ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20221103085856
Preliminary communication

Recovery of Outliers in Water Environment Monitoring Data

Jinling SONG*, Meining ZHU, Liming HUANG, Gang WANG, Dongyan JIA

Abstract: The water environment monitoring data are time sequences with outliers which depress the data quality, so outlier detection and recovery play an important role in the applications such as knowledge acquisition and prediction modelling of water environment indicators. To detect the outliers, the short-term chain comparison with the sliding window based on the time sequence characteristics is adopted. To recover outliers closer to the real data at that time, the sub-sequences are divided dynamically according to the change characteristics of the dataset, then the similarity between sub-sequences is measured by the shape distance and the outliers are recovered according to the change trend of the corresponding data in the most similar sub-sequences. The monitoring data of a water station are selected in the study. The experimental results show that the recovery method is superior to the commonly used prediction recovery method and fitting recovery method, the recovered data is smoother and the short-term trend is more obvious.

Keywords: outliers; shape distance; sub-sequence similarity; water monitoring data

1 INTRODUCTION

Data quality has become a key issue of data applications [1, 2] with its continuous deepening development. Water environment monitoring data are water environment index data automatically detected by sensors at a certain time, such as water temperature, PH, DO, electrical conductivity, NH_3-N, $KMnO_4$, TP, TN, Chla and blue-green algae, etc., which will form large-scale monitoring data over time. Analysis and processing with the large-scale monitoring data can provide important support for water environment supervision; so water pollution events can be found on time and the pollution sources can be traced back according to the changes of monitoring data; related water environment knowledge can be obtained through data mining, and water quality can be predicted in advance based on deep learning. In all applications, the data accuracy plays a vital role to the application effect. However, due to sensor failures, data transmission, reading errors and other factors in the monitoring process, outliers will appear in the water environment monitoring data, which reduce the data quality. Therefore, timely detection and repair of outliers is prerequisite to the data applications, it plays an important role to the accuracy of the later application results.

The outliers in the large-scale data are recovered with the fitting method [3, 4] and the prediction method [5-8] to improve the reliability of the data, but the outside factors such as pollution and weather greatly influence the water environment monitoring data, which lead to the error in the long-term model, and the outliers in water environment monitoring data will lead to the failure of the prediction method. The first step of outlier recovery is outlier detection [9]; traditional outlier detection techniques (based on classification [10], distance [11], clustering [12] and Information theory [13], etc.) are not sufficient, and some targeted methods are necessary for outlier detection in specific situations [14, 15].

In this study, multiple sub-sequences with ascending, descending and relatively stable trend are segmented by dividing the water environment monitoring data dynamically according to the change characteristics of the dataset. The water environment monitoring data are detected with the short-term chain comparison method and are recovered based on the similarity of the sub-sequences without any model, so that the recovered data are smoother, the short-term trend is more obvious, and the data quality is greatly improved.

2 DEFINITIONS

The monitoring indicators of the water environment are constantly changing with time. The sensors will return the monitoring data of the corresponding indicators at fixed time intervals during the online monitoring process, and the time sequence data of each indicator will be obtained. The time sequence of each indicator can be obtained according to the different monitoring water environment indicators.

Definition 1: Water environment monitoring data are the time sequences of monitoring data of a certain water environment indicator, denoted as $V = (v_1, v_2, \dots, v_n)$, where v_i $(1 \le i \le n)$ represents the monitoring data corresponding to time i.

The water environment is sensitive to external influences, and the monitoring data of each indicator will change with time. The adjacent monitoring data are similar when the time intervals are close, and the following data are similar even if big change occurs when the water environment is affected by external influences at a certain time, which means that the outliers cannot be detected simply by the change of the data. The outliers significantly deviate from other monitoring data, much bigger (or smaller) than other adjacent data within a certain period of time. Therefore, the outliers in the water environment monitoring data are such that in the k- time window of a monitoring data, the difference between them is greater or less than a certain threshold, and the number exceeds a certain amount.

