

Research Article

Research on Bridge Sensor Validation Based on Correlation in Cluster

Huang Xiaowei,^{1,2} Zhang Peng,¹ Ding Pinyi,¹ Chen Weimin,¹ and Xiong Fei³

¹Key Laboratory of Optoelectronic Technology and System, Chongqing University, Chongqing 400044, China

²Department of Information Engineering, Logistical Engineering University, Chongqing 401331, China

³College of Foreign Languages and Cultures, Chongqing University, Chongqing 400044, China

Correspondence should be addressed to Zhang Peng; zhangpeng@cqu.edu.cn

Received 21 January 2016; Revised 3 June 2016; Accepted 30 June 2016

Academic Editor: Christopher Pretty

Copyright © 2016 Huang Xiaowei et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to avoid the false alarm and alarm failure caused by sensor malfunction or failure, it has been critical to diagnose the fault and analyze the failure of the sensor measuring system in major infrastructures. Based on the real time monitoring of bridges and the study on the correlation probability distribution between multisensors adopted in the fault diagnosis system, a clustering algorithm based on k -medoid is proposed, by dividing sensors of the same type into k clusters. Meanwhile, the value of k is optimized by a specially designed evaluation function. Along with the further study of the correlation of sensors within the same cluster, this paper presents the definition and corresponding calculation algorithm of the sensor's validation. The algorithm is applied to the analysis of the sensor data from an actual health monitoring system. The result reveals that the algorithm can not only accurately measure the failure degree and orientate the malfunction in time domain but also quantitatively evaluate the performance of sensors and eliminate error of diagnosis caused by the failure of the reference sensor.

1. Introduction

In recent years, with a series of catastrophic accidents of infrastructure damage and collapse, the research of infrastructure structural health monitoring and early warning technology for the public safety has been promoted [1–3]. The structural complexity of infrastructure and the diversity of influential factors lead to typical application of multisensor systems in these monitoring systems. As such large permanent infrastructures, like bridge, dam, and nuclear power plant, are designed with a service life of decades, a severe challenge to the sensor system with service life of only a few years is posed. As the structure of monitoring systems utilizes the sensor data for evaluation of structural health, it is important that the data acquired are accurate and reliable. A faulty sensor cannot perform its function properly but instead may provide false information for evaluation, thus leading the system to produce the wrong diagnosis. This kind of false-negatives not only severely interferes with the normal work of the monitoring system and increases the system maintenance

cost but also can lead to the ignorance of the real fault information and eventually cause leakage alarm of sudden disaster accident [4, 5]. So, accurately identifying failure data and sensor fault is of important significance.

Some researches on fault diagnosis of sensor have been published. There are two main approaches: analytical redundancy and hardware redundancy. The analytical redundancy approach utilizes a mathematical model of the sensor output and the measured quantity, for example, a finite element model, and the redundancy is provided by the model [6–8]. On the premise that the analytical model has been created, the analytical redundancy approach can identify sensor fault ideally. However, since most of the structure health monitoring system is complex and the factors that affect the sensor output are various, it is hard for most of the multisensor systems to create the analytical models. The hardware redundancy approach adopts the fact that several sensors measure the same quantity. By analyzing the diversity of the redundant information, the sensor fault is identified [9]. This approach has been widely employed in the critical equipment such as

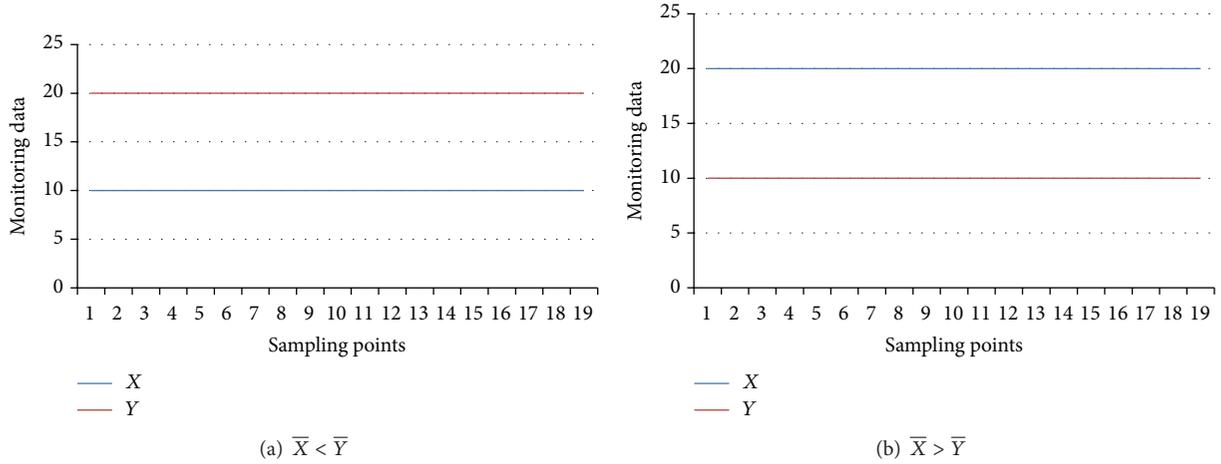


FIGURE 1: Two data vectors with the same Euclidean distance.

aircraft [10]. Due to the limitation of budget and construction, the hardware redundancy approach is almost infeasible for infrastructures such as bridges and dams.

Owing to the characteristic of the bridge structure, there are underlying nonlinear correlations between the data acquired from the different monitoring points. Thus, the data can provide “correlation redundancy.” How to identify the sensor fault by recognizing and utilizing the correlation between the data acquired by the different sensors becomes a valuable research field [11, 12]. The research on sensor fault diagnosis and validity judgment algorithm based on k -neighbor algorithm, which effectively improved the robustness of fault diagnosis algorithm [13], made a preliminary exploration in this field. This algorithm, however, still needs to first specify the target sensor to be diagnosed; if the neighboring sensor as the reference has failure or fault, it will decrease the accuracy of the diagnosis. For the structure health monitoring system including hundreds of sensors, it is necessary to explore a universal approach for sensor fault diagnosis.

