ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20240325001429
Original scientific paper

Research on Optimal Design of Civil Sensors Based on Agglomerative Hierarchical Clustering Algorithm

Xingyan CHENG*, Linyan ZHU, Yimei CHENG

Abstract: In practical engineering, the sensor measurement points will have obvious clustering characteristics. Therefore, the cohesive hierarchical clustering algorithm is introduced in this paper. Firstly, the degrees of freedom are classified according to the similarity of vibration features, and all degrees of freedom are divided into multiple clusters, and disjoint subsets are formed between the clusters to avoid the concentration of measurement points. Secondly, the hierarchical clustering algorithm is improved, and a single objective function sensor location search method is established with the minimum discomfort sensor location selection criterion and MAC pattern guarantee criterion as objective functions. Different methods are used to search the best position of acceleration sensor. Finally, the empirical performance and scalability of the proposed algorithm are verified by an example analysis. In this paper, a new branch direction of hierarchical clustering is studied, which provides a meaningful exploration for the empirical performance and scalability of balanced hierarchical clustering, and provides new possibilities for structured data analysis and mining tools.

Keywords: civil sensor optimization design; cohesive hierarchical clustering; degrees of freedom; location search; monitor the number of modes

1 INTRODUCTION

Inappropriate sensor layout will result in inappropriate information collection, information duplication, and could not effectively extract useful structural information, which will affect the accuracy of the health status assessment in the results. Theoretically, the more information about the modal parameters of the structure, the more accurate the actual state evaluation of the engineering structure will be [1]. However, if the number of sensors installed in the actual project is too large, it will lead to the high cost of arranging a large number of sensors, the cost increases, and the configuration and maintenance of sensors also require great costs, resulting in unnecessary extravagance and waste. In addition, too much sensor arrangement will capture too much response information, which will cause excessive burden on the storage and communication system in the process of transmission, and eventually lead to low transmission efficiency. The excessive processing time of the obtained response data will be too long, resulting in the delay of the structural diagnosis time, which is not conducive to timely reflecting the engineering structure information. Therefore, in engineering practice, it is necessary to analyze many factors to find the optimal position of sensor layout under the condition of reasonable number of sensors. The following problems [2, 3] should be considered to achieve discrete representation of 0 or 1 in the optimization results (where 0 means that no sensor is arranged at this position, and 1 means that sensors are arranged at this position). In this way, in the actual engineering structural problems, there may be a lot of information obtained at the arrangement points that are not of great significance to the actual health monitoring. Therefore, it is necessary to find more effective sensor arrangement measuring points to meet the actual needs. Excessive sensor arrangement will also lead to the collection of massive data, which is a great challenge for data collection and subsequent processing and analysis. It is necessary to select the appropriate sensor layout criteria to evaluate the sensor layout scheme according to the engineering practice, and select the effective sensor layout optimization method to optimize the sensor layout scheme. This paper mainly focuses on sensor optimization layout criteria and sensor optimization layout methods based on relevant criteria. The first part of this paper is introduction, the second part is related work. The third part is research and application of cohesive hierarchical clustering in civil engineering sensor optimization, The fourth part is optimal layout design of cohesive hierarchical clustering of civil sensors, the fifth part is simulation verification, and the sixth part is conclusion.

2 RELATED WORK

It focuses on the choice of sensor layout position, and establishes different sensor layout criteria. A method for determining the location of sensor layout is established [4, 5]. The core idea of this method is to make the mode vector orthogonal, so to maximize the Fisher information array. This method is the most widely used method in the field of sensor layout at present, and is the cornerstone of other sensor layout methods, with relatively mature development. This method proposed for the first time that the more sensors the better [6], too many sensors, not only redundant information will interfere with the calculation accuracy, but also reduce the calculation efficiency, a reasonable number of sensor layout and sensor layout location selection is equally important. After that, the effective independence method was improved [7]. At the same time, there would be errors between the model and the actual structure, making this theoretical method more practical. This method selects the sensor position by identifying the contribution value of each sensor position to the modal vector independence. The modal kinetic energy method is established, and the energy of the structure has great uncertainty at each position, so the layout position of the sensor should be selected on the degree of freedom containing large modal kinetic energy [8], so to obtain a higher signal-to-noise ratio. Through comparative study, it is found that modal kinetic energy method is actually an iterative operation of modal kinetic energy method. This method will first exclude the node position [9], because the node position line vector product value is small, and the defect of this method is also obvious, and the sensor is usually concentrated in a certain area. The method of adding modal vectors [10] is proposed. Its principle is to first obtain the modal matrix, then take the absolute values of the elements in each vector of the modal matrix, and then add them together. The position where the absolute values sum is the largest is selected to arrange the sensor. A modal assurance criterion is established [11]. The sensors should be positioned in such a way that the Angle between the spatial vectors is as large as possible. The modal guarantee matrix MAC [12, 13] can be obtained through pair-to-pair product calculation of modal vectors, and the sensor should be arranged in the position where the maximum non-diagonal element of the modal guarantee matrix MAC is the minimum. A method for optimizing the degrees of freedom [14, 15] is proposed, the principle of which is to consider the sensitivity of the degrees of freedom to the rate of stiffness change, and then select the sensor on these optimized degrees of freedom. This paper proposes a damage identification method that can analyze the crack length [16]. The principle is that once the structure is damaged, the structural parameters will inevitably change, and the change of structural parameters will inevitably change the structural modes, and then the crack length can be analyzed through the change of structural modes. In order to improve the efficiency of calculation, relevant scholars began to apply intelligent algorithms to the search of the optimal position of sensors. It is then slowly cooled down so that the solid particles in it become disordered when it heats up [17, 18] and ordered when it cools down. When this algorithm is used to search the optimal position of the sensor [19], it is found that whether the final search converges to the global optimal solution plays a crucial role in determining the initial parameters. Genetic algorithm is applied to the optimal position search of sensors [20, 21]. Genetic algorithm has four processes of gene inheritance, gene variation, natural selection and gene mutation, and the last remaining population is the optimal population. The effective independence method was used as the objective function [22], and the genetic algorithm was applied to the search of the optimal position of the sensor, which was widely used at that time. The Wolf pack algorithm was applied to the sensor optimal location search [23]. The multi-subgroup particle swarm optimization algorithm is applied to search the optimal position of the sensor [24] and verified by placing sensors on the Laxiwa arch dam. This kind of method can better solve the constraints of related problems, local optimal problems and other characteristics, and has been developed in the field of sensor optimization placement [25-27]. The Wolf pack algorithm is proposed to avoid the problem of premature convergence [28]. The intelligent optimization correlation algorithm uses probabilistic random search optimization strategy to solve the problem of deterministic optimization strategy [30, 31]. Recent studies have made a lot of improvements, but it is difficult to make further breakthroughs in the current relevant studies to get the final optimization results. In essence, these algorithms continuously optimize the objective function in a certain optimization direction during the optimization process, which reduces the complexity compared with the traditional methods to a certain extent, and essentially belongs to a class of random search methods. There are still many shortcomings in the layout method. Therefore, it is hoped that the characteristics of the objective function can be effectively extracted through the representation ability of the agglomerative hierarchical clustering algorithm, and an efficient optimization strategy can be used to optimize the objective function and find the optimal solution. Therefore, the cohesive hierarchical clustering optimization algorithm is introduced to optimize the sensor layout.

