Improved ISODATA Clustering Method with Parameter Estimation based on Genetic Algorithm

Kohei Arai

Graduate School of Science and Engineering
Saga University, Saga City
Japan

Abstract—**Improved ISODATA clustering method with merge and split parameters as well as initial cluster center determination with GA: Genetic Algorithm is proposed. Although ISODATA method is well-known clustering method, there is a problem that the iteration and clustering result is strongly depending on the initial parameters, especially the threshold for merge and split. Furthermore, it shows a relatively poor clustering performance in the case that the probability density function of data in concern cannot be expressed with convex function. To overcome this situation, GA is introduced for the determination of initial cluster center as well as the threshold of merge and split between constructing clusters. Through experiments with simulated data, the well-known the University of California, Irvine: UCI repository data for clustering performance evaluations and ASTER/VNIR: Advanced Spaceborne Thermal Emission and Reflection Radiometer / Visible and Near Infrared Radiometer onboard Terra satellite of imagery data, the proposed method is confirmed to be superior to the conventional ISODATA method.**

Keywords—*ISODATA clustering; nonlinear merge and split; concaveness of probability density function: PDF; remote sensing satellite imagery data; clustering; genetic algorithm: GA; nonlinear optimization*

I. Introduction

Clustering methods can be broadly divided into hierarchical and non-hierarchical clustering methods [1], [2]. Typical non-hierarchical clustering methods are the k-means method[1] and the ISODATA method[2]. In the k-means method, it is necessary to give the number of clusters and the initial cluster center in advance, and the calculation time and the obtained cluster shape will change depending on their settings. On the other hand, the ISODATA method can autonomously determine the effective number of clusters within a certain range of the set number of clusters.

The ISODATA method separates target data individuals into clusters based on the k-means method, and then divides and fuses the clusters according to a preset threshold value based on statistical indices within and between clusters, and rearranges the cluster individuals. This is a method of repeating this series of processing until the rearrangement end judgment criterion is satisfied. Therefore, although the ISODATA method has relatively high clustering accuracy, it takes a considerable amount of processing time to set these

parameters, and this determination can often be difficult. The final number of clusters, clustering accuracy and calculation time depends on these parameters. Especially in the case of clustering for multi-dimensional data such as satellite images, the number of parameters increases as the number of dimensions increases, so adjustment is extremely difficult.

Moreover, the k-means method and the ISODATA method implicitly assume that the probability density function of the target data is a convex function. That is, high clustering accuracy cannot be expected in the case of a distribution of a cluster that is a concave function in a multidimensional space.

This paper assumes that the target data is multidimensional like a satellite image, deals with the case where the probability density function of the cluster individual is concave, and improves the clustering accuracy by using the estimated optimal parameters. The author proposes a clustering method based on the ISODATA method. Genetic Algorithms (GA) [3], [4] were used for parameter optimization. There is the other alternative algorithm, sand cat swarm optimization (SCSO) [5], Grey Wolf Optimizer (GWO) algorithm [6], Moth-flame optimization algorithm [7], etc.

GA does not guarantee a global optimal solution, but it is a probabilistic optimization method that can estimate a suboptimal solution in a relatively short time [8], [9], [10]. It is effective for problems that have not been discovered and have such a large solution space that a full search is considered impossible [11], [12]. In this paper, the author first shows the effectiveness of the proposed method using simulation data in which the distribution of cluster individuals is a concave function.

The author also applied the UCI repository data set [13] and satellite image data, which are frequently used for comparative evaluation of clustering accuracy, to the proposed method and evaluated the clustering accuracy. The conventional ISODATA method and shape-independent clustering were used as the conventional clustering methods [14], [15]. It is reported here because the proposed method was superior to the conventional methods.

The following section d4escribes related research works. Then the proposed method is described followed by experiment. After that, conclusion is described together with some discussions.

---

[1] https://en.wikipedia.org/wiki/K-means_clustering
[2] https://www.harrisgeospatial.com/docs/ISODATAClassification.html

II. RELATED RESEARCH WORK

Influence due to geomorphology on context Genetic Algorithm: GA clustering is investigated [16]. On the other hand, learning processes of image clustering method with density maps derived from Self-Organizing Mapping (SOM) is also proposed [17].

Non-linear merge and split method for image clustering (closely related to the proposed method here) is proposed [18]. Meanwhile, revised pattern of moving variance for acceleration of automatic clustering is investigated [19].

