ISSN 2348 - 9928
Doi:10.46532/ijaict-2020030 Published on 05 (09) 2020

Application of Data Mining Technology under K-means Algorithm Combined with BIM Technology in Management Engineering

[1]Jun Wang and [2]Zhan Chen
[1,2] School of Computer Science and Engineering, South China University of Technology, China.
[1]junwang@scut.edu.cn

ArticleInfo
International Journal of Advanced Information and Communication Technology
(https://www.ijaict.com/journals/ijaict/ijaict_home.html)
https://doi.org/10.46532/ijaict-2020030
Received 25 April 2020; Revised form 28 July 2020; Accepted 20 August 2020; Available online 05 September 2020.

Abstract—The data mining technology of the K-means algorithm combined with BIM (Building Information Modeling) technology is applied to management engineering, which is convenient for project management personnel. Method: The K-means clustering algorithm is combined with the support vector machine algorithm. The support vector machine is used to ensure the high accuracy of the anomaly detection algorithm. The K-means clustering algorithm is used to divide the support vector machine into blocks. It also analyzes the different needs of the facility management staff, and clearly defines the content and level of detail required to build the BIM model. It not only meets the data requirements for operation and maintenance but also avoids waste caused by excessive modeling. Result: Compared with traditional support vector machines, the improved algorithm in this paper has a higher detection rate and lower false alarm rate. Also, it can shorten the detection time of large-scale data to provide an effective method for abnormal detection of sensor networks and processing of large-scale data sets. The improved method increases the detection accuracy by 8.13% and decreases the false alarm rate by 89.08%. In terms of detection time, the improved method increases by 3.82s, which is 4.67 times the traditional method. Conclusion: The structural health monitoring system can efficiently and accurately monitor the accuracy of the data. BIM can provide rich operation and maintenance data for facility management to effectively improve the efficiency of facility management.

Keywords – K-Means Algorithm; BIM; Facility Management.

1. Introduction

With the vigorous development of China's economy and science in recent years, the construction industry is also changing with the development of society. To meet the increasing demand of the urban population, China's building structure has continued to develop from single- and multi-story building structures to high-rise and even super-high-rise structures. In recent years, more and more builders have applied BIM technology to the entire life cycle of construction. The operation and maintenance phase is the longest phase in the entire life cycle of construction. And the cost of the phase accounts for more than half of the total cost of the building installation [1]. In the operation and maintenance phase of a construction project, it needs not only the operation and maintenance data for facility management but also the data generated during the design and construction phases [2]. The phase has high requirements for the integrity and accuracy of operation and maintenance data in facility management. Because the collection of operation and maintenance data in facility management involves the construction unit, the survey and design unit as well as the operation and maintenance team itself. It is difficult to obtain accurate information and has poor timeliness and incompleteness. In the traditional property operation and maintenance mode, facility managers must pass a large number of paper documents such as drawings and contracts to grasp the operation and maintenance data required for facility management. The management efficiency is low and the level is backward [3].

SHM (Structural health monitoring) refers to arranging sensor nodes around the structure so that these nodes can form a sensor network. It first uses the sensor's sensing layer to sense response data such as strain, speed, acceleration, displacement, and rotation. Then, it transmits the response data to the convergent node through a multi-hop network (sensor network) composed of each sensor node. By connecting to the Internet at the convergent node, the structural response data is transmitted to the host (base station) that the staff uses to regulate the system and analyze the structural health. Then, in the base station, the purpose of real-time detection of the structure is achieved through damage recognition technology [4,5]. Structural monitoring allows us to understand the health of the structure and the degree of damage in time. At present, the research directions of anomaly detection can be divided into three. One is to develop better signal processing methods. Two is to establish a more complete normal network behavior model. Three is to integrate artificial

ISSN 2348 - 9928
IJAICT Volume 7, Issue 9, September 2020 Doi:10.46532/ijaict-2020030 Published on 05 (09) 2020

intelligence technology into anomaly detection algorithms to imitate the human brain's inference and prediction method to detect anomalies. Many worldwide researchers have made outstanding contributions to the development of anomaly detection [6,7]. As far as the current situation in China is concerned, some of the operation and maintenance management systems have been researched and used. However, the current use of the system only stays at the data entry stage. It cannot perform a complete operation and maintenance management process. Also, it cannot perform the corresponding task allocation. The flexibility is relatively low, and the input data cannot be analyzed in an orderly manner, making the data cluttered. It can only view the data simply, and the effect cannot meet the requirements for use.