In this paper, based on the study on the correlation between the sensors in the bridge health monitoring system, a clustering algorithm based on k -medoids is proposed (Figure 2), which divides the sensors of the same type into k clusters. The sensors in the same cluster are neighbors to each other. Meanwhile, the evaluation function is defined to determine the optimal k value, minimizing the correlation between the divided clusters and maximizing the correlation between the sensors in one cluster. By studying the correlation between the sensors in one cluster, the concept and corresponding algorithm of “support rating” and “sensor validity” are proposed. The algorithm has been applied to the data analysis of sensors in actual bridge health monitoring system. The results show that the abnormal sensor can be found in multiple sensors and the fault can be located in time domain. In addition, the performance of the sensor can be evaluated quantitatively, and its change curve and performance degradation process can be described. In the case that the reference sensor has failure or fault, the failure components can be eliminated, effectively ensuring the robustness of algorithm.

2. Correlation Characteristics between Sensors

2.1. Correlation between Sensors. The bridge, with a complex structure system, reacts differently to external forces. However, due to the structural relationship of the bridge itself, a certain degree of correlation between the structural parameters of different monitoring points is a must. For the monitoring points which are more closely correlated, the data collected by the sensors show the tendency of more consistency.

Suppose that there are a total of n monitoring points in a bridge and that S is the set of all the corresponding sensors. $u, v \in S$; $T = \{ti \mid i = 1, 2, v, \dots, m\}$ is the sampling time interval; X_u and X_v are m -dimensional vectors of monitoring point u and v , which are composed of m data collected in same time interval T . Higher similarity between vector X_u and X_v indicates stronger correlation between monitor point u and v .

For two m -dimensional vectors $X = \{x_1, x_2, \dots, x_m\}$ and $Y = \{y_1, y_2, \dots, y_m\}$, their Euclidean distance is [14]

$$d(X, Y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2}. \quad (1)$$

The traditional Euclidean distance is always positive. It cannot represent the comparison of two vectors. As shown in Figure 1(a), $\bar{X} < \bar{Y}$, and on the contrary, $\bar{X} > \bar{Y}$ in Figure 1(b).

The Euclidean distance between vectors X and Y in Figures 1(a) and 1(b) calculated by (1) is the same, but they cannot reveal the comparison of X and Y . Thus, the sign is added to (1):

$$d(X, Y) = \begin{cases} \sqrt{\sum_{i=1}^m (x_i - y_i)^2}, & \bar{X} \geq \bar{Y} \\ -\sqrt{\sum_{i=1}^m (x_i - y_i)^2}, & \bar{X} < \bar{Y}. \end{cases} \quad (2)$$

Suppose that S is the set of all sensors of a bridge, in time window T_i , each sensor collects m data; in l time windows,

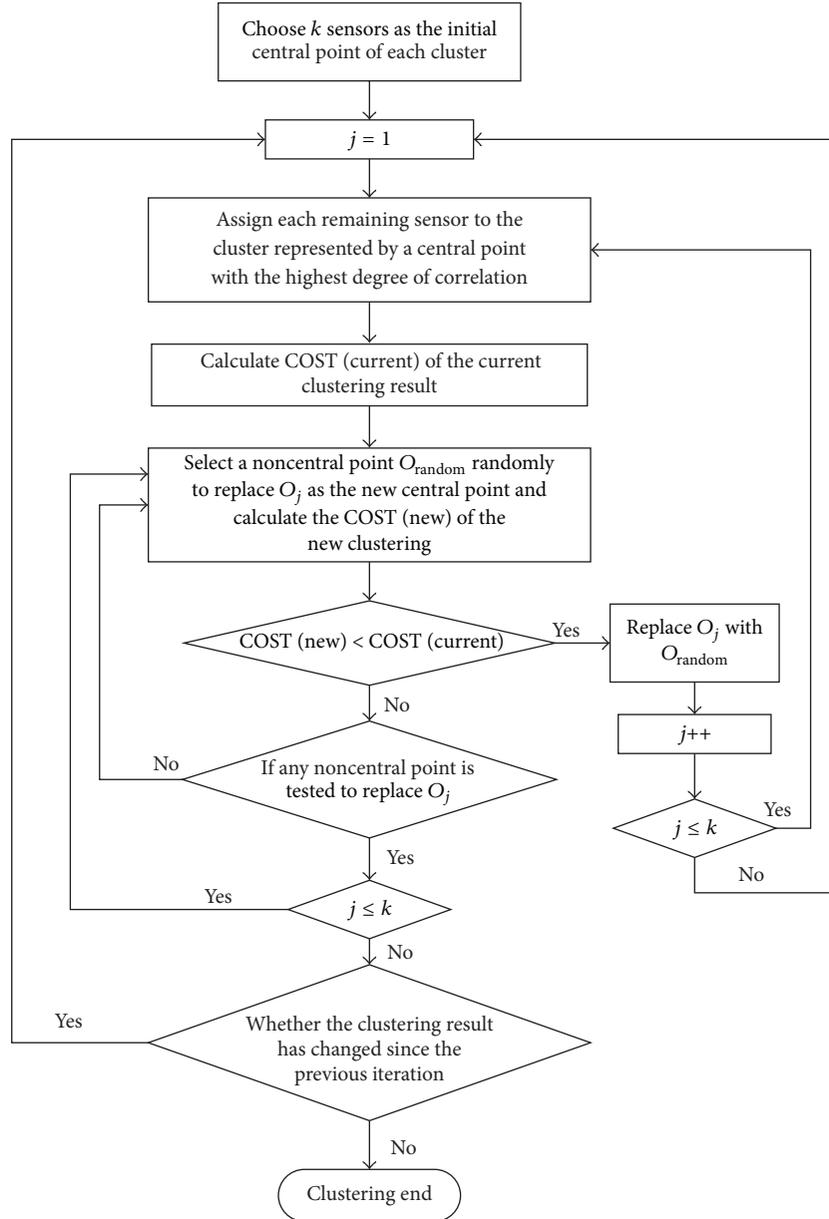


FIGURE 2: Clustering algorithm flowchart.

