Journal of Southeast University (English Edition) Vol. 40, No. 2, pp. 140 − 147 June 2024 ISSN 1003—7985

Sensor fault self-detection based on the mean shift method

Chen Hao[1] Zhu Yikai[1] Lei Bo[2] Weng Zhihai[3] Xu Hongchang[2] Wan Huaping[1]

([1]College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China)
([2]China Construction Third Engineering Bureau Co., Ltd., Wuhan 430064, China)
([3]Huzhou City Investment and Development Group Co., Ltd., Huzhou 313000, China)

Abstract: To accurately identify sensor faults caused by complex environmental conditions and ensure that structural health monitoring systems correctly perceive the structural state, a self-detection method for sensor nodes based on mean shift and sliding window techniques was proposed. The self-detection method comprises two stages, i.e., fault prescreening and fault self-detection. During the fault prescreening stage, the method rapidly identifies potentially abnormal data using the quartile method combined with the sliding window technique, significantly improving the efficiency of the method. During the fault self-detection stage, the method employs the mean shift algorithm to perform adaptive clustering of the abnormal data, effectively detecting various faults. Data from the Canton Tower were used to test the effectiveness of the method by setting four types of sensor faults, i.e., offset, drift, gain, and stuck. Then, the proposed method was compared with extremely randomized trees, random forests, support vector data description, and one-class support vector machines. Results show that the proposed method can detect the four aforementioned faults with high accuracy and computational efficiency.

Key words: sensor fault detection; mean shift; sliding window; machine learning

DOI: 10.3969/j.issn.1003 − 7985.2024.02.004

Structural health monitoring provides real-time insights into structural performance and serves as a vital technology for ensuring the safety of civil engineering structures[1-4]. The effectiveness of structural health monitoring systems relies on reliable sensor data[5]. In reality, sensors are exposed to challenging conditions, such as high temperature and corrosion[6-7]. This exposure causes various failures, such as calibration errors, battery malfunctions, and hardware defects, which produce abnormal data or false alerts[8-10]. Consequently, prompt and precise detection of sensor faults is crucial[11-12].

Sensor fault detection techniques are commonly divided into centralized and distributed methods[9]. In centralized methods, a sink node or base station first processes data from all sensor nodes and then broadcasts the diagnostic findings across the health monitoring network[13]. Lau et al.[14] introduced the centralized naïve Bayes detector method, which detects sensor faults by analyzing the end-to-end transmission time collected at the sink. Abid et al.[15] employed a centralized technique using the k-nearest neighbor and Euclidean distance. This technique ranks and clusters data from various sensor nodes to identify offset and gain faults. Warriach and Tei[16] adopted a centralized strategy with hidden Markov models, where a model developed from a training set effectively identifies offset, gain, and stuck faults. Centralized methods increase network traffic, potentially causing congestion and detection delays.

Distributed methods either function cooperatively with neighboring nodes or operate independently. Panda and Khilar[17] proposed a distributed fault detection algorithm that identifies faults by calculating the average value from neighboring nodes. Chen et al.[18] introduced a localized distributed fault detection algorithm that employs majority voting techniques to compare data from both local and neighboring sensor nodes, thereby detecting soft permanent faults. Obst[19] introduced the spatially organized distributed echo state network method, which exploits the spatiotemporal correlations among various sensors to identify multiple faults. Younis et al.[20] proposed a robust energy-efficient distributed clustering approach to detect faults through consistent communication between the cluster head nodes and the nodes within the cluster. Distributed methods reduce communication with the sink node or base station; however, communication between nodes leads to additional energy consumption and detection delays. Hence, proposing a new self-detection method for faults is imperative.

This study introduces a sliding mean shift (S-MS) fault self-detection method based on the mean shift algorithm with sliding window technology. The proposed S-MS method contains two pivotal phases, i.e., the fault prescreening phase and the fault self-detection phase. This method is insensitive to time variations and effectively detects four fault types (i.e., offset, drift, gain, and stuck) at different fault rates. This method institutes a

Received 2023-10-15, **Revised** 2023-12-21.

Biographies: Chen Hao (2000—), male, Ph.D. candidate; Wan Huaping (corresponding author), male, doctor, professor, hpwan@zju.edu.cn.

Foundation items: The National Key Research and Development Program of China (No. 2021YFF0501001), Zhejiang Provincial Natural Science Foundation (No. LR23E080003).

Citation: Chen Hao, Zhu Yikai, Lei Bo, et al. Sensor fault self-detection based on the mean shift method[J]. Journal of Southeast University (English Edition), 2024, 40(2): 140 − 147. DOI: 10.3969/j.issn.1003 −7985.2024.02.004.

fault prescreening phase by incorporating the quartile method, effectively enhancing the operational efficiency of the S-MS method.

1 Fault Taxonomy

As delineated in the introduction, a sensor malfunctions for various reasons. Faults can be stratified into permanent, intermittent, and transient categories, depending on their persistence. Permanent faults encompass software or hardware errors that perpetually yield anomalies when fully operational[21]. Mahapatro and Khilar[9] posited that a diminution in battery voltage could precipitate calibration issues, subsequently inducing sensor drift. Sensors experiencing calibration errors are categorized as having permanent faults. Ni et al.[22] explored three distinct types of calibration faults, i. e., offset, gain, and drift faults. In this study, based on the amassed data, the stuck fault is designated as a permanent fault. Subsequent sections delineate the four quintessential sensor faults, i. e., offset, drift, gain, and stuck faults, as shown in Fig. 1.

