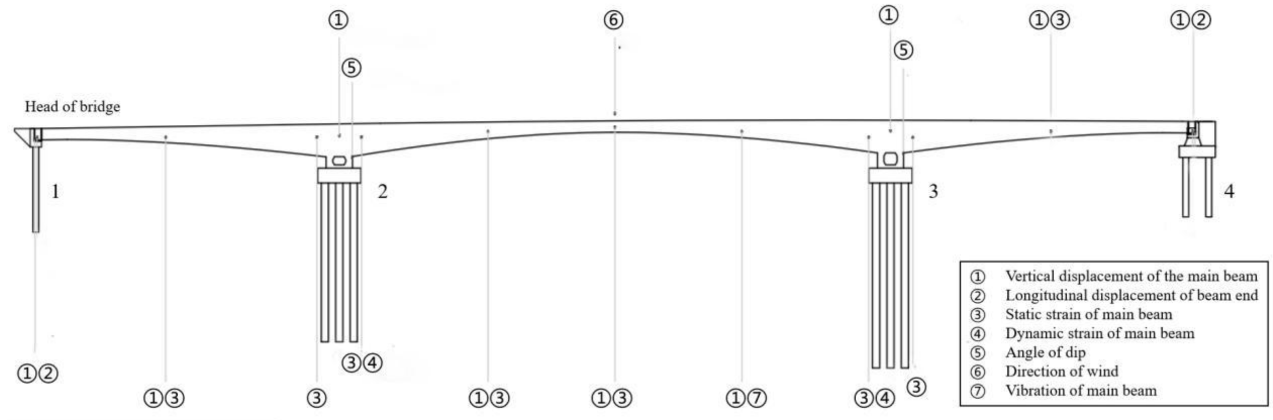

Fig. 1 Arrangement of measuring points for box girder bridge

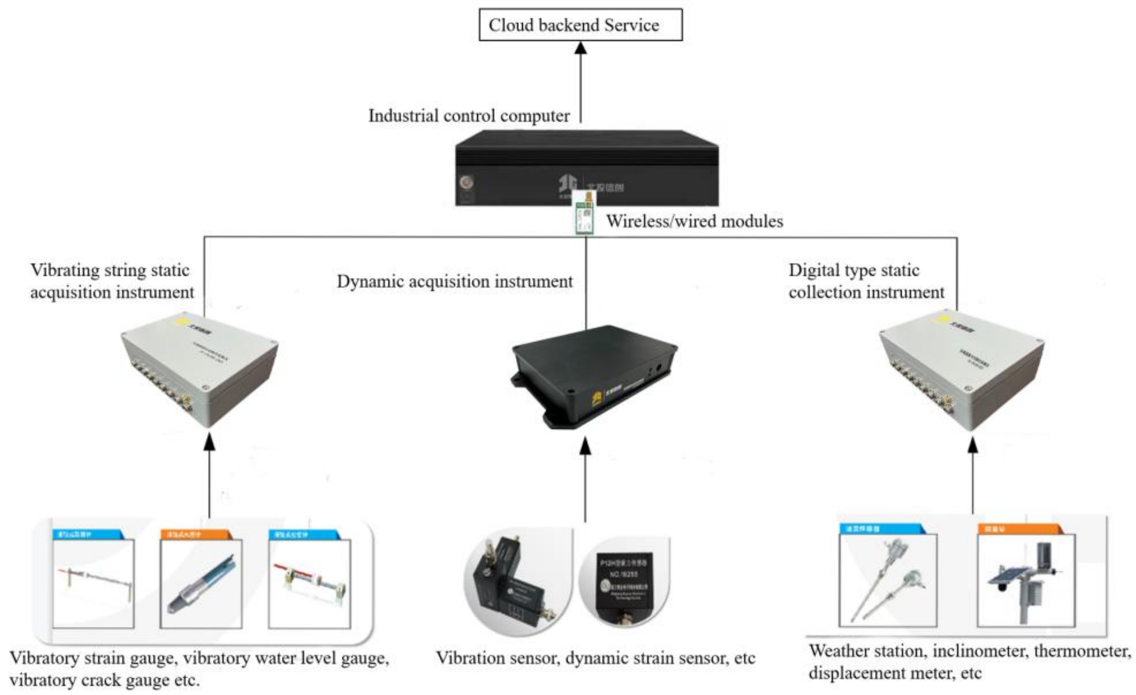

Fig. 2 Data acquisition and transmission framework

The data from two strain sensors located in the middle of section 1-2, as shown in Fig. 1, were selected for analysis in this experiment. there were collected over a period of two weeks, from December 23 to January 5, 2023, resulting in a total of 168 data points per group. The previous week's data was used for parameter tuning, while the following week's data was used for verification and comparison.

