A COMPARISION ON FUZZY C MEAN ALGORITHM AND PARTICLE SWARM OPTIMIZATION

Mr.Jitendrasinh Raulji, Mr.Dharmesh Dhangar
Head of Department,Lecturer
Department of Computer Engineering
Parul Polytechnic Institute, Vadodara , India

Abstract : Data mining provides facility that can help to finds useful patterns from large size of data. And if we talk about clustering, it is important task of data mining. Clustering is a process of grouping similar objects into a group. Object in same cluster are more similar to each other than the objects in different clusters. Fuzzy c-means (FCM) algorithm is also an important clustering technique. But the Fuzzy c-mean is noises sensitive and is easily struck at local minima. The possibilistic c-means (PCM) algorithm solves the noise sensitivity problem of FCM algorithm. The Membership function is used by the Possibilistic c-means algorithm to illustrate the degree of belonging. The component produced by the PCM communicates to a dense region in the data set. All clusters are independent of the other clusters in the PCM approach. Particle swarm optimization is a heuristic global optimization method and also an optimization algorithm, which is based on swarm intelligence. Here, In this paper I tried to compare outcomes of PCM and PSO with some dataset.

***Index Terms* - Clustering, Fuzzy c-mean, PCM, PSO, Data mining, data clustering,, Possibilistic c-means, Fuzzy possibilistic c-means.**

I. INTRODUCTION

Possibilistic fuzzy c-means algorithm : The Possibilistic fuzzy c-means algorithm uses a possibilistic type of membership function to illustrate the degree of belonging. It is advantageous that the memberships for representation feature points be as high as possible and unrepresentative points have low membership. The component produced by the PCM communicates to a dense region in the data set. All clusters are independent of the other clusters in the PCM approach.

Minimize the objective function

$$J_{PCM} = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}{}^{m} d_{ij}{}^{2} + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (1 - u_{ij})^{m}$$

The possibilistic fuzzy c-means algorithm is given below:

Step_1. Given data object X, fix C -,c,2≤c≤n,m>1,η>1 and initialize the membership function values Uij(0), 1≤i≤c;1≤j≤n, at step t, t = 0, 1, 2,…. Tmax

Step_2 Compute the cluster centers.

$$v_i^{(t)} = \frac{\sum_{j=1}^{n} u_{ij}{}^{m} x_j}{\sum_{j=1}^{n} u_{ij}{}^{m}} \quad \text{for } 1 \le i \le c$$

Step_3 Compute Euclidian distance

$$(d_{ij})^2 = ||x_j - v_i||^2 \quad \text{for } 1 \le i \le c;\ 1 \le j \le n$$

Step_4 Calculate the new u(t+1) to satisfy

$$U_{ij}^{(t+1)} = \frac{1}{1 + \left(\frac{d^2_{ij}}{\eta_i}\right)^{\frac{1}{m-1}}}$$

$$\text{for } 1 \leq i \leq c; 1 \leq j \leq n$$

Where ηi is calculated by the Eq.

Step_5 If ‖U(t+1)−U(t)‖≤C -, then stop; otherwise t = t+ 1 and return to step 2.

Particle swarm optimization: Particle swarm optimization algorithm consists of "n" particles, and each particle position stands for the potential solution in D-dimensional space. There are main three principles in which particles change its: I. to keep its inertia II. to change the condition according to its most optimist position III. to change the condition according to the swarm's most optimist position. Particle position is affected both by individual experience and near experience. When the whole particle swarm is surrounding the particle; this algorithm is called the whole PSO. If the narrow surrounding is used in the algorithm, than it is called the partial PSO. Each particle can be shown by its current speed and position, the most optimist position of each individual and the most optimist position of the surrounding. In the partial PSO, the speed and position of each particle change according the following equality.

k+1 id v = k id v +c1 k r1 (k pbestid - k id x)+ c2 k r2 (k gbestd - k id x)
x = k id x + k+1 id v