Automatic detection method for clustered micro calcification in mammogram image based on statistical texture features is proposed [20]. On the other hand, comparative study between the proposed GA based ISODATA clustering and the conventional clustering methods are conducted [21].

Image clustering method based on density maps derived from Self Organizing Mapping: SOM is proposed [22]. Meanwhile, clustering method based on Messy Genetic Algorithm: GA for remote sensing satellite image clustering is proposed [23].

Visualization of learning process for back propagation Neural Network: NN clustering is proposed [24]. On the other hand, improvement of automated detection method for clustered micro calcification base on wavelet transformation and support vector machine is attempted [25].

Image clustering method based on Self-Organization Mapping: SOM derived density maps and its application for Landsat Thematic Mapper: TM image clustering is proposed [26]. Also, comparative study between the proposed shape independent clustering method and the conventional method (k-means and the others) is conducted [27].

Genetic Algorithm: GA utilizing image clustering with merge and split processes which allows minimizing Fisher distance between clusters is proposed [28]. Also, Fisher distance-based GA clustering taking into account overlapped space among probability density functions of clusters in feature space is proposed [29].

Initial centroid designation algorithm for k-means clustering is proposed [30]. Meanwhile, Pursuit Reinforcement Competitive Learning: PRCL based online clustering with tracking algorithm and its application to image retrieval is proposed [31].

III. PROPOSED METHOD

Clustering is a method of classifying target data by collecting similar items based on the similarity or dissimilarity between target data individuals and generating groups (clusters); also called analysis. The criteria for measuring how similar the target data individuals are similarity (dissimilarity) and dissimilarity. The degree of similarity is, for example, a measure indicating that the object is more similar as the value is larger, such as the correlation coefficient.

On the other hand, the dissimilarity, which is also called the dissociation degree, is a measure indicating that the larger the value, the less similar the objects. Generally, dissimilarity (distance) is often used. The similarity defined in clustering has indices such as matching coefficient and similarity ratio, while there are many definitions of distance.

In this paper, we use local mean distance and similarity that can deal with the case where the probability density function of the target data individual is a concave function, and it is necessary for the ISODATA method by GA using the fitness function based on these. We propose a clustering method that determines the initial cluster centers as well as the thresholds for cluster partitioning and fusion.

A. Local Average Distance and Local Similarity

The author proposes the Moving Window method. Once the local range (Window) is determined, the local inter-individual distance and the similarity within the range are obtained. The Moving Window method is a method of finding the sum or average of local distances and similarities over the entire area by moving the local area little by little along each dimension (Fig. 1). In this research, the author uses a hypersphere as a window in multidimensional Euclidean space. In the figure, while moving in the local range of radius r, the distance between individuals within this range, or the sum and average of the similarities are obtained.

As shown in Eq. (1), the average value of the inter-individual distances within the local range is called the local average distance L.

$$L = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \delta_{i,j} \left| \left| l_i - l_j \right| \right| \tag{1}$$

where,

$$\delta_{i,j} = \begin{cases} 1, & if \ C_i = C_j \\ \frac{1}{2}, & if \ C_i \neq C_j \end{cases} \tag{2}$$

In addition, n is the number of individuals in the local range, 1 is the position vector of the individual, and Ck is the membership cluster of the individual of k. That is, the distance (norm) between individuals is the sum of all the distances between individuals in the local range considering 1 for the same cluster and half the weight for different classes.

Sum of local similarity is defined as follows:

The difference in distance between individuals belonging to the same cluster as the diameter within a certain local range is taken as the similarity, and the sum DS thereof is obtained as in the following formula.
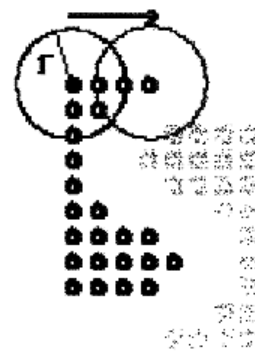

Fig. 1. Moving Window Method.

$$DS = \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \delta_{i,j}(2r - ||l_i - l_j||) \qquad (3)$$

where,

$$\delta_{i,j} = \begin{cases} 1, & if\ C_i = C_j \\ 0, & if\ C_i \neq C_j \end{cases} \qquad (4)$$

B. Determine the Size and Weight of the Local Range

The local mean distance has a role to detect inter-cluster variance. Therefore, it is desirable that the local range, that is, the window size in the Moving Window method, can cover the inter-cluster dispersion. When the window size, that is, the radius of the hypersphere is gradually increased from the minimum distance between individuals, and the total sum of the local average distances is obtained, the local average distance has a peak when the inter-cluster dispersion is covered.