1) An offset fault may occur when the sensor unit is improperly calibrated, specifically manifesting as the addition of a fixed constant to the normal data. This fault can be modeled as follows:

$$S_n^{\text{offset}} = S_n^{\text{normal}} + b \tag{1}$$

where n is the sensor node identification, S_n^{offset} is the data collected by the n-th sensor after the occurrence of an offset fault, S_n^{normal} is the regular data collected by the n-th sensor, and b is a fixed constant.

2) A drift fault occurs when the sensing data of the sensor linearly increases over time from the normal state, specifically manifesting as the addition of a linearly increasing bias term to the normal data. This fault can be modeled as follows:

$$S_n^{\text{drift}} = S_n^{\text{normal}} + b_n, \quad b_n = tb \tag{2}$$

where S_n^{drift} is the data collected by the n-th sensor after the occurrence of a drift fault, b_n is the linearly increasing bias term, t is the time.

3) A gain fault emerges when the sensing data of the sensor deviates from the expected value by changing at a rate that is a specific multiple of the normal state, specifically manifesting as normal data multiplied by a fixed constant. This fault can be modeled as follows:

$$S_n^{\text{gain}} = bS_n^{\text{normal}} \tag{3}$$

where S_n^{gain} is the data collected by the n-th sensor after the occurrence of a gain fault.

4) A stuck fault occurs when the rate of change in the sensing data of the sensor becomes zero, causing all data to manifest as a fixed value. This fault can be modeled as follows:

$$S_n^{\text{stuck}} = b \tag{4}$$

where S_n^{stuck} is the data collected by the n-th sensor after the occurrence of a stuck fault.

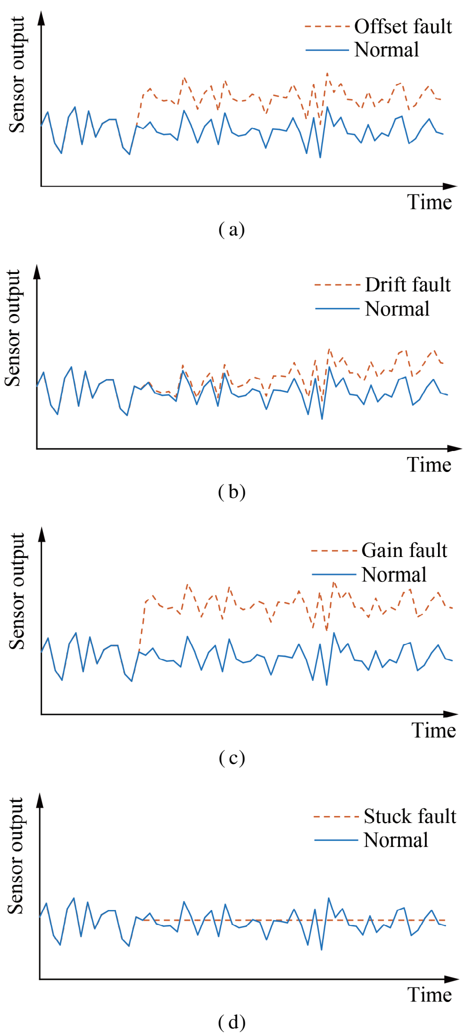

Fig. 1 Example of normal signals and faulty signals. (a) Offset; (b) Drift; (c) Gain; (d) Stuck

2 Sliding Mean Shift Method

The mean shift algorithm is a nonparametric density estimation technique that is widely employed for clustering and image processing tasks[23-24]. The mean shift algorithm neither necessitates the prespecification of the number of clusters nor requires a predefined data distribution form; therefore, it is particularly apt for solving complex, nonconvex clustering issues.

The S-MS method is an innovative approach that synergizes the mean shift algorithm with the sliding window technique and is tailored for node fault self-detection. The S-MS method principally unfolds through two critical phases, i. e., fault prescreening and fault self-detection.

2.1 Fault prescreening

As an effective method for time series analysis, the sliding window technique can break down large datasets into small and manageable parts to meet the processing needs of real-time monitoring data. During the fault prescreening phase, the sliding window technique is used to compute the median within each sliding window and the difference between consecutive medians, as follows:

$$M = \begin{cases} x_{\frac{n+1}{2}} & n = 2k+1; k = 0, 1, 2, \dots \\ \dfrac{x_{\frac{n}{2}} + x_{\frac{n+1}{2}}}{2} & n = 2k; k = 1, 2, \dots \end{cases} \tag{5}$$

$$D_{\mathrm{s}} = (d_{(1)}, d_{(2)}, \dots, d_{(n)}),\ d_i = |M_i - M_{i-1}| \tag{6}$$

where M is the median of the sliding window, d_i is the difference between the medians of consecutive sliding windows, and D_{s} is the set of differences.