3 RESEARCH AND APPLICATION OF COHESIVE HIERARCHICAL CLUSTERING IN CIVIL ENGINEERING SENSOR OPTIMIZATION

3.1 Research Framework for the Optimal Design of Civil Sensors Based on Cohesive Hierarchical Clustering

In the application of massive data mining, it is necessary to pursue both effect and efficiency. In this paper, a novel hierarchical clustering algorithm called condensed nearest neighbor clustering is proposed. The basic assumption and logic of the algorithm are different from the traditional hierarchical aggregation clustering algorithm and incremental clustering algorithm. It regards the nearest neighbor data points as the core of the cluster and is used to construct the hierarchical structure of the data iteratively. This paper explores and researches the basic framework and application of the algorithm, and how to balance the empirical performance and scalability of the algorithm. The main innovations and contributions are as follows, and the logical relationship diagram is shown in Fig. 1.

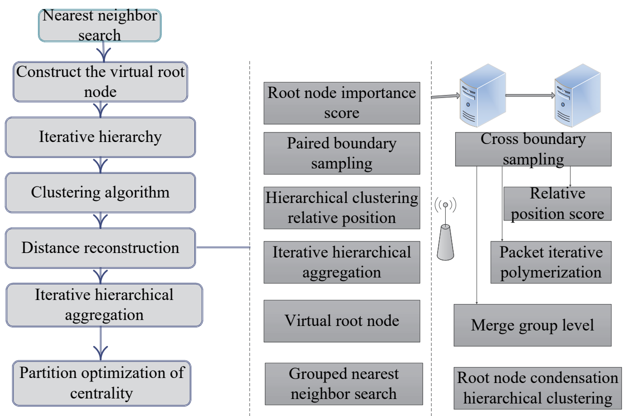

Figure 1 Schematic diagram of optimal design logic of civil sensor based on cohesive hierarchical clustering

As shown in Fig. 1, a coherent hierarchical clustering algorithm is proposed. Comparison experiments on real world data show that the clustering accuracy is better than other comparison algorithms in most cases. Aiming at the potential complexity of data features, the cohesive hierarchical clustering algorithm is optimized and extended to the community detection application of complex network data. The virtual root node is constructed as the representative node in the community, and the redundant communities are merged by calculating the intermediate centrality of the potential connecting edges between the communities, and finally the network nodes are organized into a community with a tree structure. To solve the problem of large-scale data clustering, an election tree algorithm for large-scale data parallelization is proposed. The election tree algorithm uses parallel grouping clustering and then merges root nodes to implement the clustering process. In the merging process, the relative position factor is used to realize fast cross sampling of data boundary and leaf node exchange of overlapping region, thus improving the quality of root nodes, correcting the attribution error of cluster edge nodes after group clustering, ensuring the accuracy of group clustering results on large-scale data, and realizing the clustering process with quasi-linear complexity.

3.2 Cohesive Hierarchical Clustering Model for Optimal Layout of Civil Sensors

Due to the limited number of sensors arranged on the structure, the interpolation method is used to obtain the information of unknown measuring points according to the information of known points, that is, the stress and strain values of any point are fitted by using the output response of measuring points on the structure, and the difference error is obtained by comparing with the calculated values. The error value will also change, and when the error value changes to the smallest, it indicates that the optimization scheme of the measuring point is the best. Set the target parameter t by directly collecting data x_1, ..., x_n decides that if δ represents the error of data set x and χ represents the error of target parameter t, then:

$$Y + \chi Y = \chi\left(x_1 + \delta_1, x_2 + \delta_2, ..., x_3 + \delta_3\right) \tag{1}$$

Expand the above formula according to the Taylor series:

$$\chi Y = \frac{\partial Y}{\partial x_1}\delta_1 + \frac{\partial Y}{\partial x_2}\delta_2 + ... + \frac{\partial Y}{\partial x_n}\delta_n \tag{2}$$

The optimal layout of civil sensors is equivalent to solving the two unknowns x and n in the position range according to the error minimum criterion, and obtaining the optimal layout position x vector and the number of sensors n, expressed as S. The optimization objectives are:

$$\min\left(\chi Y_{\max}\right) = \pm\left\{\left|\frac{\partial Y}{\partial x_1}\delta\right|_1 + \left|\frac{\partial Y}{\partial x_2}\delta\right|_2 + ... + \left|\frac{\partial Y}{\partial x_n}\delta\right|_n\right\} \tag{3}$$

