

Comparative Study of Particle Swarm Optimization and Fuzzy C-Means to Data Clustering

Harveen Anand
Dept. of Computer Science and Engineering
Vidyalankar Institute Of Engineering, Wadala
India

Vipul Dalal
Dept. of Computer Science and Engineering
Vidyalankar Institute of Engineering, Wadala
India

Abstract: Data has an important role in all aspects of human life and so analyzing this data for discovering proper knowledge is important. Data mining refers to find useful information (extracting patterns or knowledge) from large amount of data. Clustering is an important data mining technique which aims to divide the data objects into meaningful groups called as clusters. It is the process of grouping objects into clusters such that objects from the same cluster are similar and objects from different clusters is dissimilar. In data mining, data clustering has been studied for long time using different algorithms and everyday trends are proposed for better outcomes in this area. Particle swarm optimization is an evolutionary computational technique which finds optimum solution in many applications. Fuzzy C- means (FCM) algorithm is a popular algorithm in field of fuzzy clustering. In this paper, we present a comparative study of Particle swarm optimization and FCM to data clustering.

Keywords: Clustering, Particle Swarm Optimization, FCM, fuzzy clustering, swarm

1. INTRODUCTION

Data plays an important role in all aspects of human life and so analysing this data, extracting useful information from the data, and transforming it into understandable way is important. This is the main goal of data mining process. Clustering is one of major tasks of knowledge discovery from databases. Data clustering is the process of grouping together similar multi-dimensional data vectors into a number of clusters or bins. Data clustering is the organizing of data into specific groups, the characteristics of which are unknown. Clustering is a class of unsupervised classification method and has been applied to a wide range of problems, including exploratory data analysis, data mining, image segmentation, pattern recognition, machine learning and other diverse fields.

Data clustering has been studied for a long time by researchers with different methods. A good clustering method should produce high quality clusters. The quality of the clustering method depends on 3 components: the distance measurement used, the clustering algorithm itself (its implementation) and the ability to discover some or all the hidden patterns. Generally, clustering algorithms can be categorized into hierarchical methods, partitioning methods, Density-based methods, grid-based methods, model-based methods, constraint-based methods, frequent pattern based methods. The hierarchical clustering techniques create a hierarchical decomposition of the database .It merges some clusters in order to make a bigger cluster or divide a cluster into some clusters to make small clusters [1].On the other hand,, the partition clustering techniques partition the database into predefined number of clusters. They attempt to determine 'k' partitions that optimize a certain criterion function.

Recently, bio-inspired algorithms like particle swarm optimization (PSO), ant colony optimization (ACO) etc have found success in solving clustering problems [2].Clustering techniques based on bio-inspired algorithms have outperformed many traditional methods. This paper

focuses on comparative study of Particle Swarm Optimization algorithm and Fuzzy C-Means to data clustering.

2. FUZZY C-MEANS CLUSTERING

Fuzzy clustering is a powerful unsupervised method for the analysis of data and construction of models. It is an extension of K-Means algorithm [3].Objects on the boundaries between several classes are not forced to fully belong to one of the classes, but rather are assigned membership degrees between 0 and 1 indicating their partial membership[3].Fuzzy C-means (FCM) algorithm is most widely used algorithm for fuzzy clustering [3]. Bezdek [4] introduced Fuzzy C-means clustering method in 1981; extend from Hard C-mean clustering method [5]. Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. FCM employs fuzzy partitioning such that a data point can belong to all groups with different membership grades between 0 and 1[3].FCM is widely applied in agricultural engineering, astronomy, chemistry, geology, image analysis, medical diagnosis, shape analysis and target recognition[6].The working of fuzzy c-means algorithm is discussed using following steps.

Algorithm:

1. Initialize $U=[u_{ij}]$ matrix, $U^{(0)}$
2. At k-step: calculate the centers vectors $C^{(k)}=[c_j]$ with $U^{(k)}$

$$c_j = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m}$$

3. Update $U^{(k)}, U^{(k+1)}$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left[\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right]^{2/m-1}}$$

4. If $\|U^{(k+1)} - U^{(k)}\| < \epsilon$ then STOP; otherwise return to step 2.

Here m is any real number greater than 1,
 u_{ij} is the degree of membership of x_i in the cluster j ,
 x_i is the i th of d -dimensional measured data,
 c_j is the d -dimension center of the cluster,

This algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center. Clearly, summation of membership of each data point should be equal to one. After each iteration, membership and cluster centers are updated according to the formula.

3. PARTICLE SWARM OPTIMIZATION

Particle swarm optimization algorithm (PSO) is an evolutionary computational technique based on the movement and intelligence of swarms. PSO is a population based global optimization technique proposed by Kennedy and Eberhart [7]. The algorithm uses a number of agents (particles) that constitute a swarm moving in the search space looking for the best solution. In the context of PSO, a swarm refers to a number of potential solutions to the optimization problem, where each potential solution is referred to as a particle. Each particle has two movement characteristics, velocity and location. Each particle's movement is influenced by its local best known position and is also guided toward the best known positions in the search-space, which are updated as better positions are found by other particles. This is expected to move the swarm toward the best solutions [8].

Equation (1) and Equation (2) are the velocity and location update functions. In each generation, all particles are updated using these equations until the maximum generation count reaches or a particle gets to our termination criteria. In all cases, the global best reached so far, is the solution of our problem.

$$V_i^{k+1} = wV_i^k + c_1 \text{rand}_1(\text{pbest}_i - x_i^k) + c_2 \text{rand}_2(\text{gbest} - x_i^k) \quad --(1)$$

$$x_i^{k+1} = x_i^k + V_i^{k+1} \quad --(2)$$

Here, rand_1 and rand_2 are uniformly generated random numbers in the range [0, 1]. Acceleration constants c_1 and c_2 and inertia weight w are predefined by the user. V_i^k is the current velocity and x_i^k is the current location.