This peak is found and the radius of the hypersphere in its local range is used as the reference r_p. Experiments were performed using the simulation data shown in Fig. 2(a), (b), and (c) while changing the radius of the hypersphere. The results are summarized in Table I.

Looking at Table I, even if the radius of the hypersphere is made larger than the standard to some extent, it does not affect the clustering result, but below the standard value, it greatly affects the cluster result.

This can identify both cluster individuals if r is set so as to include the maximum inter-individual distance between different clusters, but if set shorter than this, that is, if the local range is set narrow, both cluster individuals are separated. It means that it will be difficult. Therefore, it is sufficient to set γ sufficiently long, but this is directly related to the increase in processing time (proportional to r^2), which is a problem.
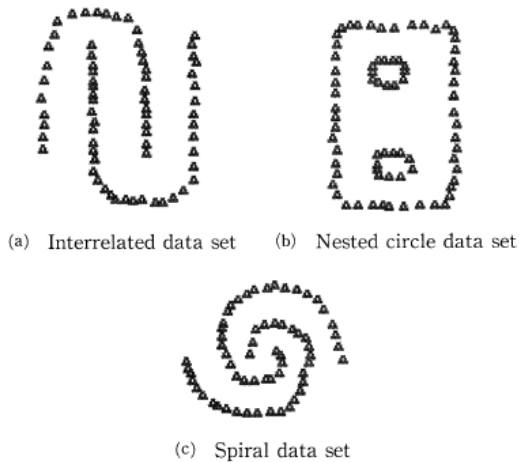

(a) Interrelated data set   (b) Nested circle data set

(c) Spiral data set

Fig. 2. 2D Simulation Data.

TABLE I. EFFECT OF WINDOW SIZE ON LOCAL AVERAGE DISTANCE

Error(%)	0.5rp	1.0rp	1.5rp	2.0rp
Interrelated	98.55	0.00	0.00	0.00
Nested circle	97.30	0.00	0.00	0.00
Spiral	96.77	96.77	0.00	0.00

Therefore, in this study, twice the reference value, $2r$, was taken as the radius of the hypersphere. On the other hand, the sum of local similarities is related to the distance between individuals within a cluster. The window size should be longer than the distance between all individuals in the cluster and their nearest neighbors, and smaller than the inter-cluster variance. When the window size is small, the clustering result has a large number of clusters due to the local average distance, and in an extreme case, all individuals themselves become clusters.

Therefore, the validity of the window size can be judged from the final number of clusters. Assuming that there is no isolated individual, it is appropriate that the window size, that is, the diameter of the hypersphere, is longer than the distance (d_{min}) between the individual farthest from other individuals and their nearest neighbors. First, clustering is performed by setting the radius of the hypersphere to d_{min}. Clustering is performed by increasing the radius of the hypersphere when the number of final clusters exceeds twice the expected number, or when isolated individuals appear. This process was repeated until the final number of clusters fell within twice the expected number.

It is important for the goodness of fit so that individuals whose distance is less than a certain threshold value belong to one cluster, that is, the sum DS of local similarities is maximized. This is equivalent to making the local average distance L as small as possible. From this point of view, it is better to make the weight M smaller, but the effective number of DS is finite, so the value of the weight must be set so that the effective value of M/L can influence the effective number of DS. Also, since maximizing DS is prioritized, it must be $M/L << DS$.

C. GA

1) Real-Coded Genetic Algorithms (RCGA): GA is an optimization algorithm that refers to the evolution of living organisms. In GA, the solution of the problem is expressed as an individual, and each individual is composed of chromosomes. Individuals evolve by selective selection, crossover, and mutation, and search for optimal solutions. Early GAs performed crossovers and mutations with binary coded bit strings of variables. This ignores the continuity of variables.

On the other hand, a GA has been proposed which uses the numerical value itself and performs crossover and mutation considering the continuity of variables. This is called Real-Coded Genetic Algorithms (RCGA). RCGA does not use the bit string like the conventional GA but expresses the individual used for the search by converting it into a real vector.