A small number of civil sensors are used to detect the general area of damage, and then the sensor arrangement measurement points are added in this area to obtain more damage information, and the damage degree and loss location of civil sensors are analyzed. It can be seen from the structure dynamics that the natural mode of the structure is a set of mutually orthogonal vectors. However, due to many factors, the orthogonality of the measured modes could not be guaranteed in practical measurement. Therefore, in the structural dynamic test, the structural modal vector must be linearly independent as much as possible, and the original modal characteristics must be retained as much as possible, which is the basic requirement for modal recognition [29]. Moreover, linear independence is particularly important when testing results validate or update finite element models. Therefore, sensor optimization is usually carried out with the maximization of modal space intersection Angle as the optimization criterion to ensure the original mode characteristics as much as possible.

$$MAC_{ij} = \frac{\delta_i^T \delta_j}{\left(\delta_i^T \delta_i\right)\left(\delta_j^T \delta_j\right)} \tag{4}$$

The Q matrix condition number is as follows:

$$Cond(Q) = \|Q\|\left\|Q^{-1}\right\| \tag{5}$$

The optimal cluster number of the cohesive hierarchical clustering algorithm is determined by the function of compactness x, overlap degree v and separation degree c. Compactness measurement:

$$C = \sum_{i=1}^{c} \frac{\sum_{j=1}^{n} c_{ij}}{\max\left\|x_j - v_i\right\|} \tag{6}$$

The compactness D_{ij} from sample X_i to class i is defined as:

$$c_{ij} = \begin{cases} x_{ij}, x_{ij} \ge x_0 \\ 0 \end{cases} \tag{7}$$

D represents the number of clusters and n represents the total number of samples. Overlap measure:

$$D = \sum_{i=1}^{c}\sum_{j=1}^{n} D_{i1,i2,j} \tag{8}$$

The overlap degree of sample X_j between class L_1 and class L_2 is:

$$D_{i1,i2,j} = \begin{cases} 1 - \left(\delta_{i1j} - \delta_{i2j}\right) \\ 0 \end{cases} \tag{9}$$

Separation measure:

$$S = \min_{i \ge 1, j \le 1}\left\|v_i - v_j\right\| \tag{10}$$

Considering the difference of data distribution, cos function index value is introduced. The following coherent-level clustering validity function is constructed:

$$cos = \frac{\min_{i \ge 1, j \le 1, i \ne j}\left\|v_i - v_j\right\| + \frac{\sum_{j=1,n} c_{ij}}{\max\left(x_j - v_i\right)}}{\sum_{i=1}^{c}\sum_{j=1}^{n} D_{i1,i2,j}} \tag{11}$$

3.3 Optimization Algorithm of Civil Sensor Based on Condensed Hierarchical Clustering

Fitness function is a criterion for evaluation and selection based on the idea of classical algorithm. Each gene corresponds to a fitness value, which promotes the evolution of genes. The development of high fitness chromosomes and the elimination of low fit chromosomes make the whole population continuously develop towards the optimal solution. The flow chart is shown in Fig. 2:

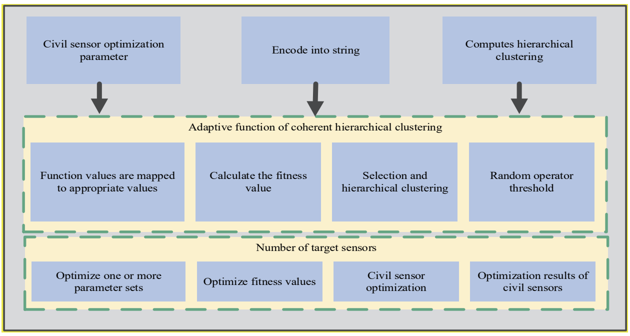

Figure 2 Flow chart of aggregation hierarchical clustering algorithm

As shown in Fig. 2, Step 1: Initialize the cluster center and set $b = 0$. Step 2: Update the classification matrix U, step 3: update the cluster center V, step 4: When the iterative operation reaches convergence, the cluster center is output, and each node in the network belongs to a different cluster and has the identity of each cluster. Otherwise, if $b = b + 1$, go to step 3 and step 5: The clustering result with the largest cos index value is the optimal clustering result, and the corresponding cluster number c value is the best cluster number, completing the optimal clustering. Set the civil sensor base station in the square area of (0, 0) 100 m × 100 m, and calculate the optimal number of cluster center nodes, as shown in Tab. 1:

Table 1 Optimal number of cluster center nodes

c	Cos / m
1	0.2659
2	0.1666
3	0.0643
4	1.2342
5	0.0186
6	0.0171
7	0.0065
8	0.0042

It can be determined from Tab. 1 that when $c = 4$, the maximum value of cos is obtained, that is, the optimal clustering number is 4. $Lt - 1$, and $Lt - 2$ is the death time of the first node and the death time of 30% node respectively. The network life cycle of cohesive hierarchical clustering is shown in Tab. 2:

Table 2 Network life cycle of cohesive hierarchical clustering

c	$Lt - 1$	$Lt - 2$
1	36	71
2	38	81
3	38	84
4	40	86
5	38	83
6	38	82
7	37	78
8	36	74

As can be seen from Tab. 2, WSNs has the longest lifetime when the number of CH nodes (Cluster Head nodes) is 4. CH nodes usually have to transmit data to distant CH nodes (relay nodes) until the data is transmitted to the base station. When the number of CH nodes in the network is too small, the distance between and within clusters in the network will increase, which will overload the energy consumption of CH nodes, and the member nodes will consume more energy. When the number of CH nodes is too large, the number of data transmission paths in the network becomes more and more, resulting in information redundancy and extra energy overhead. Therefore, it is very important for WSNs to extend the network life cycle by optimizing the number of CH nodes in a suitable range. Fig. 3 compares the performance of LEACH [24] algorithm and agglomerative hierarchical clustering algorithm in terms of network life cycle.

