Hindawi
Mobile Information Systems
Volume 2022, Article ID 7580468, 8 pages
https://doi.org/10.1155/2022/7580468

Research Article

Sample Density Clustering Method Considering Unbalanced Data Distribution

Changhui Wang

Department of Fundamental Courses, Chengdu Textile College, Chengdu, Sichuan 611731, China

Correspondence should be addressed to Changhui Wang; 2021221020127@stu.scu.edu.cn

Received 30 June 2022; Revised 23 August 2022; Accepted 6 September 2022; Published 15 September 2022

Academic Editor: Santosh Tirunagari

Copyright © 2022 Changhui Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The data distribution of the multidimensional array sensor is unbalanced in data sample collection. To improve the clustering ability of data samples, a data density clustering method of sparse scattered points and multisensor array sensor samples based on the analysis of unbalanced data distribution characteristics is proposed. The sparse scattered multisensor array network's sample data collection structure is created using the Voronoi polygon topology. By analyzing the unbalanced parameters between data classes and reconstructing the characteristic space of data sample sequence, the time series of sample data collected by sparse scattered multisensor array is reorganized, and the statistical characteristic quantity and high-order cumulant of sample data collected by sparsely scattered multisensor array are extracted. Combined with the learning algorithm of unbalanced data distribution sample feature fusion, the fuzzy clustering of sample data information flow collected by sparse scattered multisensor array elements is realized. According to the feature clustering and convergence analysis, the sparse scattered feature detection method is adopted to realize the data density clustering and data structure optimization configuration of sparse scattered multisensor array elements. The test results show that the method in this paper has good convergence, strong spectrum expansion ability, and low error rate of data clustering when collecting samples with sparse scattered points and multisensor arrays.

1. Introduction

The stability of sample data collection and transmission in a multisensor array network is proposed. The accuracy of signal transmission is the key to ensuring the stable operation of multisensor data sampling. The load of sample data collecting and transmission in multi-sensor array networks is too high, and the balance of data clustering and scheduling is constrained and limited due to the load and interference of the transmission channel in sensor networks [1]. Consequently, it is essential to create a powerful clustering model using multisensor array sample data density and sparse scattered points. Based on the resource optimization allocation model of multisensor array network, combined with the big data analysis and statistical characteristics analysis of sample data collected by sparse scattered multisensor array, the self-adaptive distribution and reliability scheduling of sample data collected by sparse scattered multisensor array are carried out [2]. It is of great significance to study the density clustering method of sample data collected by sparse scattered points and multisensor array elements to improve the stability of data detection and transmission.

The data density clustering of sparse scattered multisensor array elements is based on the analysis and reconstruction of unbalanced parameters of data distribution. The key components of the data reconstruction technique are feature selection and resampling methods, including undersampling, oversampling, and mixed sampling methods that combine undersampling and oversampling. Oversampling method balances data distribution by increasing a few kinds of samples in unbalanced data, while undersampling balances data distribution by reducing most kinds of samples in unbalanced data. In large-scale data, the undersampling method can significantly reduce the number of training data and thus improve the training speed. The Random Under Sampling (RUS) method is one of the simplest undersampling methods, as the information carried by the majority of samples is obviously ignored by the RUS technique, and it is extremely possible that some valuable samples will be eliminated during the subsequent classification procedure.

The improvement of classification idea is also an important strategy to improve the performance of unbalanced data classification. Representative methods include ensemble learning L and cost-sensitive learning [3]. In the existing research, the classification effect of the method combining resampling with the classification idea improvement strategy may be better than that of a single strategy. In reference [4], combining random undersampling and ensemble learning methods, EasyEnsemble and BalanceCascade methods are proposed. EasyEnsemble method randomly selects a number of independent subsets equal to the number of minority samples from the majority of samples, and combines them with minority samples to form a new training set, which independently trains a number of Adaptive Boosting (AdaBoost) classifiers and finally outputs the integrated classifier. The difference between BalanceCascade and EasyEnsemble is that each iteration uses the classification threshold to remove the correctly classified samples from the training set. Compared with RUS, the above two methods reduce the information loss of most kinds of samples but significantly increase the training time. Some researchers proposed RandomUnderSampling Boosting (RUSBoost) method which is based on AdaBoost. In this method, most classes were randomly undersampled in the iterative process, and a few classes formed a temporary training set. Then the weak classifier was trained by using the temporary training set and weights. RUS still uses random undersampling to balance the data distribution [5]. When the imbalance ratio is very high, it may take a lot of iterations. The configuration of sample data collected by multisensor array elements in sparse scattered points is based on the analysis of resourced parameter characteristics and clustering fusion processing. The channel transmission model of a multisensor array network is built by optimizing the distribution of virtual channels and transmission data packets. The channel equalization control is used to achieve the density clustering of sample data collected by multisensor array elements in sparse scattered points. In the traditional methods, the data density clustering methods of sparse scattered multisensor array elements mainly include the data density clustering method of multisensor array network based on neural network, the data density clustering method of sparse scattered multisensor array elements based on PCA, and the dynamic distribution method of data of sparse scattered multisensor array elements based on K-means fusion clustering [4, 6]. In reference [7], a SAC reinforcement learning-based sample data distribution method for sparse scattered point multisensor array is proposed, and a link spectrum data clustering model is established which makes the V2V link to optimize the sample data distribution for sparse scattered point multisensor array after continuous learning. However, the dynamic learning ability of this method for clustering sample data for sparse scattered point multisensor array is not good.