each sensor collects $N = m * l$ data. For any sensor P_v ($v \in S, v \neq u$), in i th time window, the Euclidean distance between the collected data and the data collected by P_u is

$$d(x_u, x_v)_i = \begin{cases} \sqrt{\sum_{j=1}^m (x_{u(m*(i-1)+j)} - x_{v(m*(i-1)+j)})^2}, & \bar{x}_u \geq \bar{x}_v \\ -\sqrt{\sum_{j=1}^m (x_{u(m*(i-1)+j)} - x_{v(m*(i-1)+j)})^2}, & \bar{x}_u < \bar{x}_v. \end{cases} \quad (3)$$

Higher correlation between monitoring point u and monitoring point v means higher similarity between data vector X_u

and data vector X_v , measured by the corresponding sensors, respectively. In different time windows, the fluctuation of $d(x_u, x_v)_i$ will be smaller. In all l time windows, the fluctuation of $d(x_u, x_v)_i$ is measured by its sample variance, $s^2(d(x_u, x_v))$ [15]:

$$s^2(d(x_u, x_v)) = \frac{1}{l-1} \sum_{i=1}^l \left(d(x_u, x_v)_i - \overline{d(x_u, x_v)} \right)^2. \quad (4)$$

Let $D(u, v) = s^2(d(x_u, x_v))$ represent the differentiation degree between data vector X_u and data vector X_v measured by sensor u and sensor v ; the smaller $D_{u,v}$ is, the smaller fluctuation of distance between X_u and X_v is. The consistency of the change tendency between X_u and X_v will be higher, which

indicates a higher correlation between monitoring point u and monitoring point v .

2.2. k -Medoid Clustering Algorithm. Suppose that S is the set of all sensors for a bridge, because of the correlation between the monitoring points, there must be some sensor clusters as well as close correlations between the data acquired by the sensors in the same cluster. k sensor clusters existing in S can be found by the strategy of clustering analysis.

First, the number of clusters " k " is determined. k sensors are chosen randomly as the central point of each cluster. And the remaining sensors are assigned to the cluster represented by a central point with the highest degree of correlation (i.e., the sample variance of the data vector distance is the smallest) according to their correlation degree with each central point. And then the noncentral point sensor is adopted to replace the existing central point repeatedly to improve the clustering quality. The clustering quality is measured by cost function COST:

$$\text{COST} = \sum_{i=1}^k \text{COST}_i, \quad (5)$$

$$\text{COST}_i = \sum D(u, v) \quad (u, v \in S_i, u \neq v).$$

COST_i represents the sum of all differentiations between the sensors in i th cluster " S_i ." The smaller COST_i is, the closer the correlation between the sensors in this cluster is. The smaller the value of COST is, the higher the quality of the current clustering result is.

In order to determine whether noncentral point " O_{random} " is a good alternative to current central point " O_j ," the following four cases are considered for each noncentral point p :

- (1) p currently belongs to cluster represented by central point " O_j ." If O_j is replaced by O_{random} as the new central point and the degree of correlation between p and O_i is the highest ($i \neq j$), p is reallocated to the cluster represented by O_i .
- (2) p currently belongs to the cluster represented by central point O_j . If O_j is replaced by O_{random} as the new central point and the degree of correlation between O_{random} and p is the highest, p is reallocated to the cluster represented by O_{random} .
- (3) p currently belongs to the cluster represented by the central point of O_i ($i \neq j$). If O_j is replaced by O_{random} as the new central point and the degree of correlation between p and O_i is the highest, the affiliation of p is unchanged.
- (4) p currently belongs to the cluster represented by central point O_i ($i \neq j$). If O_j is replaced by O_{random} as the new central point and the correlation between O_{random} and p is the highest, p is reallocated to the cluster represented by O_{random} .

The cost function will change whenever the reallocation takes place. If current central point O_j is replaced by noncentral

point O_{random} and generated new COST (new) is larger than current COST (current), it is considered that current central point O_j is acceptable. Then, nothing changed in this iteration. If new COST (new) is smaller than current COST (current), the new central point is better than O_j , so O_j is replaced.

Therefore, the clustering algorithm of sensor clusters is obtained.

2.3. Optimization of k Value. For k -medoid clustering algorithm, the key problem is the determination of k value. That is the optimal number of clusters that all the sensors are divided into. Specifically, the correlation between the clusters will be minimal, and the correlation between the sensors in the clusters will be maximal.

Let $\text{dif}(S_i, S_j)$ represent the difference between clusters S_i and S_j :

$$\text{dif}(S_i, S_j) = D(O_p, O_q), \quad (6)$$

where O_p and O_q are, respectively, the central point of clusters S_i and S_j :

$$\text{DIF} = \frac{\sum \text{dif}(S_i, S_j)}{C_k^2} \quad (7)$$

$$(i = 1, \dots, k, j = 1, \dots, k, i \neq j),$$

where $C_k^2 = k!/2(k-2)!$ is the number of pairwise combinations of k clusters. DIF is the average of the difference between each pair of clusters. A higher DIF indicates more difference among clusters of the current division.

To any cluster S_i ,

$$R_i = \frac{\sum D(u, v)}{C_n^2} \quad (u, v \in S_i, u \neq v), \quad (8)$$

where $C_n^2 = n!/2(n-2)!$ is the number of pairwise combinations of the sensors in cluster S_i . R_i is the average of the difference between each pair of n sensors in cluster S_i . A lower R_i indicates more correlation among sensors in a certain cluster.

