# Improve the Efficiency of Image Segmentation Scheme using Swarm Intelligence Techniques

**Akanksha Garg, Shiv K. Sahu**

***Abstract:*** ***Clustering analysis is a primitive exploratory approach in data analysis with little or no prior knowledge. Clustering has been widely used for data analysis and been an active subject in several research fields such as pattern recognition, information retrieval, data mining applications, bioinformatics and many others. This paper presents a particle of swarm optimization with self-optimal clustering (SOC) technique which is an advanced version of improved mountain clustering (IMC) technique. Proposed POS based SOC clustering techniques for large data. We used the POS for the selection of important parameter such as value of centroid and center, this parameter decides the selection of center point of cluster technique. The SOC clustering technique decides the cluster level wise seed and generates cluster according to their features attribute of data. The experiments also revealed the convergence property of the level fitness in Proposed. We compared our Proposed with existing clustering algorithms and shows that the results are improved.***

***Keywords:*** ***Improved Mountain Clustering, elf Optimal Clustering, Particle swarm optimization, K-means, CRM.***

## I. INTRODUCTION

Clustering is a division of data into groups of similar objects. Each group, called a cluster, consists of objects that are similar to one another and dissimilar to objects of other groups [2]. When representing data with fewer clusters necessarily loses certain fine details (akin to lossy data compression), but achieves simplification. It represents many data objects by few clusters, and hence, it models data by its clusters. Cluster analysis [3] is one of the major data analysis methods which is widely used for many practical applications in emerging areas like Bioinformatics. Clustering is the process of partitioning a given set of objects into disjoint clusters. This is done in such a way that objects in the same cluster are similar while objects belonging to different clusters differ considerably, with respect to their attributes. Data modeling puts clustering in a historical perspective rooted in mathematics, statistics, and numerical analysis. From a machine learning perspective clusters correspond to hidden patterns, the search for clusters is unsupervised learn-ing, and the resulting system represents a data concept. Therefore, clustering is unsupervised learning of a hidden data concept. Data mining applications add to a general picture three complications: (a) large databases, (b) many attributes, (c) attributes of different types. This imposes on a data analysis severe computational requirements. Data mining applications include scientific data exploration, information retrieval, text mining, spatial databases, Web analysis

**Revised Version Manuscript Received on May 16, 2017.**

**Akanksha Garg,** M.Tech Scholar, Department of Computer Science & Engineering, Lakshmi Narain College of Technology Excellence, Bhopal (Madhya Pradesh)-462021, India. E-mail: gargakankshaaug19@gmail.com

**Dr. Shiv K. Sahu,** Professor & Head, Department of Computer Science & Engineering, Lakshmi Narain College of Technology Excellence, Bhopal (Madhya Pradesh)-462021, India. E-mail: shivksahu@rediffmail.com

CRM, marketing, medical diagnostics, computational biology, and many others. Given the apparent difficulty of solving the k-means and other clustering and location problems exactly, it is natural to consider approximation, either through polynomial-time approximation algorithms, which provide guarantees on the quality of their results, or heuristics, which make no guarantees. One of the most popular heuristics for the k-means problem is Lloyd's algorithm [5], which is often called the k-means algorithm. Define the neighborhood of a center point to be the set of data points for which this center is the closest. It is easy to prove that any locally minimal solution must be centroidal, meaning that each center lies at the centroid of its neighborhood. Lloyd's algorithm starts with any feasible solution, and it repeatedly computes the neighborhood of each center and then moves the center to the centroid of its neighborhood, until some convergence criterion is satisfied. It can be shown that Lloyd's algorithm eventually converges to a locally optimal solution [5].

The rest of paper discuss as in section 2 discuss the Partition Relocation Clustering in which divide the data into sets. In section 3 discuss proposed Work which is based on SOC and POS. In section 4 discuss the experimental result and analysis with performance parameters such as GSI, SI, DI and PI. Finally discuss conclusion & future work in section 5.