Definition 2: Outliers in water environment monitoring data: Given water environment monitoring data $V = (v_1, v_2, \dots, v_n)$, the point v_i $(1 \le i \le n)$ represents the monitoring data corresponding to time i. $\eta_i^{(k)} = (v_{i-k}, v_{i-k+1}, \dots, v_{i-1}, v_{i+1}, \dots, v_{i+k-1}, v_{i+k})$ means that in the k-nearest neighbour window of point v_i, if count $\left(\sum_{j=i-k}^{i+k}(v_i - v_j > \varepsilon)\right) > \tau$ or count $\left(\sum_{j=i-k}^{i+k}(v_i - v_j > \varepsilon)\right) > \tau$ exists in $\eta_i^{(k)}$, then v_i is the outlier, ε is the difference threshold and τ is the quantity threshold.

In order to avoid misjudging the monitoring data that change greatly due to external factors as outliers, the k-nearest neighbour in Definition 2 should be a bidirectional time window, because in this case, although the gap between the monitoring data and the precursor time window is large, it is relatively close to the rear-drive time window. The nearest neighbour window k, difference threshold ε, and quantity threshold τ are important to determine whether the monitoring data v_i is outlier. Therefore, different k, ε, τ could be selected to detect the outliers in water environment monitoring data according to different water environment monitoring indicators. The selection of k, ε, τ should refer to the relevant parameters of hydrodynamics (such as river velocity, pollutant attenuation coefficient, diffusion coefficient, etc.), and be determined comprehensively in combination with the results of manual testing and expert experience.

The recovery of outliers in water environment monitoring data is essentially to find the data that are closest to the true data corresponding to the outliers, and replace the outliers in the original sequences with them, so as to obtain the recovered data sequences.

Definition 3: The recovery sequence of water environment monitoring data: Given water environment monitoring data $V = (v_1, v_2, \ldots, v_n)$, the point v_i $(1 \le i \le n)$ represents the monitoring data corresponding to time i. If v_i is outlier, and the recovered ones are v_i', then the sequence $V = (v_1, v_2, \ldots, v_i', \ldots, v_n)$ is the recovered sequence of water environment monitoring data, where v_i' should be as close as possible to the true data of v_i.

3 OUTLIER DETECTION METHODS OF WATER ENVIRONMENT MONITORING DATA

According to definition 2, given k, ε, τ, the difference between each monitoring data v_i in time sequences V and its k-nearest neighbours can be calculated one by one through the sliding window mode, and the number of points c whose differences exceeding the threshold ε are recorded. When all comparisons are completed, v_i can be determined as outlier if c is greater than τ, and the position should be recorded. The description of the outlier detection algorithm is as follows:

Algorithm 1: Outlier detection algorithm for water environment monitoring data

Input: Time sequences $V = (v_1, v_2, \ldots, v_n)$, the size k of nearest neighbour window, the difference threshold ε, the number threshold τ
Output: Outlier location set *out L*
1. out $L = \varnothing$;
2. for each $v_i \in V$, do
3. $c1 = 0$; //difference quantity of greater than ε
4. $c2 = 0$; //difference quantity of less than $-\varepsilon$
5. for each v_j in k-nearest neighbour time window, do
6. if $(v_i - v_j > \varepsilon)$ $c1 = c1+1$; //number of greater than ε
7. if $(v_i - v_j < -\varepsilon)$ $c2 = c2+1$; //number of less than ε
8. end for
9. if ($c1 > \tau$ or $c2 > \tau$) out L = out $L \cup i$;
10. end for
11. return out L;

4 OUTLIER RECOVERY METHOD OF WATER ENVIRONMENT MONITORING DATA

According to the purpose of outlier recovery, how to make the recovery data closer to the real data at that time is a key issue. The characteristics of water environment monitoring data show that the data change curves of sub-sequences in different time periods have high similarity. If the sub-sequence that is most similar to those where outliers are located can be determined, the outliers can be recovered according to the corresponding trend in those sub-sequences, and it is verified that the recovered data are very close to the true data.

The following steps are required to recover the outliers of water environment monitoring data: The water environment monitoring data are divided into several sub-sequences, then the similarity of the outlier sub-sequence with other sub-sequences are calculated and the most similar ones are selected, and the outliers are recovered according to the change trend of the corresponding data in the most similar sub-sequences.

4.1 Dynamic Division Method of Sub-Sequence

The data segments of the water environment monitoring data in a relatively short period of time are in obvious order: they are generally ascending, descending or in a relatively stable state, but the scale is not fixed. Therefore, the data sequence should be dynamically divided according to the changing characteristics of the data instead of on a fixed scale.