The degree of agreement of the parameter value can be described by the standard deviation of the sensor health status score (SDS) and the bridge health status score (SDB). In the normal operating state of the sensor, the data fluctuation is small, and the corresponding score standard deviation is smaller. According to the pre-experiment, the value of the parameter w_acc as shown in Table 1 may cause the method to divide by zero or lose the dimensionality reduction features, so the value of w_acc should be selected first. The mean square error is the smallest when the value of w_acc is 0.85, as shown in Table 2.

In general, as the value of d decreases, the number of clusters increases accordingly. The number of clusters reflects their dimensional characteristics, each cluster is more likely to capture the original details in the original data. Reducing the number of clusters means that the data is merged more, which means that some detail is lost, even though its dimensionality reducing properties are better. The clustering evaluation effect with d=0.6 is more appropriate and corresponds to the actual situation, as shown in Table 3.

Table 2. Effect of the value of w_acc on the mean square deviation of the sensor health status score and the bridge health status score when d = 0.6

Value	SDS	SDB
w_acc ≤ 0.7	--	--
w_acc = 0.75	0.1424	0.0343
w_acc = 0.8	0.1420	0.0351
w_acc = 0.85	0.1060	0.0248
w_acc = 0.9	0.1521	0.0778
w_acc ≥ 0.95	--	--

Table 3. Effect of the value of d on the mean square deviation of the sensor health status score and the bridge health status score when w_acc = 0.85

Value	SDS	SDB
d=0.4	0.1709	0.0408
d=0.5	0.1149	0.0249
d=0.6	0.1060	0.0249
d=0.7	0.1075	0.0242
d=0.8	0.0738	0.0439
d≥ 0.9	--	--

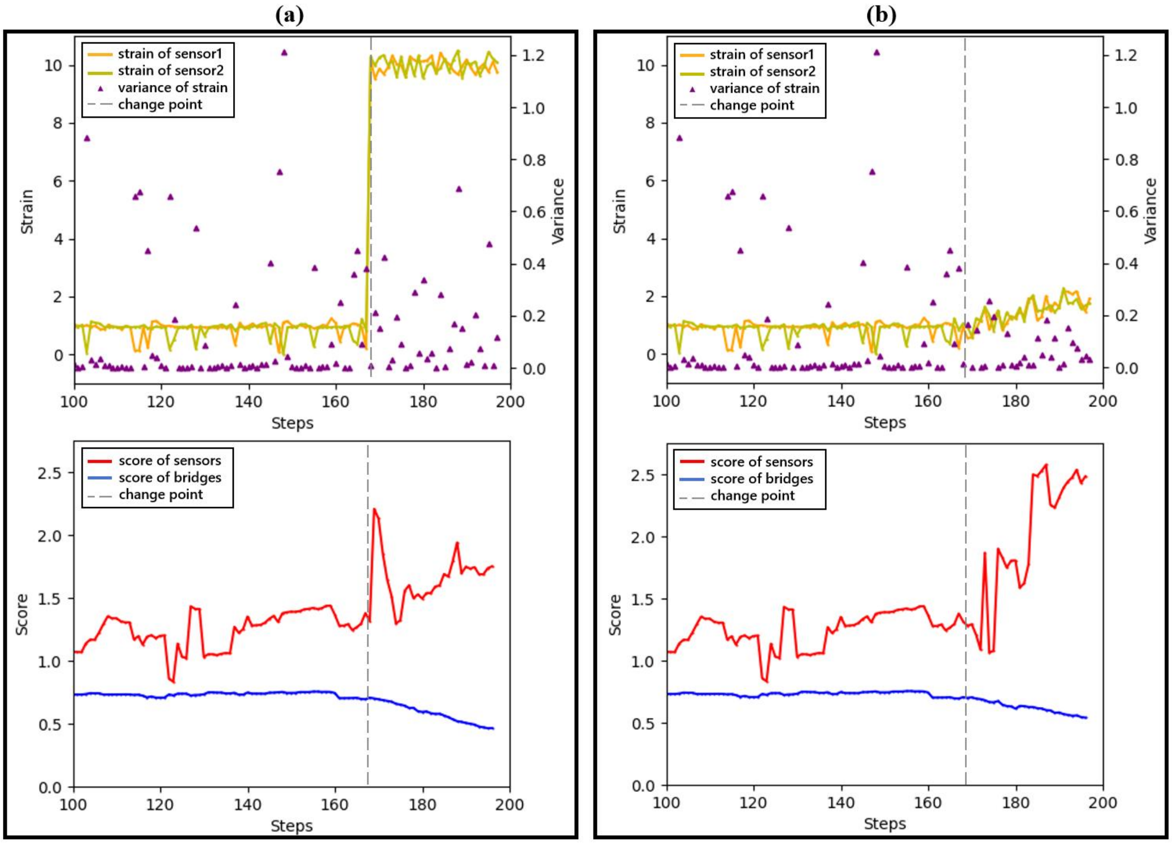

Fig. 3 Sensor state score and bridge state score responses for two different input cases