In order to solve the above-given problems, this paper proposes a data density clustering method based on the analysis of unbalanced data distribution sample characteristics. First, the sample data collecting topology of the sparse scattered multisensor array network is established using the Voronoi polygon topology. The sparse scattered feature detection approach is then used in accordance with the feature clustering and convergence analysis. It is done to optimize the data structure and density clustering of the sample data obtained from the sparse scattered multisensor array. Finally, the simulation test analysis shows that this method has superior performance in improving the clustering ability of data density of samples collected by multisensor array elements with sparse scattered points.

2. Topology Model and Data Structure Analysis of Multisensor Network

2.1. Multisensor Element Network Topology Model. Firstly, the model of the topological structure of the multisensor array network is created, and the Voronoi polygon topology is used to produce the topological structure of the sample data collection of the sparsely distributed multisensor array network. It is assumed that the multisensor array network consists of N main networks, and $U = \{u_1, u_2, \ldots, u_n\}$ is used to represent the communication data clustering set of secondary users to be connected, and $C = \{a_1, a_2, \ldots, a_m\}$ represents the channel structure parameters of sample data collection of sparse scattered multisensor array. The network topology of multisensor elements is established by gathering the spectrum parameters of samples collected by sparse scattered points and multisensor elements. These sparse scattered points and multisensor elements are combined with spectrum bandwidth analysis, also the method of dynamic judgment and channel fusibility is adopted to optimize the clustering of samples collected by sparse scattered points and multisensor elements. When the clustering center parameter curve of samples collected by sparse scattered points and multisensor elements is 0/1, the output spectrum bandwidth of samples collected by sparse scattered points and multisensor elements meets the clustering convergence in N-dimensional space, $0 \leq w_{kj} \leq 1$. According to the service, it should be switched to the reserved spectrum characteristic quantity under the condition that the communication in the sparse scattered point multisensor array acquisition system meets the service quality requirements. The maximum membership $\sum_{j=1\ldots D} w_{kj} = 1 (1 \leq k \leq K)$ of the sparse scattered point multisensor array acquisition, sample data are obtained by taking the blocking rate as the constraint parameter. The distribution link model of the sparse scattered point multisensor array acquisition sample data are established, as shown in Figure 1.

According to the distribution chain of sample data collected by sparse scattered multisensor array elements shown in Figure 1, let $Cen = [Cen_k]_K$ represent the statistical characteristic quantity of the number of primary users and the number of secondary users in the multi-sensor array element network. In the process of data clustering, incremental scheduling is adopted to obtain the segmented set at the H layer where the requested link sends the synchronization request time series:

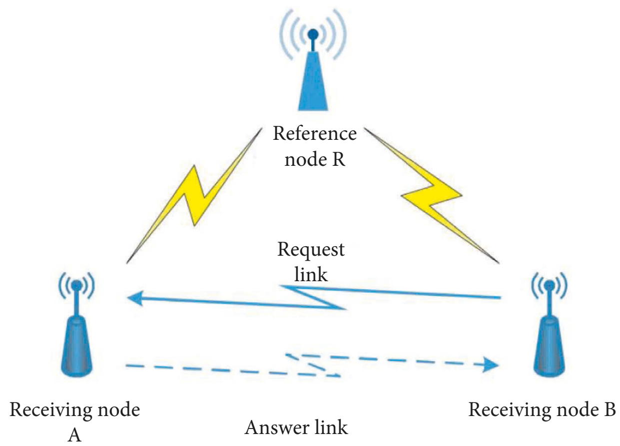

Figure 1: Schematic diagram of sample data distribution link for multisensor array acquisition.

$$\mathrm{Cen}_k = \{\mathrm{Cen}_{k1}, \mathrm{Cen}_{k2}, \cdots, \mathrm{Cen}_{kD}\}. \quad (1)$$

The beacon code is used to record the beacon in each period, and the hierarchical center of kth dispersion of the master node is obtained. Let $U = [u_{ki}]_{K\times N}$ represent the membership matrix of sample data collected by sparse scattered multisensor array elements, where u_{ki} represents the time stamp of sample data collected by sparse scattered multisensor array elements, and $x_i\,(1 \le i \le N)$ belongs to the beacon code of kth. $0 \le u_{ki} \le 1$ at the end of a complete synchronization period. It obtains the offset time series $Q = \{q_1, q_2, \ldots, q_i, \ldots, q_m\}$ and $P = \{p_1, p_2, \ldots, p_j, \ldots, p_n\}$ of sample data collected by sparse scattered multisensor array elements through the forwarding response link, $W = [w_{kj}]_{K\times D}$ is used to represent the joint distribution subspace weight matrix of the sample data collected by sparse scattered multisensor array elements. Therefore, the time series reorganization of the sample data collected by sparse scattered multisensor array elements is realized by analyzing the unbalanced parameters between data classes and reconstructing the characteristic space of the data sample sequence.