In this study, the sequences are dynamically divided based on the maximum distance threshold φ, that is, the data is considered to be in a stable state when the maximum distance of the data segment is less than the threshold φ, and the data segment that continues to rise (decrease) is not restricted by the threshold φ (not be divided when the maximum distance exceeds the threshold), so that the change rule of each sub-sequence curve is ascending, descending or relatively stable. The dynamic division method cannot only reduce the number of sub-sequences but also preserve the trend of each sub-sequence to the greatest extent.

In order to indicate the trend of the sub-sequences, the variable "flag" is set. The sub-sequence is in the ascending state when "flag" is 1, descending when flag is –1, and in the stable state when flag is 0. The difference of adjacent data $(v_i - v_{i-1})$ and the current flag can tell the trend of the data sequences. The dynamic division of sub-sequences can be realized with the maximum distance threshold φ.

The dynamic division steps of water environment monitoring data are as follows: (1) assign an initial value to "flag" according to the initial trend of the data sequences; (2) perform the following operations for each monitoring data v_i: calculate the difference e_i with the previous data, and calculate the maximum distance of the sub-sequences. If the maximum distance does not exceed the threshold φ and e_i opposite to the signs of "flag", the "flag" should be 0. If the maximum distance is greater than the threshold φ, the "flag" and e_i can tell the overall sub-sequence curve and changes, and the division is recorded in the sub-sequences. The specific sub-sequence dynamic division algorithm is described as follows. Outliers have been ignored so that they can be included in the sub-sequences.

Algorithm 2: Dynamic sub-sequence division algorithm of water environment monitoring data

Input: time sequences $V = (v_1, v_2, \ldots, v_n)$, outlier marker set $out\ L$, maximum distance threshold φ of sub-sequence segments
Output: Sub-sequence marker set $SubL$
1. Initialize $SubL = \varnothing$, $Subseq = 1$; // $Subseq$ records the start mark of each sub-sequence, Sub stores the start mark and status of all sub-sequences
2. Initialize $minV = v_1$, $maxV = v_1$, $flag = 0$; //Store the minimum, maximum, and sub-sequence state respectively
3. if $(v_2 - v_1 > 0)$ $flag = 1$; //Initial trend is ascending
4. if $(v_2 - v_1 < 0)$ $flag = -1$; //Initial trend is descending
5. for each v_i in $\{v_2, \ldots, v_{n-1}\}$, do
6.if $i \notin outL$ // v_i is not outlier
7. $e_i = v_i - v_{i-1}$;
8. $minV = \min(minV, v_i)$;
9. $maxV = \max(maxV, v_i)$;
10. if $(maxV - minV > \varphi)$ //Divide sub-sequences when the maximum distance exceeds the threshold
11. if $(e_i{*}flag < 0)$ //The curve state changes
12. $SubL = SubL \cup (Subseq, i - 1, flag)$;
13. $Subseq = 0$; //Sub-sequence has been divided
14. end if
15. else if $(e_i{*}flag < 0)$ $flag = 0$; //Sub-sequence trend changes under the maximum distance threshold
16. end if
17. if $(Subseq == 0)$
18. $Subseq = i$; //Start mark of the next sub-sequence
19. $maxV = minV = v_i$; //Set initial values
20. if $(v_{i+1} - v_i > 0)$ $flag = 1$; //Set initial state of the next sub-sequence
21. if $(v_{i+1} - v_i < 0)$ $flag = -1$;
22. end if
23. end if
24. end for
25. return $SubL$;

4.2 The Algorithm of Sub-Sequences' Similarity

The water environment monitoring data can be divided into multiple sub-sequence segments with algorithm 2, and similarity of the divided sub-sequences can be compared. Since the similarity of sub-sequences is curve similarity, that is, sub-sequences are similar in shapes (similar trends), and are closest in distance. The similarity of the dynamic range of trends cannot be distinguished with simple Euclidean distance, so the shape distance is adopted to measure the similarity between two sub-sequences. The shape distance takes the sum of the products of the mode distance and the amplitude change distance of each point as the distance between the sub-sequences, and the calculation formula is shown in Eq. (1). Among them, V_1, V_2 are time sequences, M_{1x}, M_{2x} are the mode of the x-th point of time sequences V_1, V_2 respectively, the range is $\{-3, -2, -1, 0, 1, 2, 3\}$, representing the modes of {acceleration descending, uniform descending, decelerating descending, unchanged, decelerating ascending, uniform ascending, accelerating ascending}; A_{1x}, A_{2x} are the changing of the x-th amplitude of vibration ($A_x = v_{x+1} - v_x$,).