The research content of this paper is to develop an anomaly detection algorithm suitable for the sensor network of structural health monitoring systems for the purpose of simple, efficient, energy-saving and accurate detection. The K-means clustering algorithm is combined with the support vector machine algorithm. The support vector machine is used to ensure the high accuracy of the anomaly detection algorithm. The K-means clustering algorithm is used to divide the support vector machine into blocks. It also analyzes the different needs of the facility management staff, and clearly defines the content and level of detail required to build the BIM model. It not only meets the data requirements for operation and maintenance but also avoids waste caused by excessive modeling. This paper has the following innovations. First, the K-means clustering method is applied to the operation and maintenance management system so that the system can perform more efficient data analysis. Second, BIM technology is applied to the anomaly detection of buildings and combined with the operation and maintenance management system.

2. Methodology

K-means clustering method

Clustering Analysis is to divide the high similarity data in the dataset into the same class so that the similarity between the data in the same class is as high as possible. And it makes the difference between the data of different classes as large as possible to form two or more classes. Clustering Analysis is a research project that has been active in the field of data mining. Much statistical software such as S-plus, SPSS, and SAS have included many cluster analytical tools such as K-means and K-medoids. The main requirements of data mining for Clustering Analysis are scalability, high-dimensional applicability, effectiveness, anti-noise, and insensitivity to input order [8,9]. Scalability means that the clustering algorithm needs to handle both limited small sample data and massive data over one million. High-dimensional practicability requires that the clustering algorithm not only performs well when processing low-dimensional data but also easily copes with high-dimensional data with dozens of dimensions [10]. Anti-noise refers to the ability to resist data sets that contain a lot of noise in reality [11]. Some clustering algorithms are very sensitive to the input order of the data set. Different input orders may cause large differences in clustering results. This is due to the imperfection of the clustering algorithm, and it is necessary to develop a clustering algorithm that is not sensitive to the input order. The comparison results of various clustering algorithms are shown in Tab. 1.

Table 1. The result of clustering algorithm comparison

Arithmetic	Efficiency	Memory space	Sensitivity to data
K-means	General	General	Sensitivity
CLARANS	Lower	Higher	Insensitivity
BIRCH	High	Higher	Insensitivity
DBSCAN	General	High	Sensitivity
CURE	Higher	General	Insensitivity

According to the comparison results, K-means has extremely high data sensitivity and the memory space is not high. Although the efficiency is general, K-means is most suitable for the system based on the algorithm and system requirements. At the same time, the K-means algorithm is simple to implement and is a very widely used clustering algorithm. Therefore, this system uses the K-means algorithm.

K-means has strong versatility and is independent of other applied disciplines. Therefore, it is widely used in the fields of pattern recognition, machine learning, spatial database, biology, and marketing [12]. In 1967, J.B.Mac Queen first proposed the K-Means clustering algorithm. Its principle is simple and easy to understand, and the clustering process is easy to implement. It is one of the most influential and most commonly used clustering algorithms in various industries such as engineering, medicine, agriculture, and computers. And it belongs to a clustering algorithm based on partition. K-means clustering is a dynamic and continuous iterative process. The measurement of sample points and the similarity between sample points and various clusters usually uses the Euclidean distance equation. That is, the closer the distance between two sample points, the greater the similarity. The ultimate goal of clustering is to make data with high similarity as close as possible to the same class. And it makes the similarity between classes as low as possible (that is, the distance is as small as possible). K-means clustering first determines the number of clusters to be divided as K and selects K cluster centers. Then according to the distance (similarity) between each sample point and each cluster center, the assignment of each sample point is determined. And each sample point is divided into the class to which the nearest cluster center belongs. The average of each class is calculated as the new cluster center. When the cluster center of each class no longer changes or the difference between the two cluster centers is less than a certain threshold specified in advance, the clustering process stops. The formation of K classes is complete. Taking one-dimensional data with n data points including $x_1, x_2 \ldots\ldots x_n$ as an example (the same applies to high-dimensional data), it is supposed the selected one-

