

Clustering of Data Using Particle Swarm Optimization and Fuzzy C-Means for Dataset

Mukul Pandey¹, Ekta Thaokar², Priti Choudhary³, Pratiksha Telang⁴

1(Department of Computer Engineering, SRPCE, Nagpur

2 (Department of Computer engineering, SRPCE, Nagpur

3 (Department of Computer engineering, SRPCE, Nagpur

4 (Department of Computer engineering, SRPCE, Nagpur

ABSTRACT:

Clustering is the process of dividing the data points into a number of groups such that data points in the same groups are same or unique characteristic and other with different characteristic. In simple words, the aim is to separate groups with similar attribute, property and assign them into clusters. Clustering is used to represent the large datasets by a fewer number of prototype. Particle swarm optimization (PSO) algorithm is a new population artificial intelligence algorithm and has good performance on optimization. Fuzzy c-means (FCM) is a data clustering technique in which a dataset is grouped into n clusters with every data point in the dataset belonging to every cluster to a certain degree in terms of 0 and 1. For example, a certain data point that lies close to the centre of a cluster will have a high degree of belonging or membership to that cluster and another data point that lies far away from the centre of a cluster will have a low degree of belonging or membership to that cluster. The data clustering are used in both algorithm and find out which algorithm give better result. The Fuzzy C-Means clustering algorithms are also used for the data clustering. In this paper we find the outcome result of fuzzy c-mean algorithm at a similar dataset and compare these results to the PSO clustering algorithm.

Keyword: -Particle Swarm Optimization(PSO), Fuzzy C-Means Clustering

I INTRODUCTION

Particle Swarm Optimization (PSO) is a populace based imaginary enhancement system created by Dr.Eberhart and Dr. Kennedy [1] in 1995, enlivened by social conduct of winged creature running or fish tutoring. The framework is introduced with populace of arbitrary arrangement and scans for optima by refreshing ages. The PSO likewise decides how to refresh the speed of a molecule. All particles refreshes its new speed in view of current speed at which it move and the best position it has investigated and furthermore in view of the worldwide best position assessed by swarm [2]. The PSO calculation depends on manmade brainpower which sets aside less opportunity to tackle advancement issue, the Streamlining issue is composed as condition for effectively fathom by PSO. The PSO procedure at that point is iterated a settled number of times or until the point that a base blunder in light of wanted execution record is accomplished. It has been demonstrated that this straightforward model can manage troublesome enhancement issues proficiently. The PSO is a manmade brainpower calculation which is utilized to tackle the perplexing issue, in this paper we utilize the dataset and plot the protest capacity of that information which is process with the help of PSO Algorithm for discovering the ideal outcomes on the dataset. The PSO calculation was utilized as a part of numerous looks into on the bunching subject. The essential

PSO calculation, which limits a target work $f(x)$ of a variable vector x defined on n dimensional space, utilizes a swarm of m particles. Every molecule 'I' of the swarm is related with a position [5] in a nonstop n -dimensional hunt space. Correspondingly, the speed is likewise a n -dimensional vector. Indicating with x_k^I 'I' and v_k^I 'I' separately the position and speed of molecule 'I' at cycle k of the PSO calculation.

III METHODOLOGY

A. FUZZY C-MEANS CLUSTERING (FCM)

Fuzzy C-mean (FCM) is a calculation which is utilized for bunching which enables one bit of information to have a place with at least two bunches. This strategy (created by Dunn in 1973 and enhanced by Bezdek in 1981) is as often as possible utilized as a part of example acknowledgment. Fuzzy C-Means (FCM) is an information grouping system in which a dataset is gathered into n bunches with each datum point in the dataset having a place with each bunch of an information to a specific degree. It begins with an underlying conjecture for the group focuses, which are expected to check the mean area of each bunch. FCM allocates each datum point a participation review for each

group of these dataset. By iteratively refreshing the bunch focuses and the enrollment grades for every datum point, this calculation (FCM) iteratively moves the group focuses to the correct area inside an informational index. The all cycles is really in light of limiting a target work that speaks to the separation from any given information point to a bunch focus weighted by that information point's participation review. Let consider as iris dataset which is examine with Fuzzy c-mean (FCN) calculation.

1. LOAD DATA:-

Any kinds of information is stacked which is as .dat file. The dataset is acquired from the information record 'iris.dat'. This dataset was gathered and contains arbitrary number of tests of blossoms having a place with three sorts of iris blooms which is setosa, versicolor, and virginica. For every one of the kinds of blossoms, 50 perceptions for sepal length, sepal width, petal length, and petal width are recorded. The information is stacking utilizing Fuzzy and subsequent to stacking the information plot these information in the dimensional shape which is either 2D and 3D, however in this paper we utilize the 2 Dimensional plotting.

2. PLOT DATA IN 2-D:-

The information to be grouped is 4-dimensional that is in this information there are four elements are available in the information and speaks to sepal length, sepal width, petal length, and petal width. From every one of the three gatherings two qualities (for instance, sepal length versus sepal width) of the blossoms are plotted in a 2-dimensional plot. In the wake of plotting the information there is have to setup the parameters which imply that distinguish the quantity of bunches, parcel framework and discover the criteria for the base and the greatest estimation of emphases.