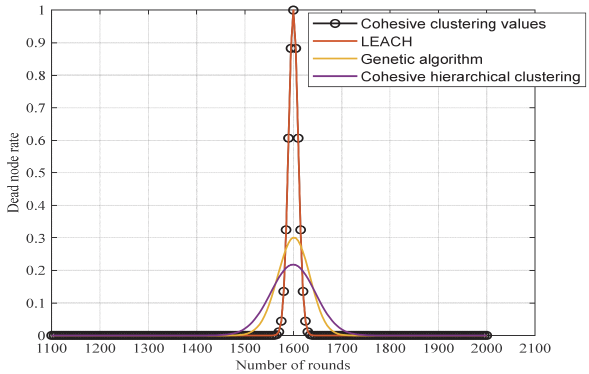

Figure 3 Number of dead network nodes

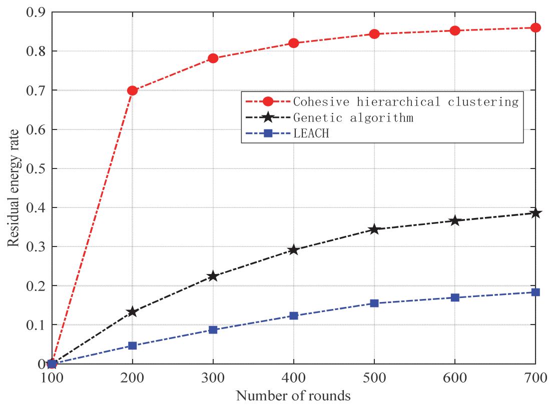

Figure 4 Residual energy of the network

As can be seen from Fig. 3 and Fig. 4, compared with LEACH algorithm [24], the energy consumption of agglomerative hierarchical clustering algorithm is more stable, which makes the energy consumption of the network more uniform and improves the performance of the network. With fixed clustering method, when the energy consumption of clusters is faster, nodes in the clusters will die together. The cohesive hierarchical clustering algorithm significantly improves the network life cycle.

4 OPTIMAL LAYOUT DESIGN OF COHESIVE HIERARCHICAL CLUSTERING OF CIVIL SENSORS

When two sensors obtain similar mode values in the same direction, one sensor can be selected to represent the vibration state information in the same direction. The aggregation hierarchical clustering algorithm is used to automatically aggregate the degrees of freedom, and the measurement points with less linear support for the target mode are removed step by step by step, and the optimal arrangement of the sensor is finally obtained. The process of this method is shown in Fig. 5, and the specific algorithm is described as follows.

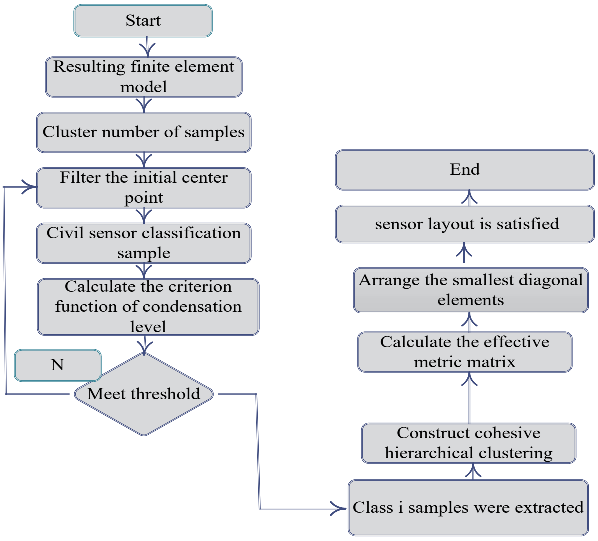

Figure 5 Flow chart of optimizing layout of civil sensors

Step 1: Determine the degree of freedom, calculate the mode work of each order of the structure, and select the important mode number.

Step 2: Taking the modal values of important modal vibration as the sample attributes, the cohesive hierarchical clustering algorithm is used to automatically classify each degree of freedom, and the standard measure function is used as the criterion until the criterion function meets the convergence conditions.

Step 3: For each cluster sample, the effective independence method is used to calculate the effective measurement of the sensor to be selected. According to the effective measurement size, the measurement point corresponding to the minimum effective independent contribution value is deleted, and it is repeated until the required number of measurement points is reached, so as to obtain more structural vibration information with as few sensors as possible. Modal confidence factor MAC (Multiple Atomic Chain) is a good tool for evaluating the intersection Angle of modal vectors [12], which can reflect the correlation between two spatial vectors. MAC is located between 0 and 1, where a value of 0 means that the two vectors are orthogonal, and a value of 1 means that the two vectors are completely related. Therefore, the larger the non-diagonal elements of MAC, the greater the correlation of vectors, and the lower the ability to reflect the characteristics of the original structure. The smaller the non-diagonal elements of MAC, the smaller the correlation of the vector, and the better it reflects the original structure characteristics. The optimized sensor arrangement can also use the mode expansion method to construct the response at the unmeasured points from the known finite points. Through the numerical analysis of the finite element software, the modal modes of the broken frame are calculated as a set of theoretical data, so as to obtain the effect value at any point of the structure. The modal displacements at known measuring points were extracted, and the modal displacements at non-measuring points were obtained by cubic spline interpolation fitting as a set of experimental data. The mean square error of these two groups of data is used to evaluate the advantages and disadvantages of the three methods in sensor optimization layout, and the mean square error can be expressed as:

$$\delta = \sum_{i=1}^{n} \frac{\frac{1}{\delta_i}(Q_{ij}^{m} - Q^{s}{}_{ij})^2}{n} \tag{12}$$

The three optimization methods are optimized by gradually deleting the measurement points with the smallest diagonal element and the smallest modal energy of the information matrix until the number of sensors meets the requirements. Tab. 3 shows the sensor ordering results of the three optimization methods when all measurement points are retained, that is, sensors are set in all degrees of freedom directions (The number of directions and angles that can move independently in space, expressed by x and y).