## II. PARTITIONING RELOCATION CLUSTERING

The data partitioning algorithms, which divide data into several subsets. Because checking all possible subset systems is computationally infeasible, certain greedy heuristics are used in the form of iterative optimization [2]. This means different relocation schemes that iteratively reassign points between the k clusters. Unlike traditional hierarchical methods, in which clusters are not revisited after being constructed, relocation algorithms can gradually improve clusters. With appropriate data, this results in high quality clusters. One approach to data partitioning is to take a conceptual point of view that identifies a cluster with a certain model whose unknown parameters have to be found. More specifically, probabilistic models assume that the data comes from a mixture of several populations whose distributions and priors we want to find. Corresponding algorithms are described in the subsection Probabilistic Clustering. One clear advantage of probabilistic methods is the interpretability of the constructed clusters [4]. Having concise cluster representation also allows inexpensive computation of intra-clusters measures of fit that give rise to a global objective function (see log-likelihood below).

Another approach starts with the definition of objective function depend-ing on a partition. As we have seen (subsection Linkage Metrics), pair-wise distances or similarities can be used to compute measures of inter and intra-cluster relations. In iterative improvement approach such pair-wise computations would be too expensive. For this reason, in this approach a cluster is associated with a unique cluster representative. Now the computation of an objective function becomes linear in N (and in a number of clusters k << N). Depending on how representatives are constructed, partitioning relocation algorithms are subdivided into k-medoids and k-means methods. K-medoid is the most appropriate data point within a cluster that represents it. Representation by k-medoids has two advantages: it presents no limitations on attributes types and the choice of medoids is dictated by the location of a predominant fraction of points inside a cluster and, therefore, it is insensitive to the presence of outliers. In k-means case a cluster is represented by its centroid, which is a mean (usually weighted average) of points within a cluster. This works conveniently only with numerical attributes and can be negatively affected by a single outlier.

## III. PROPOSED ALGORITHM

In this paper proposed the swarm intelligence techniques as Particle Swarm optimization with the self optimal clustering techniques and improved mountain clustering. The self-optimal clustering technique faced a problem of index generation and validation of data control. For the validation of data used swarm based optimization technique. The family of swarm intelligence gives better optimal value of index for the process of cluster generation. In the continuity of chapter discuss the partition clustering, particle of swarm optimization, proposed algorithm and proposed model.

Particle Swarm Optimization (PSO) is a swarm-based intelligence algorithm [7] influenced by the social behavior of animals such as a flock of birds finds a food source or a school of fish protecting them from a predator. A particle in PSO is analogous to a bird or fish flying through a search (problem) space. The movement of each particle is coordinated by a velocity which has both magnitude and direction. Each particle position at any instance of time is influenced by its best position and the position of the best particle in a problem space. The performance of a particle is measured by a fitness value, which is problem specific. The PSO algorithm is similar to other evolutionary algorithms. In PSO, the population is the number of particles in a problem space. Particles are initialized randomly. Each particle will have a fitness value, which will be evaluated by a fitness function to be optimized in each generation. Each Particle knows its best position pbest and the best position so far among the whole gathering of particles gbest. The pbes tof a particle is the best result (fitness value) so far reached by the particle, whereas gbest is the best particle in terms of fitness in an entire population. In each generation the velocity and the position of particles will be updated as in Eq4.1 and 4.2, respectively. The heuristic optimizes the cost of task-resource mapping based on the solution given by particle swarm optimization technique.

$$v_i^{K+1} = \omega v_i^k + c1rand1 \times \left(pbseti - x_i^k\right) + c2rand2 \times \left(gbest - x_i^k\right) \dots (4.1)$$

$$x_i^{K+i} = x_i^k + v_i^{k+1} \dots \dots (4.2)$$

The clustering process to partition X into k cluster with weights for both views and individual variables is modeled as minimization of the following objective function

$$p(U,Z,V,W) = \sum_{l=1}^{k}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j\in Gi}^{1} uiwtvjd(xij,zij) + n\sum_{j=1}^{m} vjlog(vj) + \lambda\sum_{t=1}^{T} wtlog(wt) \dots \dots (1)$$