ISSN 2348 - 9928
IJAICT Volume 7, Issue 9, September 2020 Doi:10.46532/ijaict-2020030 Published on 05 (09) 2020

dimensional sample set is $\{x_1, x_2, x_3 ...... x_n\}$, and its Euclidean distance is as shown in Eq. (1). The flowchart of the clustering algorithm is shown in Fig. 1.

$$d = \sqrt{\sum_{k=1}^{n} (x_k - y_k)^2} \quad (1)$$

Where: d represents the Euclidean distance between two samples. The meaning of each attribute in the dataset is different, and the contribution to the cluster is also different. Thus, it is necessary to consider the weight of the attribute. The contribution of each attribute to the cluster is expressed by the size of the weight. It is added to the equation for calculating the Euclidean distance by using the weighting method of the variation coefficient. The coefficient of variation in a dataset is equal to the standard deviation of all the data in this data set divided by the absolute value of its mean. It is supposed that there are n data including $x_1, x_2 ...... x_n$ in the dataset and p attribute. These are shown in Eq. (2), (3), (4).

$$S_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2} \quad \text{i=1,2, …, n} \quad (2)$$

$$v_i = S_x / |\overline{x}| \quad \text{i=1,2, …, p} \quad (3)$$

$$w_i = \frac{v_i}{\sum_{i=1}^{p} v_i} \quad \text{i=1,2, …, p} \quad (4)$$

Where: v_i represents the coefficient of variation. w_i represents the weight of each attribute. Weighted Euclidean distance is shown in equation (5).

$$d(x_i, x_j) = \sqrt{w_1 (x_{i1} - x_{j1})^2 + w_2 (x_{i2} - x_{j2})^2 + \cdots w_p (x_{ip} - x_{jp})^2} \quad (5)$$

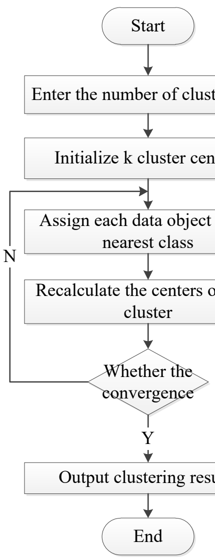

Fig 1. The flowchart of the K-means clustering algorithm

BIM facility management method

BIM is a software technology that is gradually applied throughout the life of construction. BIM can provide an accurate 3D digital representation of buildings visually and dimensionally. It is also a database with the ability to track attribute data for individual components in a building model. The BIM model describes the 3D geometry and properties of the actual facility. BIM provides building geometry and is also a structured database that provides non-graphical data on building component details. Facilities components in BIM can be sorted, counted, and queried. BIM can be connected to the component data in the model. It provides a wide range of functions including data extraction, cost estimation, space and asset management, and energy analysis. Also, it can manage the relationships of all objects in the model, and their other characteristics, to quickly query and change data.

BIM has the function of parameterization. Components in the model have attributes and parameters. These define the relationship with other components. For example, an object of a door will be attached to or associated with an object of a wall. BIM technology has changed the delivery process of the project, made the collaboration of all parties more efficient, realized data sharing, reduced costs, and improved efficiency.

COBie (Construction Operations Building Information Exchange) is an international standard for building asset management. It transforms the previous delivery format based on paper information into an international open standard delivery format. It is difficult to extract information for the delivery of paper materials. Compared with it, COBie has made the delivery process of operation and maintenance data more standardized and efficient. It effectively reduces the loss of operation and maintenance data during transmission and the cost of operation and maintenance data transmission [13]. The organizational framework of COBie data is shown in Fig. 2.

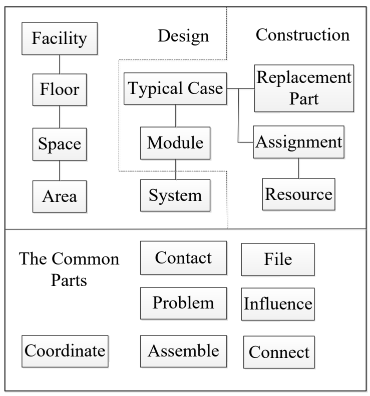

Fig 2. The organizational framework of COBie data

BIM is a technology that integrates all stages, participants, and business systems in the industrial construction process [14]. The data of various professional business systems are integrated into the BIM model. Relatively comprehensive and complete operation and maintenance data of facility management can be collected through BIM [15]. In addition, after integrating BIM and FM, operation and maintenance data can be obtained from the BIM model and entrees into the operation and maintenance system. The important operation and maintenance data, such as space, area, and facility type, can be obtained from BIM without re-entering manually. It not only avoids the cost of manually inputting operation and maintenance data but also generates higher quality operation and maintenance data. The flowchart of BIM management is shown in Fig. 3.

