

Retraction

Retracted: Audit Analysis of Abnormal Behavior of Social Security Fund Based on Adaptive Spectral Clustering Algorithm

Complexity

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

Copyright © 2024 Complexity. 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.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] Y. Wu, Y. Chen, and W. Ling, "Audit Analysis of Abnormal Behavior of Social Security Fund Based on Adaptive Spectral Clustering Algorithm," *Complexity*, vol. 2021, Article ID 4969233, 11 pages, 2021.

Research Article

Audit Analysis of Abnormal Behavior of Social Security Fund Based on Adaptive Spectral Clustering Algorithm

Yan Wu,¹ Yonghong Chen,² and Wenhao Ling³ 

¹School of Business, Xi'an University of Finance and Economics, Shaanxi, Xi'an 710100, China

²School of Public Administration, Xi'an University of Finance and Economics, Shaanxi, Xi'an 710061, China

³Institute of Local Government and Societal Governance, Henan University, Kaifeng, Henan 475004, China

Correspondence should be addressed to Wenhao Ling; lwh@henu.edu.cn

Received 7 April 2021; Revised 27 April 2021; Accepted 10 May 2021; Published 19 May 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Yan Wu et al. 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.

Abnormal behavior detection of social security funds is a method to analyze large-scale data and find abnormal behavior. Although many methods based on spectral clustering have achieved many good results in the practical application of clustering, the research on the spectral clustering algorithm is still in the early stage of development. Many existing algorithms are very sensitive to clustering parameters, especially scale parameters, and need to manually input the number of clustering. Therefore, a density-sensitive similarity measure is introduced in this paper, which is obtained by introducing new parameters to transform the Gaussian function. Under this metric, the distance between data points belonging to different classes will be effectively amplified, while the distance between data points belonging to the same class will be reduced, and finally, the distribution of data will be effectively clustered. At the same time, the idea of Eigen gap is introduced into the spectral clustering algorithm, and the verified gap sequence is constructed on the basis of Laplace matrix, so as to solve the problem of the number of initial clustering. The strong global search ability of artificial bee colony algorithm is used to make up for the shortcoming of spectral clustering algorithm that is easy to fall into local optimal. The experimental results show that the adaptive spectral clustering algorithm can better identify the initial clustering center, perform more effective clustering, and detect abnormal behavior more accurately.

1. Introduction

The development history of the social security fund market is less than 20 years, but it has developed rapidly and has become an indispensable part of the national economic life [1, 2]. Theoretical and empirical studies have shown that, in a market dominated by individual investors, investment decisions lack sufficient rationality due to limited information acquisition capabilities. From the perspective of the domestic securities market, individual investors are obviously speculative, often fast in and out for short-term combat, are susceptible to market sentiment, and often blindly follow the trend, causing disorderly volatility in my country's securities market, leading to significant market price fluctuations. Divorced from the actual situation of the operation of the national economy, it has failed to effectively guide resource allocation and industrial structure optimization, and it is

also difficult to use market forces to evaluate the benefits of listed companies [3, 4]. Therefore, the detection of abnormal behaviors of social security funds is beneficial to the development of the fund market [1, 5].

Cluster analysis is a statistical method for studying sample classification, as well as a method of data mining, which can effectively realize the exploration of the internal connections between things [6, 7]. The fundamental purpose of the clustering algorithm is to automatically classify the given samples into corresponding classes under certain standards. Manual classification of data has great limitations in practical applications [8]. Clustering belongs to the category of unsupervised classification. In the process of classification, the degree of similarity between the given classification objects is distinguished according to some characteristics of the objects themselves. Clustering algorithms have received more and more attention from

researchers and are currently widely used in many fields. With various abnormal behaviors in the fund market in the past two years, many researchers have used clustering algorithms to detect such abnormal behaviors [9, 10]. The spectral clustering algorithm is a relatively new research direction in clustering algorithms, which uses the properties of the feature vector in the similarity matrix of the data set to cluster the data set. The idea of the spectral clustering algorithm is to use the similarity between data points to cluster the data set. This method can also be applied to the cluster analysis problem in the nonmeasure space [11].

Many methods based on spectral clustering have achieved many good results in the practical application of clustering, but the study of spectral clustering algorithms is still in the early stage of development. Many existing algorithms are very important for clustering parameters, especially scale parameters. It is sensitive and requires manual input of the number of clusters. Therefore, this paper proposes an adaptive spectral clustering algorithm. The main contributions of the algorithm are as follows:

- (1) This paper introduces a density-sensitive similarity measure, which is obtained by introducing new parameters to deform the Gaussian function. Under this metric, the distance between data points belonging to different classes will be effectively enlarged, while the distance between data points belonging to the same class will be reduced, and finally the distribution of data will be effectively clustered.
- (2) The idea of the intrinsic gap is introduced into the spectral clustering algorithm, and the intrinsic gap sequence is constructed on the basis of the Laplacian matrix to solve the problem of the number of initial clusters.
- (3) The advantage of the artificial bee colony algorithm's strong global search ability is used to make up for the shortcoming of the spectral clustering algorithm that it is easy to fall into the local optimum. At the same time, in order to prevent the premature phenomenon of artificial bee colony algorithm, its location search formula is improved.

The organizational structure of this article is as follows. The second part shows the relevant research content of this article. The third part shows the research algorithm of this article. The fourth part is the experimental results. The fifth part is the conclusion.

2. Related Knowledge

At this stage, the computer network has basically realized globalization, and the huge network system facilitates the exchange and transmission of fund information. Facing the increasingly complex network environment, it is of great significance to accurately detect the abnormal behavior of fund investment [12, 13]. In this regard, experts and scholars apply clustering algorithms to detection technology, relying on unsupervised and semisupervised methods to quickly

and accurately identify abnormal and nonabnormal behaviors, which is also one of the current research hotspots [14–17].