In this equality, k id v and k id x stand for separately the speed of the particle "i" at its "k" times and the d-dimension quantity of its position; k pbestid represents the d-dimension quantity of the individual "i" at its most optimist position at its "k" times. k gbestd is the d-dimension quantity of the swarm at its most optimist position. The speed of the particle created at its each direction is confined between -vdmax, and vdmax. If the number of vdmax is too big, the solution is far from the best, if the number of vdmax is too small, the solution will be the local optimism; c1 and c2 represent the speeding figure, regulating the length when flying to the most particle of the whole swarm and to the most optimist individual particle. If the figure is too small, the particle is probably far away from the target field, if the figure is too big, the particle will maybe fly to the target field suddenly or fly beyond the target field. The proper figures for c1 and c2 can control the speed of the particle's flying and the solution will not be the partial optimism. Usually, c1 is equal to c2 and they are equal to 2; r1 and r2 represent random fiction, and 0-1 is a random number. In local PSO, instead of persuading the optimist particle of the swarm, each particle will pursuit the optimist particle in its surrounding to regulate its speed and position. Formally, the formula for the speed and the position of the particle is completely identical to the one in the whole PSO.

II. Study Area

The major thrust area is Data Mining. Data mining commonly involves four classes of tasks or techniques Classification, Clustering, Association Rule Mining, and Regression. Among these all the Clustering is a task of assigning a set of objects in to groups called clusters. Here, we work on clustering to improve and justifying comparisons of two different algorithms..

III. Aim and Objective of the Study

This study aims to compares two different algorithm named Possibilistic fuzzy c-means algorithm and Particle swarm optimization algorithm. Here, we have try to analyzed some outcomes of both algorithm with different dataset.

IV. Result

Instances (n, c, d)	PCM		
	Worst	Average	Best
Iris (150, 3, 4)	70.89	68.41	68.89
Cancer (683, 2, 9)	2175.7	2123.3	2256.8
Wine (178, 3, 13)	11182.3	11789.6	11472.8

Instances (n, c, d)	FPSO		
	Worst	Average	Best
Iris (150, 3, 4)	68.62	68.30	68.22
Cancer (683, 2, 9)	2631.2	2694.7	2694.5
Wine (178, 3, 13)	11230.3	11628.5	12147.0

V. Acknowledgment

With the cooperation of my guide, I am highly indebted to Asst. Prof. ANKUR N SHAH, for his valuable guidance and supervision regarding my topic as well as for providing necessary information

REFERENCES

[1] Bassam M. El-Zaghmouri, Marwan A. Abu-Zanona“ Fuzzy C-Mean Clustering Algorithm Modification and Adaptation for Applications ”,World of Computer Science and Information Technology Journal (WCSIT),year 2012.
[2] Introduction to machine learning. Cambridge: MIT Press Bezdek, J. (1974). Fuzzy mathematics in pattern classification. Ph.D. thesis. Ithaca, NY:Cornell University Alpaydin, E. (2004).
[3] Ming-Chuan Hung and Don-Lin Yang ” An Efficient Fuzzy C-Means Clustering Algorithm”, IEEE,2001.
[4] Tejwant Singh, Mr. Manish Mahajan,” Performance Comparison of Fuzzy C Means with Respect to Other Clustering Algorithm”, IJARCSSE ,VOLUME 4,2014.
[5] Deepak Kumar Niware Dr. Setu Kumar Chaturvedi“Web Usage Mining through Efficient Genetic FuzzyCMeans”,(IJCSIS) International Journal of Computer Science and Information Security,Vol. 11, No. 9,year 2013.
[6] Rui Xu, Student Member, IEEE and Donald Wunsch II, Fellow, IEEE” Survey of Clustering Algorithms” IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 16, NO. 3,2005..
[7] Timothy C. Havens, Senior Member, IEEE, James C. Bezdek, Life Fellow,IEEE, Christopher Leckie,” Fuzzy c-Means Algorithms for Very Large Data”, IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 20, NO. 6,2012.