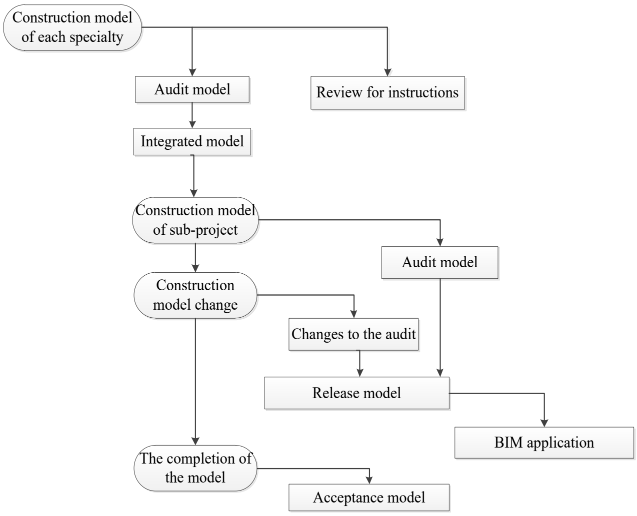

Fig 3. The flowchart of BIM management

Anomaly detection algorithm based on block support vector machine

Support Vector Machine (SVM) is a general algorithm based on the theory of structural risk minimization. It overcomes the problems of “dimension disaster” and “over-learning” and belongs to convex quadratic programming. At present, the support vector machine is in a booming stage and is widely used by researchers in various industries as an emerging force. The support vector machine can be used as classifiers or regression analysis. It is a very effective method for processing high-dimensional, small samples, and inseparable data, with strong generalization.

The support vector machine has good generalization, which can solve the problems of “dimension disaster” and “over-learning”. But there are also many problems that have not been completely overcome. For example, when applying a support vector machine algorithm, there is no unified method for how to choose a proper kernel function. There is no theoretical guidance in this regard, and it is almost based on experience. In addition, the support vector machine can handle high-dimensional data well, but its computational complexity depends only on the number of samples in the dataset. The number of samples in the actual problem is often massive, which leads to high calculation complexity and low efficiency.

The support vector machine is usually classified into two types: linear support vector machine and nonlinear support vector machine. The linear support vector machine is a classification algorithm in two-dimensional space. The non-linear support vector machine refers to using a kernel function to map data into a high-dimensional space when it encounters linear inseparable data.

The support vector machine has good classification performance and is good at handling high-dimensional and indivisible small sample data. Therefore, it is widely used in network anomaly detection. But its complexity depends on the amount of data to be processed. When it is used to process massive amounts of network data, problems such as long training time and low detection efficiency occur. Thus, this paper uses the block support vector machine for network anomaly detection, which can shorten training time and improve detection efficiency.

In the block support vector machine, the block means that before training the support vector machine, the K-means clustering algorithm is first used to divide the training dataset into K clusters. And the support vector machine is trained in each cluster to form K support vector machine classifiers. When detecting, the Euclidean distance between the K cluster centers and each detection data is first calculated, and each detection data is assigned to the cluster closest to it. The support vector machine trained in the cluster is used to perform anomaly detection on this detection data. It makes K support vector machines work together to reduce the number of data that each support vector machine needs to process. This can reduce the training time for large-scale data, and improve the speed of network anomaly detection on the premise of ensuring detection accuracy.

The computer model used in this test is Macbook Pro. The operating system is OS X Yosemite 10.10.4. The memory is 8GB, and the processor is 2.7GHz Intel Core i5. Because other versions of MATLAB still have some vulnerabilities, which are not perfected. A newer version of Matlab2014b is selected as the platform for testing. The experimental algorithm flow is shown in Fig. 4.