Spectral clustering algorithms are derived from the theory of spectrograms. Based on different segmentation methods, researchers have proposed some more classic spectral clustering algorithms. PF algorithm [18], as the original prototype of spectral clustering, has been extensively studied in the field of machine learning. Wang et al. [19] proposed a canonical cut set to divide the graph and proposed the famous SM algorithm. The clustering result of the SM algorithm is also significantly higher than that of the PF algorithm. Xu et al. [20] proposed the famous SLH algorithm. After the algorithm constructs the similarity matrix according to the multipath canonical cut set criterion, the data of eigenvalues and eigenvectors is determined by the number of cluster inputs. If the number of clusters is 3, then the number of feature vectors is 3. Although its consideration is comprehensive, its high complexity affects the calculation speed and the efficiency is poor, but it still achieves a good clustering effect. Ding et al. [21] proposed the NJW algorithm. The algorithm is the same as the SLH algorithm. The number of eigenvalues is determined by the number of clustering groups after obtaining the eigenvectors. Shiotsuka et al. [22] proposed the Mcut algorithm. This algorithm deeply studies the vector and uses this vector as a criterion as a feature vector. The clustering results obtained tend to be in a balanced state, and it takes a long time for larger data sets. Zhu et al. [23] proposed the MS algorithm. Similar to the SLH algorithm and the NJW algorithm, the selection method of the feature vector is similar. Contrary to Mcut, it has a poor clustering effect when the data set is small or when the segmented image is required to be small.

At present, many scholars have carried out extensive research on spectral clustering, mainly focusing on the determination of the number of clustering groups, the construction of feature vectors, and the construction of similarity matrices.

Aiming at the adaptive problem of the number of clustering groups, literature [24] converts large data sets into small data sets, uses correlation to merge the grouped data, and after fully studying its principle, proposes a new spectral clustering algorithm; the algorithm can automatically obtain the number of clustering groups. Literature [25] calculates the difference between adjacent eigenvalues after arranging the eigenvalues by calculating the eigenvectors and eigenvalues of the data similarity matrix. If the previous feature value minus the current feature value is the largest, the number of current feature values counted is the number of cluster groups. Literature [26] first studied the distribution of data, estimated the proportional parameter of each data point based on this, and proposed the STSC algorithm, but the calculation consumes too many resources and runs slowly.

Aiming at the problem of feature vector selection and similarity matrix construction, literature [27] sets a threshold based on the feature requirements of the data and runs the NJW algorithm multiple times. The number of runs is manually set, and the adaptation of the threshold

parameter a is obtained, which eliminates the interference of manual input on the construction of the similarity matrix, and improves the clustering effect. Literature [28] uses the rough clustering method to cluster the constructed left and right singular vectors and proposes rough spectral clustering, which is successfully applied in text data mining. Literatures [29–31] proposed that the existing feature vectors do not fully reflect the characteristics of the data itself, and these features themselves are also difficult to extract. Therefore, some control information is added to the feature vector selection of the spectral clustering algorithm, and a semisupervised spectral clustering feature vector selection algorithm is proposed. Some supervision information is added to the algorithm, and the feature vector is selected on this basis.

3. Adaptive Spectral Clustering Algorithm

3.1. Spectral Clustering Algorithm. The spectral clustering algorithm is based on the division of spectral graphs, and its essence is to transform the clustering problem into the optimal segmentation problem of graphs [32–35]. The spectral clustering algorithm regards the data samples as the vertices in the graph, which is represented by the set J , the vertices are connected by edges, and the edges are represented by the set B . Assign a weight value w according to the similarity between samples, and an undirected weighted graph $T = (J, B)$ based on the similarity of the samples can be obtained. In this way, in the graph T , the clustering problem is transformed into the optimal graph segmentation problem on the graph T , so that the internal similarity of each subgraph after division is very high, and the similarity between different subgraphs is very low.

Let D_i denote the i -th data sample point, and let $\text{dis}(d_i, d_j)$ denote the distance between d_i and d_j . This distance is usually expressed by the European distance $\|d_i - d_j\|$, and σ is expressed as a scale parameter. We can get the similarity matrix:

$$S_{ij} = e^{-\left(\text{dis}(d_i, d_j) / 2\sigma\right)^2}. \quad (1)$$

The distribution of surrounding data can be effectively reflected by the degree of the data point. The degree matrix is a diagonal matrix composed of diagonal elements with all degree values, expressed as

$$\text{Dui}_{ij} = \sum S_{ij}. \quad (2)$$

The noncanonical Lagrangian matrix is expressed as

$$\text{Fei}_{ij} = \text{Dui}_{ij} - S_{ij}. \quad (3)$$

The canonical Laplace matrix is expressed as

$$\text{Fei}_{ij} = \text{Dui}_{ij}^{-1/2} S_{ij} \text{Dui}_{ij}^{-1/2}. \quad (4)$$

3.2. Artificial Bee Colony Algorithm. The artificial bee colony algorithm is an artificial intelligence bionic algorithm with strong searchability, which has the advantages of fewer

control parameters, simple and easy to implement algorithm, and strong robustness. In the artificial bee colony algorithm, the members of the algorithm are nectar, lead bee, follow bee, and scout bee, among which scout bee is also called hire bee. Among the four members, the nectar source and the leading bee have a one-to-one correspondence in number. The three types of bees search in their respective search neighborhoods and compare their search results continuously to obtain the optimal solution. “Swing dance” is an important way for these three bees to transmit information. Figure 1 is a schematic diagram of the artificial bee colony algorithm.

The realization process of the artificial bee colony algorithm is as follows.

3.2.1. Initialize Parameters. Suppose that the total number of nectar sources is n , the maximum number of loops of the algorithm is max_m , the maximum number of iterations is max_it , the maximum number of searches is max_s , the number of bees is collected m , and the number of honey bees staying in a nectar source is recorded, and $m = 0$. By using formula (5), we can randomly generate n honey source information.

$$I_{ij} = \min_i \times \text{rand}(\text{max_j} - \min_j) + \min_j. \quad (5)$$

In the formula, $j \in \{1, 2, \dots, m\}$, rand is a random number between (0, 1), and max_j and \min_j are the upper and lower limits of the j -th dimension.

3.2.2. Lead the Bee Stage. Lead the bee to search for the local nectar source location based on the greedy selection strategy in the nectar source neighborhood. In the process of searching for a nectar source, if a new and better nectar source is found, the lead bee will compare the nectar source with the best nectar source searched in history. If $F(J_{ij}) > F(I_{ij})$, select the nectar source and pass the formula (6) to update the position of the nectar source, change the position of the nectar source from I_{ij} to J_{ij} , and lead the bee position to update the position of the food source by the following formula:

$$J_{ij} = f_{ij} + \lambda(f_{ij} - f_{kj}). \quad (6)$$

In the formula, k is generated randomly. J_{ij} represents the new nectar source position generated near f_{ij} , $\lambda \in (-1, 1)$ is a random number, and the new nectar source position is constrained by λ to be generated near the original nectar source f_{ij} .