Therefore, the conventional GA chromosomal locus composition replaces the bit string with a real vector. At this time, the multidimensional vector centered on the initial cluster is used as the real vector. In this research, the initial cluster center optimization is performed based on RCGA.

2) Fitness function: The author defines a fitness function that maximizes the ratio of the sum of inter-cluster variance to the sum of intra-cluster variance. As for the clustering result, the cluster is configured to be optimal when the probability density function of the target data individual can be regarded as a convex function. On the other hand, if the fitness functions are defined as in Eq. (5) using the sum of local similarities and the local average distance, an optimal cluster configuration is possible even in the case of a concave function distribution.

$$F = \sum DS + M / \sum L \qquad (5)$$

where *F* is the goodness of fit and *M* is the weight. The local mean distance represents the intra-cluster variance in the local range. The smaller this value, the smaller is the intra-cluster variance. If all individuals in the local range belong to individual clusters, the local mean distance is the minimum.

When this value is used as the goodness of fit, the individuals within the range tend to belong to different clusters, and it is possible to perform clustering so that the individuals at both ends of the local area with large inter-individual variance belong to different clusters.

The sum of local similarities is related to the intra-cluster variance in the local range. The larger this value, the smaller the variance within the cluster in the local range. When all individuals in the local range belong to one cluster, the sum of local similarities becomes maximum. When this value is used as the goodness of fit, the number of clusters tends to decrease, and it is possible to perform clustering so that individuals with close distances belong to one cluster.

By using Eq. (5), it is possible to generate a fitness function that does not depend on the shape of the cluster by balancing the sum of the local average distance and the local similarity and considering the intra-cluster variance and inter-cluster variance at the same time.

3) Crossover: Since RCGA does not code, a crossover operator specialized for this is required. Typical examples of this include blend crossover (BLX-α) and unimodal normal distribution crossover (UNDX). BLX-α determines offspring as follows.

   a) Let α, b be the two parents in the real-valued vector space.

   b) Calculate the intersection [A, B].

$$A = \min (a, b) - \alpha | a-b | \qquad (6)$$

$$B = \max (a, b) + \alpha | a-b | \qquad (7)$$

where α is a coefficient parameter.

   c) Determine the offspring from the interval [A, B] with uniform random numbers.

That is, in the conventional GA, unlike the crossover method in which some or all of the loci of chromosomes are replaced, the vector obtained by subtracting the coefficient times the inter-individual distance from the smaller individual vector to the larger individual vector is added. It is a method

of determining a crossing vector according to a uniform random number generated in the range up to it.

4) Mutation method: Mutations were given by normal distribution. That is, it is a method of changing the real number vector of an individual according to the normal distribution according to the probability of the mutation to be set.

5) Selection method: The selection method of RCGA is the tournament method, and the elite strategy to enhance the convergence performance of RCGA is adopted.

6) RCGA Specific method and parameter setting: RCGA is more efficient because there are many real-valued parameters in ISODATA method. In this study, RCGA is used to optimize the thresholds of initial cluster centers and division and fusion in the ISODATA method. The specific method is shown below.

- Set the number of RCGA populations to 30 and the number of generations to 300.

- Since RCGA is used, real coding is constructed by constructing a multidimensional vector from real values of each parameter and performing crossover and mutation on the vector.

The composition of the vector of chromosomes is as follows,

$$C (S, M, A_1, A_2, ..., A_m) \qquad (8)$$

where S is the division threshold, M is the fusion threshold, and A_m is the m-th coordinate of the initial cluster center.

In this way, the chromosome is defined by Eq. (8), and RCGA is performed to find the optimal thresholds for division and fusion of the initial cluster center and ISODATA.

- Set the tournament selection size to 3.

- Use the BLX-α method as the crossover method. The value of α is set to 0.5 and the crossover probability is set to 70%, in order to prevent falling into a local solution and also considering the speed of convergence.

- The mutation method uses the normal distribution mutation method and sets σ to 0.5. Here, if the mutation probability is set to 5% or less, the local solution is likely to fall, and if the mutation probability is increased, the efficiency of GA is deteriorated, so it is set to 5%.

In order to compare the clustering results, we perform clustering of 100 sets of initial cluster centers randomly determined by uniform random numbers and calculate the maximum likelihood result and average value. The author also does some hierarchical clustering.