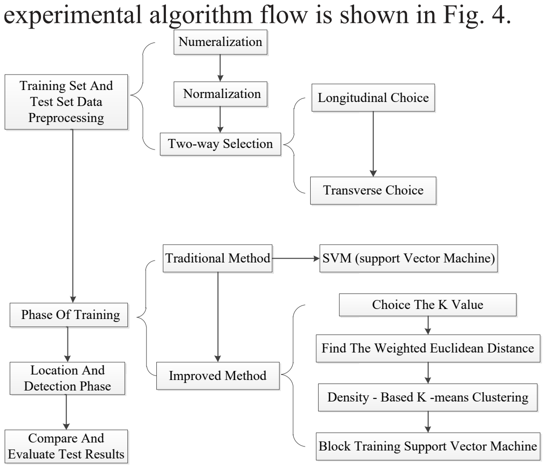

Fig 4. The flowchart of the experimental algorithm

3. Results

Experimental results of anomaly detection systems

Through improved K-means clustering, the training sample dataset is divided into 4 classes. In each class of data, two classes of classification support vector machines are trained to obtain four SVM classifiers. The complexity and time consumed for training and detection of support vector machine classifiers mainly depend on the amount of data to be processed. Therefore, a support vector machine is divided into several small classifiers for training and detection. It is helpful to reduce the computational complexity of each support vector machine. Since several

ISSN 2348 - 9928
Doi:10.46532/ijaict-2020030 Published on 05 (09) 2020

classifiers work at the same time, it can reduce the training time and improve the detection efficiency.

In the data preprocessing phase, the anomaly data in the training sample data is labeled as 1 and the normal data is labeled as 0. These have been stored in the vector label. Then, the sum function is applied to sum the labels of the four classes of data respectively. Since the sum of the first class of data label is 0, it proves that the first class of data is normal data. The sum of the last three classes of data labels is less than the data capacity of the corresponding class. It indicates that the last three classes include normal data and anomaly data, respectively. Therefore, the support vector machine classifiers for the last three classes of data are needed to train. In the subsequent testing process, the test data positioned as the first class can be directly judged as normal data.

The traditional support vector machine and block support vector machine are used to train the training data samples after bidirectional selection. The test data samples after the bidirectional selection are detected, and the detection precision and false alarm rate of the traditional and improved methods are calculated. Then, tic and toc functions in Matlab are used to calculate the time of the detection process to compare the advantages and disadvantages of these two detection methods. The specific detection result is shown in Fig. 5.

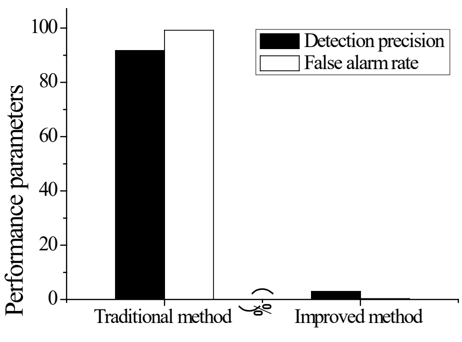

Fig 5. The detection result graph of sample data

From the test results, for the 10,000 training data and 5000 test data, the detection precision of the improved method is 99.24%, and the false alarm rate is 0.38%. The traditional method of support vector machine has a detection precision of 91.68% and a false alarm rate of 3.03%. After calculation, the detection precision of the improved method increases by 8.13%, and the false alarm rate decreases by 89.08%. In terms of detection time, the improved method increases by 3.82s, which is 4.67 times the traditional method. However, it is acceptable for the measurement process with a small amount of data to not cause an excessive increase in time.

Since the clustering results of the K-means clustering algorithm are very sensitive to the selection of the initial center, random changes of the initial center may have a great impact on the clustering results. This paper proposes a density-based initial center selection method, which avoids the impact of large initial center volatility on the results. It also overcomes the phenomenon that the initial centers are too concentrated or too discrete when randomly selected, and strives to maintain a reasonable distance between each initial center. The former random selection is replaced by a more scientific method that makes the initial center more stable, improving the accuracy of the K-means clustering algorithm.

The kddcup.data_10_percent dataset is used as training data, and the corrected dataset is used as test data. The traditional and improved methods are respectively detected. To compare and prove the effect of adding bidirectional selection processing in the data preprocessing phase, the data that has undergone numerical and normalized processing is not subjected to bidirectional feature selection. And the traditional methods described above are directly used for training and detection. The detection result is shown in Fig. 6.