4. PSO CLUSTERING

In context of clustering, a single particle represents N cluster centroid vectors. That is, each particle X_i is constructed as follows:

$$X_i = (m_{i1}, m_{i2}, \dots, m_{iN_c})$$

where m_{ij} refers to the j -th cluster centroid vector of the i -th particle in cluster C_{ij} . [9]. Therefore, a swarm represents a

number of candidate clustering for the current data vectors. The aim of PSO is to find the particle position that results in best evaluation of a given fitness function. The particle swarm optimization based clustering algorithm is as follows[1]:

Algorithm:

1. Initialize each particle with random cluster centroid
 2. Do
 - a. For each Particle
 - i. For Each Data Vector Z_p in Data set
 - ii. Calculate $D(Z_p, m_{ij})$ for all cluster centroids.
 - iii. Assign Z_p to a Cluster C_i according to this condition
 - b. $D(Z_p, m_{ij}) = \min_{r=1,2,\dots,N_c} \{d(Z_p, m_{ir})\}$
 - c. Calculate Fitness of Particle.
 - d. Update the Global Best ($gbest$) and Local Best ($lbest$).
 - e. Update the Particle Velocity and Location using Equation (1) and (2).
3. While Termination Criteria is satisfied.

$$J_c = \frac{\sum_{i=1}^{N_c} \left[\frac{\sum_{r=Z_p=C_{ij}} d(Z_p, m_{ir})}{|C_{ij}|} \right]}{N_c}$$

5. ADVANTAGES AND DISADVANTAGES OF PSO AND FCM

Advantages of PSO

1. PSO algorithm is based on intelligence and it is applied on both scientific research and engineering [10].
2. PSO accepts the real number code and that is decided directly by the solution. Calculation in PSO is simpler and efficient in global search [11].
3. PSO algorithm has no mutation and overlapping calculation. The search can take place by the speed of the particle. Most of optimist particle can be able to transmit the information onto other particles during the development of several generations, and the speed of researching is faster[11].
4. It is insensitive to initial conditions [10].

Disadvantages of PSO

1. Low Convergence [10].
2. Weak local Search ability.
3. PSO cannot work on the problem of non co-ordinate system like solution of energy field

and moving rules for the particles in the energy field [11].

4. The method easily suffers from the partial optimism, which causes the less exact at the regulation of its speed and direction [11].

Advantages of FCM

1. Converges [3].
2. Unsupervised [3].
3. Easy to implement.

Disadvantages of FCM

1. Depends on cluster centre initialization [3].
2. Sensitivity to the initial guess (speed, local minima) [3].
3. Sensitivity to noise and one expects low (or even no) membership degree for the outliers (noisy points).

6. CONCLUSION AND FUTURE WORK

From the study of PSO and FCM clustering we found that FCM which depends on initial condition may cause the algorithm to converge to suboptimal solution while on other hand, PSO is less sensitive to initial condition due to its suboptimal population-based nature. So particle swarm optimization is more likely to find near optimal solution. During the study, it was found that in future research, more improvement can be done in area of initialization of algorithm, improved fitness function, combining the advantages with other clustering algorithms to improve the efficiency and performance of clustering.

7. REFERENCES

- [1] Ehsan Toreini, Maryam Mehrnejad "Clustering data using particle swarm optimization using a new fitness" Proceedings IEEE International Conference, 2011.
- [2] A.A.Esmin, D.L.pereira, F.P.A. De Araujo, "Study of different approach to clustering data by using particle swarm optimization algorithm" 2008 IEEE Congress on Evolutionary Computation (CEC 2008).
- [3] R.Suganya, R.Shanthi" Fuzzy C- Means Algorithm- A Review" International Journal of Scientific and Research Publications, Volume 2, Issue 11, November 2012 1 ISSN 2250-3153.
- [4] J.C.Bedzek, "Pattern recognition with Fuzzy Objective Function Algorithm" ,New York, Plenum Press, 1981.
- [5] S. Ghosh, S. K. Dubey, "Comparative analysis of K-Means and Fuzzy C-Means" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No.4, 2013.
- [6] Y. Yong, Z. Chongxun and L. Pan, "A Novel Fuzzy C-Means Clustering Algorithm for Image Thresholding", Measurement Science Review, vol. 4, no.1, 2004.
- [7] J. Kennedy and RC Eberhart, "Particle Swarm Optimization," Proc.IEEE International Joint Conference on Neural Networks, vol. 4, pp. 1942-1948, 1995.
- [8] Yallamati Prakasarao, Arumalla Nagaraju, A.Veerawamy, "Experimental of Data Clustering using particle swarm optimization", IJESR/Aug 2012/vol-2/Issue-8/Article No-16/864-870.
- [9] Van der Merwe, D.W. and AP.Engelbrecht, "Data Clustering using particle swarm optimization", in Evolutionary Computation, 2003.
- [10] Pritesh Vora, Bhavesh Oza" A Survey on K-Means and Particle Swarm Optimization" (IJISME) 2013, pp 24-26.
- [11] Qinghai B," The Analysis of Particle Swarm Optimization Algorithm", in CCSE, February 2010, vol.3.
- [12] Sunita Sarkar, Arindam Roy, Bipul S." Application of Particle Swarm Optimization in Data Clustering", IJCA (0975-8887), Vol 65-No 25, March 2013.
- [13] Han J: Kamber, M (2001)."Data Mining: Concepts and Techniques",Morgan Kaufmann, San Francisco.
- [14] Jain,A.K; Murty,M N Flynn,P.J.(1999),"Data Clustering: a review" ACM Computing Survey 31(3):264-323.