D. Improved ISODATA

As mentioned above, the ISODATA method is a method in which target data individuals are divided by constructing cluster boundaries by the hyperplane (Voronoi division[3]) by

---

[3] https://en.wikipedia.org/wiki/Voronoi_diagram

the k-means method. That is, the probability density function of the cluster is implicitly assumed to be a convex function, and if the probability density function of the target data individual is a concave function, accurate cluster division cannot be expected. When the probability density function is concave, the ISODATA method may be able to deal with the division and fusion process, but there is a possibility that clusters that are correctly classified by division and fusion once will be destroyed by rearrangement.

E. Reduction of Calculation Amount

In the proposed method, the initial cluster center is estimated in advance by the real-valued GA (RCGA), so that the clustering result can obtain an almost optimal cluster result without relying on the modification of the cluster center by repeating the ISODATA method. That is, the clustering result is hardly affected even if the number of iterations of the ISODATA method is reduced to some extent. In the experiment, the number of repetitions of the ISODATA method was set to 4.

In the Moving Window method, the computational complexity increases exponentially as the number of dimensions increases. In the proposed method, we decided to reduce the amount of calculation by moving each individual in order, rather than moving gradually along each dimension.

IV. EXPERIMENTS

A. Dataset of Data used

The UCI repository dataset is a data archive published by knowledge discovery researchers at the University of California, Irvine (University of California, Irvine). It can be accessed by anyone on the web page[4]. In this research, Iris, Wine, New thyroid, and Ruspini dataset of R, and fossil dataset of Chernoff are used for the experiment.

1) Iris data set The Iris data set is data of three types of iris flowers. The total number of individuals in the dataset is 150, including 50 individuals for each type. The data in this dataset is four-dimensional. They are the width and length of sepal, the width and length of petal, and the unit is cm. The Iris dataset is one of the best-known datasets for clustering.

2) Wine dataset: The Wine Dataset is the chemical analysis data for three Italian wines. The total number of individuals in the dataset is 178, and the number of individuals of each type is 59, 71, 48, respectively. The data in this dataset is 13 dimensions. There is a large difference in the range of data in each dimension. There are dimensions where all numbers are less than or equal to 1 and dimensions that include up to 1,000. Therefore, clustering Wine datasets is considered difficult. In this study, the normalized (Min: 0, Max: 1) Wine dataset was used.

3) Ruspini data set: The Ruspini Data Set is included in R and S-plus. It is four-dimensional data with four categories. The total number of individuals is 75, including 23, 20, 17, and 15 individuals.

4) New thyroid data set: The New thyroid data set is 5D data on infectious diseases in UCI. The total population is 215. The number of categories is 3, category 1 includes 150 individuals, category 2 includes 35, and category 3 includes 30.

5) Fossil data set: The fossil data set is 6-dimensional data on three types of limestone by Chernoff. The total number of individuals is 87, and each category includes 40,34,13 individuals.

These five datasets are standard datasets that are often used to compare clustering methods. Experiments were performed on these data sets. The results are summarized in Table II.

It can be seen from Table II that the error is 20.23% lower in the proposed method compared to the ISODATA method when the initial cluster centers are set randomly. It is also shown that the error of the proposed method is much smaller than that of the k-means method (ICCDA) and the single linkage method.

B. Satellite Imagery Data used

Clustering was performed using a part (Fig. 3) near Kashima city in Japan extracted from satellite images of the Saga area taken by ASTER / VNIR: Visible and Near Infrared Radiometer on December 7, 2004. The sampling area shown in the figure was set, and 30×30 identical cluster image data individuals were set. Clusters of these individuals correspond to four cluster types: sea, plants, freshwater, and urban areas. The dimension of the image data individual is 3, and the maximum value of each dimension is 255. The image is shown in Fig. 4.