$$D(V_1, V_2) = \sum_{x=1}^{k} | M_{1x}, M_{2x} | * | A_{1x}, A_{2x} | \tag{1}$$

It can be seen from Eq. (1) that the smaller the shape distance, the greater the similarity of the two sub-sequences. The reciprocal of $D(V_i, V_j)$ is taken as the similarity of the sub-sequence, as shown in Eq. (2). In the case of inconsistent sub-sequence scales, a rolling method can be adopted to calculate the similarity between the short sequence and each segment with equal length in the long sequence, and the maximum similarity is the final similarity of the two sub-sequences.

$$\mathrm{sim}(V_i, V_j) = 1/D(V_i, V_j) \tag{2}$$

4.3 Outlier Recovery Method

The foundation of recovering outliers in water environment monitoring data is the sub-sequences with the greatest similarity. Therefore, the sub-sequence V_i where v_i is located should be found when recovering outlier v_i, and the similarity between V_i and other sub-sequences should be calculated according to Eq. (2), and the sub-sequence V_j with the greatest similarity should be selected. Since only sub-sequences with the same overall trend will have a higher similarity, the sub-sequences with the same trend as V_i should be calculated according to the trend of each sub-sequence stored in Algorithm 2. At last, the outlier v_i can be recovered according to the change trend of the sub-sequence V_j, that is, $v_i = v_{i-1} + (v_j - v_{j-1})$. The outlier recovery algorithm for water environment monitoring data is described as follows:

Algorithm 3: Outlier recovery algorithm for water environment monitoring data

Input: Time sequence V, divided sub-sequence set $SubL$, outlier set $outL$
Output: Recovered time sequences V
1. Initialize $simseq = \varnothing$; //Store most similar sub-sequences
2. Initialize $seqS = 0$, $maxS = 0$; //Store the similarities
3. for each $i \in outL$, do//Recover each outlier
4. V_i =get sub-sequence in V containing v_i base on $SubL$;
5. for each sub-sequence $V_j \neq V_i$ in $SubL$,do
6. if (the trend of V_j is same with V_i)
7. $seqS = \mathrm{sim}(V_i, V_j)$;
8. if $(seqS > maxS)$
9. $maxS = seqS$;
10. $simseq = V_j$;
11. end if
12. end if
13. end for
14. v_j =t he value corresponding to v_i in $simseq$;
15. v_{j-1} = the previous value of v_{i-1} in $simseq$;
16. $v_i' = v_{i-1} + (v_j - v_{j-1})$;
17. V = replace outliers v_i with v_i' in V;
18. end for

5 EXPERIMENTAL ANALYSIS

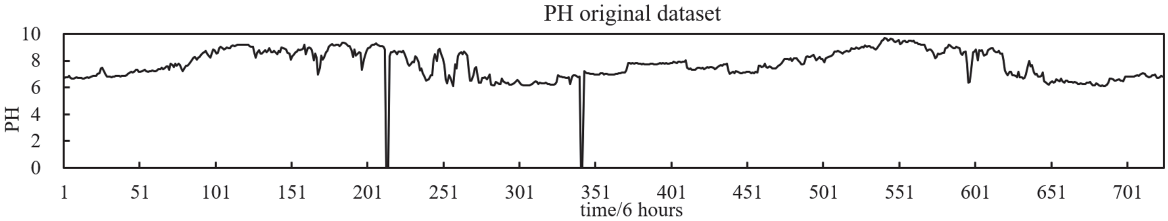
In order to verify the effectiveness of the proposed outlier's recovery method of water environment monitoring data, the water environment monitoring data of a water station is selected for experiments, and the outlier recovery results in this study are analyzed and discussed.