2.2. Analysis of Data Structure of Sensor Array Sampling. In the cluster of sample data collected by multisensor elements in sparse scattered points, the interactive parameter synchronization package is constructed first. $S = (U, A, V, f)$ is the statistical distribution set of reference node identification (ID), where U represents the period number, which is collectively referred to as the domain of sample data distribution of sparse scattered points and multisensor elements. A means sending timestamp (ST). The model parameters of a sample data configuration of sparse scattered multisensor array elements are obtained as$V = \cup_{a\in A} V_a$, where V_a is reference node ID and reference node period number information, and the output loss time distribution set is $a \in A$. In the range of $f\colon U \times A \longrightarrow V$, the frequency drift and phase offset of sample data collected by sparse scattered multisensor array elements are adaptively estimated. The data clustering

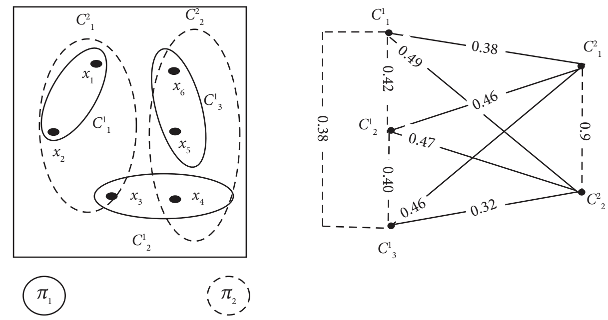

Figure 2: Analysis of data structure model.

sets $\forall u \in U, \forall a \in A$, in the same period of the same reference node are obtained by the method of model parameter estimation, and the output frequency offset is $f(u, a) \in V_a$. The sample data acquisition topology of a sparse scattered multisensor array network is established using the Voronoi polygon topology, and the data structure model is examined. According to the sample data structure analysis of sparse scattered multisensor array, the data distribution imbalance control algorithm is used for optimization control [9] as shown in Figure 2.

According to the data structure distribution in Figure 2, the statistical feature quantity and high order cumulant of the sample data flow collected by sparse scattered points and multisensor array elements are extracted. In the process of resource allocation, the self-sparse structure is constantly adjusted according to the accumulated historical data, and the fuzzy clustering processing of the sample data information flow collected by sparse scattered points and multisensor array elements is realized by combining the learning algorithm of unbalanced data distribution and sample feature fusion [10].

3. Sample Density Clustering Optimization

3.1. Data Clustering Feature Extraction Considering Data Imbalance. Bowyer–Watson algorithm is used to construct the sample data configuration model of sparse scattered point multisensor array. Combining with the topological structure of sparse scattered point multisensor array sampling data clustering, in a limited universe, the obtained channel gain is described as $S = (U, A, V, f)$ in sequence, if $A = C \cup D$, $D = \{d\}$, and $C \cap D = \Phi$, the transmitting power of the transmitter PT of sparse scattered point multisensor array sampling data are C and D in sequence, assuming that the distributed metadata feature quantity of the i-th channel of network M is $S = (U, C \cup D)$. The statistical features of the sample data stream collected by sparse scattered multisensor array elements are extracted by low-frequency wireless spectrum data clustering, and the allocation decision table is obtained. When $a \in A$ is satisfied, the disturbance component of the sample data collected by sparse scattered multisensor array elements in the main link is $\mathrm{IND}(A - \{a\}) = \mathrm{IND}(A)$, the sum of SU disturbances of any network is added, and the frequency division multiplexing method is adopted [11]. Get the linear eigenvalues of the clustering characteristic distribution attribute a of the sparse scattered point multisensor array acquisition sample data in the closed interval A. The clustering hierarchical function of the sparse scattered point multisensor array acquisition sample data in the backhaul base station is $S = (U, A)$. For$P \subseteq A$, there is $\mathrm{IND}(P) = \mathrm{IND}(A)$. Distribute the superimposed data of the access subframe to the users of the base station and get the statistical characteristic quantity of the sparse scattered point multisensor array acquisition sample data fusion as follows:

$$\min \sum_{p_j \in N_k(p_i)} \left(\left(p_j - \overline{p}\right) . t_{\mathrm{comm}}(X) n_i - g\left(u_j, v_j\right)\right)^2, \tag{2}$$

where $t_{\mathrm{comm}}(X)$ is the resource size of the three-level backhaul layer base station of the multisensor array network sample data collection and transmission channel, p_j is the frequency band information of the access subframe, n_i is the interference component of the sparse scattered point multisensor array acquisition sample data output channel, and $g(u_j, v_j)$ is the frequency component of the access sparse scattered point multisensor array acquisition sample set of subset D. Each sample with correct classification is removed as follows:

$$\pi_q = \left\{C_1^q, C_2^q, \ldots, C_K^q\right\}, \tag{3}$$

where the statistical feature of the nearest neighbor sample points is $G = \langle V, W^c \rangle$, where K is the reliability clustering parameter of searching the nearest neighbor parameters. Based on the nearest neighbor parameter method and the sequential decision mechanism, the remaining transmissible load π_{b_s} of the V link K is obtained which satisfies $b_s \neq b_t$, and then the joint feature point of the unbalanced data distribution sample is the shared neighbor of $C_{k_q}^{b_t}$, $C_{k_i}^{b_s}$ and $C_{k_i}^{b_s}$. Therefore, the fuzzy statistical characteristic quantity $S = (U, C \cup D)$ of agent K of sparse scattered point multisensor array element collection sample is obtained. For $U = \{u_1, u_2, \ldots, u_n\}$ as the universe, the density peak clustering algorithm is adopted, and the variation attribute of sparse scattered point multisensor array element collection sample is $C = \{a_1, a_2, \ldots, a_m\}$. Under the multiobjective evolutionary constraint, the fuzzy decision attribute of sparse scattered point multisensor array element collection sample is obtained and the control instruction set of data clustering satisfies $C \cap D = \Phi$. Combining with the data distribution imbalance sample feature fusion number, the neural network is trained back to obtain the best data clustering. The information entropy of the sample data collected by sparse scattered multisensor array elements are $\xi_{c_1}^{d_2} = 3/5$, $\xi_{c_2}^{d_2} = 2/5$,$\xi_{c_3}^{d_2} = 2/5$, $\max g_{c_1}(d_2) = 6/5$, $\max g_{c_2}(d_2) = 3/8$, and $\max g_{c_3}(d_2) = 1/10$. Therefore, the method based on multiobjective evolution is adopted to obtain the clustering feature extraction results of sample data collected by sparse scattered points and multisensor array elements. The adaptive configuration of sample data collected by sparse scattered points and multisensor array elements is realized [12].

3.2. Adaptive Adjustment and Output of Data Clustering Center. Feature clustering and convergence analysis suggest that the approach of sparse scattered feature detection be used to modify the dynamic resource allocation [13]. By analyzing the unbalanced parameters between data classes and reconstructing the feature space of data sample sequence, the time series of sample data collected by sparse scattered multisensor array elements is reorganized, and the statistical feature quantity and high-order cumulant of sample data stream collected by sparse scattered multisensor array elements are extracted. The distribution set of sparse points is taken as SDF. The output sample data collected by sparse scattered multisensor array elements is fixed for channel switching mode, and through idle communication spectrum conversion, the optimized model parameter of output sample data are collected by sparse scattered multisensor array elements area $\xi_{c_1}^{d_3} = 1$, $\xi_{c_2}^{d_3} = 1$,$\xi_{c_3}^{d_3} = 1$, $\max g_{c_1}(d_3) = 7/4$, $\max g_{c_2}(d_2) = 3/8$, $\max g_{c_3}(d_2) = 7/4$. Combined with the evaluation results of sample data return efficiency collected by sparse scattered multisensor array elements, the dynamic spectrum allocation strategy is adopted. Considering that the correlation between sample points is not only unbalanced with the data distribution of neighboring points, the frequency division multiplexing mechanism is introduced, and the calculation formula of the cross matrix of neighboring parameter clustering is obtained as follows:

$$\begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ \cdots \\ d_N \end{bmatrix} = \begin{bmatrix} d_{11} & d_{12} & d_{13} & \cdots & d_{1N} \\ d_{21} & d_{22} & d_{23} & \cdots & d_{2N} \\ d_{31} & d_{32} & d_{33} & \cdots & d_{NN} \\ \cdots & & & & \\ d_{N1} & d_{N2} & d_{N3} & \cdots & d_{NN} \end{bmatrix} * \begin{bmatrix} w_{i1} \\ w_{i2} \\ w_{i3} \\ \cdots \\ w_{iN} \end{bmatrix}, \tag{4}$$

where $d_{ij}$is the Euclidean distance of each sample point, and w_{ij} is the information entropy of the sample data collected by the i-th sampling node. Considering that there may be outliers in the data set, these outliers will affect the selection of parameter K. The balanced control parameters of sparse scattered points and multisensor array elements are obtained by using the unbalanced scheduling between classes. At the node I, the characteristic sequence of data collection and clustering is denoted as $(w_{1,j}, w_{2,j}, \ldots, w_{tj})$, where t is the statistical characteristic quantity of detection time. The neighborhood of each sample point is obtained by the parameter K search algorithm, which is denoted as$\Pi = \{\pi_1, \pi_2, \ldots, \pi_B\}$, and a resource optimization allocation model is constructed. The sparse scattered feature detection method is used to realize the data density clustering and data structure optimization configuration of the sample data collected by sparse dispersed points and multisensor array