TABLE II. CLUSTERING ERROR OF THE FOUR CLUSTERING METHODS FOR FIVE DATASETS OF UCI REPOSITORY

Error (%)	Proposed	k-means	ISODATA	Single linkage
Iris	10.67	10.67	21.33	32.00
Wine	4.49	5.62	34.65	61.24
Ruspini	0.00	0.00	12.43	0.00
Fossil	4.60	4.60	34.42	13.79
New thyroid	13.02	13.95	31.63	29.77
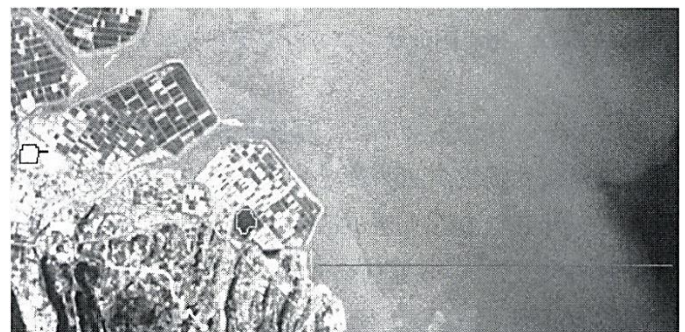
Fig. 3. Satellite Images Around Kashima City, Saga Prefecture in Japan (Band: 1,2,3 N).

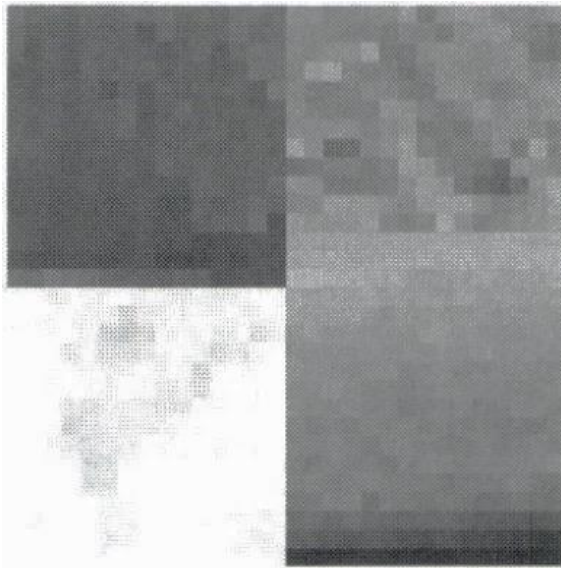[4] http://mlearn .ics.uci.edu/MLRepository.html

Fig. 4.   Sampling Image Representing a Cluster.

The proposed method and the conventional ISODATA method are applied to the satellite image, and the cluster results are compared. In the conventional ISODATA method, the result based on the Ward method is used for the initial cluster center. The results are shown in Table III and Fig. 5. As shown in Fig. 5 of the clustered results of the proposed method and ISODATA, there are some clustering errors at the bottom right of the clustered result of ISODATA while there is no error for the proposed method.

As can be seen from this table, it is shown that the proposed method is also effective for satellite data in which the probability density function of the target data individual is a concave function. Also, from the center of the original image in Fig. 6(a), a part of 32 * 32 pixels include the above four categories. Fig. 6(a) is extracted to perform clustering by the proposed method and the conventional ISODATA method. Applied and compared the cluster results. The results of clustering are shown in Fig. 6(b) and 6(c).

TABLE III.   RESULTS OF EXPERIMENTS USING SATELLITE IMAGES

Method	Proposed	ISODATA
Error (%)	0.00	3.89
Elapsed time (s)	25567.445	1.134

(a) Proposed   (b) ISODATA
Fig. 5.   Clustered Results of Proposed Method and ISODATA.
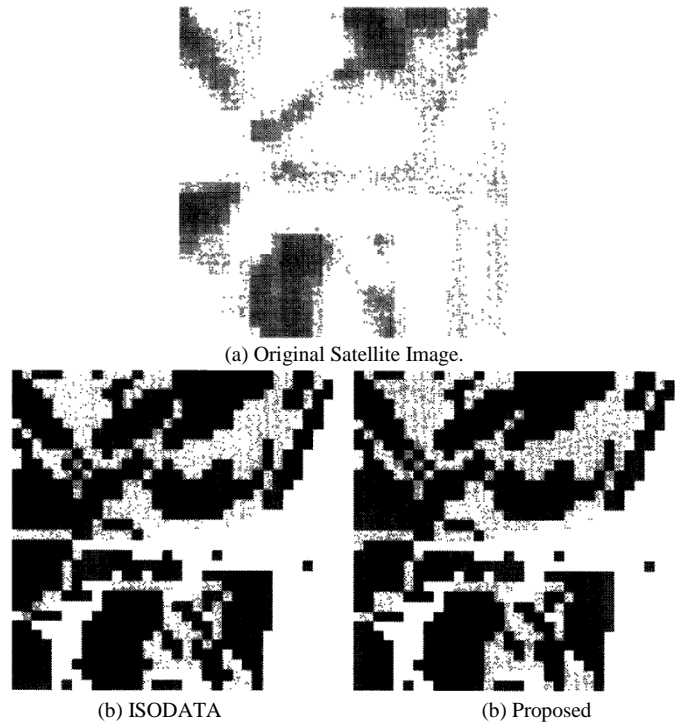
(a) Original Satellite Image.