5.1 Data Preparation

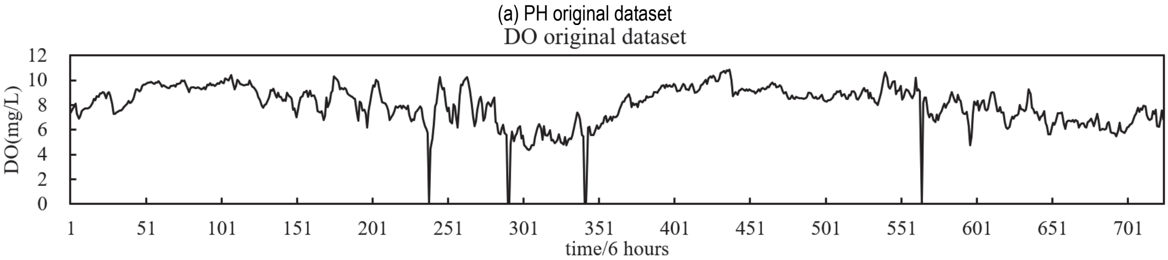
Two different water quality indicators, PH and DO (unit: mg/l), which are monitored online, are selected for algorithm verification. The monitoring data of the water station in the past 5 years are adopted. The time interval of monitoring data is 6 hours. There are about 7300 points for each water quality indicator, and some of the data are intercepted for experimental display. The original data curves of the two indicators are shown in Fig. 1. It can be seen from Fig. 1 that the overall trend of the dataset curves of pH value and DO is smooth, but there are also some obvious suspicious abnormal points (missing values and outliers, etc.), which will affect the later data analysis performance if they are not recovered in time.

5.2 Algorithm Effectiveness

The effectiveness of the outlier detection algorithm is analyzed for the two sets of datasets. Due to the difference of PH and DO, the nearest neighbour window size k, the difference threshold ε, and the number threshold τ should be different in the outlier detection algorithm. After comparing analysis in many experiments, the k, ε, τ in the outlier detection algorithm are set as follows: $k = 3$, $\varepsilon = 0.5$, $\tau = 4$ in the pH dataset; $k = 3$, $\varepsilon = 0.9$, $\tau = 3$ in the DO dataset, and 11 outliers are detected in the PH dataset, and 21 outliers in the DO dataset. There is no FP (False Positive) or FN (False Negative) data. Fig. 2 shows the execution result of the outlier detection algorithm. The data points represented by the circles are the outliers that deviate significantly from their neighbour nodes. The results show that the outlier detection algorithm proposed in this study can successfully detect the outliers that deviate significantly from the neighbour nodes from the water environment monitoring dataset, and retain the data fluctuation caused by the change of the water environment to the greatest extent.