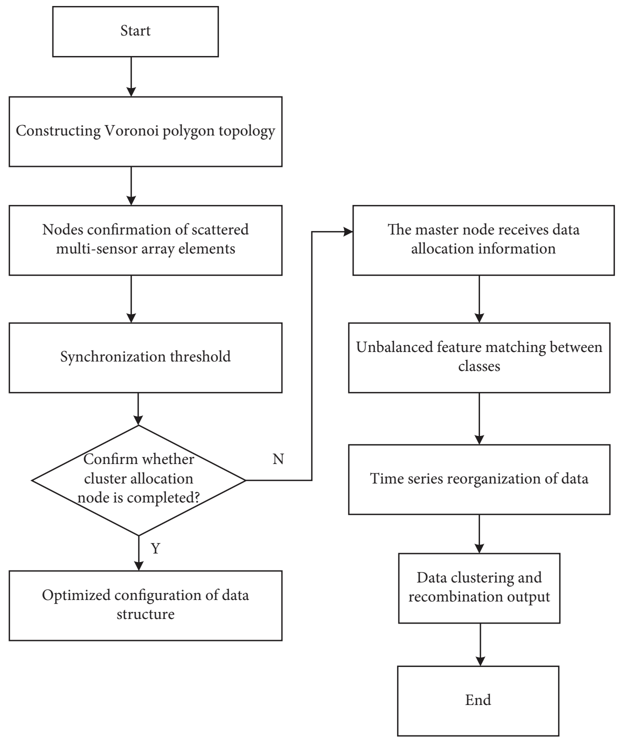

Figure 3: Implementation process of algorithm.

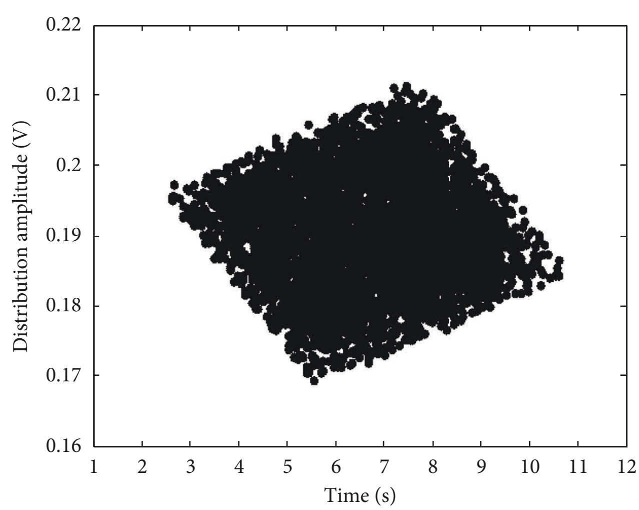

Figure 4: Original time series of sample data collected.

elements based on the clustering and convergence analysis of those samples. The optimized clustering function is as follows:

$$J_{ESSC}(U,V,W,X) = \sum_{k=1}^{K}\sum_{i=1}^{N} u_{ki}^{m} \sum_{j=1}^{D} w_{kj}\left(Cen_{kj} - x_{ij}\right)^{2} + \gamma \sum_{k=1}^{K}\sum_{j=1}^{D} w_{kj} \log w_{kj-\eta} \sum_{k=1}^{K}\left(\sum_{i=1}^{N} u_{ki}^{m}\right)\sum_{j=1}^{D} w_{kj}\left(Cen_{kj} - \overline{Cen}_{j}\right),$$
$$\text{s.t. } 0 \le u_{ki} \le 1, \sum_{k=1}^{K} u_{ki} = 1, 1 \le i \le N; 0 \le w_{kj} \le 1, \sum_{j=1}^{D} w_{kj} = 1, 1 \le k \le K, \tag{5}$$