Subject to

$$\begin{cases} \sum_{l=1}^{k} ui.\, l = 1, ui, l \in (0,1), 1 \le i \le n \\ \sum_{t=1}^{T} wt = 1, 0 \le wt \le 1, \quad \dots..(2) \\ \sum_{j\in Gi} vj = 1, 0 \le vj \le 1, 1 \le t \le T, \end{cases}$$

Where

U is a n×k portion matrix whose element ui,l are binary where ui,l=1 indicates that object I is allocated to cluster l;
Z={Z1,Z2,…….,Zk} is a set of k vectors representing the centers of the k clusters
W={W1,W2,……..Wt} are T weight for T view
V={v1,v2,……vm} are m weight for m variable
d(xij,zlj) is a distance or dissimilarity measure on the jth variable between the ith object and the center of the lth cluster. if the variable is numerical , then
$d(xij ,zlj)=(xij-zlj)^2$…….(3)
if the variable is categorical, then
d(xij,zlj)={0 (xi.j=zl.j)……..(4)
1 (xi.j=/zl.j)

The first term in (1) is the sum of the within cluster dispersions, the second and third terms are two negative weight entropies. two positive parameters are control the strength of cluster.

## IV. EXPERIMENTAL RESULT ANALYSIS

In this paper we perform experimental process of modified SOC with POS. The proposed method implements in matlab 7.14.0 and tested with very reputed data set from UCI machine learning research center. In the research work, we have measured GSI, PI, SI, DI and Elapsed time rate. To evaluate these performance parameters I have used five datasets from UCI machine learning repository [27] namely Iris, Glass identification, Diabetes, Cleveland and Ecoli data set.


Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

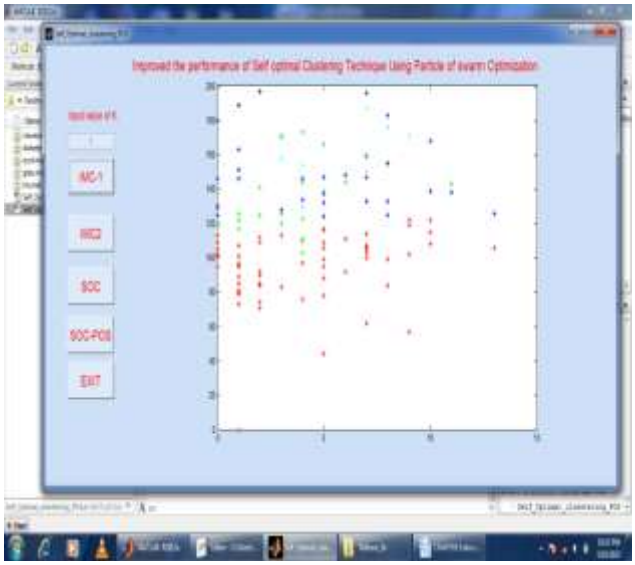

**Figure 1: Shows that the Diabetes dataset with method IMC-1 using for the input value of k is 1.**

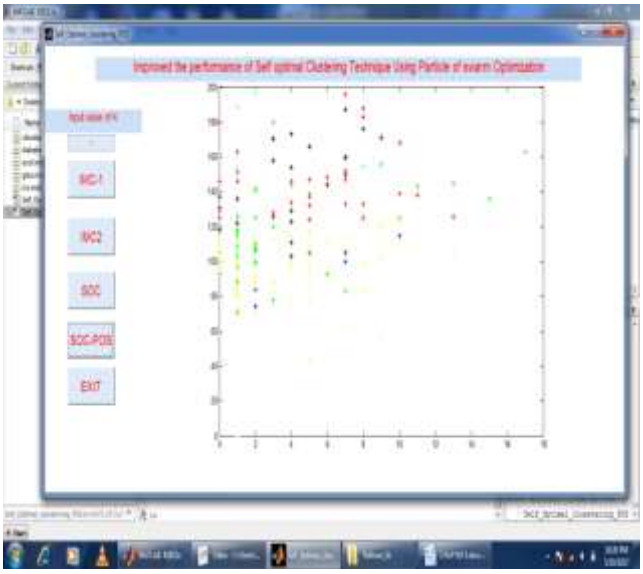

**Figure 2: Shows that the Diabetes dataset with method SOC-POS using for the input value of k is 1.**

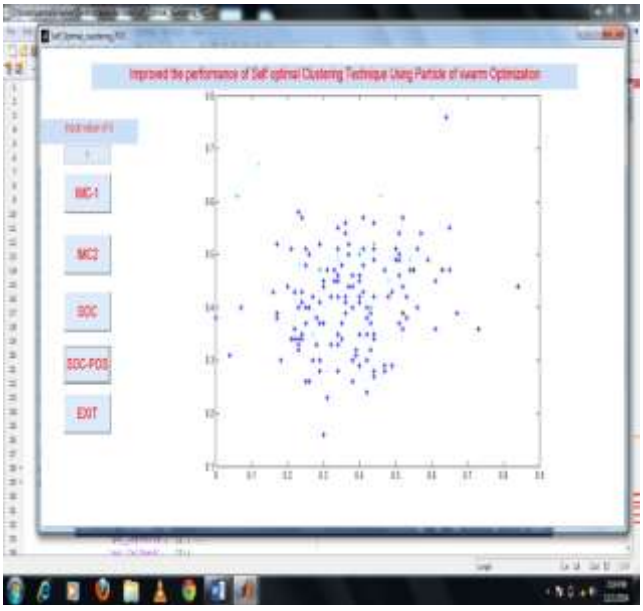

**Figure 3: Shows that the Cleveland dataset with method SOC-POS using for the input value of k is 5.**

**Table 1: Shows that the performance evaluation for all clustering techniques with the input value is 2, for the Cleveland dataset.**

| Cluster Value | Result Techniques | GSI | PI | SI | DI |
|---|---|---|---|---|---|
| 2 | IMC-1 | 1.34 | 0.87 | 0.17 | 0.15 |
| | IMC-2 | 1.36 | 0.90 | 0.18 | 0.16 |
| | SOC | 1.38 | 0.94 | 0.20 | 0.17 |
| | SOC-POS | 1.40 | 1.03 | 0.26 | 0.23 |

**Table 2: Shows that the performance evaluation for all clustering techniques with the input value is 3, for the Diabetes dataset.**

| Cluster Value | Result Techniques | GSI | PI | Elapsed TIME |
|---|---|---|---|---|
| 3 | IMC-1 | 3.740 | 2.163 | 26.144 |
| | IMC-2 | 3.760 | 2.978 | 38.058 |
| | SOC | 3.780 | 2.901 | 26.042 |
| | SOC-POS | 3.800 | 2.387 | 29.144 |

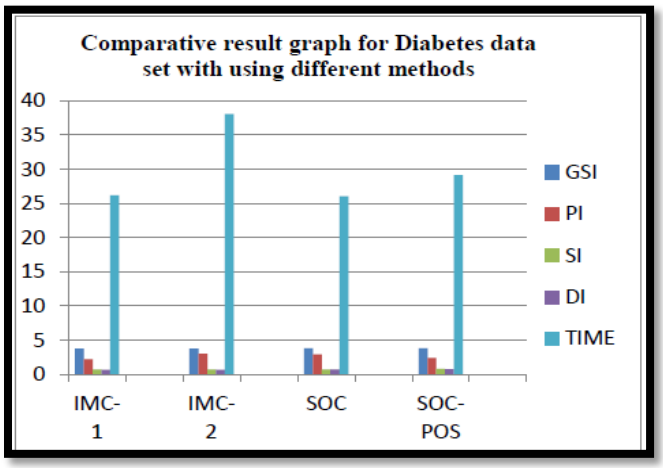


**Figure 4: Shows that the comparative result for diabetes dataset using clustering techniques with the input value is 3.**

