Analysis of Intergrade Variables In The Fuzzy C-Means And Improved Algorithm Cat Swarm Optimization (FCM-ISO) In Search Segmentation

Jeprolel Saragih1, Opim Salim Sitompul2, and Zakaria Situmorang3

1, 2, 3 Department of Computer Science, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia, 20155

Email: jepoezipper@gmail.com1, opim@usu.ac.id2, zakarias65@yahoo.com3

Abstract. One of the techniques known in Data Mining namely clustering. Image segmentation process does not always represent the actual image which is caused by a combination of algorithms as long as it has not been able to obtain optimal cluster centers. In this research will search for the smallest error with the counting result of a Fuzzy C Means process optimized with Cat swarm Algorithm Optimization that has been developed by adding the weight of the energy in the process of Tracing Mode. So with the parameter can be determined the most optimal cluster centers and most closely with the data will be made the cluster. Weigh inertia in this research, namely: (0.1), (0.2), (0.3), (0.4), (0.5), (0.6), (0.7), (0.8) and (0.9). Then compare the results of each variable values inertia (W) which is different and taken the smallest results. Of this weighting analysis process can acquire the right produce inertia variable cost function the smallest.

1. Introduction

The rapid development of technology makes it easier in data processing. Data that can be processed not only data in the form of text, but can be a multimedia component. In digital image processing, there is an important process that is often used is the process of segmentation. Given the importance of the segmentation process as an initial processor, a segmentation method that can perform accurate object separation [1] is required. Each pixel of a segmented result is usually unattainable that does not conform the actual image. This is due to several factors, namely when shooting is not good. So there are images that blur, lack of light and there is noise. Search for membership value on conform the actual images segmentation is not achieved because of the undersegmentation. In the first case, pixels be owned by to the same object are classified as be owned by different clusters. A single object can be represented by two or more clusters. Or in other cases the opposite happens, the pixels be owned by different clusters are appertained as be owned by the same cluster [2]. For segmentation can be done by clustering, which is clustering part of the way in data mining that is used to enter data into the corresponding group with no deep knowledge of the group [3]. Use of clustering method to explore different groupings of data. The purpose of this clustering is to optimize (the small fitness function of previous research) fitness function specified in the clustering process, which generally attempts to minimize the variation between clusters [4]. Image-based clustering segmentation can be applied to any images as data set. To solve the problem of segmentation image, there are several clustering algorithms that can be used and Fuzzy C-Means one of the popular algorithms used in fuzzy clustering techniques. In his research [5] proposed a new algorithm in optimization techniques that mimic cat
behavior named Cat Swarm Optimization (CSO). This algorithm has a number of advantages in solving optimization problems compared to previous techniques such as Particle-Swarm-Optimization (PSO) or PSO with Weighting Factor.

Initially this algorithm is still used to solve unconstrained minimization problem. The results of the model evaluation developed show that the model developed with CSO produces images with better quality than GA [7]. A study [8] implemented FCM algorithm for clustering with pixel-test data from digital images. The result is better than the K-Means algorithm, but in the experiment still yields a considerable cluster validity of 1.1. And in 2015 conducted research on ICSO [9] In this research the algorithm will yield index of average cluster validity which is small enough, but not optimal (small cluster validity index) according to them. Weighing the value of inertia (w) into the Cat Swarm Optimization method makes ICSO more effective for Clustering problems [10].

2. Literature Review

2.1. Algoritma Clustering (Clustering Algorithm)

This method divides the data under clusters and data that has the same characteristics that are grouped into the same cluster and cluster data has different characteristics are congregate under another group. Basically clustering is a method to search and classify data that have similarity between one data with other data. Clustering is one method of Data Mining which is unsupervised, meaning that this method is applied without any training (training) and without any teacher (teacher) and no need target output. Two kinds of grouping methods used in collecting data contained in Data Mining is namely Hierarchical Clustering and Non Hierarchical Clustering [14].