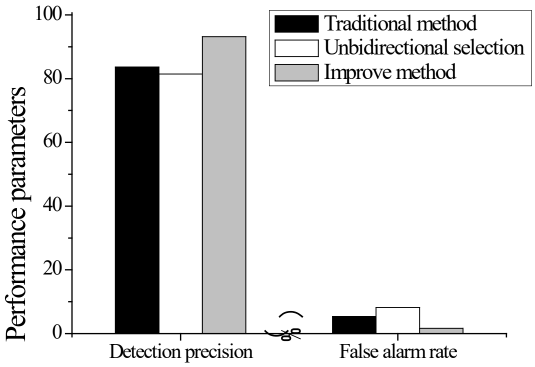

Fig 6. The data detection result of kddcup.data_10_percent

Experimental results of operation and maintenance management platform

Docking BIM with the data of the integrated energy management system can realize accurate measurement of energy consumption data (large-scale equipment such as lighting, HVAC, elevators, water supply, and drainage) in the project and real-time energy consumption monitoring. At the same time, BIM and the integrated energy management system are combined with other intelligent subsystems and related data of information collection. The data includes indoor and outdoor environments, people stream, vehicle stream, event information (large conferences and high-standard reception) for comprehensive energy optimization analysis.

Because BIM provides accurate 3D models of the building, it can help analyze and plan emergency measures and rescue route requirements. BIM technology can provide multiple functions for disaster prevention. A simulation of the evacuation of personnel is carried out to obtain the evacuation time. The exit channel and possible blocking points are analyzed. The technology can also view the range of video surveillance shooting angles. The operation and maintenance platform combines the BIM model to collect statistics on the time, duration, affected floor, loss, and cause of previous disasters. Dangerous sources are monitored by equipment sensors set up on the floor. Once there is a fire hazard such as high equipment temperature,

Corresponding author at: Jun Wang, South China University of Technology, China.

the system will issue an early warning on this. Then, the management staff will review it, issue a maintenance work order, and notify the maintenance staff to execute the work order. Maintenance staff will check according to the position of the located equipment. The management of disaster data will effectively reduce the possibility of disasters as well as the waste of human and financial resources.

Before the disaster occurs, the rich information and the disaster simulation software of the BIM model can be used to simulate the disaster process. Then, corresponding preventive measures and emergency plans for evacuation and rescue support after disasters can be formulated. When a disaster occurs, the BIM model can provide rescuers with data on the location of emergency situations such as fire facilities distribution and safe access locations. It is convenient for rescuers to analyze the situation at the disaster site, and take effective response measures in time to improve the effectiveness of emergency actions.

4. Discussion

In the structural health monitoring system, the sensor network bears the important mission of transmitting structural response signals. However, abnormal conditions will inevitably occur during the long-term service of the sensor network. It results in disordered response data received by the base station. The results of structural monitoring and damage recognition are inaccurate. This paper designs an anomaly detection algorithm of sensor networks based on the block support vector machine. The traditional methods are improved in feature selection, K-means clustering and support vector machine. The similarity of the two data in the cluster is determined by the traditional Euclidean distance. The weighted Euclidean distance defines the weight of each attribute according to the contribution (sensitivity) of each attribute to the cluster. Then, the weight is added to the traditional Euclidean distance equation to distinguish each different attribute. This process is more targeted to high-dimensional data.

The complexity and the training time of the support vector machine only depend on the features of the amount of data in the dataset. Then, the method of block support vector machine is adopted. It refers to splitting a large support vector machine into K small support vector machines, and the training and detection of these support vector machines are performed simultaneously in their respective clusters. It saves training time, reduces complexity, and improves detection efficiency. After preprocessing including bidirectional selection, the small sample data in the kddcup.data_10_percent the dataset is used to perform the test of traditional and improved methods. The detection results show that the improved detection method has higher detection precision and a lower false alarm rate, but the detection time is relatively long. The key to applying BIM to facility management is to solve the problem of "information faults" and realize the complete and accurate transmission of operation and maintenance data. Through research on the definition of facility management and the content of operation and maintenance, the operation and maintenance data of facility management are summarized and classified. According to the concept and functional characteristics of BIM, the convenience and advantages brought by BIM technology to the collection, management, transmission, and retrieval process of operation and maintenance data are clarified.