Table 3 Optimization results of all test points are retained for comparison

Degree of freedom	Independent method		Motion energy		Cohesive hierarchical clustering	
	Calculated value	Sort	Calculated value	Sort	Calculated value	Sort
1x	0	21	0	21	0	21
2x	0.0786	18	0.1423	19	0.008	18
3x	0.1143	16	0.1153	18	0.014	16
4x	0.1352	15	0.1765	16	0.0208	16
5x	0.2435	13	0.2435	12	0.0567	12
6x	0.3507	5	0.3601	6	0.1182	6
7x	0.3565	4	0.3872	4	0.1427	5
8x	0.3276	7	0.3256	8	0.1236	8
9x	0.1876	14	0.1754	17	0.0269	16
1y	0	22	0	21	0	21
2y	0.4523	2	0.3657	5	0.1603	4
3y	0.312	8	0.3125	10	0.0843	10
4y	0.3125	12	0.3232	9	0.0897	9
5y	0.4216	3	0.4786	2	0.1896	2
6y	0	23	0	23	0	20
7y	0.1879	15	0.2145	13	0.0368	14
8y	0.3254	7	0.3367	7	0.1245	5
9y	0.3783	4	0.4125	4	0.1678	3

According to the optimization results of the three methods when 8 sensors are deployed, modal guarantee criterion, mean square error minimum criterion and Fisher information matrix criterion are respectively used to evaluate. The MAC values of the three methods were calculated, as shown in Tab. 4 and Fig. 6. It can be seen that the total mean square error of MAC value and the correlation of modal vector are also the smallest in the cohesive hierarchical clustering algorithm, and the layout method can reflect the structural characteristics better than the other two methods.

Table 4 Comparison of modal assurance criteria of the three methods

Optimization method	MAC				MAC mean square error
	Order 1, 2 modes	Order 2, 3 modes	Order 3, 4 modes	Order 4, 5 modes	
Motion energy	0.003	0.0025	0.0073	4.87e-3	0.0257
Independent method	0.0094	3.87e-6	0.0127	0.0043	0.0078
Cohesive hierarchical clustering	2.9103e–7	0.0002	3.6e–4	0.0068	0.0046

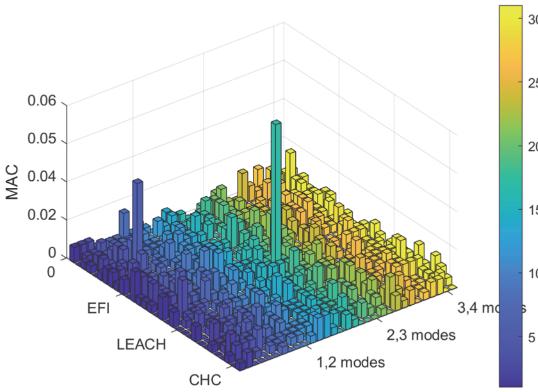

Figure 6 Comparison of MAC values of the three methods

As can be seen from Tab. 5 and Fig. 7, the minimum total mean square error value obtained by the cohesive hierarchical clustering algorithm is 0.0063, which indicates that among the three methods, the fitting results from eight measurement points are the closest to the ANSYS numerical simulation results. The layout effect of the cohesive hierarchical clustering algorithm is better than that of the motion energy method and the effective independent method.

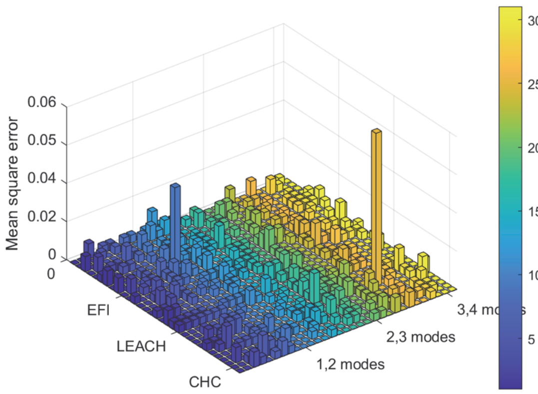

Figure 7 Comparison of the minimum mean square error of the three methods

Table 5 Comparison of the minimum mean square error values of the three methods

Optimization method	Mean square error of all modes				Total mean square error
	Order 1,2 modes	Order 2,3 modes	Order 3,4 modes	Order 4,5 modes	
Motion energy	0.0236	0.0456	0.0092	0.0247	0.0271
Independent method	0.0132	0.0089	0.0036	0.0064	0.0076
Cohesive hierarchical clustering	0.0085	0.0013	0.0052	0.0022	0.0063

5 SIMULATION VERIFICATION

In order to better identify the damage, this paper selects the layout index and the MAC minimum as the objective function of optimizing the layout of civil sensors, and determines the optimal location of the sensors through the cohesive hierarchical clustering algorithm. The structural monitoring system has the advantages of high recognition accuracy and strong ability to reflect the original structure information. Only acceleration sensors are deployed, and the number of sensors is 3 - 20. The calculation results are shown in Tab. 6.

Table 6 Calculation results of dual-target sensors with different numbers

Number of acceleration sensors	Layout index	MAC minimum	Calculation time / s
3	186.34	0.9312	652.95
4	193.93	0.9321	663.36
5	178.56	0.8831	667.85
6	191.31	0.8624	671.28
7	183.56	0.8348	672.74
8	185.98	0.8199	673.89
9	179.82	0.9135	676.39
10	183.34	0.8223	678.38
11	183.98	0.8163	681.56
12	181.04	0.8415	683.69
13	184.78	0.8226	687.27
14	183.25	0.8252	691.91
15	187.54	0.7418	706.82
16	184.28	0.8034	707.45
17	184.43	0.8051	712.58
18	184.89	0.8166	724.81
19	195.58	0.8478	725.85
20	189.62	0.7446	734.75

It can be seen from Tab. 6 that the optimization calculation time of civil sensors in the cohesive hierarchical clustering algorithm is related to the number of acceleration sensors. The more the number of acceleration sensors are deployed, the longer the calculation time. At the same time, it can be seen that the layout index of the optimal layout position and the minimum value of MAC are different when different number of acceleration sensors are deployed, which is very crucial for the selection of the number of acceleration sensors. The final objective function values of all decision variables under different number of acceleration sensors optimized by the cohesive hierarchical clustering algorithm for civil sensors are shown in Fig. 8.