2.2. Clustering Based Image Segmentation

The Clustering method specifically used to compare the similarity of the value of a pixel to another. Methods in clustering based segmentation include iteration method, K-Means, Fuzzy C-means, Kohonen neural network, and various other clustering techniques. One excellent method used for image segmentation is Fuzzy C-Means clustering. Fuzzy C-Means is an improvised k-means algorithm with Fuzzy Set Theory by applying a measure of membership, in which one pixel image can be owned by multiple clusters [15].

2.3. Fuzzy C-Means Algorithm

Fuzzy C-Means (FCM) clustering is a fuzzy grouping model so that data can be a member of all clusters specified with different membership levels between 0 and 1. The probability of the existence of a data to a cluster is seen from its measure of membership or it can be said to be more likely to cluster closer to the value of the data. The basic concept of FCM, first is to establish the centre point of cluster that will tag the around location for any cluster. In the initial conditions the centre point of the cluster is still not accurate. Each data has a membership measure for each cluster [16]. By repairing the cluster center and membership value of each data repeatedly, it can be found that the point of the cluster will go to the right location. Looping is done to minimize the objective function that describes the distance from the location of the data against the center point of a cluster is given weight by the levels of similarity data point from the set of escapes [7].
3. Research Methodology

The proposed algorithm in Figure 1 can be described in the steps below:

1) Determine the FCM-ICSO parameters: Clustering the number of clusters, maximum iterations, minimum tolerance, seeking memory (SM) and seeking range (SR). The number of clusters is as much as 3 clusters.

2) Generating the initial solution of the cluster center randomly. Calculate the result of Cost Function by using the Xie Beni Index equation.

3) Evaluate the cluster center based on the value of the Cost Function.

4) Generating the new center point of the cluster based on the result of the Cost Function is the minimum Cost Function value at the center of the previous cluster.

5) Update the result of Cost Function. Stop Criteria if iteration has done Cost Cost Function 10 times iteration or cost function passes the minimum limit of 0.001.

6) Weighting Inertial Variables ([0.1], [0.2], [0.3], [0.4], [0.5], [0.6], [0.7], [0.8], [0.9]

7) Update Cluster Center with tracing mode.

8) Criteria to stop if iteration has been doing Cost Function search for 10 times iteration

9) The output of a system is the minimum Cost Function value which is the most optimum cluster center.

4. Experimental Results

4.1 Fuzzy C-Means Performance

The result is all the V vectors that are the center point of the cluster in the initial iteration. With \( m > 1 \) to give weighting to every measure of membership. Then, the Output generated is the value of belonging matrix \( U \) representing the membership function of the observed data. Up to a certain iteration limit or a difference of objective function that is lower than the specified error value. The hope is that every data goes in the right cluster (\(| P_t - P_{t-1} | < \zeta \) or \( t > \max \text{Iterasi} \))

4.2 Testing with Fuzzy C-Means

Testing process on both algorithm use same parameter, that is: smallest error expected \( \xi = 0.001 \), for weight = 2. Test image taken is horse.jpg, flower.jpg and boat.jpg. Then take a random sample of 4x4 pixels and the data set of each pixel using the pixel region image tool on Matlab R2012b as the data attribute that will be used as the data set. Furthermore, \( \mu \) random and normalized random data are generated by dividing the \( \mu \) value by the total \( \mu \) to obtain the results according to \(| P_t - P_{t-1} | < \zeta \) or \( t > \max \text{Iteration} \).

### Table 1. The Centre point of Cluster

| C | j1            | j2            | j3            |
|---|---------------|---------------|---------------|
| 1 | 150,8173073   | 136,970567    | 79,86408755   |
| 2 | 161,6204215   | 153,0618237   | 89,69781211   |
| 3 | 190,7245784   | 92,93978696   | 27,91236388   |
4.3 Inertial Variable Weighted Analysis on Tracing Mode Process
The optimal cluster search center search is performed with some parameters SM = 4, SR = 0.4, CD = {0, 1}, Minimum tolerance = 0.0001, Number of Clusters = 3. The stopping criteria specified earlier is if the cost function ≤ 0.001 or maximum 10 iteration is met. Inertial weights 0.8 produce the smallest cost functions. The explanation of the comparison of the weighting values of variables given in the FCM-ICSO algorithm can be explained in Figure 2.