To identify potential outliers or abrupt changes, the quartile method is employed to calculate the upper threshold value of the differences between the medians of consecutive sliding windows and functions as the activating mechanism for the mean shift algorithm throughout the fault self-detection phase, as follows:

$$k = q(n-1) + 1 \tag{7}$$

$$Q(D_{\mathrm{s}}, i) = \begin{cases} D_{\mathrm{s}}[f] & k \in \mathbf{Z} \\ D_{\mathrm{s}}[f] + (k-f)(D_{\mathrm{s}}[c] - D_{\mathrm{s}}[f]) & k \notin \mathbf{Z} \end{cases} \tag{8}$$

$$G_\lambda = Q_3 + 4I_{\mathrm{QR}},\ I_{\mathrm{QR}} = Q_3 - Q_1 \tag{9}$$

$$\mathrm{anomaly}(i) = \begin{cases} \text{True} & d_i > G_\lambda \\ \text{False} & \text{otherwise} \end{cases} \tag{10}$$

where k is the index value, q is the quartile value, I_{QR} is the interquartile range, and G_λ is the upper threshold value for the difference between the medians of consecutive sliding windows.

The fault prescreening stage aims to identify suitable thresholds to preliminarily determine whether the data are abnormal. Eq. (10) indicates that when the difference between the medians of consecutive sliding windows surpasses the upper threshold, i. e., when $d_i > G_\lambda$, a node fault has transpired during the fault prescreening phase, subsequently initiating the mean shift algorithm for fault detection.

2.2 Fault self-detection

During the fault self-detection phase, the S-MS method incorporates the sliding window technique, multiplies the mean value of the sliding window variance by a correction coefficient, and employs the product as a threshold for detecting stuck faults, as follows:

$$\sigma_i = \sqrt{\frac{1}{w}\sum_{j=i-w+1}^{i}(x_j - \mu_i)^2},\quad \mu_i = \frac{1}{w}\sum_{j=i-w+1}^{i} x_j \quad i = w, w+1, \dots, n \tag{11}$$

$$G_{\mathrm{std}} = \alpha \frac{1}{n}\sum_{i=1}^{n}\sigma_i \tag{12}$$

$$s_{\mathrm{tuck}} = \begin{cases} \text{True} & \sigma_i < G_{\mathrm{std}} \\ \text{False} & \text{otherwise} \end{cases} \tag{13}$$

where μ_i is the average value of the i-th sliding window, σ_i is the variance of the i-th sliding window, α is the correction coefficient, and G_{std} is the detection threshold. Eq. (13) indicates that when the variance of the sliding window falls below the detection threshold, the S-MS method determines that a stuck fault has occurred in the node.

In addition, the S-MS method utilizes the mean shift algorithm to detect other fault types. The primary steps are as follows:

1) Using the mean shift algorithm, the initial cluster center, denoted as $C_{\mathrm{initial}} = (x_0, y_0, z_0)$, is derived from a standard dataset.

2) When the n-th dataset is normal, the mean shift algorithm computes a new initial cluster center, denoted as $C_{\mathrm{new\_initial}} = (x_i, y_i, z_i)$, supplanting the preceding initial cluster center $C_{\mathrm{initial}} = (x_0, y_0, z_0)$. Should the n-th dataset manifest anomalies, the algorithm will determine two or more cluster centers, denoted as $C_1 = (x_1, y_1, z_1)$ and $C_2 = (x_2, y_2, z_2)$.

3) Compute the Euclidean distance between the two cluster centers and the new initial cluster center, denoted as D_1 and D_2.

4) Determine the new initial cluster center based on the computed distance and subsequently assign a label to each cluster (i. e., normal or abnormal).

$$(C_{\mathrm{new\_initial}}, C_{\mathrm{anomaly}}) = \begin{cases} (C_1, C_2) & D_1 \leqslant D_2 \\ (C_2, C_1) & D_1 > D_2 \end{cases} \tag{14}$$

Eq. (14) indicates that, by comparing the Euclidean distances between the two cluster centers and the new initial cluster center, the S-MS algorithm identifies the more distant cluster center as abnormal and substitutes the new initial cluster center with the nearer cluster center.

3 Methodological Validation

3.1 Description of the dataset and evaluation metrics

The Canton Tower has a total height of 600 m. The antenna mast and the primary structure of the tower are 146 and 454 m high, respectively. As shown in Fig. 2, the principal structure is equipped with 20 acceleration sensors, each with a sampling frequency of 50 Hz. To validate the effectiveness of the S-MS method, this study utilizes 60 000 actual acceleration measurements recorded over a 20-min period from Sensor 1 on June 23, 2011[25-26].

To assess the S-MS method, the dataset described previously was partitioned chronologically into 40 groups.

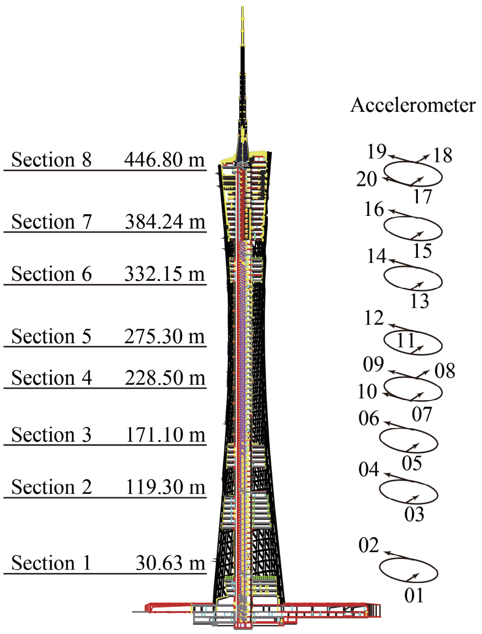

Fig. 2 Deployment of accelerometers on the Canton Tower

Each group contains 1 500 acceleration readings, corresponding to data accumulated by the acceleration sensor over a 30-s span. A total of 800 groups are prepared, each exhibiting varying fault rates (i. e., 50%, 40%, 30%, 20%, and 10%) and distinct fault types (i. e., offset, drift, gain, and stuck).