It can be seen from Fig. 8 that under different number of acceleration sensors, the final result of the optimization search of civil sensors through the agglomerative hierarchical clustering algorithm is indeed the optimal solution, and the two objective function values of layout index and MAC minimum are comprehensively considered.

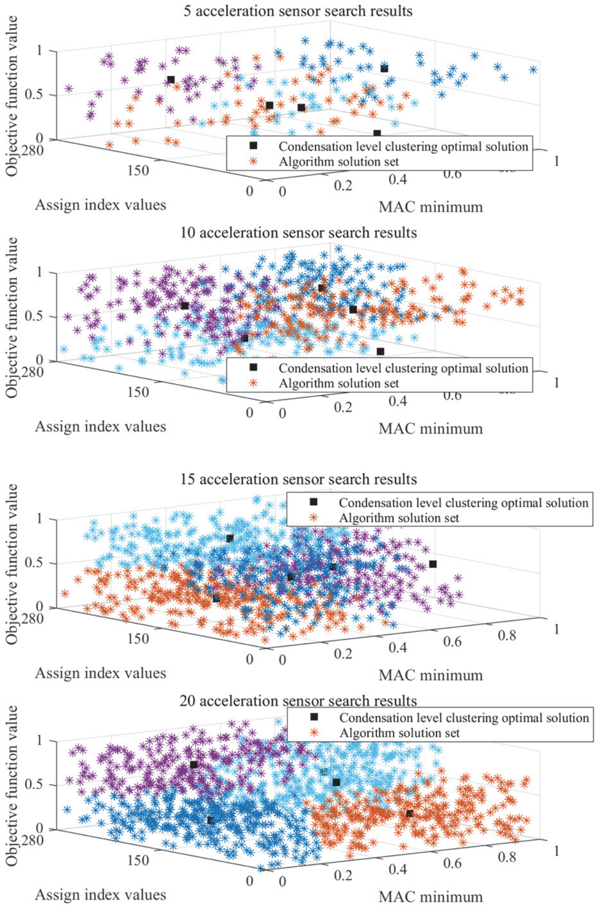

Figure 8 Dual-target sensor search results

With layout index, MAC minimum and cost as objective functions, the sensor layout optimization with three objective functions is carried out. The number of acceleration sensors can be selected from 3 - 20. The calculation results of optimized layout of civil engineering sensors are shown in Fig. 9:

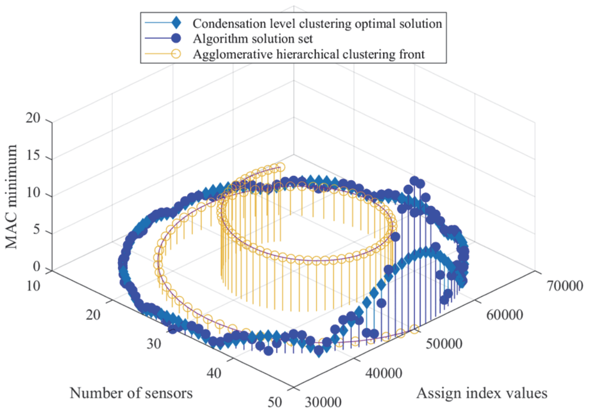

Figure 9 Calculation results of two types of sensors with three targets

As can be seen from Fig. 9, when the number of equivalent strain sensors is 25, the layout index value is 37535.73, and the minimum MAC value is 0.1395, that is, when the number of acceleration sensors is 2 and the number of strain sensors is 35, the corresponding solution is the optimal solution of the cohesive hierarchical clustering of the three objective functions, which is the optimal optimization of the three objective functions.

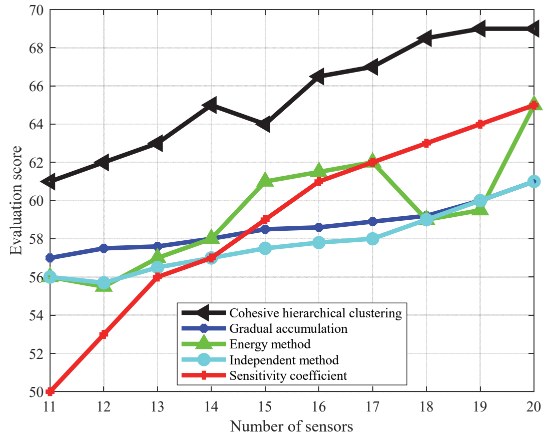

Figure 10 Comprehensive evaluation of civil sensor optimization

Fig. 10 shows the comprehensive scores of the five methods under the comprehensive evaluation system, which is a comprehensive evaluation from five aspects. If all evaluation indicators reach full marks, the comprehensive evaluation score is 100 points, which is not realistic. In engineering, it is impossible to arrange sensors on all degrees of freedom, and all comprehensive evaluation scores cannot reach full marks. As can be seen from the figure above, although the method presented in this paper cannot achieve all the optimal results in terms of a single evaluation index, the comprehensive evaluation score is still the highest with the increase of the number of sensors. The results show that the proposed method can guarantee the response, orthogonality, damage sensitivity and so on, and can obtain a good comprehensive evaluation effect. In order to evaluate the practicability of the cohesive hierarchical clustering method proposed in this paper and verify the practicability of this method, the average acceleration frequency response function of 29 sensors is calculated as shown in Fig. 11.