Figure 1. Cost Function comparison graph with different Inertia (W) weighting

Figure 1 shows the cost value generated at iteration 10, where iteration 10 is the maximum limit of iteration. So from the graph above can be seen that the smallest cost function on multiplication variable inertia (W) with value 0.8. Where the cost function is 0.339254076. The speed of displacement of cluster center in tracing mode is -64.27624303-55.98142603 and -27.38984305. Then each attribute test data is done update the center position of the cluster.

4.6 Assignment of inertia variables to produce smaller cost functions

By multiplying the 0.8 inertial variable in the tracing mode, the smallest error is found. But the researcher searches the lowest error point by narrowing the range of variables that will be multiplied. Because the lowest cost function graph is in the range of 0.7 to 0.8, the researcher narrows down the smallest cost function finding by determining the range of inertia variable values from 0.77 to 0.87.

Figure 2. Cost Comparison Cost Function graph with more optimal Inertia (W) variable value
5. Conclusion

Determination of proper inertia weighting can result in a smaller Cost Function that is 0.334034524. In this study weighing 0.8 (w) on the process of tracing the Cat Swarm Optimization algorithm mode. Based on the graph of inertial multiplication results between 0.7 and 0.9 there is a chart down and rise again. Then tested again find the lowest cost function value with inertia weighting 0.77 to 0.88. There is a smaller cost value on the multiplication of inertia weight of 0.81 with cost function = 0.334034524. The result of the combination of algorithm of FCM-ICSO was successful with both generate value cost smaller function so that more accurate in looking for members of the centre cluster.

6. References

[1] Gonzalez R C and Woods R E 2002. Digital Image Processing, 2nd. (New Jersey: Prentice Hall).
[2] Gonzales R C 2001. Digital Image Processing. Prentice Hall: New Jersey.
[3] Santosa B and Mirsa K N 2009. Isacsoc: A coca. Scientific Research and Essays7(49), pp. 4176-4185.
[4] Berkhin, P. 2009. Survey of Clustering Data Mining Techniques. Grouping Multidimensional Data. pp 25-71.
[5] Chu S-C and Tsai P-W 2007 cibotboc. Ijoiciac31: 163-173.
[6] Liu Y and Shen Y 2010. dwcso. Jocit58.
[7] Halder A, Pramanik and Kar A 2011 Dynamic Image Segmentation using Fuzzy C-Means based Genetic Algorithm. J. Comp Appl 286.
[8] Hadi I and Sabah M 2015 Improvement Cat Swarm Optimization for Efficient Motion Estimation. IJhybinftech81 pp279-294.
[9] Kumal Y and Sahoo G 2015 aicsoafc. CiidmVol1pp 187-197.
[10] Oliver AMunoz X, Batlle J, Pacheco L, and Freixenet J 2006 isoisusr. AQtrieeic2:315–320.
[11] Chuang K-S, Tzeng H-L, Chen S, Wu J and Chen T-J 2006 Fcmwsiise. Cmig309–15.
[12] Oliveira J V D and Pedrycz W 2007 Afc. Bldcpd.
[13] Xie, X. L. & Beni, G. A. 1991. Validity Measure For Fuzzy Clustering. IEEE Trans Pattern Anal MachIntellVol. 13(8): 841-847
[14] Yang M S 1993 Asofc. Mathl. Comput. Modelling Vol1811pp1-16.
[15] Shah, B., Shah, S., & Kosta. 2012. Novel Improved Fuzzy C-Means Algorithm For MR-Image Segmentation. International Journal of Soft Computing and Engineering 2(3):355-356.
[16] Herditomo., Sunaryo., & Naba, A. 2014. Penerapan Metode Hybrid Fuzzy C-Means dan PSO untuk segmentasi geografis. Jurnal Electrics Electronics Communication Controls Informatics Systems8(1): 27-32.