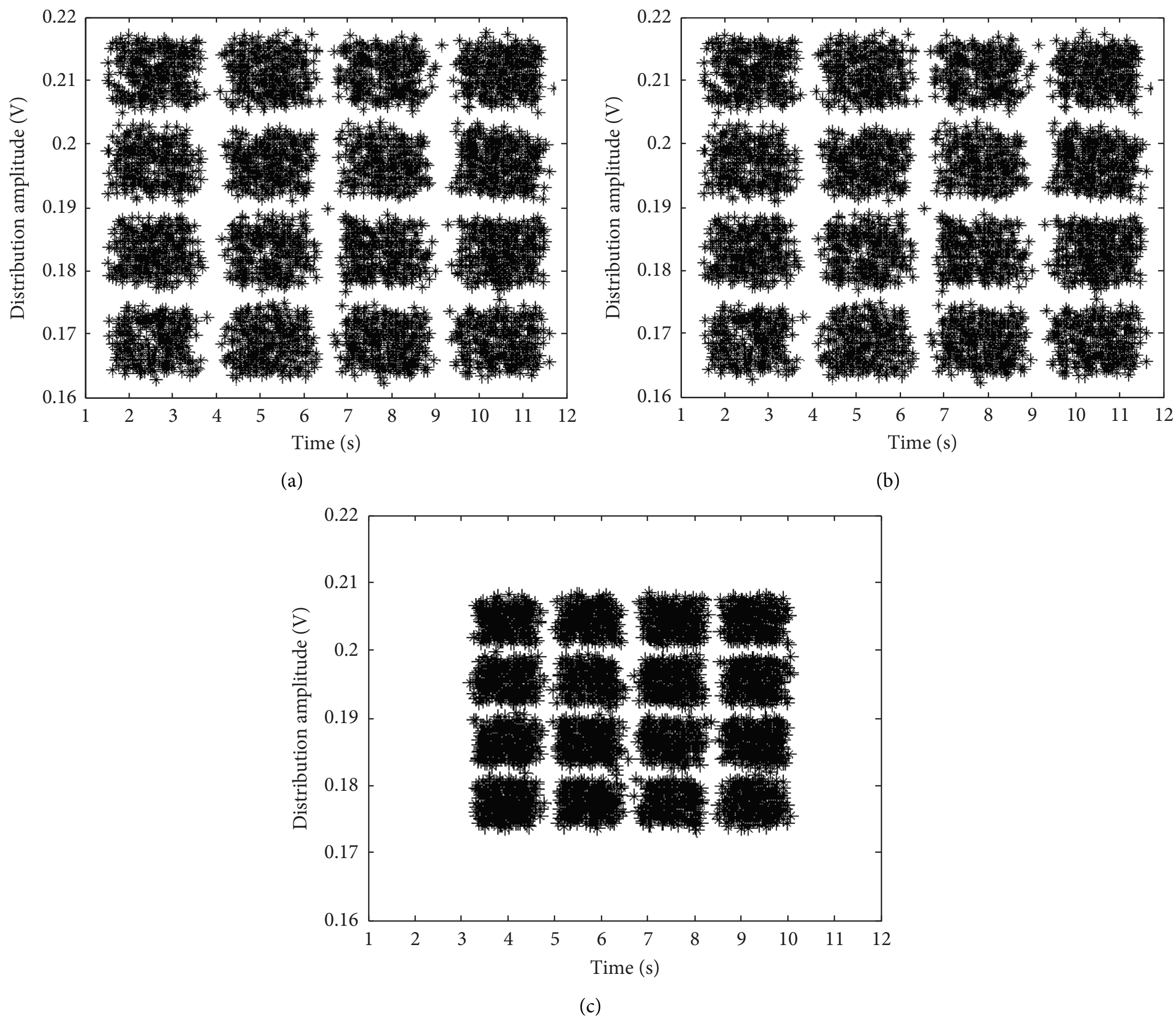

Figure 5: Output result of data clustering. (a) K-means method. (b) FCM method. (c) Methods of this paper.

where $\overline{\text{Cen}}$ is the average value of adjacent order of cluster center of sample data collected by multisensor array elements of each sparse scattered point; E is sample data which collected for sparse scattered multisensor array elements; γ represents the boundary area of most kinds of sample clusters; $\sum_{i=1}^{N} u_{ki}^{m}$ represents the dispersion degree of regional markers; w_{kj} represents the wrong area; N represents the size of resource allocation; K represents the fuzzy weighting coefficient of sample data clusters collected by sparse scattered multisensor array elements. To sum up, according to feature clustering and convergence analysis, sparse scattered feature detection method is adopted [14]. The process of data density clustering and data structure optimization of sparse scattered point multisensor array acquisition samples is realized. The process of data density clustering of sparse scattered point multisensor array acquisition samples is improved as shown in Figure 3.

4. Simulation and Result Analysis

Matlab simulation experiment is used to verify the application performance of this method in data density clustering of sparse scattered multisensor array [15] The parameter are set accordingly. The multisensor array network's K channel number is set to 200, its bandwidth to 40 kHz, its iteration count to 500, and its network benefit to 0.56. According to the above-given parameter settings, the data density clustering simulation of sparse scattered multisensor array is carried out, and the original time series of sparse scattered multisensor array data are given as shown in Figure 4.

Taking the time series of sample data collected by sparse scattered points and multisensor array elements in Figure 4 as the research object, this method is used to realize the density clustering of sample data collected by sparse scattered

Table 1: Error comparison of data clustering.