Let

$$R = \frac{\sum_{i=1}^k R_i}{k}, \quad (9)$$

where R is the average of R_i of k clusters of the current division. A lower R indicates more correlation in each cluster of the current division.

Therefore, the evaluation function is defined as follows:

$$\text{EVA} = \frac{\text{DIF}}{R}. \quad (10)$$

The effect of the current division is improved as the value of EVA rises. Therefore, we can firstly select different k values for clustering by rule of thumb, calculate each EVA value, and finally select k with the highest EVA value as the final number of clusters.

3. Support Rating and Validity

Through using k -medoid clustering algorithm, the set of all sensors S can be divided into k sensor clusters. Each sensor belongs to one cluster, and high degree of correlation can be found between the sensors in the same cluster. That is to say, the sensors in the same cluster are mutually the neighboring sensors.

The support rating refers to the reliability of the data measured by the target sensor in some time window. it is supported by the data measured by some neighboring sensor. The validity of the data measured by the target sensor is supported by all other λ neighboring sensors in the same cluster. To ensure the validity of data from a target sensor, the weighted sum of λ support ratings must be taken into consideration.

In different time windows, the data vector distance between two neighboring monitoring points is varying. However, in the case that the damage to the bridge structure is not severe, the data vector distance between two monitoring points with close correlation will be fluctuated within the reasonable range. If so, it can be considered that the sample vector distances of two monitoring points in different time windows will fluctuate randomly around the expectation value.

Hypothesis 1 (this hypothesis is verified by using K. Pearson χ^2 method in Section 5). The data vector distance of any two closely correlated monitoring points u and v in different time windows obey normal distribution; that is, $d(x_u, x_v) \sim N(E(d(x_u, x_v)), \sigma_{uv}^2)$, where $E(d(x_u, x_v))$ is the value of expectation of $d(x_u, x_v)$ and σ_{uv}^2 is its variance. When sample size is large enough, $\bar{d}(x_u, x_v) \approx E(d(x_u, x_v))$ and $s^2(d(x_u, x_v)) \approx \sigma_{uv}^2$, where $\bar{d}(x_u, x_v)$ is the sample mean of the data vector distances in different time windows and $s^2(d(x_u, x_v))$ is the sample variance of data vector distances in different time windows.

According to normal distribution, the sample value is more likely to occur as it is closer to the mathematical expectation of the mean value. Therefore, it can be considered that the closer $\bar{d}(x_u, x_{v_j})_i$ (i is the time window number) is to $E(d(x_u, x_{v_j}))$, the higher the probability of reliability of the data of monitoring point u measured in this time window supported by the neighboring monitoring point v_j is. That is, the validity of sensor S_u supported by its neighbor monitoring point v_j . According to the distribution probability of normal distribution, the probability of $|d(x_u, x_{v_j})_i - E(d(x_u, x_{v_j}))| > 3\sigma_{uj}$ is about 0.3% [16]. Generally, the sample point with the error of less than 3σ is acceptable, compared to the expected mean value. That is, in case of $|d(x_u, x_{v_j})_i - E(d(x_u, x_{v_j}))| \leq 3\sigma_{uj}$, it can be considered that the validity of sensor S_u is supported by its neighboring monitoring point v_j in the time window i .

In case of $z_i = |d(x_u, x_{v_j})_i - E(d(x_u, x_{v_j}))|$, in i th time window, the validity of sensor S_u supported by its neighboring

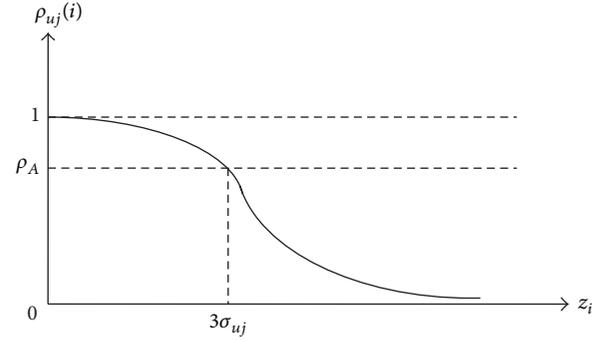


FIGURE 3: Characteristic of $\rho_{uj}(i)$.

monitoring point v_j , that is, the support rating of v_j for S_u in the time window is

$$\rho_{uj}(i) = f(z_i), \quad (0 \leq \rho_{uj}(i) \leq 1, 0 \leq z_i \leq \infty). \quad (11)$$

Formula (10) meets the following requirements:

In case of $z_i = 0$,

$$\rho_{uj}(i) = 1,$$

In case of $z_i \rightarrow \infty$,

$$\rho_{uj}(i) \rightarrow 0 \quad (12)$$

In case of $z_i = 3\sigma_{uj}$,

$$\rho_{uj}(i) = \rho_A.$$

In this equation, ρ_A is the confidence threshold of $\rho_{uj}(i)$; in case of $\rho_{uj}(i) \geq \rho_A$, it can be considered that the validity of sensor S_u is supported by its neighboring monitoring point v_j in i th time window; otherwise, the validity of S_u is not supported by v_j .

$\rho_{uj}(i)$ is the monotonically decreasing function of z_i ; when $d(x_u, x_{v_j})_i$ is closer to $E(d(x_u, x_{v_j}))$, $\rho_{uj}(i)$ is closer to 1. In case of $d(x_u, x_{v_j})_i = E(d(x_u, x_{v_j}))$, $\rho_{uj}(i) = 1$, and vice versa, $\rho_{uj}(i)$ is closer to 0.

In order to allow (11) to meet the requirements of (12), let

$$\rho_{uj}(i) = 2 - \frac{2}{1 + e^{-\beta z_i}}, \quad (13)$$

where $\beta = \ln((2 - \rho_A)/\rho_A)/3\sigma_{uj}$.

Figure 3 shows the characteristic curve of $\rho_{uj}(i)$.