In this study, three performance metrics, i. e., accuracy (A), precision(P), and F_1 score, were used to evaluate the detection capabilities of the S-MS method. A measures the proportion of samples that are correctly classified, as expressed in Eq. (15). P measures the ratio of the true positives to the total predicted as positives, as expressed in Eq. (16). The F_1 score is the harmonic mean of precision and recall, as expressed in Eq. (17).

$$A = \frac{T_{\mathrm{P}} + T_{\mathrm{N}}}{T_{\mathrm{P}} + T_{\mathrm{N}} + F_{\mathrm{P}} + F_{\mathrm{N}}} \tag{15}$$

$$P = \frac{T_{\mathrm{P}}}{T_{\mathrm{P}} + F_{\mathrm{P}}} \tag{16}$$

$$F_1 = \frac{T_{\mathrm{P}}}{T_{\mathrm{P}} + (F_{\mathrm{P}} + F_{\mathrm{N}})/2} \tag{17}$$

where T_{P} is the true positive, which means that the actual condition and predicted result are abnormal; F_{N} is the false negative, which means that the actual condition is abnormal but the predicted result is normal; T_{N} is the true negative, which means that the actual condition and the predicted result are normal; and F_{P} is the false positive, which means that the actual condition is normal but the predicted result is abnormal.

3.2 Comparison with other sensor fault detection methods

To comprehensively assess the effectiveness of the proposed S-MS method, multiple existing sensor fault detection methods are used for comparison. The extremely randomized trees (ET) method for fault detection was reconstructed, according to Saeed et al. [27]. Similarly, the random forest (RF) method was adopted, as described by Noshad et al. [28]. In addition, a comparative analysis was conducted using two unsupervised methods, i. e., support vector data description (SVDD) and one-class support vector machine (OC-SVM).

For the ET and RF methods, the initial 10 groups of abnormal data were utilized as a training set to construct models, predicting data from Groups 11 to 15 with varying fault types and rates. Similarly, for the SVDD and OC-SVM methods, the initial 10 groups of normal data were used as a training set to construct models, predicting data from Groups 11 to 15 with varying fault types and rates. The S-MS method is an adaptive density clustering approach that identifies anomalies through the distances between cluster centroids. The S-MS method does not require extensive data to train the model and needs only a small amount of normal data to obtain the initial normal cluster centroid $C_{\mathrm{initial}} = (x_0, y_0, z_0)$. For the S-MS method, the first group of normal data was employed as a training set, predicting data from Groups 11 to 15.

Fig. 3 illustrates that, for offset faults, the S-MS, ET, and RF methods achieve the highest accuracy rates, outperforming the SVDD and OC-SVM methods. Specifically, the accuracy rates of the S-MS method at Indices 2, 4, and 5 are comparable to those of the RF method, exceeding the ET method by 0.2%. However, in Indices 1 and 3, the accuracy rates of the S-MS method are smaller than those of both the ET and RF methods because the training sets used for these two methods contain more representative samples. Moreover, both methods are more sensitive to the characteristics of offset faults. For drift faults, except at Index 4, the accuracy rates of the S-MS, ET, and RF methods surpass those of the SVDD and OC-SVM methods. Notably, the S-MS method achieves accuracy rates at Indices 1, 3, 4, and 5 that are comparable to or higher than those of the ET and RF methods but are lower at Index 2. For gain faults, the S-MS, ET, and RF methods outperform the SVDD and OC-SVM methods, although the S-MS method has slightly lower accuracy rates than the ET and RF methods. For stuck faults, the S-MS method is particularly superior, whereas the other methods perform moderately.

Fig. 4 illustrates that, for offset faults, the S-MS, ET, and RF methods have higher precision values, outperforming the SVDD and OC-SVM methods. The precision values of the S-MS method at Indices 4 and 5 are comparable to those of the ET and RF methods but are slightly lower at Indices 1, 2, and 3. For drift faults, the S-MS, ET, and RF methods surpass the SVDD and OC-SVM methods, with the S-MS method showing superior results,