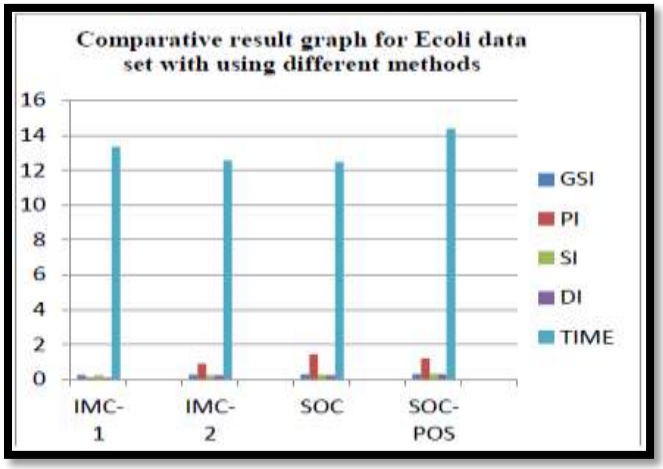


**Figure 5: Shows that the comparative result for Ecoli dataset using clustering techniques with the input value is 5.**


*Published By:*
*Blue Eyes Intelligence Engineering*
*& Sciences Publication*

## V. CONCLUSION AND FUTURE WORK

In this paper the Proposed algorithm can compute fitness for level and individual variables simultaneously in the clustering process. With the two types of fitness, compact level and important variables can be identified and effect of low-quality level and noise variables can be reduced. Therefore, Proposed can obtain better clustering results than individual variable weighting clustering algorithms from multi-level data. We used four real-life data sets to investigate the properties of two types of fitness in Proposed. We discussed the difference of the fitness between Proposed and SOC, IMC based algorithms. Our empirical result shows that our proposed algorithm shows better result in comparison of IMC and SOC algorithm. The POS algorithm takes more time for the selection of estimated value of P. The values of M influence the cluster quality during level of data. In future we used optimization technique for self-selection of optimal cluster for large data.