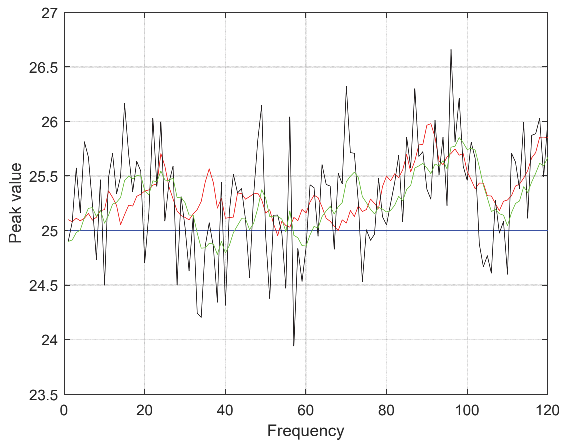

Figure 11 Frequency response function of acceleration

6 CONCLUSION

In this paper, an optimization arrangement method of civil sensors based on aggregation hierarchical clustering is proposed. Each degree of freedom is automatically classified, and the measurement point with the highest modal resolution is selected from each degree of freedom by an effective independent method as the actual measurement position. The cohesive hierarchical clustering algorithm is established, and the optimal arrangement of multi-objective single-type sensors and multi-objective two-type sensors is carried out respectively for the planar beam structure. The numerical examples show that the multi-objective cohesive hierarchical clustering algorithm can be well applied to the sensor optimization layout of the three objective functions of layout index value, MAC minimum value and cost. At the same time, by comparing the layout results of a single type sensor and multiple types of sensors, it is verified that the layout of multiple types of sensors at the same time can get better results than the layout of a single type of sensor. The numerical results show that the cohesive hierarchical clustering method is superior to the motion energy method and the independent method, which can better reflect the characteristics of the original structure and provide more structural vibration information. The proposed method has high iterative search efficiency and can provide a new way for the sensor optimization of spacecraft with complex structures. In this paper, only a few factors are considered to determine the initial number of sensors, and the initial number of sensors should be considered comprehensively in combination with various factors.

Acknowledgments

The research was supported by the Key scientific research project plan of Henan Higher Education Institutions (24B580003) and Kaifeng science and technology project (2401012). The authors also would like to thank the Dr. Sun Hanzheng's studio, School of Civil Engineering and Transportation Engineering, Yellow River Conservancy Technical Institute.

7 REFERENCES

[1] Zhou, S., Xu, Z., & Liu, F. (2017). Method for Determining the Optimal Number of Clusters Based on Agglomerative Hierarchical Clustering. *IEEE Transactions on Neural Networks and Learning Systems, 28*(99), 3007-3017. https://doi.org/10.1109/TNNLS.2016.2608001

[2] Al-Dabooni, S. & Wunsch, D. (2018). Model Order Reduction Based on Agglomerative Hierarchical Clustering. *IEEE Transactions on Neural Networks and Learning Systems, 9*(6), 31-45. https://doi.org/10.1109/TNNLS.2018.2873196

[3] Li, W., Sun, W., & Li, B. (2023). Flexibility transformation strategy of thermal power units for typical scenario considering wind and solar consumption based on generalized Benders decomposition algorithm. *Electrical Engineering, 105*(6), 3423-3432. https://doi.org/10.1007/s00202-023-01890-8

[4] Wang, X., Yang, Y., & Chen, M. (2020). AGNES-SMOTE: An Oversampling Algorithm Based on Hierarchical Clustering and Improved SMOTE. *Scientific Programming, 2020*(2), 61-69. https://doi.org/10.1155/2020/8837357

[5] Su, B. & Xie, N. (2018). Research on safety evaluation of civil aircraft based on the grey clustering model. *Grey systems: theory and application, 8*(1), 110-120. https://doi.org/10.1108/GS-10-2017-0034

[6] Wei, Z., Liang, C., & Tang, H. (2021). Research on Vehicle Scheduling of Cross-Regional Collection Using Hierarchical Agglomerative Clustering and Algorithm Optimization. *Journal of Physics: Conference Series, 2083*(4), 42022-42034. https://doi.org/10.1088/1742-6596/2083/4/042022

[7] Liu, H., Fen, L., & Jian, J. (2018). Overlapping Community Discovery Algorithm Based on Hierarchical Agglomerative Clustering. *International Journal of Pattern Recognition and Artificial Intelligence, 32*(03), 1850008. https://doi.org/10.1142/S0218001418500088

[8] Xiujie, Z., Xiaoxia, G., & Hu, Z. (2018). Clustering-based multi-objective estimation of distribution algorithm and its application. *Systems Engineering and Electronics, 7*(3), 28-47.

[9] Yarahmadi, A M. (2022). Modelling the Energy Consumption of Driving Styles Based on Clustering of GPS Information. *Modelling, 3*(3), 385-399. https://doi.org/10.3390/modelling3030025

[10] Li, H. & Tong, J. (2019). A novel clustering algorithm for time-series data based on precise correlation coefficient matching in the IoT. *Mathematical Biosciences and Engineering, 16*(6), 6654-6671. https://doi.org/10.3934/mbe.2019331

[11] Liu, S., Zou, G., & Zhang, S. (2020). A clustering-based method of typical architectural case mining for architectural innovation.*Journal of Asian Architecture and Building Engineering, 2020*(6), 121-139.