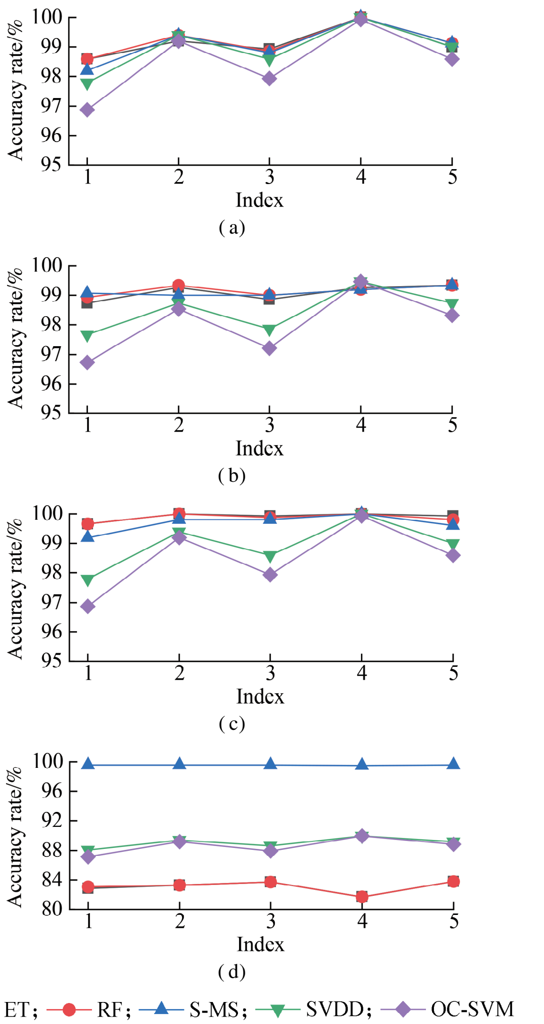

Fig. 3 Accuracy rate of different methods. (a) Offset; (b) Drift; (c) Gain; (d) Stuck

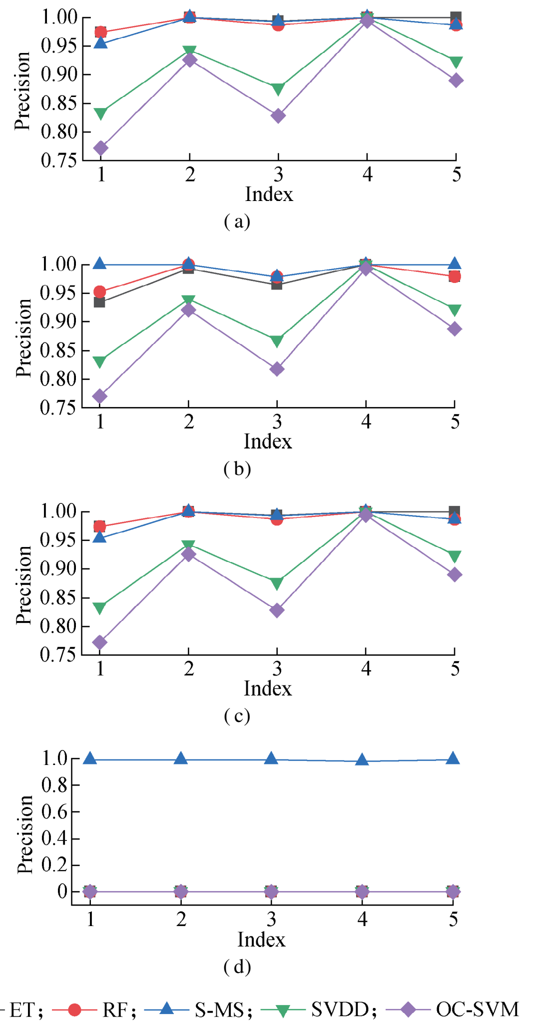

Fig. 4 Precision of different methods. (a) Offset; (b) Drift; (c) Gain; (d) Stuck

exceeding the ET and RF methods by approximately 0.03. For gain faults, the accuracy rates of the SVDD and OC-SVM methods fluctuate considerably, whereas those of the S-MS, ET, and RF methods maintain higher consistency. The S-MS, ET, and RF methods perform better than the SVDD and OC-SVM methods, and the S-MS method achieves the highest precision values at Indices 2, 3, and 4, exceeding the ET and RF methods by 0.01. For stuck faults, the S-MS method is particularly superior, whereas the other methods perform poorly.

Fig. 5 illustrates that, for offset faults, the S-MS, ET, and RF methods achieve the highest F_1 scores, outperforming the SVDD and OC-SVM methods. The F_1 scores of the S-MS method at Indices 2, 4, and 5 are comparable to those of the RF method and are approximately 0.6% higher than those of the ET method but slightly lower at Indices 1 and 3. For drift faults, the F_1 scores of the S-MS, ET, and RF methods are higher than those of the SVDD and OC-SVM methods. Notably, the F_1 scores of the S-MS method at Indices 1 and 3 match those of the RF method and are approximately 1% better than those of the ET method but are lower at Indices 2, 4, and 5. For gain faults, the F_1 scores of the S-MS method exceed those of the SVDD and OC-SVM methods but are lower than those of the ET and RF methods. Compared with those of the SVDD and OC-SVM methods, the F_1 scores of the S-MS, ET, and RF methods exhibit higher stability. For stuck faults, the S-MS method performs well, whereas the other methods perform poorly.

4 Performance Analysis

4.1 Evaluation of the computational efficiency

To comprehensively evaluate the computational efficacy of the S-MS method, the method was tested on a system equipped with an Intel Core i9 CPU and 16 GB of RAM. For the S-MS method, the first group of normal data was utilized as a training set to construct the model, predicting data from Groups 11 to 40 with varying fault types and rates. The test set covered five fault rates (i.e., 50%, 40%, 30%, 20%, and 10%) and four fault types (i.e., offset, drift, gain, stuck) for a total of 600 groups.