## REFERENCES


1. Nishchal K. Verma, Abhishek Roy "Self-Optimal Clustering Technique Using Optimized Threshold Function" IEEE SYSTEMS JOURNAL, IEEE 2013. Pp 1-14.
2. Pavel Berkhin "A Survey of Clustering Data Mining Techniques" Pp 1-59.
3. K. A. Abdul Nazeer, M. P. Sebastian "Improving the Accuracy and Efficiency of the k-means Clustering Algorithm" WCE 2009. Pp 1-6.
4. Hae-Sang Park, Chi-Hyuck Jun "A simple and fast algorithm for K-medoids clustering" Expert Systems with Applications, 2009. Pp 3336–3341.
5. Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, Angela Y. Wu "A local search approximation algorithm for k-means clustering" Elsevier B.V. All rights reserved, 2004. Pp 89-112.
6. LUO Xin "Chinese Text Classification Based on Particle Swarm Optimization" 4th National Conference on Electrical, Electronics and Computer Engineering, NCEECE 2015, Pp 53-59.
7. Ramachandra Rao Kurada, Dr. K Karteeka Pavan, Dr. AV Dattareya Rao "A Preliminary Survey On Optimized Multiobjective Metaheuristic Methods For Data Clustering Using Evolutionary Approaches" International Journal of Computer Science & Information Technology (IJCSIT) Vol 5, No 5, October 2013. Pp 58-78.
8. Nishchal K. Verma, Payal Gupta, Pooja Agrawal and Yan Cui "MRI Brain Image Segmentation for Spotting Tumors Using Improved Mountain Clustering Approach" 2011.
9. N. K. Verma, P. Gupta, P. Agarwal, M. Hanmandlu, S. Vasikarla, and Y. Cui, "Medical image segmentation using improved mountain clustering approach," in Proc. 6th Int. Conf. ITNG, Las Vegas, NV, USA, 2009, pp. 1307–1312.
10. Rui Xu, and Donald Wunsch "Survey of Clustering Algorithms" IEEE Transactions On Neural Networks, VOL. 16, NO. 3, MAY 2005. Pp 645-678.
11. Yixin Chen, James Z. Wang, and Robert Krovetz "CLUE: Cluster-Based Retrieval of Images by Unsupervised Learning" IEEE transactions on image processing, vol. 14, no. 8, august 2005. Pp 1187-1201.
12. N. K. Verma, A. Roy, and S. Gupta, "Color segmentation using improved mountain clustering technique version-2," in Proc. 2nd IEEE Int. Conf. Intell. Human Comput. Interact., Allahabad, India, 2011, Pp 536–542.
13. Jennifer Erxleben, Kelly Elder and Robert Davis "Comparison of spatial interpolation methods for estimating snow distribution in the Colorado Rocky Mountains" Hydrol. Process. 2002. Pp 3627–3649.
14. Singh Vijendra, Kelkar Ashwini, Sahoo Laxman, "An effective clustering algorithm for data mining", Proc. of the 2010 International Conference on Data Storage and Data Engineering, pp.250–253, 2010.
15. Yu Jin, Qian Feng, Qi Rongbin, "Improvement of stochastic particle swarm optimization by succession strategy", Communications of the Systemics and Informatics World Network, Vol.3, pp.155–159, 2008.
16. N.R. Pal, K. Pal, J.C. Bezdek et al., "A possibilistic fuzzy C-Means clustering algorithm", IEEE Trans. Fuzzy Systems, Vol.13, No.4, pp.517–530, 2005.
17. Lv Zehua, Jin Hai, Yuan Pingpeng, Zou Deqing, "A fuzzy clustering algorithm for interval-valued data based on Gauss distribution functions", Acta Electronica Sinica, Vol.38, No.2, pp.295–300, 2010.
18. C.L. Sun, J.C. Zeng, J.S. Pan "An improved vector particle swarm optimization for constrained optimization problems", Information Scieces, Vol. 181, 2011. Pp. 1153–1163.
19. Mathew, Juby, and R. Vijayakumar. "Scalableparallel clustering approach for large data usinggenetic possibilistic fuzzy c-means algorithm", 2014 IEEE International Conference on Computational Intelligence and Computing Research,2014.
20. RM Suresh, K Dinakaran, P Valarmathie, "Model based modified k-means clustering for microarray data", International Conference on Information Management and Engineering, Vol.13, pp 271-273, 2009, IEEE.
21. C. Escudero "Classification of Gene Expression Profiles: Comparison of k-means and expectation maximization algorithms", IEEE Computer Society, 2008, pp. 831-836.
22. Manpreet Kaur, Usvir Kaur " A Survey on Clustering Principles with K-Means clustering Algorithms Using Different Methods in Detail" International Journal of Computer Science and Mobile Computing, Vol-2, 2013. Pp 327-331.
23. K. Kameshwaran, K. Malarvizhi "Survey on Clustering Techniques in Data Mining" IJCSIT: International Journal of Computer Science and Information Technologies, Vol-5, 2014. Pp 2272-2276.
24. S. Suganya, Rose Margaret "Image Segmentation Using Two Weighted Variable Fuzzy K Means" International Journal of Computer Applications Technology and Research Volume 2, 2013. Pp 270-276.
25. Geng Li, Stephan Gunnemann, Mohammed J. Zaki "Stochastic Subspace Search for Top-K Multi-View Clustering" ACM, 2013. Pp 1-6.
26. Xiaowen Dong, Pascal Frossard, Pierre Vandergheynst ,Nikolai Nefedov "Clustering on Multi-Layer Graphs via Subspace Analysis on Grassmann Manifolds" 2013. Pp 1-13.
27. https://archive.ics.uci.edu/ml/datasets.html




*Published By:*
*Blue Eyes Intelligence Engineering*
*& Sciences Publication*