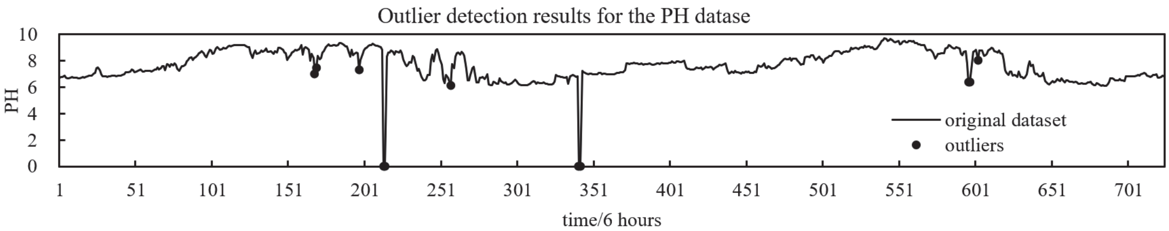


(a) PH original dataset

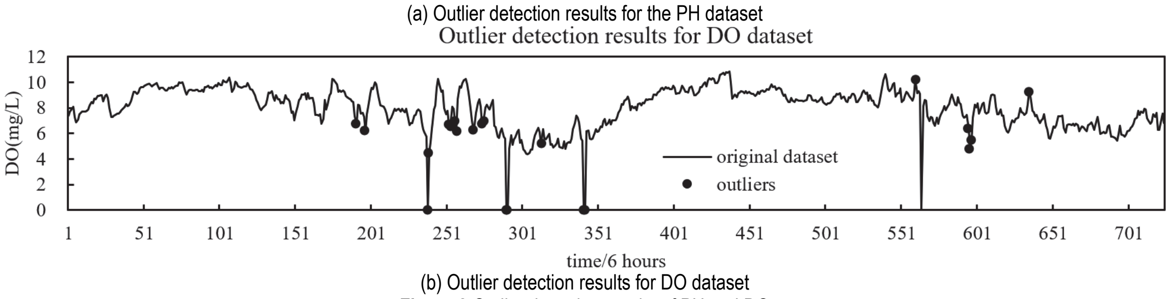


(b) DO original dataset

Figure 1 Original dataset of PH and DO



(a) Outlier detection results for the PH dataset



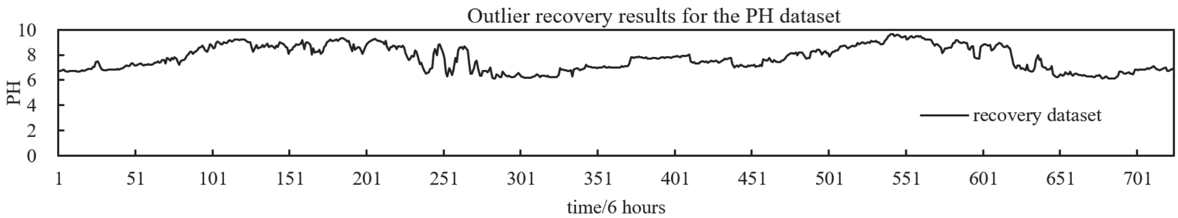
(b) Outlier detection results for DO dataset

Figure 2 Outlier detection results of PH and DO

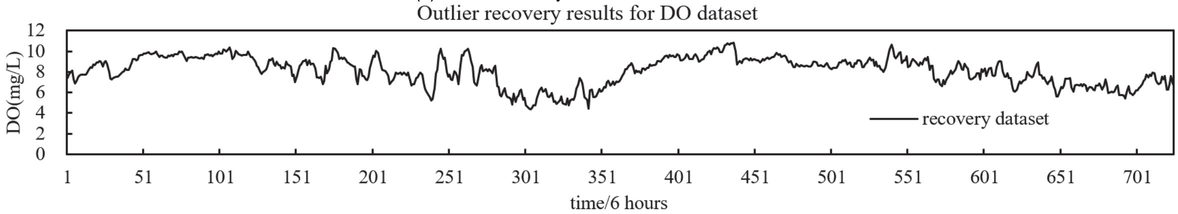
Then, the outlier recovery algorithm is evaluated for the two datasets. In the sub-sequence division stage, the maximum distance threshold set by the PH dataset is $\varphi = 0.3$, which is divided into 101 sub-sequences, and the maximum distance threshold set by the DO dataset is $\varphi = 0.8$, which is divided into 104 sub-sequences with moderate scales, indicating that the method is effective. In the outlier recovery stage, the sub-sequence with the greatest similarity can be found for the sub-sequences where the outliers are located. The recovery effect of the outliers is shown in Fig. 3 and it can be seen that the recovered curve becomes relatively smooth, and there are no obvious points that deviate from the neighbour nodes. Further the repair value distortion is tested using the fitting degree R^2 shown in Eq. (3). The R^2 of PH repair dataset is 0.8725, and the R^2 of DO repair dataset is 0.9042. The fitting degrees of the two repair datasets have reached a high level when with missing values in the original dataset, indicating that the distortion of the repaired data is very low. The analysis indicates that the recovery method based on the similarity of sub-sequences is feasible, and the outlier recovery method in this study is effective.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$

where y_i is the original value, $\hat{y}_i$ is the repair value, $\bar{y}$ is the average of the original values.