3.2.3. Follow the Bee Phase. The leading bee transmits the nectar source information it carries to the follower bee in the form of “swing dance,” and the follower bee selects a higher-quality nectar source to follow according to the principle of roulette. Calculate the probability p_i selected by the bee by the following formula. F_i represents the fitness value of the nectar source F_i . The larger the fitness value of F_i , the higher

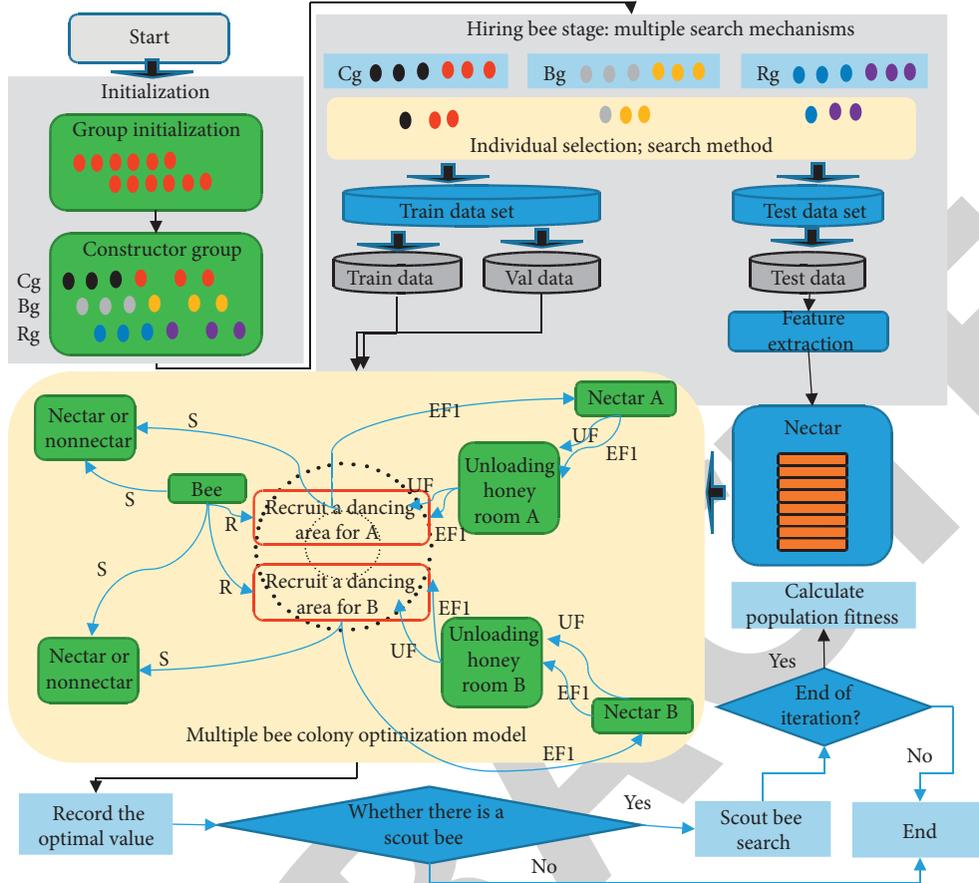


FIGURE 1: Artificial bee colony algorithm.

the probability of being selected by the bee. F_i is generated by the formula (8).

$$p_i = \frac{f_i}{\sum f_i}, \quad (7)$$

$$F_i = \begin{cases} \frac{1}{1 + f(F_i)}, \\ \frac{1}{1 + \|f(F_i)\|}. \end{cases} \quad (8)$$

3.2.4. Reconnaissance Bee Stage. If the nectar source F_i has not changed after mining, the nectar source is abandoned, a new nectar source is generated from equation (6), and a new search is continued. At this time, the lead bee attached to the nectar source F_i becomes a scout bee.

3.2.5. Record the Optimal Solution. Compare the fitness values of the current n feasible solutions, select the optimal solution, and judge whether \max_it is greater than \max_m at this time, and if it is greater than \max_m , the algorithm ends.

Because the artificial ant colony algorithm is a parallel algorithm in nature, the search process of each ant is independent of each other and communicates only through pheromone.

Therefore, the artificial ant colony algorithm can be regarded as a distributed multiagent system. It starts independent solution search at multiple points in the problem space at the same time, which not only increases the reliability of the algorithm but also makes the algorithm have a strong global search capability.

3.3. Overview of Density-Sensitive Similarity Measures. In general, clustering is an unsupervised machine learning process. Using some prior knowledge of the data set can improve the effectiveness of clustering. The most important thing is the consistency assumption of the data set, that is, local consistency and global consistency.

- (1) Local consistency: adjacent data points have higher similarity in spatial position
- (2) Global consistency: data points located on the same structure have a higher similarity

For example, in Figure 2, there are two types of points, point a belongs to one of them, and points $b, c, d,$ and e belong to the other. Local consistency is reflected in that the similarity between point d and points b and e is higher than the similarity between point d and points f and c . Global consistency is reflected in that the similarity between point c and point d is higher than the similarity between point c and point a . However, in this example, the traditional Euclidean distance can only reflect the local consistency of the data, not the global

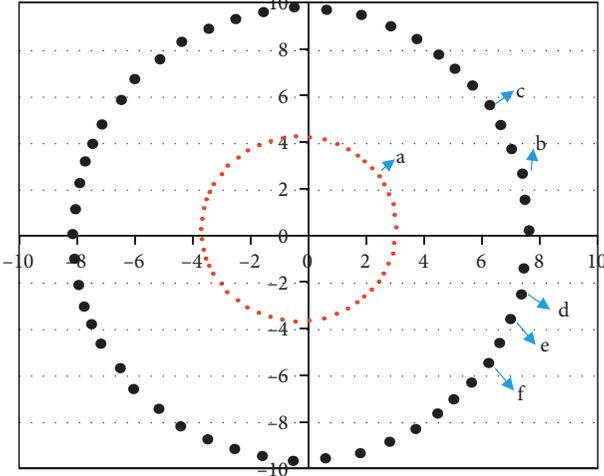


FIGURE 2: Data points on different trends.

consistency of the data. Assuming that in Figure 2, points c and f belong to the same class, and point a belongs to another class. Then, we expect the similarity between c and f to be greater than the similarity between c and a , but the point that is under the Euclidean distance measure c is closer to point a .

Based on the above problems, we design a similarity measure that can satisfy both local consistency and global consistency density-sensitive similarity measures. This metric can shorten the distance between data points in the same class and, at the same time, enlarge the distance between data points in different classes and effectively describe the actual distribution of data points, so as to achieve a good clustering effect.