In i th time window, $\gamma_u(i)$, the validity of sensor S_u is the weighted sum of support ratings of all λ neighboring monitoring points in the same cluster:

$$\gamma_u(i) = \sum_{j=1}^{\lambda} \omega_j \rho_{uj}(i), \quad (14)$$

where weight ω_j will meet the following requirements:

$$(1) \sum_{j=1}^{\lambda} \omega_j = 1.$$

- (2) For the monitoring point having closer correlation with monitoring point u , its weight is larger. That is, the smaller the value of $s^2(d(x_u, x_{v_j}))$ is, the larger ω_j is.

Thus, ω_j is defined as follows:

$$\omega_j = \frac{\prod_{p=1}^{\lambda} \varepsilon_p, (1 \leq p \leq \lambda, p \neq j)}{\sum_{h=1}^{\lambda} \left(\prod_{p=1}^{\lambda} \varepsilon_p, (1 \leq p \leq \lambda, p \neq h) \right)}, \quad (15)$$

where $\varepsilon_j = s^2(d(x_u, x_{v_j}))$.

4. Fault Diagnosis Method

Based on the definition of support rating and validity, the algorithm of sensor fault diagnosis and validity measurement is presented:

- (1) The data are selected as the samples when the system is in good working condition and the sensor performance has no recession; the reasonable time window length is set, through using k -medoid clustering algorithm, the set of all sensors S is divided into k sensor clusters.
- (2) For each sensor, on the basis of taking other sensors in the same cluster as neighboring sensors, sample mean $\bar{d}(x_u, x_{v_j})$ and sample variance $s^2(d(x_u, x_{v_j}))$ of the data vector distance between it and each nearest neighboring monitoring point v_j are obtained.
- (3) For the diagnostic data, the support rating of each neighboring monitoring point v_j within each time window for target sensor S_u ($\rho_{uj}(i)$) is calculated with (13). The validity of S_u is calculated in accordance with (14). In the case of larger sample size, $\bar{d}(x_u, x_{v_j}) \approx E(d(x_u, x_{v_j}))$, $s^2(d(x_u, x_{v_j})) \approx \sigma_{uj}^2$; so the mean value of the sample is used to replace the mathematical expectation of mean value; the sample variance is used to replace the variance for calculation.

When $\gamma_u(i) < \rho_A$, it is considered that sensor u has serious fault, so it needs examination, repair, and replacement. At this time, the support rating of u for any neighboring sensor v in the same cluster is $\rho_{vu}(i) < \rho_A$. As $\rho_{vu}(i)$ is only one of λ components of the validity of $v(\gamma_v(i))$, $\gamma_v(i)$ is still bigger than ρ_A . In order to eliminate the influence of faulty sensor on the validity evaluation of the neighboring sensor, the validity of other neighboring sensors in the same cluster can be calculated after the faulty sensor is removed.

5. Verification

Located at the central zone of main urban area of Chongqing, Caiyuanba Yangtze River Bridge is a key bridge to connect between Yuzhong district in the north and Nanan district in the south (Figure 4). It is the half-through thrust-free steel box girder tied-arch bridge. The main arch is nearly 100 meters in height with the main bridge span of 420 meters. The



FIGURE 4: A picture of the Caiyuanba Grand Bridge.

main girder is a two-story truss structure: the upper concrete bridge deck is a highway bridge deck with daily traffic flow of 60 thousand vehicles; the lower part is an urban light rail train passage with single-line weight of 400 tons. The span and height of the main arch are among the best in the world. The combined structure of steel box girder in the upper part and concrete Y structure in the lower part, the combined use of highway transportation and city light rail, the ultralong rail joints, and the whole assembly process are original designs around the globe.

The Caiyuanba Bridge health monitoring system includes five classes, eight kinds and seven subsystems of sensors concerning strain, cable force, vibration, deflection, expansion joint displacement, and three-dimensional deformation and temperature. They also monitor the strain, displacement, vibration, cable force, temperature and other structural status parameters, and environmental parameters of some key parts such as main girder, main arch, Y structure, pier, sling, and tie rod. Health monitoring system analyzes the monitoring data and implements the safety assessment and early warning of the bridge damage.

The deformation of main girder deflection mainly directly reflects the mobile load condition of the bridge as well as the linear change of longitudinal girder under mobile load condition, which is one of the most important parameters of bridge structure. The Caiyuanba Bridge health monitoring system adopts the connected-pipe photoelectric deflection sensor system to capture the deflection change of main girder, setting 15 deflection monitoring points, respectively, on the downstream and upstream sides of the steel truss girder (Figure 5). Therefore, there are 30 photoelectric deflection sensors in total. Thus, the deflection data are acquired once per hour.

Strain is an important parameter reflecting the condition of the bridge. When considering the layout of measuring points, the focus is on the stress and fatigue condition of truss girder, and the main arch as main bearing components as well as the bridge pier and Y structure. A total of 20 strain monitoring cross sections are set for the total bridge, each of which has 4 measuring points (Figure 6); in total, there are 80 strain sensors. As for deflection and strain acquisition, the data are acquired once per hour on a regular basis.

5.1. Verification of Clustering Algorithm. The data collected at the sampling time interval between September 1, 2013, and April 30, 2014, is taken as the sample; in addition, the data ($N = 5832$) are collected at each monitoring point.