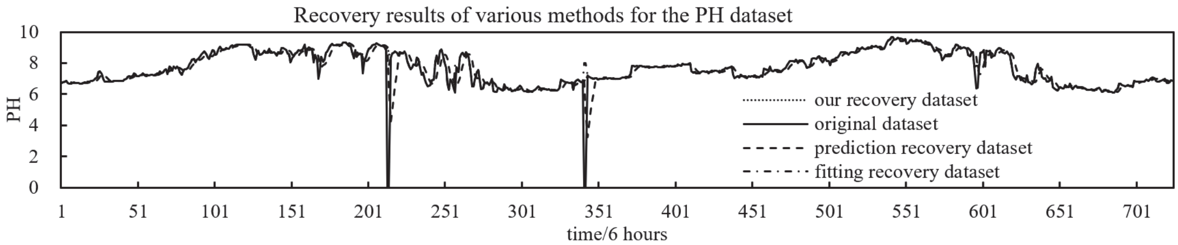


(a) Outlier recovery results for the PH dataset

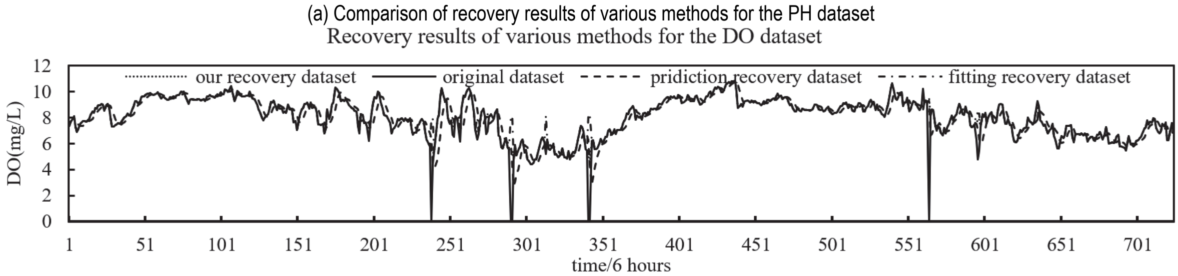


(b) Outlier recovery results for DO dataset

Figure 3 Outlier recovery results of PH and DO



(a) Comparison of recovery results of various methods for the PH dataset



(b) Comparison of recovery results of various methods for the DO dataset

Figure 4 Comparison of recovery results of various methods

5.3 Comparative Experiment

In order to further verify the effectiveness of our outlier recovery method, the method is compared with the outlier recovery method on the experimental dataset based on Bi-GRU prediction [8] and fitting [3]. The results of the three methods are shown in Fig. 4. It can be seen from Figure 4 that there is a certain deviation between the original dataset and the recovered dataset obtained based on the prediction method, especially the prediction results of the original missing values deviate from the original curve, indicating that the prediction method is significantly affected by the missing values. Compared with the original curve, there are many bulges in the recovery data based on the fitting method. The reason is that the fitting polynomial only reflects the long-term trend of the original data, and the specific fitting value is quite different from the original data. The recovery dataset obtained by our method overlaps the original dataset. The regions where the outliers are located are smoother, and the short-term trend of the data is more obvious.

Then the fitting degree of repair datasets was calculated respectively. The R^2 of PH dataset repaired by fitting-based method is 0.7482 and DO dataset is 0.7665, the R^2 of PH dataset repaired by prediction-based method is 0.7822 and DO dataset was 0.8193. Compared with the fitting-based repair method, the R^2 of our method are increased by 16.6% and 13.8% respectively. The R^2 are increased by 15.6% and 10.4% respectively relative to prediction-based repair method. It shows that our repair dataset has better fitting degree and least distortion to the original dataset.

Through the comparative experiments, the prediction-based and fitting-based repair methods are not applicable to the water environment monitoring data. Our method repairs the outlier with a small cost and retains the original data characteristics, which has effectively improved the data quality of water environment monitoring data.

6 CONCLUSIONS

To deal with the outliers in water environment monitoring data, we detect with the short-term chain comparison method, and recovery according to the similarity of the sub-sequences. The comparative experiments using PH and DO monitoring data from a water station show that the R^2 of our method is increased by 16.6% and 13.8% respectively compared with the fitting-based method, and is increased by 15.6% and 10.4% respectively relative to prediction-based method, which illustrates that our repaired datasets have the lowest distortion in three methods, and the data quality is improved effectively. Following researches will focus on improving the proposed recovery algorithm, detecting abnormal sub-sequences in the water environment monitoring data, and mining and analyzing the recovered water environment monitoring data.

Acknowledgment

The work is supported by S&T Program of Hebei (Grant No. 21370103D and 21373301D) and Research on Social Sciences Development in Hebei Province (Grant No. 20210201445).

7 REFERENCE

[1] Fan, W. (2015). Data quality: From theory to practice. *ACM Sigmod Record*, *44*(3), 7-18. https://doi.org/10.1145/2854006.2854008

[2] Gao, J., Ismail, N., & Gao, Y. (2022). Computer big data analysis and predictive maintenance based on deep learning. *Ingénierie des Systèmes d'Information*, *27*(2), 349-355. https://doi.org/10.18280/isi.270220

[3] Guo, Z. (2020). Fast outlier detection algorithm based on local density and connectivity. *Scientific Journal of Intelligent Systems Research*, *2*(1), 18-27.