Define the following similarity measures with adjustable density:

$$S_{ij} = e^{-\left(\alpha \text{dis}(d_i, d_j) / 2\sigma^2\right)} - 1. \quad (9)$$

In equation (9), $\text{dis}(d_i, d_j)$ represents the Euclidean distance between data points d_i and d_j . The variable α is the density parameter. For simple data sets, α generally takes a natural number greater than 1. When the data set is complex and its probability distribution function is not convex, the value of α can be smaller, generally set to 0.2.

The given data point corresponds to a vertex set J of an undirected weighted graph $T = (J, B)$. Since the density-sensitive similarity measure does not satisfy the triangular inequality, it cannot be directly used to construct the similarity matrix. Therefore, we redefine a distance measure based on this measure.

Use $L = \{L_1, L_2, \dots, L_n\}$ to represent the path between the connecting points L_1 and L_n with n vertices on the graph, where $L_k \in J$, $(L_k, L_{k+1}) \in B$. Use L_{ij} to represent the set of all paths between the data point pair d_i and d_j , and define the distance sensitive to the density between d_i and d_j :

$$D_{-}S_{ij} = \min \sum (L_k, L_{k+1}). \quad (10)$$

3.4. Adaptive Spectral Clustering Algorithm. The traditional spectral clustering method requires manual input of the

number of clusters before clustering. However, in practical applications, the value of the number of clusters k cannot be determined in advance under normal circumstances. At the same time, the calculation of similarity in traditional spectral clustering algorithms is also greatly affected by parameter values. Aiming at these two problems, this paper proposes an adaptive spectral clustering algorithm by introducing density-sensitive similarity measures and an artificial ant colony algorithm.

The algorithm uses the characteristics of the size of the intrinsic gaps in the cluster to automatically determine the number of initial clusters and then uses the orthogonal feature vector to classify the data. Suppose that, in an ideal state, there are k separable classes for a given data set S . For the normalized similarity matrix, there is a conclusion: the first k largest eigenvalues of the matrix are 1. At the same time, the $k+1$ eigenvalue is less than 1, and the actual distribution of the k clusters determines the size of the difference between the two eigenvalues. The more obvious the distribution, the greater the difference in eigenvalues, and conversely, the smaller the difference. At the same time, this difference is defined as the intrinsic gap. According to the perturbation theory of the matrix, the larger the value of the intrinsic gap, the more stable the nature of the subspace formed by the selected k eigenvectors.

The idea of the Eigen gap is developed based on the matrix disturbance theory. Based on the obtained Laplace matrix, the idea of eigenvalues λ is arranged in descending order, that is, $\lambda_1 > \lambda_2 > \dots > \lambda_n$. The gap sequence G represents the difference between the k and $k+1$ eigenvalues, that is, $T = \lambda_k - \lambda_{k+1}$. Among them, the larger the intrinsic gap, the more stable the subspace constructed by the selected k feature vectors. Usually, the number of clusters in the original data set is determined according to the first maximum value of the gap sequence in this card. The initial number of clusters k is generated as shown in

$$k = \max(T_1, \dots, T_n). \quad (11)$$

In the follow-up phase, the follower bee generates a new nectar source in its vicinity according to equation (6) and makes comparison choices. But in formula (6), F_{ij} represents the nectar source near f_{ij} that is superior to f_{ij} in the previous nectar source. Since the parameter λ constrains F_{ij} near f_{ij} , there is a lack of overall search and comparison. Therefore, when the nectar source location is updated by formula (6), the algorithm is easy to fall into the local optimum during operation. This article will introduce the concept of a globally optimal solution into this formula. Experiments show that the improved position update formula (12) has a strong purpose and directionality in the search process, and the algorithm has a fast convergence speed, which makes it easy to jump out of the local optimal phenomenon.

$$F_{ij} = f_{ij} + \lambda(f_{ij} - f_{kj}) + \alpha(f_{\text{best},j} - f_{ij}), \quad (12)$$

where F_{ij} is the location of the new nectar source near f_{ij} . Both k and j are random numbers generated by random formulas. The algorithm implementation process is shown in Figure 3.

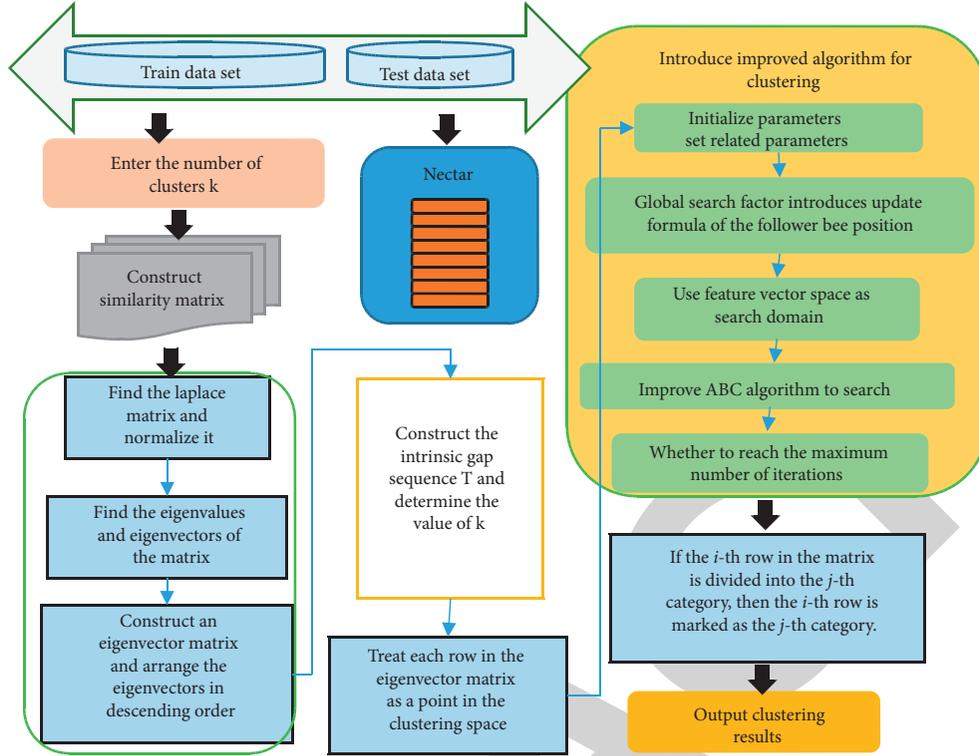


FIGURE 3: Algorithm flow.

4. Results and Discussion

4.1. Clustering Criterion Function. In order to verify the rationality of the initial clustering center selected by the algorithm, the experiment uses the first clustering criterion function value after the initial clustering center is selected to judge.