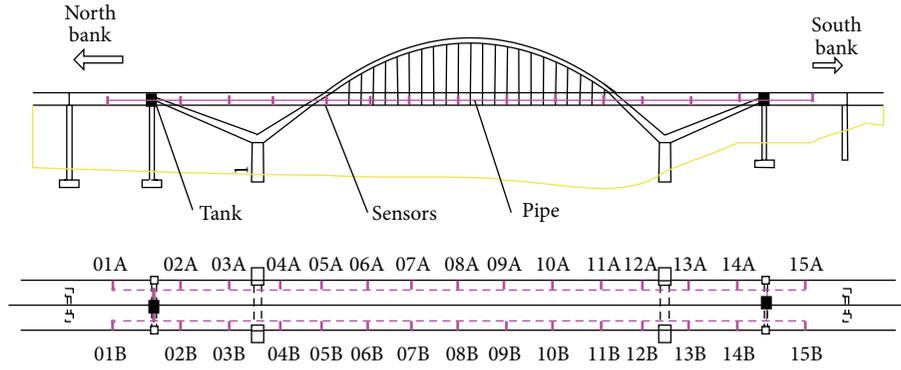


FIGURE 5: Layout of deflection checkpoints.

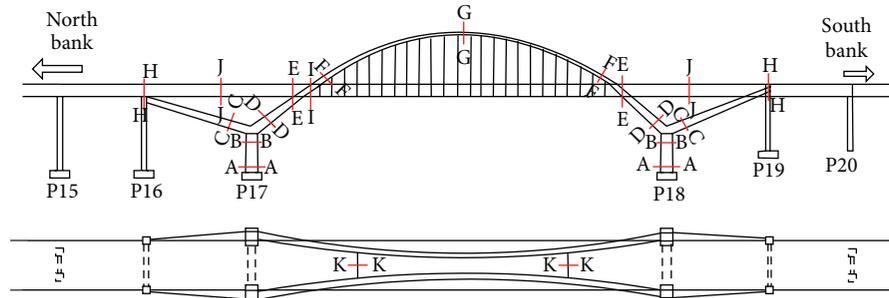


FIGURE 6: Layout of strain checkpoints.

TABLE 1: EVA of different k value.

k	6	7	8	9	10
EVA	40.7	51.6	57.3	49.4	38.6

The time window length is 24 hours; the number of samples at each monitoring point is $m = 24$; vector distance $d(x_u(i), x_v(i))_{i=1}^N$ between the acquisition data vectors of any sensor u, v ($u, v \in S$) in each time window and their sample variance $s^2(d(x_u, x_v))$ are, respectively, calculated. k -medoid clustering algorithm is used to cluster 30 deflection sensors. Generally, if the number of the sensors is n , then k , the number of the clusters, should be at interval $[2, n/2]$. Furthermore, according to the analysis of the bridge structure, set $k = 6, 7, 8, 9, 10$, and the sensors are clustered, respectively, by using k -medoids clustering algorithm; moreover, the evaluation function EVAs are calculated, respectively. The results are also listed in Table 1.

Therefore, in case of $k = 8$, the clustering result is optimal. The result is shown in Table 2.

Similarly, the data collected in the sampling time interval between September 1, 2013, and April 30, 2014, is taken as the sample. k -medoid clustering algorithm is used to cluster 80 strain sensors. According to the analysis of the bridge structure, when $k = 18, 19, 20, 21$, and 22, the sensors are clustered (Table 3), respectively; moreover, the evaluation function EVAs are calculated. The result shows that when $k = 20$, the clustering result is optimal, which is just equal to the number of strain monitoring cross sections. Furthermore,

TABLE 2: Clustering result when $k = 8$.

Cluster 1	05A 11B11A 05B
Cluster 2	14A14B 15B15A
Cluster 3	02A 02B01A
Cluster 4	03A 03B04A 04B 01B
Cluster 5	06A06B 10A 10B
Cluster 6	07A07B 09B 08B
Cluster 7	08A09A
Cluster 8	12A 12B13A 13B

TABLE 3: EVA of different k value.

k	18	19	20	21	22
EVA	67.6	103.7	127.4	119.8	98.2

four sensors for each cross section are in the same cluster. It can be seen that such clustering algorithm can better reflect the structural correlation between sensors.

5.2. Verification of Hypothesis 1. According to the idea of k -medoid clustering algorithm, there is close correlation between the sensors in the same cluster. The test method of K. Pearson χ^2 is used to test Hypothesis 1 [15].

For example, sensors 06A and 10A are neighboring sensors in the same cluster. The sample mean of the data vector distances of the 242 time windows from September 1, 2013, to April 30, 2014, is $\overline{d(x_{06A}, x_{10A})} = 8.32$. Meanwhile,

TABLE 4: K. Pearson χ^2 verification of Hypothesis 1.

Interval	m_i	p_i	np_i	$\frac{(m_i - np_i)^2}{np_i}$
$-\infty \sim 3.32$	6	0.019	4.598	0.427
$3.32 \sim 4.32$	7	0.029	7.018	0.000
$4.32 \sim 5.32$	15	0.057	13.794	0.105
$5.32 \sim 6.32$	24	0.096	23.232	0.025
$6.32 \sim 7.32$	33	0.136	32.912	0.000
$7.32 \sim 8.32$	40	0.163	39.446	0.008
$8.32 \sim 9.32$	38	0.163	39.446	0.053
$9.32 \sim 10.32$	34	0.136	32.912	0.036
$10.32 \sim 11.32$	25	0.096	23.232	0.135
$11.32 \sim 12.32$	14	0.057	13.794	0.003
$12.32 \sim 13.32$	8	0.029	7.018	0.137
$13.32 \sim +\infty$	4	0.019	4.598	0.078
Sum	242	1	242	1.008

the sample variance is $s^2(d(x_{06A}, x_{10A})) = 5.72$. Thus, the hypothesis is $H_0: d(x_{06A}, x_{10A}) \sim N(8.32, 5.72)$, and its verification is shown as follows.

According to the probability density, $\varphi(x) = (1/\sqrt{5.72 \times 2\pi})e^{-(x-8.32)^2/11.44}$, p_i , the probability of $d(x_{06A}, x_{10A})$ at different time intervals and the diversity χ^2 are calculated.