[4] Liu, W. (2019). Traffic flow prediction based on local mean decomposition and big data analysis. *Ingénierie des Systèmes d'Information*, *24*(5), 547-552. https://doi.org/10.18280/isi.240513

[5] Zheng, S., Zhu, Y., Li, D., Cao, Z. J., Deng, Q., & Phoon, K. (2021). Probabilistic outlier detection for sparse multivariate geotechnical site investigation data using Bayesian learning. *Geoscience Frontiers*, *12*(1), 425-439. https://doi.org/10.1016/j.gsf.2020.03.017

[6] Kulanuwat, L., Chantrapornchai, C., Maleewong, M., Wongchaisuwat, P., Wimala, S., Sarinnapakorn, K., & Boonya-aroonnet, S. (2021). Anomaly detection using a sliding window technique and data imputation with machine learning for hydrological time series. *Water*, *13*(13), 1862. https://doi.org/10.3390/w13131862

[7] Ye, F., Liu, Z., Liu, Q., &Wang, Z. (2020). Hydrologic time series anomaly detection based on flink. *Mathematical Problems in Engineering*, *2020*, 3187697. https://doi.org/10.1155/2020/3187697

[8] Guo, K. P. (2021). *Design and implementation of air quality prediction system based on GCN-LST*. University of Chinese Academy of Sciences.

[9] Govindaraj, M., Kaliappan, S., & Swaminathan, G. (2022). Outlier detection of functional data using reproducing kernel Hilbert space. *Instrumentation Mesure Métrologie*, *21*(4), 145-150. https://doi.org/10.18280/i2m.210404

[10] Sakr, M., Atwa, W., & Keshk, A. (2021). Genetic-based summarization for local outlier detection in data stream. *International Journal of Intelligent Systems and Applications*, *13*(1), 58-68. https://doi.org/10.5815/ijisa.2021.01.05

[11] Shao, M., Qi, D., & Xue, H. (2021). Big data outlier detection model based on improved density peak algorithm. *Journal of Intelligent & Fuzzy Systems*, *40*(4), 6185-6194. https://doi.org/10.3233/JIFS-189456

[12] Yu, W., He, Y., & Qin, H. (2022). A new outlier detection algorithm based on observation-point mechanism. *Journal of Shenzhen University Science and Engineering*, *39*(3), 355-362. https://doi.org/10.3724/SP.J.1249.2022.03355

[13] Ando, S. (2007). Clustering needles in a haystack: An information theoretic analysis of minority and outlier detection. *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, 13-22. https://doi.org/10.1109/ICDM.2007.53

[14] Xiao, H., Guan, D., Zhao, R., Yuan, W., Tu, Y., & Khattak, A. M. (2020). Semi-supervised time series anomaly detection model based on LSTM autoencoder. *International Conference on Big Data and Security*, 41-53. https://doi.org/10.1007/978-981-16-3150-4_4

[15] Zhan, P., Hu, Y., Zhang, Q., Zheng, J., & Li, X. (2018). Feature-based dividing symbolic time series representation for streaming data processing. *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*, 817-823. https://doi.org/10.1109/ITME.2018.00184

Contact information:

Jinling SONG
(Corresponding author)
School of Mathematics and Information Technology,
Hebei Agricultural Data Intelligent Perception and Application Technology Innovation Center,
Hebei Normal University of Science & Technology,
Qinhuangdao 066004, China
Hebei Key Laboratory of Ocean Dynamics, Resources and Environments,
Qinhuangdao 066004, China
E-mail: songjinling99@126.com

Meining ZHU
School of Mathematics and Information Technology,
Hebei Agricultural Data Intelligent Perception and Application Technology Innovation Center,
Hebei Normal University of Science & Technology,
Qinhuangdao 066004, China

Liming HUANG
School of Business Administration,
Hebei Normal University of Science & Technology,
Qinhuangdao 066004, China

Gang WANG
School of Mathematics and Information Technology,
Hebei Agricultural Data Intelligent Perception and Application Technology Innovation Center,
Hebei Normal University of Science & Technology,
Qinhuangdao 066004, China

Dongyan JIA
School of Mathematics and Information Technology,
Hebei Agricultural Data Intelligent Perception and Application Technology Innovation Center,
Hebei Normal University of Science & Technology,
Qinhuangdao 066004, China