If the function value is smaller, it means that the selection of the initial cluster center is more reasonable and closer to the true cluster center. In addition, due to the decrease of the function value of the clustering criterion, the quality of the clustering is improved, and the algorithm is more efficient. The experimental results are shown in Figure 4.

It can be seen from Figure 4 that when $k=2$, the clustering criterion function value of the improved algorithm is significantly higher than that of the contrast algorithm. However, starting from $k=3$, the clustering criterion function value of the improved algorithm has dropped significantly. As the number of clusters k increases, the clustering criterion function values of the comparison algorithm and the improved algorithm tend to be parallel to each other, and there is always a certain gap. This is because the improved algorithm pays more attention to the distribution of the initial cluster centers. When $k=2$, the improved algorithm selects the least tight data point as the second initial cluster center. Due to the small number of clusters, the selection of the second initial cluster center is far from the true cluster center, resulting in an increase in the value of the clustering criterion function. However, the improved algorithm takes into account the distribution of

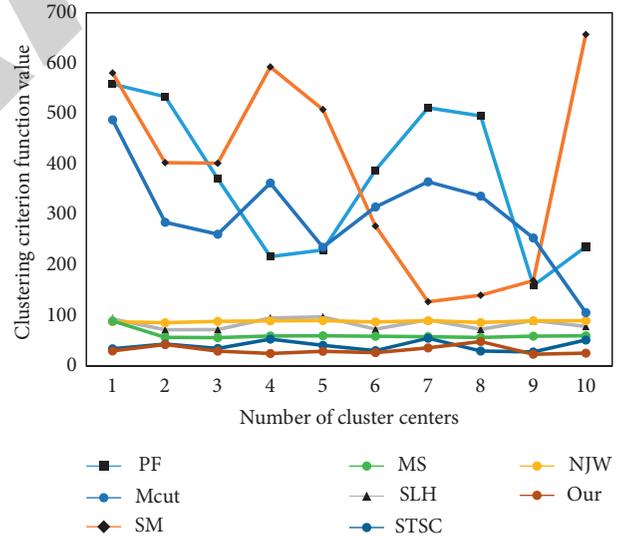


FIGURE 4: Clustering criterion function.

real cluster centers and pays more attention to the uniform distribution. When the number of clusters k increases, the uniformly distributed data points are more in line with the distribution of the optimal clustering center, which effectively reduces the clustering criterion function value of the improved algorithm, thereby improving the clustering quality. It can be seen from Figure 4 that when $k=8$, the algorithm in this paper can achieve the best performance. Therefore, the initial number of clusters in this paper is selected as 8.

4.2. Analysis of Convergence Time. In order to verify the execution efficiency of the algorithm, the experiment uses the convergence time to judge. If the convergence time is shorter, the algorithm runs faster and the execution efficiency is higher. In addition, because the convergence time is reduced, the processing efficiency of the algorithm is increased, and the effect of clustering is improved. The experimental results are shown in Figure 5.

It can be seen from Figure 5 that as the number of clusters k increases, the convergence time of the improved algorithm is the fastest. This is because the improved algorithm takes into account the distribution of the true clustering centers and selects the initial clustering centers from the data points with a low degree of compactness, which greatly reduces the amount of data for distance calculations, and greatly improves the algorithm's performance while ensuring the clustering effectiveness. Although, when the number of clusters k increases, the initial cluster center selection process of the two algorithms becomes similar, and the convergence time gradually approaches, the convergence time of the improved algorithm is still lower than that of the comparison algorithm.

4.3. Performance Analysis of Clustering Algorithm. The clustering results of the improved new algorithm in this paper are given below and compared with the clustering results of the traditional spectral clustering method (STSC). The experimental results are shown in Figure 6, where Figures 6(a) and 6(c) are the clustering results of the STSC algorithm. Figures 6(b) and 6(d) are the results of the algorithm in this paper. Figure 6(a) contains three irregularly shaped data sets. It can be seen from the classification results that, for a simple data set, both the STSC algorithm and the algorithm in this paper have achieved ideal clustering results. Figures 6(c) and 6(d) show two concentric circular data sets. The traditional spectral clustering algorithm got the wrong classification result for the intersection of the parabola, while the algorithm in this paper got the correct classification result.

From the above experimental analysis, we can see that for some simple data sets with obvious classification, both the STSC algorithm and the algorithm in this paper can get the correct classification results. However, for some complex data sets, such as concentric circles, the clustering results of the STSC algorithm have large errors. Because the algorithm in this paper introduces a density-adjustable similarity measure, the similarity between different types of data points is reduced, and the ideal clustering result is obtained, and the number of clusters is accurately calculated. The STSC algorithm must manually input the number of clusters and use Euclidean distance as the similarity measure, which does not accurately and effectively reflect the actual clustering distribution of the data, so the clustering effect is relatively poor.

To verify the clustering quality of the algorithm, the experiment uses the number of iterations to judge. If the number of iterations is less, it proves that the initially selected cluster center is closer to the real cluster center, and

the selection result is more reasonable. In addition, due to the reduction in the number of iterations, the accuracy of clustering increases, and the algorithm is more efficient. The experimental results are shown in Figure 7.

It can be seen from Figure 7 that the improved algorithm takes into account the distribution of real cluster centers and pays more attention to data points with low tightness. When the number of clusters k increases, the data points with low tightness are closer to the clusters of the new cluster, which effectively reduces the number of iterations of the improved algorithm, resulting in an increase in the number of iterations. The convergence time is shown in Figure 8.

It can be seen from Figure 8 that as the number of clusters k increases, the convergence time of the improved algorithm is significantly less than that of the contrast algorithm. This is because the improved algorithm proposed in this paper guarantees that there is a significant distance between the selected data points, so that the selected data points belong to different clusters to the greatest extent while ensuring the efficiency of the algorithm execution, which improves the execution efficiency of the algorithm, without losing a good initial clustering center; this also significantly reduces the convergence time of the algorithm, accelerates the convergence of the algorithm, and improves the execution efficiency of the algorithm.

4.4. Abnormal Behavior Analysis. We injected outliers into the data set to generate a synthetic data set to evaluate the effect of anomaly detection. The specific detection effect is shown in Figure 9.