(b) ISODATA   (b) Proposed
Fig. 6.   Original Satellite Image and the Clustered Results of the Conventional ISODATA and the Proposed Method.

In comparison with the ISODATA method, which does not form a cluster corresponding to this distribution, it can be seen that the proposed method forms an appropriate cluster.

It can be seen that the proposed method also gives better results for satellite image data than the conventional ISODATA method, especially for classification of fresh water.

V.   CONCLUSION

From the results of clustering experiments using simulation, UCI repository dataset and ASTER / VNIR images, the proposed method was superior to the conventional method in all cases. It was found that the proposed method has a higher degree of separation than the conventional method even when the probability density function of the clustering target data individual is a concave function, and a good clustering result is obtained. Therefore, the proposed method overcomes the problem that accurate clustering cannot be expected when it is difficult to set the parameters of the ISODATA method (initial cluster center and threshold of division / fusion: split / merge) and when the probability density function of the ISODATA method is concave.

Although the calculation time of the proposed method is longer than that of the conventional method by the amount of parameter setting by the real-valued genetic algorithm, the cluster accuracy is significantly improved, and it is effective when the cluster accuracy is important.

VI.   FUTURE RESEARCH WORKS

Further research works are required for validation of the proposed clustering method with the other satellite imagery data. Furthermore, the alternative optimization algorithms of

GA such as sand cat swarm optimization (SCSO), Grey Wolf Optimizer (GWO) algorithm, Moth-flame optimization algorithm, etc. have to be tried for the parameter selection of the proposed clustering based on the modified ISODATA clustering.

REFERENCES

[1] R.J. Williams, Simple statistical gradient-follow -ing algorithms for connectionist reinforcement learning, Machine Learning, 8, 3-4, 229-256,1992.

[2] C.Yi-tsuu, Interactive Pattern Recognition, Marcel Dekker Inc., New York and Basel, 1978.

[3] John H.Holland, Adaptation in Natural and Artifical Systems, Ann Arbor, The University of Michigan Press, 1975.

[4] John H.Holland, Adaptation in Natural and Artifical Systems, Massachusetts Institue of Tech -nology, 1992.

[5] Seyyedabbasi, A., Kiani, F. Sand Cat swarm optimization: a nature-inspired algorithm to solve global optimization problems. Engineering with Computers (2022).

[6] Seyyedabbasi, A., Kiani, F. I-GWO and Ex-GWO: improved algorithms of the Grey Wolf Optimizer to solve global optimization problems. Engineering with Computers 37, 509–532 (2021).

[7] Seyyedabbasi, A., Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm, Knowledge-Based Systems, Volume 89, November 2015, Pages 228-249, 2015.

[8] Kohei Arai, Basic Theory of Pattern Recognition, Academic Book Publishing, 1999.

[9] Ono Isao, Sato Hiroshi, Kobayashi Shigenobu, Function optimization by real-valued GA using unimodal normal distribution crossover UNDX, Journal of Japan Society for Artificial Intelligence, 14, 6, 1999.

[10] Takagi Mikio, Shimoda Y., Arai Kohei, Image Analysis Handbook, The University of Tokyo Press, pp.641-684, 1991.

[11] L.J. Eshleman and J.P. Scha. er, Real-Coded GA and internal schemata, Foundations of Genetic Algorithms, 2, 187, 1993.

[12] E.G.M. de Lacerda, Model Selection of RBF Networks via Genetic Algorithms, Pernambuco Fedoral University Informatics Center (Report), 2003.

[13] UCI Repository, http://www.sgi.com/tech/mlc/db/.

[14] S Bandyopadhyay, An automatic shape indepen -dent clustering technique, Machine Intelligence Unit, January 06 Pattern Recognition, Vol.37, No.1, 2004.