Fig. 6 shows the total time taken by the S-MS method

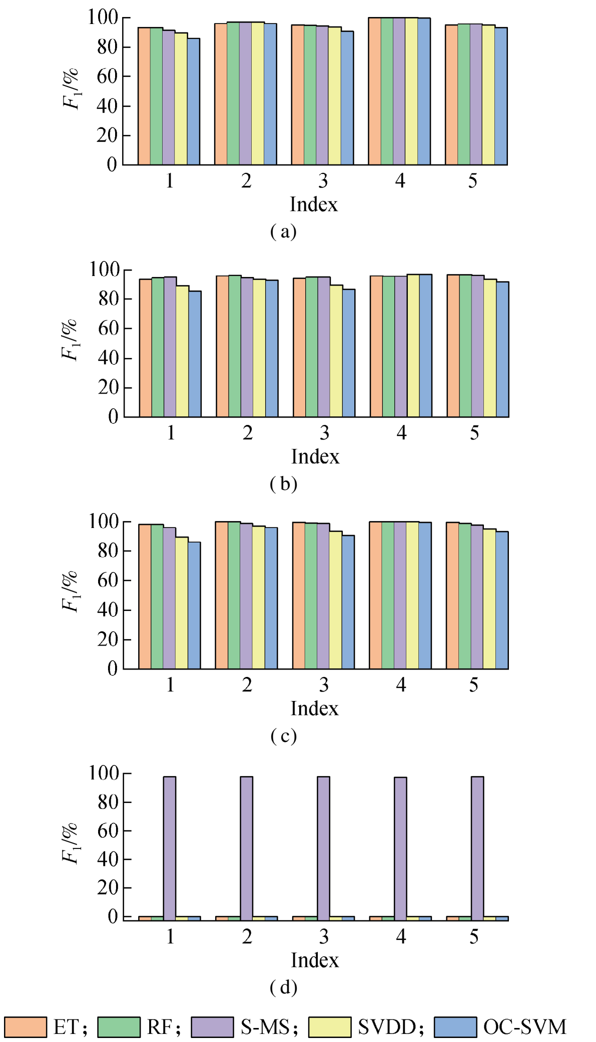

Fig. 5 F_1 score of different methods. (a) Offset; (b) Drift; (c) Gain; (d) Stuck

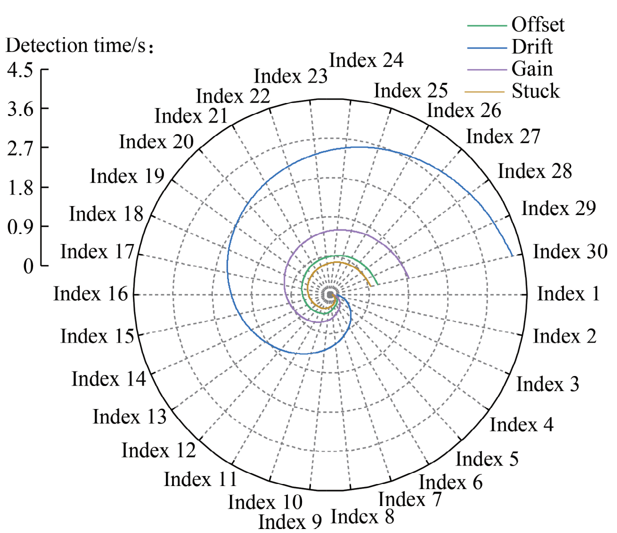

Fig. 6 Detection time of the S-MS method

to detect faults across abnormal data groups, reflecting the combined detection time for all groups. The combined detection time for all 30 abnormal data groups was 900 s. In this scenario, where every group has faults, the S-MS method efficiently identifies all 30 abnormal data groups within a time range of 1.0-4.5 s, constituting only 0.11% to 0.5% of the total sensor data collection time.

4.2 Impact of the fault rates on the performance of the method

Subsequently, this study examined the impact of different fault rates on the performance of the method. Only the first group of normal data was used as the training set, with predictions made for faults in Groups 11 to 40.

Fig. 7 shows distinct shifts in the F_1 score of the S-MS method across groups with various fault rates, indicating that, within a single abnormal data group, an increased fault rate aligns with a higher F_1 score of the S-MS method. For diverse abnormal data groups, the F_1 score ranges from 93.71% to 99.73%. The contrasts in F_1 scores for groups with different fault rates are significant. Overall, the S-MS method consistently maintains high F_1 scores across various abnormal data groups, exhibiting only a slight effect on time variations. In addition, as the fault rate increases, the S-MS method achieves higher F_1 scores.

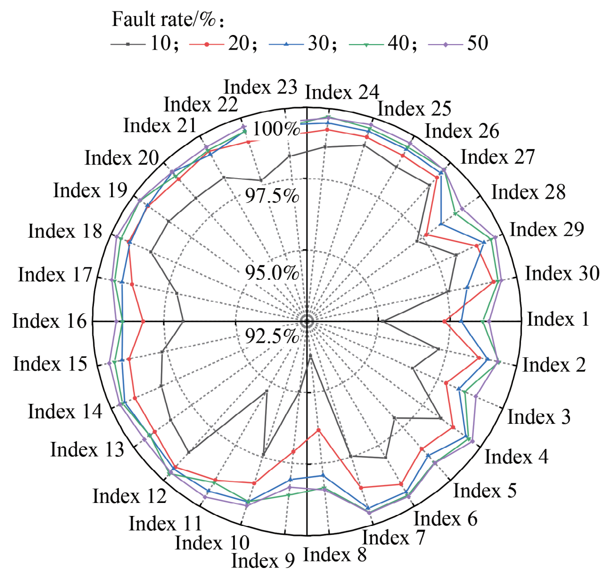

Fig. 7 F_1 score of the S-MS method across groups with various fault rates

5 Conclusions

1) The proposed S-MS method contains two pivotal phases, i.e., the fault prescreening phase, which employs the quartile method combined with the sliding window technique, and the fault self-detection phase, which is grounded in the mean shift method. The proposed method is insensitive to time variations and effectively detects four fault types (i.e., offset, drift, gain, and stuck) at different fault rates. As the fault rate increases, the S-MS method exhibits better performance.