In Table 4, m_i is the frequency of $d(x_{06A}, x_{10A})$ at different intervals, n is the total number of time windows, and $\chi^2 = \sum_{i=1}^l ((m_i - np_i)^2 / np_i) = 1.008$. Because the number of intervals is $l = 12$, and the number of parameters estimated by the sample data is 2, the free degree of χ^2 is $9 = l - 2 - 1$. Under the condition with a significant level of 0.05 and a free degree of 9, $\chi_{0.05}^2(9) = 16.9$, $\chi^2 = \sum_{i=1}^l ((m_i - np_i)^2 / np_i) = 1.008 < 16.9$. Consequently, the hypothesis H_0 is acceptable.

In total, 44 pairs of sensor combinations in 8 clusters are tested. A total of 43 tests are conducted under the condition with a significant level of 0.05. It can be considered that Hypothesis 1 proves right.

5.3. Verification of Fault Diagnosis Algorithm. After the division of sensor clusters is determined, the support degree between the neighboring sensors in the cluster can be calculated according to (13), and the validity of each sensor is calculated according to (14).

Taking the deflection sensor cluster 1 as an example, the data collected for 1440 times in 60 time windows in the interval from June 1, 2014, to July 30, 2014, are used in the test, which is shown in Figure 7.

According to (12) and Figure 3, if the value of confidence threshold ρ_A is too high, the curve of $\rho_{uj}(i)$ will be too flat when $z_i \leq 3\sigma_{uj}$ and too steep when $z_i > 3\sigma_{uj}$. On the contrary, if ρ_A is too low, the curve of $\rho_{uj}(i)$ will be too steep when $z_i \leq 3\sigma_{uj}$ and too flat when $z_i > 3\sigma_{uj}$. Generally speaking, ρ_A is set in the interval $[0.6, 0.75]$.

In this test, the confidence threshold value is set: $\rho_A = 0.7$. On this basis, the support ratings between various sensors in

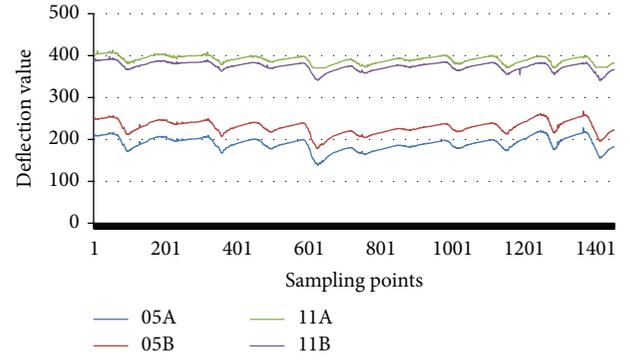


FIGURE 7: Consecutive historical data acquired by sensors in deflection cluster 1.

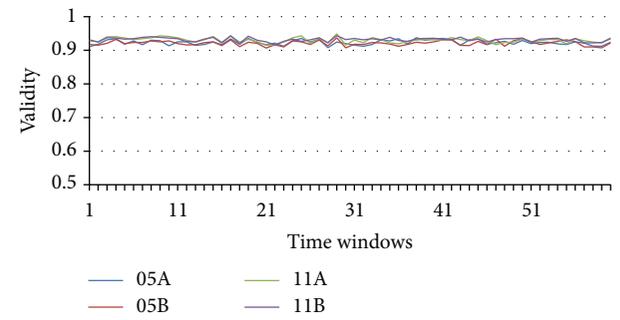


FIGURE 8: Validity curve of sensors in deflection cluster 1.

the cluster are calculated, and then the validity of each sensor is obtained too.

As shown in Figure 8, the validity of each sensor is larger than 0.9; the sensors can be considered to be in good working condition.

The data collected by strain sensors in cluster 3 from June 1, 2014, to July 30, 2014, is shown in Figure 9.

According to (13) and (14), the support ratings between various sensors in the cluster are calculated, and then the validity of each sensor is obtained.

It can be seen from Figure 10 that the validity of strain sensor N11.B began to have a significant decline in late July, even down to $\rho_A = 0.7$ in mid-June. It can be caused by the failure of strain sensor N11.B or the damage of the bridge structure. After the on-site inspection, the sensor's failure was confirmed.

At this time, as the data measured by sensor N11.B become unbelievable, the degree of support for other sensors in strain cluster 3 is also no longer reliable. Affected by it, validities of N11.A, N11.C, and N11.D decline to some degree. Therefore, before the validities of N11.A, N11.C, and N11.D are calculated, the support rating component of N11.B for its target sensor will be removed in (14). The obtained results are shown in Figure 11. After the support rating component of the failed sensor is eliminated, the validity of other sensors in the cluster resumes to be more than 0.9 again.

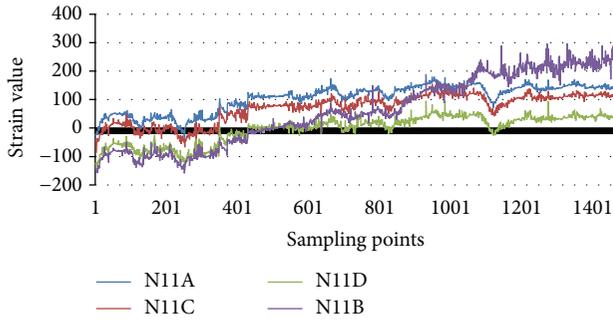


FIGURE 9: Consecutive historical data acquired by sensors in strain cluster 3.

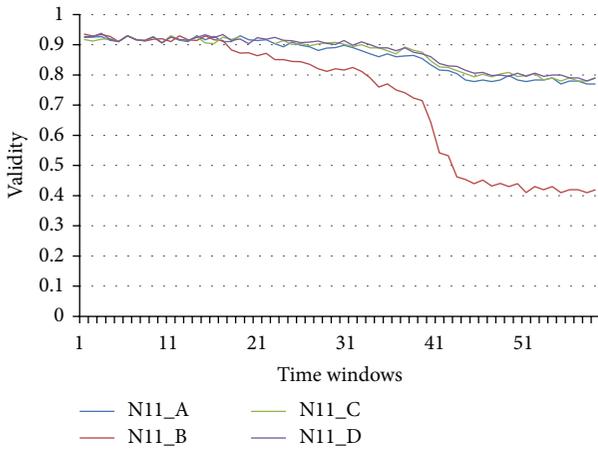


FIGURE 10: Validity curve of sensors in strain cluster 3.