[15] Kohei Arai and Ali Ridho Barakba, Method for shape independent clustering in case of numerical clustering together with condensed clustering, Proc. of the SCI symposium 2004.

[16] Akira Yoshizawa and Kohei Arai, Influence due to geomorphology on context genetic Algorithm Clustering, Transaction of the Japanese Geomorphological Union, Vol.22, No.4, 23-28, 2001.

[17] Kohei Arai, Learning processes of image clustering method with density maps derived from Self-Organizing Mapping(SOM), Journal of Japanese Society of Photogrammetry and Remote Sensing, 43, 5, 62-67, 2004.

[18] Kohei Arai, Non-linear merge and split method for image clustering, Journal of Japanese Society of Photogrammetry and Remote Sensing, 43, 5, 68-73, 2004.

[19] Ali Ridho Barkbah and Kohei Arai, Revised pattern of moving variance for acceleration of automatic clustering, Electric and Electronics Polytechnics in Surabaya EEPIS Journal, 9, 7, 15-21 (2004).

[20] Kohei Arai, I.N.Abdullah, H.Okumura, Automatic detection method for clustered micro calcification in mammogram image based on statistical texture features, International Journal of Advanced Computer Science and Applications, 3, 5, 7-11, 2012.

[21] Kohei Arai, X.Q. Bu, Comparative study between the proposed GA based ISODATA clustering and the conventional clustering methods, International Journal of Advanced Computer Science and Applications, 3, 7, 125-131, 2012.

[22] Kohei Arai, Image clustering method based on density maps derived from Self Organizing Mapping: SOM, International Journal of Advanced Computer Science and Applications, 3, 7, 102-107, 2012.

[23] Kohei Arai, Clustering method based on Messy Genetic Algorithm: GA for remote sensing satellite image clustering, International Journal of Advanced Research in Artificial Intelligence, 1, 8, 6-11, 2012.

[24] Kohei Arai, Visualization of learning process for back propagation Neural Network clustering, International Journal of Advanced Computer Science and Applications, 4, 2, 234-238, 2013.

[25] Kohei Arai, Indra Nugraha Abdullah, Hiroshi Okumura, Rie Kawakami, Improvement of automated detection method for clustered micro calcification base on wavelet transformation and support vector machine, International Journal of Advanced Research in Artificial Intelligence, 2, 4, 23-28, 2013.

[26] Kohei Arai, Image clustering method based on Self-Organization Mapping: SOM derived density maps and its application for Landsat Thematic Mapper image clustering, International Journal of Advanced Research in Artificial Intelligence, 2, 5, 22-31, 2013.

[27] Kohei Arai, Cahya Rahmad, Comparative study between the proposed shape independent clustering method and the conventional method (k-means and the others), International Journal of Advanced Research in Artificial Intelligence, 2, 7, 1-5, 2013.

[28] Kohei Arai, Genetic Algorithm utilizing image clustering with merge and split processes which allows minimizing Fisher distance between clusters, International Journal of Advanced Research in Artificial Intelligence, 2, 9, 7-13, 2013.

[29] Kohei Arai, Fisher distance based GA clustering taking into account overlapped space among probability density functions of clusters in feature space, International Journal of Advanced Research in Artificial Intelligence, 2, 11, 32-37, 2013.

[30] Ali Ridho Barakbah, Kohei Arai, Centronit: Initial Centroid Designation Algorithm for K-Means Clustering, EMITTER: International Journal of Engineering Technology, 2, 1, 50-62, 2014.

[31] Kohei Arai, Pursuit Reinforcement Competitive Learning: PRCL Based Online Clustering with Tracking Algorithm and Its Application to Image Retrieval, International Journal of Advanced Research on Artificial Intelligence, 5, 9, 9-16, 2016.

AUTHORS' PROFILE

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 also was with National Space Development Agency of Japan from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science on April 1990. He was a councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is a Science Council of Japan Special Member since 2012. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Science Commission "A" of ICSU/COSPAR since 2008 then he is now award committee member of ICSU/COSPAR. He wrote 55 books and published 680 journal papers as well as 450 conference papers. He received 66 of awards including ICSU/COSPAR Vikram Sarabhai Medal in 2016, and Science award of Ministry of Mister of Education of Japan in 2015. He is now Editor-in-Chief of IJACSA and IJISA. http://teagis.ip.is.saga-u.ac.jp/index.html