[12] Prayoga, N. R., Fahrudin, T. M., & Kamisutara, M. (2020). Unsupervised Twitter Sentiment Analysis on The Revision of Indonesian Code Law and the Anti-Corruption Law using Combination Method of Opinion Word and Agglomerative Hierarchical Clustering. I*nternational journal of engineering and technology, 8*, 200-220. https://doi.org/10.24003/emitter.v8i1.477

[13] Guo, H. & Hong, H. (2019). Research on Filtering Algorithm of MEMS Gyroscope Based on Information Fusion. *Sensors*, *19*(16), 3552-3565. https://doi.org/10.3390/s19163552

[14] Zhang, Y., Yaqiong, L. V., & Lei, T. U. (2019). Intelligent Logistics Supplier Selection Based On Improved Agglomerative Hierarchical Clustering Algorithm. *IEEE, 2019*, 690-712. https://doi.org/10.1109/INDIN41052.2019.8972146

[15] Mao, Y., Mwakapesa, D. S., & Wang, G. (2021). Landslide Susceptibility Modelling based on AHC-OLID Clustering Algorithm. *Advances in Space Research*, *2021*(2), 184-197.

[16] Tso, W. W., Demirhan, C. D., & Heuberger, C. F. (2020). A hierarchical clustering decomposition algorithm for optimizing renewable power systems with storage. *Applied Energy, 270*, 270-285. https://doi.org/10.1016/j.apenergy.2020.115190

[17] Zhao, Q., Lin, Q., & Chen, S. (2021). Research on Signal Coverge of Localizer Based on Deygout Algorithm. *2021 IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT), 2021*, 118-122. https://doi.org/10.1109/ICCASIT53235.2021.9633466

[18] Wang, B. X., Zhao, W. S., & Wang, D. W. (2021). Optimal Design of Planar Microwave Microfluidic Sensors Based on Deep Reinforcement Learning. *IEEE sensors journal, 21*(24), 27441-27449. https://doi.org/10.1109/JSEN.2021.3124294

[19] Lakshmanan, P. & Venugopal, G. (2022). Design of a Dynamic Demand Response Model Through Intelligent Clustering Algorithm Based on Load Forecasting in Smart Grid. *Elektronika ir Elektrotechnika, 28*(3), 28-43. https://doi.org/10.5755/j02.eie.30596

[20] Garikapati, P., Balamurugan, K., & Latchoumi, T. P. (2021). A Cluster-Profile Comparative Study on Machining AlSi7/63% of SiC Hybrid Composite Using Agglomerative Hierarchical Clustering and K-Means. *Silicon, 13*, 961-972. https://doi.org/10.1007/s12633-020-00447-9

[21] Kamis, N. H., Chiclana, F., & Levesley, J. (2018). Geo-uninorm Consistency Control Module for Preference Similarity Network Hierarchical Clustering Based

Consensus Model. *Knowledge-Based Systems, 162*, 103-114. https://doi.org/10.1016/j.knosys.2018.05.039
[22] Khalid, A., Hameed, A., & Tahir, M. F. (2022). Estimation of genetic divergence in wheat genotypes based on agro-morphological traits through agglomerative hierarchical clustering and principal component analysis. *Cereal Research Communications, 51*(1), 217-224. https://doi.org/10.1007/s42976-022-00287-w
[23] Deng, X., Tan, Z., & Tan, M. (2021). A clustering-based climatic zoning method for office buildings in China. *Journal of Building Engineering, 42*, 102778. https://doi.org/10.1016/j.jobe.2021.102778
[24] Liu, N., Xu, Z., & Zeng, X. J. (2021). An Agglomerative Hierarchical Clustering Algorithm for Linear Ordinal Rankings.*Information Sciences, 557*, 170-193. https://doi.org/10.1016/j.ins.2020.12.056
[25] Tsumoto, S., Kimura, T., & Hirano, S. (2021). Comparison of Clustering Strategies using a Partition Tree. *2021 IEEE International Conference on Imaging Systems and Techniques (IST), 2021*, 1-6. https://doi.org/10.1109/IST50367.2021.9651356
[26] Guo, G., Yu, K., & Wang, H. (2020). Quantum Hierarchical Agglomerative Clustering Based on One Dimension Discrete Quantum Walk with Single-Point Phase Defects. *Computers, materials, and continuum, 65*(2), 1397-1409. https://doi.org/10.32604/cmc.2020.011399
[27] Bodrunova, S. S., Orekhov, A. V., & Blekanov, I. S. (2020). Topic Detection Based on Sentence Embeddings and Agglomerative Clustering with Markov Moment. *Future Internet, 12*(9), 144. https://doi.org/10.3390/fi12090144
[28] Jo, T. (2020). Semantic string operation for specializing AHC algorithm for text clustering. *Annals of Mathematics and Artificial Intelligence, 88*, 1083-1100. https://doi.org/10.1007/s10472-019-09687-x
[29] Oey, E., Marpaung, A. B., & Sofyan, M. I. (2019). Analysing salesmen itinerary with agglomerative hierarchical clustering and vehicle routing algorithm - a case study of a confectionery supplier in Indonesia. *International Journal of Industrial and Systems Engineering, 31*(3), 178-192. https://doi.org/10.1504/IJISE.2019.10019920
[30] Zhang, Q., Fang, S., & Ye, H. (2024). Research on Clustering of Material Inspection Data for Highway Construction Projects Based on Agglomerative Hierarchical Clustering. *IOP Publishing Ltd., 5*(14), 12001-12014.
[31] Li, T., Rezaeipanah, A., & Din, E. (2022). An ensemble agglomerative hierarchical clustering algorithm based on clusters clustering technique and the novel similarity measurement. *J. King Saud Univ. Comput. Inf. Sci., 34*(6), 3828-3842. https://doi.org/10.1016/j.jksuci.2022.04.010

Contact information:

Xingyan CHENG
(Corresponding author)
School of Civil Engineering and Transportation Engineering,
Yellow River Conservancy Technical Institute, Kaifeng, 475004, China
E-mail: cxykf999@163.com

Linyan ZHU
Shenzhen ShenshuiZhaoye Engineering Consulting Co., Ltd,
Shenzhen 518000, China

Yimei CHENG
School of Water Conservancy,
North China University of Water Resources and Electric Power,
Zhengzhou, 450000, China