Iterations	K-means	FCM	This method
10	0.284	0.441	0.199
20	0.277	0.474	0.197
30	0.285	0.416	0.194
40	0.279	0.480	0.183
50	0.258	0.462	0.181
60	0.285	0.473	0.167
70	0.289	0.440	0.116
80	0.274	0.435	0.109
90	0.283	0.412	0.102
100	0.299	0.448	0.096

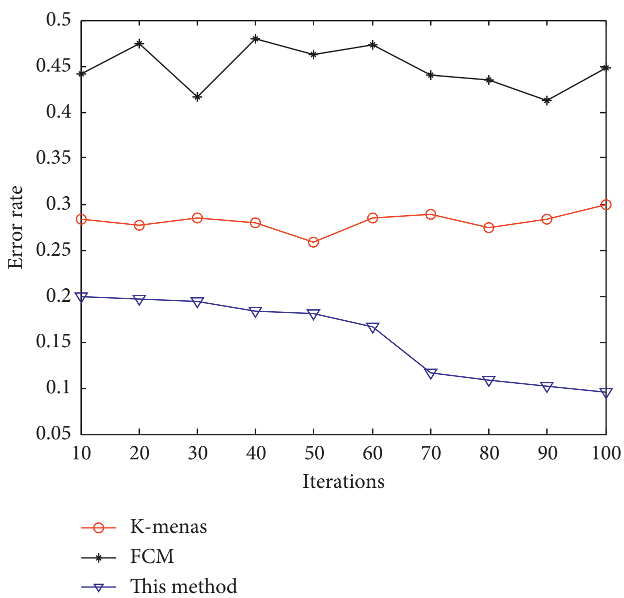

Figure 6: Comparison of BER of data clustering.

Table 2: Experimental data set.

Dataset name	Number of attributes	Number of samples
Brest cancer	9	285
Iris	4	150
Adult	15	356

points and multisensor array elements, and the data clustering results of each frequency band are shown in Figure 5.

It is clear from the analysis of Figure 5 that the distribution convergence of various approaches is evaluated when sparse scattered multisensor array samples are data density clustered by using this method, which has great spectrum adaptability and good interclass equilibrium classification. Taking the bit error rate of data clustering as the test index, the comparison results are shown in Table 1 and Figure 6. From the analysis of Figure 6, it can be seen that the bit error rate by a distance of sparse scattered multisensor array samples in this method is low and the convergence is good.

In order to test the clustering effect of the algorithm in this paper, brest cancer, iris, and adult in the UCI database are selected as the test data sets. The experimental data sets

Table 3: Comparison of different clustering algorithms.

Data set	Clustering algorithm	Convergence time/s
Brest cancer	K-means algorithm	40
	FCM algorithm	35
	Algorithm in this paper	23
Iris	K-means algorithm	70
	FCM algorithm	68
	Algorithm in this paper	45
Adult	K-means algorithm	123
	FCM algorithm	114
	Algorithm in this paper	66

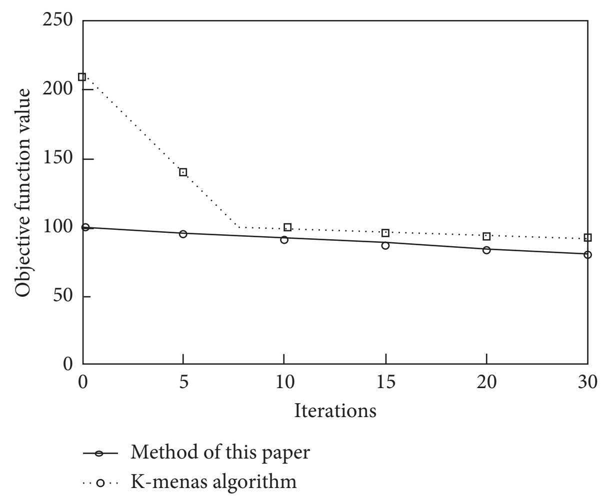

Figure 7: Comparison of the stability of the two algorithms on iris data set.

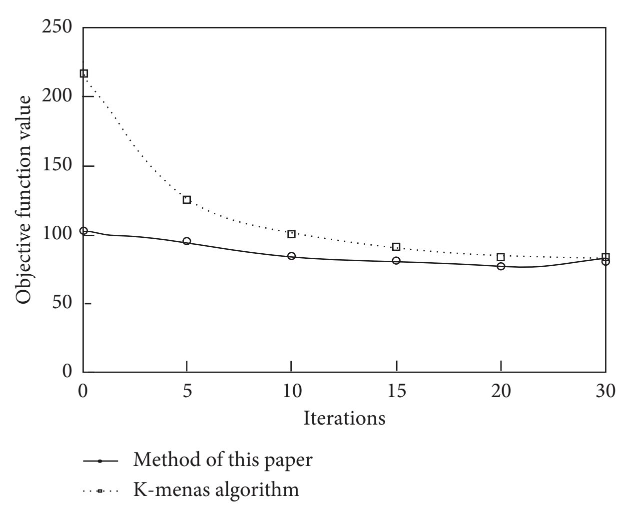

Figure 8: Comparison of the stability of the two algorithms on adult data set.

are shown in Table 2. In this experiment, the parameters are set as follows: the fuzzy index is 2, the maximum number of iterations is 50, and the initial value of iterations is 1.

The breast cancer, iris, and adult datasets in Table 2 were tested for 30 times using the k-means algorithm, FCM algorithm, and the algorithm in this paper, respectively, and the mean value of convergence time obtained from each experiment was statistically compared. The comparison data are shown in Table 3.