2) The proposed S-MS method achieves sensor fault

detection accuracies of over 98% for four fault types (i. e., offset, drift, gain, and stuck), with an average of 99. 37%; precision of over 0. 86 for four fault types (i. e., offset, drift, gain, and stuck), with an average of 0. 97; and F_1 scores of over 91% for four fault types (i. e., offset, drift, gain, and stuck), with an average of 96. 85%. Compared with other sensor fault detection methods (i. e., ET, RF, SVDD, and OC-SVM), the proposed S-MS method exhibits the best sensor fault detection performance in terms of accuracy rate, precision, and F_1 score.

3) The proposed S-MS method exhibits high computational efficiency, with fault detection durations accounting for only 0. 11% to 0. 5% of the total sensor data collection time, indicating that sensors equipped with this method can promptly self-detect anomalies based on the monitoring data. Moreover, aided by the fault prescreening phase, the S-MS method rapidly evaluates the datasets and significantly decreases the detection times, further enhancing its efficiency.

References

[1] Ghoshal A, Sundaresan M J, Schulz M J, et al. Structural health monitoring techniques for wind turbine blades [J]. *Journal of Wind Engineering and Industrial Aerodynamics*, 2000, **85**(3): 309 – 324. DOI: 10. 1016/S0167-6105(99)00132-4.

[2] Ko J M, Ni Y Q. Technology developments in structural health monitoring of large-scale bridges[J]. *Engineering Structures*, 2005, **27**(12): 1715 – 1725. DOI: 10. 1016/j. engstruct. 2005. 02. 021.

[3] Spencer B F Jr, Ruiz-Sandoval M E, Kurata N. Smart sensing technology: Opportunities and challenges [J]. *Structural Control and Health Monitoring*, 2004, **11**(4): 349 – 368. DOI: 10. 1002/stc. 48.

[4] Wan H P, Ni Y Q. Bayesian multi-task learning methodology for reconstruction of structural health monitoring data [J]. *Structural Health Monitoring*, 2019, **18**(4): 1282 – 1309. DOI: 10. 1177/1475921718794953.

[5] Hu J, Guo J, Ou J P, et al. Measurement of wind field characteristics at a long-span suspension bridge [J]. *Journal of Southeast University* (*English Edition*), 2011, **27** (3): 328 – 334. DOI: 10. 3969/j. issn. 1003-7985. 2011. 03. 020.

[6] Liu X L, Huang Q, Ren Y, Fan Y H, et al. Extraction of cable forces due to dead load in cable-stayed bridges under random vehicle loads [J]. *Journal of Southeast University* (*English Edition*), 2015, **31**(3): 407 – 411. DOI: 10. 3969/j. issn. 1003-7985. 2015. 03. 019.

[7] Wang H, Li A Q, Guo T, et al. Comparable study on typhoon and strong northern wind characteristics of the Runyang Suspension Bridge based on field tests [J]. *Journal of Southeast University* (*English Edition*), 2009, **25**(1): 99 – 103. DOI: 10. 3969/j. issn. 1003-7985. 2009. 01. 021.

[8] Mao J X, Wang H, Spencer B F Jr. Toward data anomaly detection for automated structural health monitoring: Exploiting generative adversarial nets and autoencoders [J]. *Structural Health Monitoring*, 2021, **20**(4): 1609 – 1626. DOI: 10. 1177/1475921720924601.

[9] Mahapatro A, Khilar P M. Fault diagnosis in wireless sensor networks: A survey[J]. *IEEE Communications Surveys & Tutorials*, 2013, **15**(4): 2000 – 2026. DOI: 10. 1109/SURV. 2013. 030713. 00062.

[10] Zidi S, Moulahi T, Alaya B. Fault detection in wireless sensor networks through SVM classifier[J]. *IEEE Sensors Journal*, 2018, **18**(1): 340 – 347. DOI: 10. 1109/JSEN. 2017. 2771226.

[11] Magalhães F, Cunha Á, Caetano E. Vibration based structural health monitoring of an arch bridge: From automated OMA to damage detection[J]. *Mechanical Systems and Signal Processing*, 2012, **28**: 212 – 228. DOI: 10. 1016/j. ymssp. 2011. 06. 011.

[12] Mao J X, Wang H, Feng D M, et al. Investigation of dynamic properties of long-span cable-stayed bridges based on one-year monitoring data under normal operating condition[J]. *Structural Control and Health Monitoring*, 2018, **25**(5): e2146. DOI: 10. 1002/stc. 2146.

[13] Muhammed T, Shaikh R A. An analysis of fault detection strategies in wireless sensor networks[J]. *Journal of Network and Computer Applications*, 2017, **78**: 267 – 287. DOI: 10. 1016/j. jnca. 2016. 10. 019.