As shown in Figure 9, as the number of clusters k increases, the detection rate of the algorithm in this paper gradually increases. The improved algorithm proposed in this paper first selects the points with the largest and smallest compactness, so as to ensure that the two initial cluster centers belong to different clusters. In addition, the initial cluster centers are randomly selected to ensure that the selected data points are evenly distributed to the greatest extent. When the number of clusters k increases, the uniform distribution of data points is more consistent with the distribution of the optimal clustering center, thereby improving the clustering performance and the effect of anomaly detection. Finally, despite the slight increase in execution time, the detection effect of the improved algorithm is still better than the comparison algorithm.

DE algorithm has strong global convergence ability and robustness and does not need to rely on the characteristic information of the problem. It is suitable for solving some optimization problems in complex environments that cannot be solved by conventional mathematical programming methods. Therefore, this article will compare the use of DE optimization, no DE optimization, no artificial ant colony algorithm, and artificial ant colony algorithm. The experimental results are shown in Figure 10. It can be seen from Figure 10 that the artificial ant colony algorithm used in this paper has the best performance.

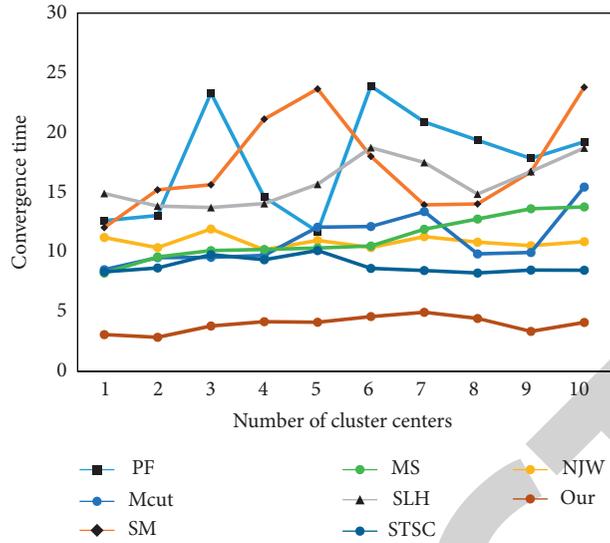


FIGURE 5: Convergence time.

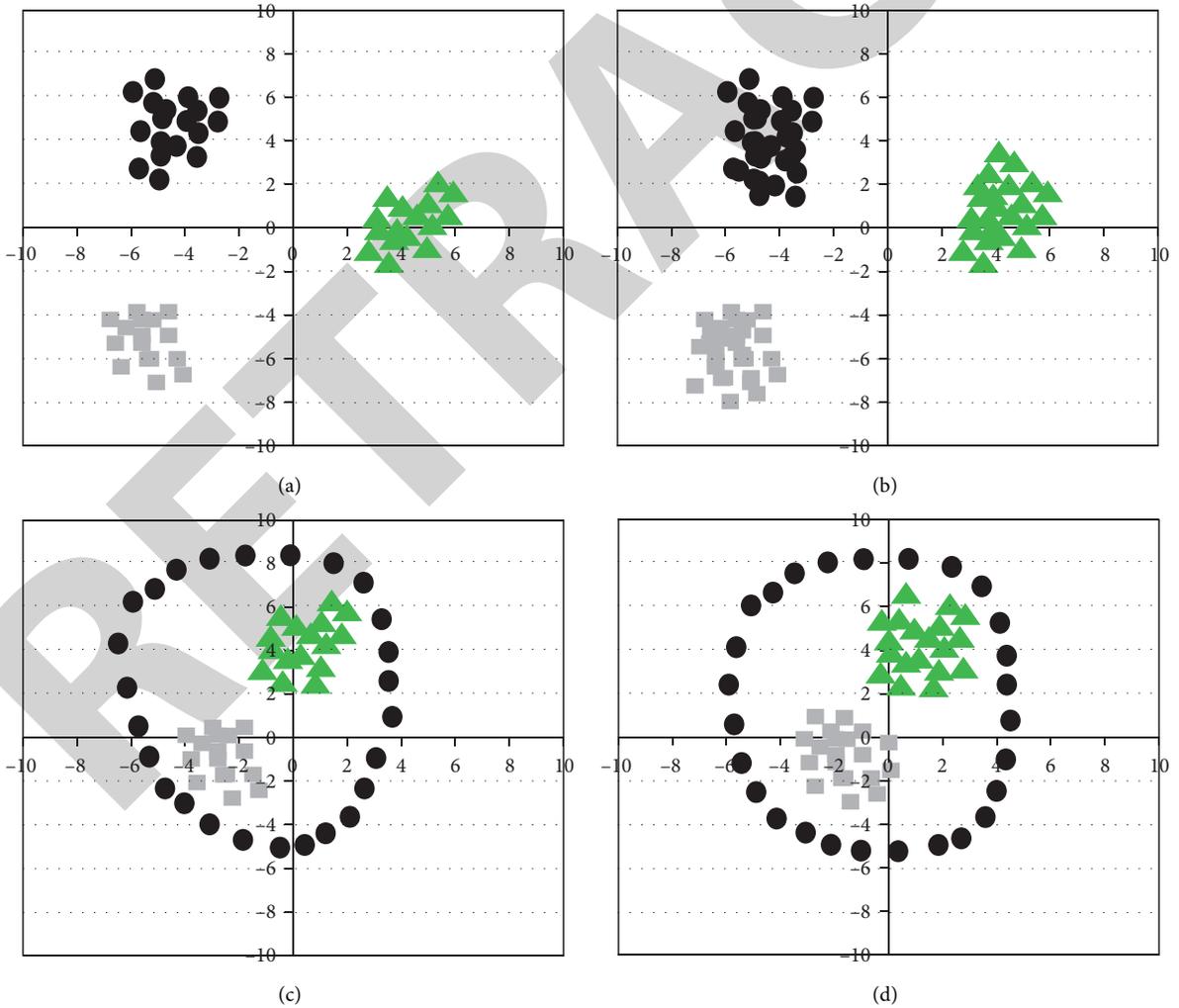


FIGURE 6: Comparison of clustering results between SC algorithm and this algorithm.

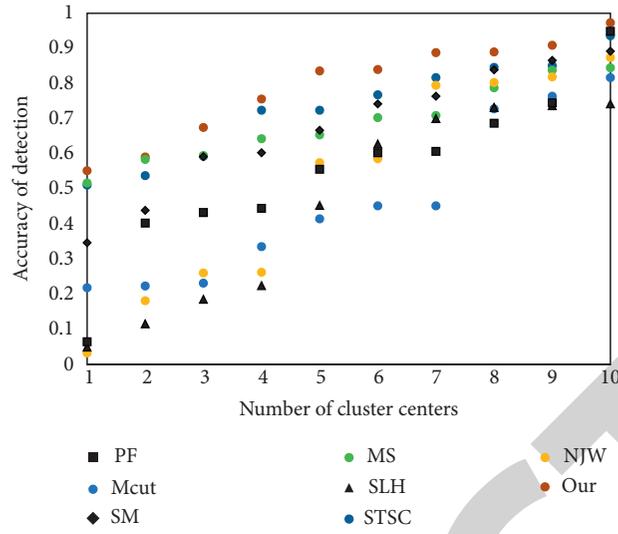


FIGURE 7: Analysis of the number of iterations and clustering performance.