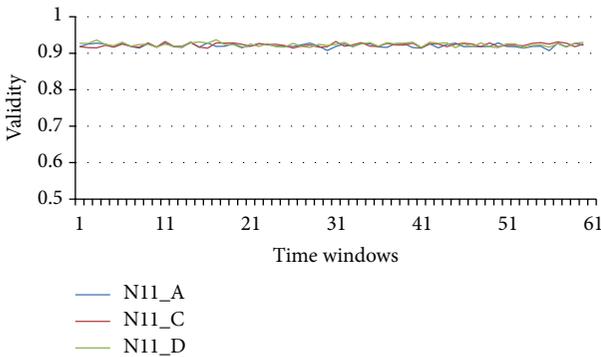


FIGURE 11: Validity curve of sensors with component of failed sensor eliminated.

6. Conclusion

Based on the correlation of the data collected by sensors, the sensors in the bridge monitoring system can be divided into multiple sensor clusters through k -medoid clustering algorithm. The data collected by sensors in the same cluster are closely correlated. With the evaluation function, the number of clusters k can be evaluated and optimized to ensure the

optimal division (most correlated inside the cluster; most sparse between the clusters).

According to the results of experiments on various bridge sensors in the actual bridge health monitoring system, the validity of target sensor can be obtained by combining the support ratings of neighboring sensors for the target sensor. Therefore, we can measure the performance of sensor and describe its change track and performance of decline process as well as locate the faults of sensors in the time domain. No need to specify the target sensor to be diagnosed in advance helps effectively avoid the diagnostic error caused by the failure of the reference sensor and the robustness of the algorithm. When the validity of sensor is calculated, the accuracy and robustness in quantitative evaluation on sensor performance are further improved owing to eliminating the support component of failure sensor.

Competing Interests

The authors declare that they have no competing interests.

References

- [1] L. Xu, Z. M. Qu, J. Q. Zheng, and Y. Liu, "Concrete cracks monitoring of a practical bridge by using structural health monitoring technique," *Key Engineering Materials*, vol. 648, pp. 1–8, 2015.
- [2] J. Li, H. Hao, K. Fan, and J. Brownjohn, "Development and application of a relative displacement sensor for structural health monitoring of composite bridges," *Structural Control & Health Monitoring*, vol. 22, no. 4, pp. 726–742, 2015.
- [3] L. Liu, X. W. Huang, W. M. Chen et al., "Design of short message sender in bridge health monitoring-control alarm system," *Chinese Journal of Scientific Instrument*, vol. 29, no. 4, pp. 483–486, 2008.
- [4] R. C. Tennyson, A. A. Mufti, S. Rizkalla, G. Tadros, and B. Benmokrane, "Structural health monitoring of innovative bridges in Canada with fiber optic sensors," *Smart Materials and Structures*, vol. 10, no. 3, pp. 560–573, 2001.
- [5] S. Jang and B. F. Spencer, "Structural health monitoring for bridge structures using smart sensors," NSEL Report NSEL-035, 2015.
- [6] G. Schreier and P. M. Prank, "Fault tolerant ship propulsion control: sensor fault detection using a nonlinear observer," in *Proceedings of the European Control Conference (ECC '99)*, pp. 4586–4591, IEEE, August–September 1999.
- [7] A. F. De Loza, J. Cieslak, D. Henry, J. Dávila, and A. Zolghadri, "Sensor fault diagnosis using a non-homogeneous high-order sliding mode observer with application to a transport aircraft," *IET Control Theory & Applications*, vol. 9, no. 4, pp. 598–607, 2015.
- [8] P. Goupil, "Oscillatory failure case detection in the A380 electrical flight control system by analytical redundancy," *Control Engineering Practice*, vol. 18, no. 9, pp. 1110–1119, 2010.
- [9] J. Kullaa, "Detection, identification, and quantification of sensor fault in a sensor network," *Mechanical Systems & Signal Processing*, vol. 40, no. 1, pp. 208–221, 2013.
- [10] J. Marzat, H. Piet-Lahanier, F. Damiogéot, and É. Walter, "Fault diagnosis for nonlinear aircraft based on control-induced redundancy," in *Proceedings of the 1st Conference on Control and*

- Fault-Tolerant Systems (SysTol '10)*, pp. 119–124, Nice, France, October 2010.
- [11] S. Hu, W. M. Chen, P. Zhang, and Y. Zhang, “Research of continuous invalid data identification based on multi-informaton correlation analysis,” *Chinese Journal of Scientific Instrument*, pp. 72–74, 2008.
 - [12] V. Reppa, M. M. Polycarpou, and C. G. Panayiotou, “Decentralized isolation of multiple sensor faults in large-scale interconnected nonlinear systems,” *IEEE Transactions on Automatic Control*, vol. 60, no. 6, pp. 1582–1596, 2015.
 - [13] X. W. Huang, W. M. Chen, P. Zhang, and Sh. R. Hu, “Research on deflection sensor validity based on K -neighbor algorithm,” *Chinese Journal of Scientific Instrument*, vol. 33, no. 5, pp. 1090–1095, 2012.
 - [14] J. W. Han and M. Kamber, *Data Mining Concept and Techniques*, Morgan Kaufmann, 2nd edition, 2006.
 - [15] R. H. Sun, H. Y. Yi, and Q. S. Liu, *Mathematical Statistics*, Chongqing University Press, 2000.
 - [16] H. Yan, Q. S. Liu, and B. Zh, *Mathematical Statistics*, Higher Education Press, 2004.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