3. PARAMETERS SETUP:-

For Fuzzy c-mean grouping calculation the parameters required are number of bunches, the estimation of segment lattice, emphases ought to be greatest and least number of change are characterized and set. The quantity of figuring bunches are distinguish the setup parameters, in which iterative process are done and create beginning groups to process the most extreme achieved cycle for finding the exact esteem.

3. COMPUTE CLUSTERS:

Fuzzy C-Means (FCM) clustering is an iterative procedure. In this procedure first beginning fluffy parcel framework is created. In each progression of the cycle, the focuses of the group and the participation review point are refreshed and the target work is limited to locate the best area for the bunches. The procedure stops when the most extreme number of emphases is come to, or when the target work change between two back to back cycles is not as much as the base measure of change determined.

B. PARTICLE SWARMOPTIMIZATION(PSO)

The PSO calculation essentially utilizes the target capacity to assess its competitor arrangements, and works upon the resultant wellness esteems. particles are move with a specific speed to discover the ideal arrangement of the issue, in this paper the PSO calculation are utilized as the Fitness work where the wellness Function are utilized as a target capacity of The PSO calculation. With the assistance of these wellness work there are fitting a bend as indicated by the Datasets. In spite of the fact that the wellness work is utilized for wellness esteem estimation in PSO, it is additionally utilized for dissecting the magnificence of the bunches framed.

The molecule Swarm streamlining gives diverse incentive at an alternate time for a specific dataset in this paper we consider the iris dataset on which the information bunching are performed. Based on the trial results delivered by analysts Shi and Eberhart, the molecule swarm improvement gives two esteems as result which is neighbourhood best and worldwide best.

Amid the inquiry procedure, the molecule progressively changes its situation as per two highlights, in particular their own best position and the worldwide best position. The c1 and c2 esteems are utilized to deal with the P1 and P2 esteems. Each molecule is refreshed by P1 and P2 amid its cycles. The speed and area of ith element changes utilizing to the accompanying conditions:

$$V = w * v + c1 * rand1 * (p1 - x) + c2 * rand2 * (p2 - x)$$

The wellness work utilized as a part of the PSO is given as takes after:-

$$\text{Min } f(x) = 100 * (x(1) ^2 - x(2)) ^2 + (1 - x(1)) ^2$$

IV EXPERIMENTAL RESULTS

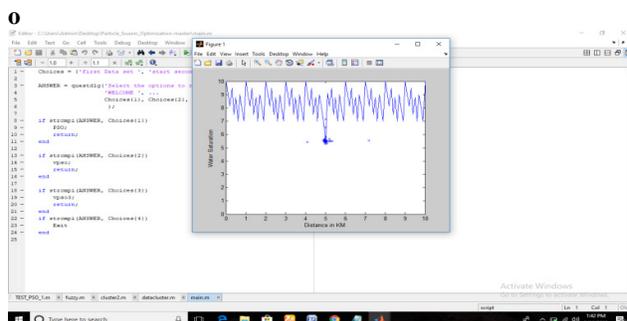


Fig 1: Output after applying on dataset using PSO

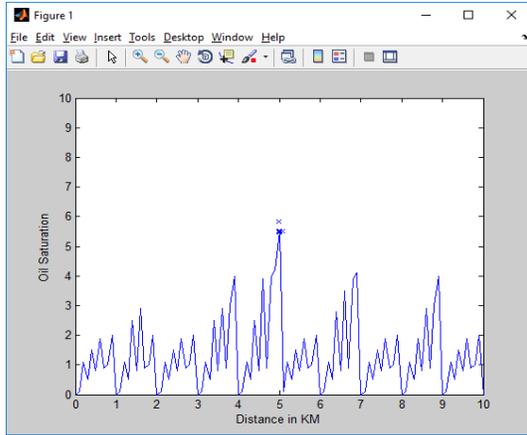


Fig 2: Output after applying on dataset using PSO

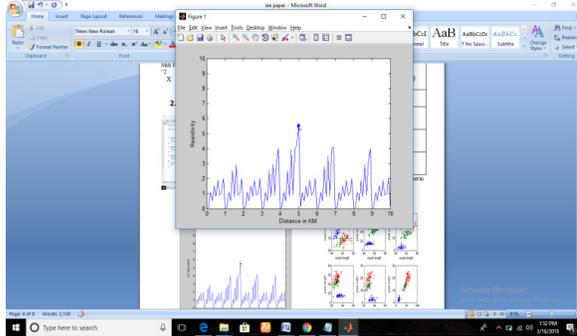


Fig 3: Output after applying on dataset using PSO

Here a data is plotted according to the fitness function. And as we know that if existence of oil there is four factors is responsible which are: distance, Resistivity, water saturation and oil saturation. Where **Water saturation α 1/resistivity** and the **oil saturation α Resistivity**. Oil has the high resistivity.

But now we take different dataset and analyze what time required for each algorithm for every dataset.