5. Conclusion

Compared with traditional support vector machines, the improved algorithm in this paper has a higher detection rate and a lower false alarm rate. Also, it can shorten the detection time of large-scale data, and provide an effective method for anomaly detection of sensor networks and processing of large-scale datasets. The improved method increases the detection precision by 8.13% and decreases the false alarm rate by 89.08%. In terms of detection time, the improved method increases by 3.82s, which is 4.67 times the traditional method. The author's research in structural monitoring technology, sensor network, anomaly detection, and Matlab software is only at the initial phase, and understanding of related fields is not comprehensive enough. Therefore, the work that can be done in the anomaly detection of the sensor network of the structural monitoring system is limited. There are still many aspects that need to be further improved. The conclusion is drawn by combining the author's practical experience and the learning of a large number of BIM data in the operation and maintenance phase. Compared with previous studies, the innovation of this paper lies in the design and implementation of a stable new operation and maintenance management system. Also, the system can deliver tasks flexibly and controllably, reducing the operation and maintenance workload. On the basis of this system, the BIM method is combined for anomaly detection during operation and maintenance. Due to the limited research time, author's experience and learning level, the research on the operation and maintenance data of BIM-based facility management from several points has limitations.

References

[1] Li H, He H, Wen Y. Dynamic particle swarm optimization and K-means clustering algorithm for image segmentation. Optik, 126 (2015), 24, pp. 4817-4822.

[2] Gupta S, Kumar R, Lu K, et al. Local search methods for k-means with outliers. Proceedings of the VLDB Endowment, 10 (2017), 7, pp. 757-768.

[3] Slamet C, Rahman A, Ramdhani M A, et al. Clustering the Verses of the Holy Qur'an using K-Means Algorithm. Asian Journal of Information Technology, 15 (2016), 24, pp. 5159-5162.

[4] Qin J, Fu W, Gao H, et al. Distributed $ k $-means algorithm and fuzzy $ c $-means algorithm for sensor networks based on multiagent consensus theory. IEEE transactions on cybernetics, 47 (2016), 3, pp. 772-783.

[5] Botía J A, Vandrovcova J, Forabosco P, et al. An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. BMC systems biology, 11 (2017), 1, pp. 47.

[6] Frandsen P B, Calcott B, Mayer C, et al. Automatic selection of partitioning schemes for phylogenetic analyses using iterative k-means clustering of site rates. BMC evolutionary biology, 15 (2015), 1, pp. 13.

[7] Xing K, Hu C, Yu J, et al. Mutual privacy preserving $ k $-means clustering in social participatory sensing. IEEE

ISSN 2348 - 9928
IJAICT Volume 7, Issue 9, September 2020 Doi:10.46532/ijaict-2020030 Published on 05 (09) 2020

Transactions on Industrial Informatics, 13 (2017), 4, pp. 2066-2076.

[8] Cohen-Addad V, Klein P N, Mathieu C. Local search yields approximation schemes for k-means and k-median in euclidean and minor-free metrics. SIAM Journal on Computing, 48 (2019), 2, pp. 644-667.

[9] Liu H, Wu J, Liu T, et al. Spectral ensemble clustering via weighted k-means: Theoretical and practical evidence. IEEE transactions on knowledge and data engineering, 29 (2017), 5, pp. 1129-1143.

[10] Friggstad Z, Rezapour M, Salavatipour M R. Local search yields a PTAS for k-means in doubling metrics. SIAM Journal on Computing, 48 (2019), 2, pp. 452-480.

[11] Ferrandez S M, Harbison T, Weber T, et al. Optimization of a truck-drone in tandem delivery network using k-means and genetic algorithm. Journal of Industrial Engineering and Management (JIEM), 9 (2019), 2, pp. 374-388.

[12] Luarn P, Lin H W, Chiu Y P, et al. The Categorising Characteristics of Facebook Pages: Using the K-Means Grouping Method. International Journal of Business and Management, 11 (2016), 2, pp. 60.

[13] Malav A, Kadam K, Kamat P. Prediction of heart disease using k-means and artificial neural network as Hybrid Approach to Improve Accuracy. International Journal of Engineering and Technology, 9 (2017), 4, pp. 3081-3085.

[14] Haut J M, Paoletti M, Plaza J, et al. Cloud implementation of the K-means algorithm for hyperspectral image analysis. The Journal of Supercomputing, 73 (2017), 1, pp. 514-529.

[15] Shakeel P M, Baskar S, Dhulipala V R S, et al. Cloud based framework for diagnosis of diabetes mellitus using K-means clustering. Health information science and systems, 6 (2018), 1, pp. 16.