[14] Lau B C P, Ma E W M, Chow T W S. Probabilistic fault detector for Wireless Sensor Network[J]. *Expert Systems with Applications*, 2014, **41**(8): 3703 – 3711. DOI: 10. 1016/j. eswa. 2013. 11. 034.

[15] Abid A, Kachouri A, Guiloufi A B F, et al. Centralized KNN anomaly detector for WSN[C]//2015 *IEEE 12th International Multi-Conference on Systems, Signals & Devices* (*SSD*15). Mahdia, Tunisia, 2015: 1 – 4. DOI: 10. 1109/SSD. 2015. 7348091.

[16] Warriach E U, Tei K. Fault detection in wireless sensor networks: A machine learning approach[C]//2013 *IEEE 16th International Conference on Computational Science and Engineering*. Sydney, Australia, 2014: 758 – 765. DOI: 10. 1109/CSE. 2013. 116.

[17] Panda M, Khilar P M. Energy efficient distributed fault identification algorithm in wireless sensor networks[J]. *Journal of Computer Networks and Communications*, 2014, **2014**: 1 – 16. DOI: 10. 1155/2014/323754.

[18] Chen J R, Kher S, Somani A. Distributed fault detection of wireless sensor networks[C]//*Proceedings of the* 2006 *Workshop on Dependability Issues in Wireless Ad Hoc Networks and Sensor Networks*. Los Angeles, CA, USA, 2006: 65 – 72. DOI: 10. 1145/1160972. 1160985.

[19] Obst O. Distributed fault detection in sensor networks using a recurrent neural network[J]. *Neural Processing Letters*, 2014, **40**(3): 261 – 273. DOI: 10. 1007/s11063-013-9327-4.

[20] Younis O, Fahmy S, Santi P. An architecture for robust sensor network communications[J]. *International Journal of Distributed Sensor Networks*, 2005, **1**(3/4): 305 – 327. DOI: 10. 1080/15501320500330786.

[21] Barborak M, Dahbura A, Malek M. The consensus problem in fault-tolerant computing[J]. *ACM Computing Surveys*, 1993, **25**(2): 171 – 220. DOI: 10. 1145/152610. 152612.

[22] Ni K, Ramanathan N, Chehade M N H, et al. Sensor network data fault types[J]. *ACM Transactions on Sensor Networks*, 2009, **5**(3): 1－29. DOI: 10.1145/1525856.1525863.
[23] Fukunaga K, Hostetler L. The estimation of the gradient of a density function, with applications in pattern recognition[J]. *IEEE Transactions on Information Theory*, 1975, **21**(1): 32－40. DOI: 10.1109/TIT.1975.1055330.
[24] Xiao C X, Liu M. Efficient mean-shift clustering using Gaussian KD-tree[J]. *Computer Graphics Forum*, 2010, **29**(7): 2065－2073. DOI: 10.1111/j.1467-8659.2010.01793.x.
[25] Ni Y Q, Xia Y, Liao W Y, et al. Technology innovation in developing the structural health monitoring system for Guangzhou New TV Tower[J]. *Structural Control and Health Monitoring*, 2009, **16**(1): 73－98. DOI: 10.1002/stc.303.
[26] Ni Y Q, Xia Y, Lin W, et al. SHM benchmark for high-rise structures: A reduced-order finite element model and field measurement data[J]. *Smart Structures and Systems*, 2012, **10**(4/5): 411－426. DOI: 10.12989/sss.2012.10.4_5.411.
[27] Saeed U, Jan S U, Lee Y D, et al. Fault diagnosis based on extremely randomized trees in wireless sensor networks [J]. *Reliability Engineering & System Safety*, 2021, **205**: 107284. DOI: 10.1016/j.ress.2020.107284.
[28] Noshad Z, Javaid N, Saba T, et al. Fault detection in wireless sensor networks through the random forest classifier[J]. *Sensors*, 2019, **19**(7): 1568. DOI: 10.3390/s19071568.

基于均值漂移方法的传感器故障自检测

陈　昊[1]　朱一凯[1]　雷　波[2]　翁志海[3]　徐鸿昌[2]　万华平[1]

([1] 浙江大学建筑工程学院, 杭州 310058)
([2] 中建三局集团有限公司, 武汉 430064)
([3] 湖州市城市投资发展集团有限公司, 湖州 313000)

摘要:为准确识别复杂环境导致的传感器故障,确保结构健康监测系统能正确感知结构状态,提出了一种基于均值漂移和滑动窗技术的传感器节点自检测方法. 该方法包含故障预筛选和故障自检测 2 个阶段. 在故障预筛选阶段,结合四分位数法和滑动窗口技术快速识别出潜在的异常数据,有效地提升了方法运行效率;在故障自检测阶段,结合均值偏移算法对异常数据实施自适应聚类,有效地检测了多种故障. 利用广州塔实测数据验证所提方法的有效性,设置偏移、漂移、增益和卡滞 4 种故障来模拟传感器故障,并与极端随机树、随机森林、支持向量数据描述和单类支持向量机方法进行对比. 结果表明,该方法能够检测出 4 种故障,并具有较高的正确率和计算效率.

关键词:传感器故障检测;均值漂移;滑动窗口;机器学习
中图分类号:TU317