By analyzing the data in Table 3, the k-means algorithm and FCM algorithm have a long convergence time during the experiment, while the convergence time of this algorithm is significantly reduced. Therefore, this algorithm shows a good clustering effect in the experiment. The k-means algorithm is compared with the clustering algorithm in this paper in terms of convergence speed. Iris and adult data sets in the UCI database are used during the experiment. The algorithm in this paper is executed by changing input parameters and 50 experiments are conducted. The results are shown in Figures 7 and 8.

Figures 7 and 8 show that, for two different types of data sets, when using this algorithm and the k-means algorithm to test the stability of the algorithm, the convergence speed of this algorithm is significantly faster than that of the k-means algorithm although the objective function values of this algorithm and k-means algorithm are very close at the end and the number of iterations is less with good stability. Therefore, this algorithm is better than the k-means algorithm in convergence and stability.

5. Conclusions

In this paper, an effective clustering model of sample data density collected by sparse scattered points and multisensor elements is established. The self-adaptive distribution and reliability scheduling of sample data collected by sparse scattered points and multisensor elements are carried out by optimizing the resource allocation model of the multisensor elements network. A data density clustering method based on the characteristic analysis of unbalanced data distribution samples with sparse scattered points and multisensor array elements is proposed. The convergence control of resource allocation is carried out by using the data distribution imbalance sample feature fusion score and the resource optimal allocation model is constructed. According to the feature clustering and convergence analysis, the sparse scattered feature detection method is adopted to realize the data density clustering and data structure optimal allocation of samples collected by sparse scattered multisensor array elements. The simulation results show that this method has good interclass equilibrium classification and high convergence accuracy improving the optimal allocation and scheduling ability of sparse scattered sample data of multisensor array network.

Data Availability

The data used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] Liu, J. Wang, and Y. Zhengguo, "Imbalanced data classification algorithm based on ball cluster partitioning and undersampling with density peak optimization," *Journal of Computer Applications*, vol. 42, no. 5, pp. 1455–1463, 2022.

[2] X. Zheng, C. Ren, Y. Yang, Z. Gong, X. Chen, and Z. Hao, "QuickDSC: clustering by quick density subgraph estimation," *Information Sciences*, vol. 581, pp. 403–427, 2021.

[3] S. Jian, D. Li, and Y. Yu, "Research on taxi operation characteristics by improved DBSCAN density clustering algorithm and K-means clustering algorithm," *Journal of Physics: Conference Series*, vol. 1952, no. 4, Article ID 042103, 2021.

[4] J. Yang, Y. Zhang, H. Jiang, and X. Zhu, "Detection method of physical-layer impersonation attack based on deep Q-network in edge computing," *Journal of Computer Applications*, vol. 40, no. 11, pp. 3229–3235, 2020.

[5] C. Rosa, D. Helmut, A. H. Pablo et al., "Probing the existence of a rich galaxy overdensity at z= 5.2," *Monthly Notices of the Royal Astronomical Society*, vol. 502, no. 3, pp. 4558–4575, 2021.

[6] Q. Liu, J. W. Zhai, Z. Z. Zhang et al., "A survey on deep reinforcement learning," *Chinese Journal of Computers*, vol. 41, no. 1, pp. 1–27, 2018.

[7] Z. Xue and H. Wang, "Effective density-based clustering algorithms for incomplete data," *Big Data Mining and Analytics*, vol. 4, no. 3, pp. 183–194, 2021.

[8] S. Hashida, Tamura K, and T. Sakai, "Classifying tweets using convolutional neural networks with multi-channel distributed representation," *IAENG International Journal of Computer Science*, vol. 46, no. 1, pp. 68–75, 2019.

[9] Y. Ling and X. Zhang, "An Improved K-Means Algorithm Based on Multiple Clustering and Density," in *Proceedings of the ICMLC 2021: 2021 13th International Conference on Machine Learning and Computing*, Shenzhen China, February-March 2021.

[10] N. Wu, K. Wang, L. Wan, and N. Liu, "A source number estimation algorithm based on data local density and fuzzy C-means clustering," *Wireless Communications and Mobile Computing*, vol. 2021, no. 5, pp. 1–7, Article ID 658785, 2021.

[11] A. Laith, H. A. Khaled, A. Mohamed, S. Mohammad, and A. Maryam, "Enhanced Flow Direction Arithmetic Optimization Algorithm for mathematical optimization problems with applications of data clustering," *Engineering Analysis with Boundary Elements*, vol. 138, pp. 13–29, 2022.

[12] Y. F. Yang, H. Zhou, and B. Zhou, "Data clustering based on the modified relaxation Cheeger cut model," *Computational and Applied Mathematics*, vol. 41, no. 1, pp. 55–62, 2022.

[13] A. Kaur and Y. Kumar, "Neighborhood Search Based Improved Bat Algorithm for Data clustering," *Applied Intelligence*, pp. 1–35, Springer, Berlin, Germany, 2022.

[14] P. Kaviti, "R-tree based density clustering for multispectral images," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 4, pp. 4328–4335, 2020.