The data returned by the bridge strain sensor in the second week and the variance of the two sensors is shown in Fig. 3. The two sensors are located close to each other, which indicates a high degree of correlation. Here, two different simulation experiments are designed to simulate different inputs and observe the response. The burst fracture situation was simulated, and the input value was added to 10 in one step, as shown in Fig. 3 (a). As well as the simulated case of a sensor offset case, the input value was added to 2 in 30 steps, as shown in Fig. 3 (b). The change is made starting from the 169-th data point, the change point in Figure 3, and the response of the score output to the change is observed.

The simulation results are shown in Fig. 3. The scoring strategy is highly sensitivitive to changes in the health of sensors. As shown in Fig. 3(a), the ratings respond to instantaneous changes and decay to the normal value in the short term. In Fig. 3(b), the score shows a continuous upward trend when the sensor receives a continuous and gradually changing ramp signal. At the same time, when the sensor synchronization changes, the absolute value of the score increases correspondingly, reflecting a certain correlation. In the state change of bridge health, comparing the plots of Fig.3 (a) and (b), the reduction of the score increases accordingly when the input value changes more, indicating that the slope can effectively show the changes in the health status of the bridge, and the scoring system has the ability to mitigate external disturbances to some extent. This confirms the availability of the scoring strategy.

4. Summary

This study analyzed the shortcomings of the existing BHM system in damage warning and the analysis strategy and evaluation method of sensor health status and bridge health status based on Gaussian mixture model clustering are proposed. The evaluation results under different parameters are compared with the analysis of two weeks of bridge strain sensor data, and the results show the availability of this strategy.

In the next research plan, the algorithm and the evaluation model will be optimized, all sensor data will be added for a comprehensive evaluation, and the reliability of the method will be verified based on long-term data.

Acknowledgements

We acknowledge the support from the Guangxi Key R&D projects: Development of bridge and tunnel emergency monitoring system based on φ-OTDR (AB22035047), (AB23026038).

References

[1] Wang Lingbo, Wang Qiuling, Zhu Zhao, et al. Research status and prospect of bridge health monitoring technology. China Journal of Highways, 2021, 34(12): 25-45.

[2] Yu Chuanjin, Chen Qian, Liang Aixia, et al. Research progress of big data of bridge engineering in 2020. Journal of Civil and Environmental Engineering, 2021, 43(S1): 321-326.

[3] Daneshvar H M, Gharighoran A, Zareei A S, et al. Early damage detection under massive data via innovative hybrid methods: application to a large-scale cable-stayed bridge. Structure and Infrastructure Engineering, 2020, 17(7): 1-19.

[4] Zhang S, Liu Y. Damage detection of bridges monitored within one cluster based on the residual between the cumulative distribution functions of strain monitoring data. Structural Health Monitoring, 2020, 19(6): 1764-1789.

[5] Locke W, Sybrandt J, Redmond L, et al. Using drive-by health monitoring to detect bridge damage considering environmental and operational effects. Journal of Sound Vibration, 2020, 468.

[6] Zhan Jiawang, Yan Yuzhi, Zhang Fei, et al. Damage Alarming Method for Railway Simply-Supported Steel Truss Girder Bridge Based on Wavelet Energy Entropy. China Academy of Railway Science Corporation Limited, 2018, 39(3): 24-30.

[7] He Shuanhai, Li Yuan, Ren Wei, et al. Analysis and Prediction of Long-term Deflection of Large Span Continuous Rigid Bridge. Journal of Wuhan University of Technology, 2014, 36(4): 113-115.

[8] Ni Y Q, Wang Y W, Zhang C. A Bayesian approach for condition assessment and damage alarm of bridge expansion joints using long-term structural health monitoring data. Engineering Structures, 2020, 212:110520.

[9] Vagnoli M, Prescott R R, Andrws J. A Bayesian belief network method for ridge deterioration detection. Journal of Risk and Reliability, 2021, 235(3): 338-355.

[10] Huang Yujun, Li Shiyun, Zhang Jinquan. Application of information technology in bridge health monitoring system. Guangdong Communication Technology, 2023, 43(03): 60-63.

[11] Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking. computer vision and pattern recognition, IEEE Computer Society Conference on. IEEE, 1999.