Sr.no.	Datasets	PSO execution time (insec)	FCM execution time (in sec)
1.	Iris	07	02
2.	Dataset1	09	03

3.	Dataset2	05	01
4.	Dataset3	8.5	2.5

Table: -Time required (in second) for Dataset in FCM and PSO algorithm.

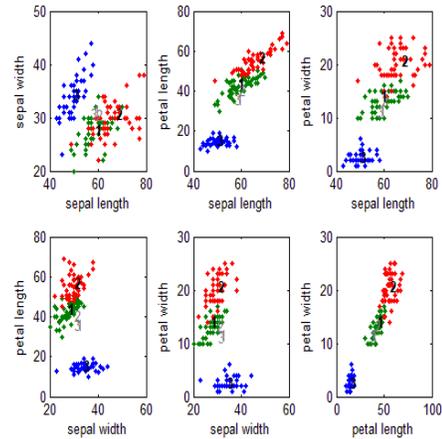


Fig4:Output after applying FCM on Iris dataset.

V CONCLUSION

In this work, two clustering algorithms are used on same dataset and compare the result of each of the algorithm that outcome of PSO with the result of Fuzzy C-Mean clustering algorithm after an experimental result we find that the PSO algorithm is slower as compared to the fuzzy c-mean clustering algorithm. In the Fuzzy C-Mean clustering algorithm we use the iris dataset and the four parameters of the iris dataset are considered and converted into 2D. With respect of time the Fuzzy C-Means solve the problem in very few seconds as compared to the particle swarm optimization. As shown in table 1 FCM takes less time than PSO.

REFERENCES

- [1]. Optimization based Routing Protocols in Mobile Ad-Hoc Networks". International Journal of Computer Applications (0975 – 8887) Volume 119 No.10, June 2015
- [2]. DONG Chaojun, QIU Zulian," Particle Swarm Optimization Algorithm Based on the Idea of Simulated Annealing(2006) "IJCSNS International Journal of Computer Science and Network Security, VOL.6

- [3]. Jerome E. Onwunalu, Louis J. Durlofsky, "Application of a particle swarm optimization algorithm for determining optimum well location and type". *Compute Geosci* (2010).
- [4]. P Jaganathan, S Jaiganesh ,A Particle Swarm Optimization based fuzzy c means approach for efficient web document clustering,P Jaganathan et.al / *International Journal of Engineering and Technology (IJET)*.
- [5]. Thomas A. Runkler, Christina Katz. Fuzzy clustering by particle swarm optimization [C]. 2006 *IEEE International Conference on Fuzzy Systems*. 2006: 601-608.
- [6]. Cai WL, Chen SC, Zhang DQ. Fast and robust fuzzy cmeans clustering algorithms incorporating local information for image segmentation. *Pattern Recognition*, 2007, 40(3): 825-833.
- [7]. PU PB, WANG G, LIU TA. Research of improved fuzzy c-means algorithm based on particle swarm optimization. *Computer Engineering and Design*, 2008, 29(16): 42774279
- [8]. QiangNiu , Xinjian Huang , An Improved Fuzzy C-means Clustering Algorithmbased on PSO,*JOURNAL OF SOFTWARE*, VOL. 6, NO. 5, MAY 2011
- [9]. P Jaganathan , S Jaiganesh ,A Particle Swarm Optimization based fuzzy c means approach for efficient web document clustering,*International Journal of Engineering and Technology (IJET)*
- [10]. Optimization based Routing Protocols in Mobile Ad-Hoc Networks". *International Journal of Computer Applications* (0975 – 8887) Volume 119 No.10, June 2015
- [11]. DONG Chaojun, QIU Zulian," Particle Swarm Optimization Algorithm Based on the Idea of Simulated Annealing(2006) "*IJCSNS International Journal of Computer Science and Network Security*, VOL.6
- [12]. Jerome E. Onwunalu, Louis J. Durlofsky, "Application of a particle swarm optimization algorithm for determining optimum well location and type". *Compute Geosci* (2010).
- [13]. P Jaganathan, S Jaiganesh ,A Particle Swarm Optimization based fuzzy c means approach for efficient web document clustering,P Jaganathan et.al / *International Journal of Engineering and Technology (IJET)*.
- [14]. Thomas A. Runkler, Christina Katz. Fuzzy clustering by particle swarm optimization [C]. 2006 *IEEE International Conference on Fuzzy Systems*. 2006: 601-608.
- [15]. Cai WL, Chen SC, Zhang DQ. Fast and robust fuzzy cmeans clustering algorithms incorporating local information for image segmentation. *Pattern Recognition*, 2007, 40(3): 825-833.
- [16]. PU PB, WANG G, LIU TA. Research of improved fuzzy c-means algorithm based on particle swarm optimization. *Computer Engineering and Design*, 2008, 29(16): 42774279
- [17]. QiangNiu , Xinjian Huang , An Improved Fuzzy C-means Clustering Algorithmbased on PSO,*JOURNAL OF SOFTWARE*, VOL. 6, NO. 5, MAY 2011
- [18]. P Jaganathan , S Jaiganesh ,A Particle Swarm Optimization based fuzzy c means approach for efficient web document clustering,*International Journal of Engineering and Technology (IJET)*