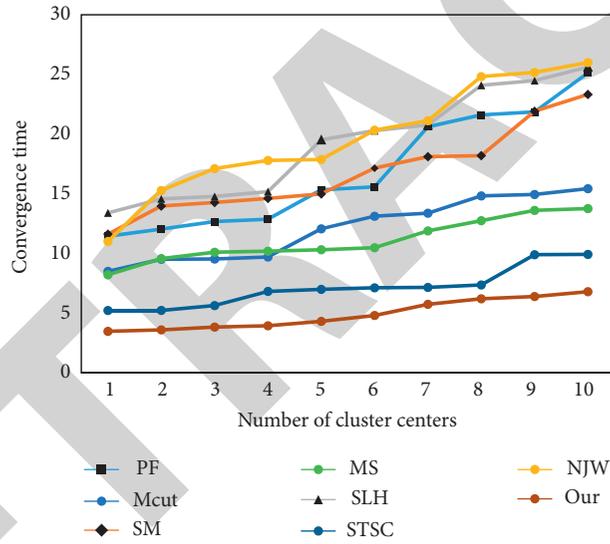


FIGURE 8: Convergence time.

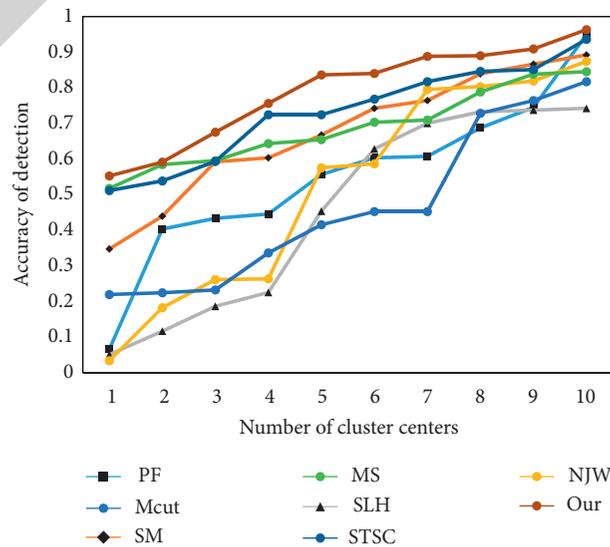


FIGURE 9: Detection effect of abnormal behavior.

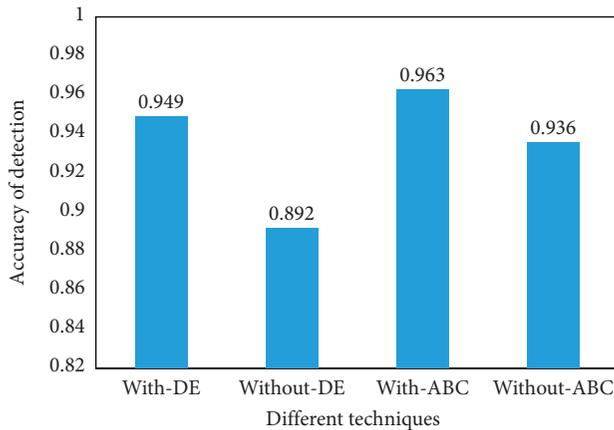


FIGURE 10: Detection effect of different optimization techniques.

5. Conclusion

This paper first introduces a density-sensitive similarity measure, which is obtained by introducing new parameters to deform the Gaussian function. Secondly, in view of the defect that the number of clusters cannot be automatically determined, this paper quotes the idea of the Eigen gap, constructs an Eigen gap sequence, calculates the first maximum value, and determines the number of clusters k , which solves the problem that the spectral clustering algorithm has the problem of insufficient sensitivity of cluster centers. Finally, in order to improve the global search ability of the algorithm, it is combined with the artificial bee colony algorithm with strong global searchability. In order to further enhance the algorithm's global searchability, the global search factor is introduced into the artificial bee colony algorithm, which effectively improves the artificial bee colony algorithm's tendency to fall into local optimality, appear premature, and slow the convergence speed of the algorithm. Experimental results show that the improved algorithm in this paper has a good improvement in stability and optimization and can achieve a good clustering effect. However, the improved algorithm in this paper also has limitations, such as the problem that the algorithm takes a long time in the clustering process. Therefore, how to combine the advantages of artificial bee colony algorithm and spectral clustering algorithm and reduce the time complexity of the algorithm will be the problem to be studied and solved in the next step.

Data Availability

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

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] W. Han and L. Longzhu, "Analysis of the operation mechanism of American social security fund," in *Proceedings of the 2019 International Conference on Economic Management and Cultural Industry (ICEMCI 2019)*, pp. 26–29, Atlantis Press, Paris, France, 2019.
- [2] Y. Gan, "Governance, legitimacy, and decision-making capability of the Chinese national social security fund—against the backdrop of international comparison," *Journal of Chinese Governance*, pp. 1–22, 2021.
- [3] Y. N. Abdi and D. M. Minja, "Influence of corporate governance compliance programs and performance of state corporations in Kenya: the case of national social security fund (NSSF)," *International Academic Journal of Human Resource and Business Administration*, vol. 3, no. 2, pp. 360–383, 2018.
- [4] N. Liu and A. Zhang, "The impact of social security fund & insurance fund on corporate innovation," in *Proceedings of the 2016 13th International Conference on Service Systems and Service Management (ICSSSM)*, pp. 1–6, IEEE, Piscataway, NJ, USA, 2016.
- [5] K. Murari, "Risk-adjusted performance evaluation of pension fund managers under social security schemes (national pension system) of India," *Journal of Sustainable Finance & Investment*, pp. 1–15, 2020.
- [6] E. Ahlqvist, P. Storm, A. Käräjämäki et al., "Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables," *The Lancet Diabetes & Endocrinology*, vol. 6, no. 5, pp. 361–369, 2018.
- [7] A. Shensa, J. E. Sidani, M. A. Dew, C. G. Escobar-Viera, and B. A. Primack, "Social media use and depression and anxiety symptoms: a cluster Analysis," *American Journal of Health Behavior*, vol. 42, no. 2, pp. 116–128, 2018.
- [8] L. D. Bos, L. R. Schouten, L. A. Van Vught et al., "Identification and validation of distinct biological phenotypes in patients with acute respiratory distress syndrome by cluster analysis," *Thorax*, vol. 72, no. 10, pp. 876–883, 2017.
- [9] W. Liang, K. C. Li, J. Long et al., "An industrial network intrusion detection algorithm based on multifeature data clustering optimization model," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 2063–2071, 2019.
- [10] D. R. Domínguez, R. P. Díaz Redondo, A. F. Vilas, and M. B. Khalifa, "Sensing the city with instagram: clustering geolocated data for outlier detection," *Expert Systems with Applications*, vol. 78, pp. 319–333, 2017.
- [11] D. Huang, C. D. Wang, J. S. Wu et al., "Ultra-scalable spectral clustering and ensemble clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 6, pp. 1212–1226, 2019.
- [12] X. Outlier, "Detection algorithms in data mining," in *Proceedings of the International Symposium on Intelligent Information Technology Application*, IEEE, Piscataway, NJ, USA, 2009.
- [13] J. Wen, J. Yang, B. Jiang et al., "Big data driven marine environment information forecasting: a time series prediction network," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 1, p. 1, 2020.
- [14] M. T. Law, R. Urtasun, and R. S. Zemel, "Deep spectral clustering learning," *Proceedings of Machine Learning Research*, vol. 70, pp. 1985–1994, 2017.
- [15] Z. Zhou and A. A. Amini, "Analysis of spectral clustering algorithms for community detection: the general bipartite

- setting,” *Journal of Machine Learning Research*, vol. 20, no. 47, pp. 1–47, 2019.
- [16] X. Meng, “Feature selection and enhanced krill herd algorithm for text document clustering,” *Computing Reviews*, vol. 60, no. 8, p. 318, 2019.
- [17] L. M. Abualigah, A. T. Khader, and E. S. Hanandeh, “Hybrid clustering analysis using improved krill herd algorithm,” *Applied Intelligence*, vol. 48, pp. 1–26, 2018.
- [18] F. M. Bianchi, D. Grattarola, and C. Alippi, “Spectral clustering with graph neural networks for graph pooling,” 2020, <https://arxiv.org/abs/1907.00481>.
- [19] Y. Wang, L. Wu, X. Lin, and J. Gao, “Multiview spectral clustering via structured low-rank matrix factorization,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 10, pp. 4833–4843, 2018.
- [20] Y. Xu, L. Shi, F. Huang et al., “Recognition of Stores’ relationship based on constrained spectral clustering,” in *Proceedings of the Proceedings of the 2019 3rd International Conference on Innovation in Artificial Intelligence*, Association for Computing Machinery, New York, NY, USA, pp. 111–115, March 2019.
- [21] S. Ding, L. Cong, Q. Hu, H. Jia, and Z. Shi, “A multiway p-spectral clustering algorithm,” *Knowledge-Based Systems*, vol. 164, pp. 371–377, 2019.
- [22] D. Shiotsuka, K. Matsushima, and O. Takahashi, “Crack detection using improved spectral clustering considering effective crack features,” in *Proceedings of the Proceedings of the 2018 International Conference on Robotics, Control and Automation Engineering*, Association for Computing Machinery, New York, NY, USA, pp. 181–185, December 2018.
- [23] X. Zhu, S. Zhang, W. He et al., “One-step multi-view spectral clustering,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 10, pp. 2022–2034, 2018.
- [24] L. Dong, S. Liu, and H. Zhang, “A method of anomaly detection and fault diagnosis with online adaptive learning under small training samples,” *Pattern Recognition*, vol. 64, pp. 374–385, 2017.
- [25] S. K. S. L. Preeth, R. Dhanalakshmi, R. Kumar et al., “An adaptive fuzzy rule based energy efficient clustering and immune-inspired routing protocol for WSN-assisted IoT system,” *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–13, 2018.
- [26] S. G. Phillip, H. Aihara, C. C. Thompson et al., “775 duodenal ESD and sutured defect closure across a lumen apposing metal stent,” *Gastrointestinal Endoscopy*, vol. 87, no. 6, pp. AB116–AB117, 2018.
- [27] Y. Zhou, S. Yu, R. Sun et al., “Topological segmentation for indoor environments from grid maps using an improved NJW algorithm,” in *Proceedings of the 2017 IEEE International Conference on Information and Automation (ICIA)*, pp. 142–147, IEEE, Macao, China, July 2017.
- [28] R. Rajamohamed and J. Manokaran, “Improved credit card churn prediction based on rough clustering and supervised learning techniques,” *Cluster Computing*, vol. 21, no. 1, pp. 65–77, 2018.
- [29] X. Ma, B. Wang, and L. Yu, “Semi-supervised spectral algorithms for community detection in complex networks based on equivalence of clustering methods,” *Physica A: Statistical Mechanics and its Applications*, vol. 490, pp. 786–802, 2018.
- [30] W. Bo, Z. B. Fang, L. X. Wei, Z. F. Cheng, and Z. X. Hua, “Malicious URLs detection based on a novel optimization algorithm,” *IEICE Transactions on Information and Systems*, vol. E104.D, no. 4, pp. 513–516, 2021.
- [31] B. Bai, Z. Guo, C. Zhou, W. Zhang, and J. Zhang, “Application of adaptive reliability importance sampling-based extended domain PSO on single mode failure in reliability engineering,” *Information Sciences*, vol. 546, pp. 42–59, 2021.
- [32] W. Wang, F. Xia, H. Nie et al., “Vehicle trajectory clustering based on dynamic representation learning of internet of vehicles,” *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, 2020.
- [33] B. Li, R. Liang, W. Zhou, H. Yin, H. Gao, and K. Cai, “LBS meets blockchain: an efficient method with security preserving trust in SAGIN,” *IEEE Internet of Things Journal*, p. 1, 2021.
- [34] J. Chen, C. Du, Y. Zhang, P. Han, and W. Wei, “A clustering-based coverage path planning method for autonomous heterogeneous UAVs,” *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–11, 2021.
- [35] J. Yang, Y. Zhao, J. Liu et al., “No reference quality assessment for screen content images using stacked autoencoders in pictorial and textual regions,” *IEEE Transactions on Cybernetics*, pp. 1–13, 2020.