K-Medoids Clustering Technique using Bat Algorithm

Monica Sood

Assistant Professor
Lovely Professional University,
Punjab, India

Shilpi Bansal

M.tech (IT)
Lovely Professional University,
Punjab, India

ABSTRACT

Clustering is one of the data analysis methods that are widely used in data mining. In this method, we partitioned the data into different subset which is known as cluster. Cluster analysis is the data reduction toll for classifying a "mountain" of information into manageable meaningful piles. This method is vast research area in the field of data mining. In this paper, a partitioning clustering method that is K-Medoids algorithm is used with Bat algorithm. We proposed a new algorithm based on the echolocation behaviour of bats to know the initial value to overcome the K-Medoids issues. In this algorithm, we can find the initial representative object easily with the help of using Bat algorithm. They provide us better cluster analysis and we can achieve efficiency. This paper introduces the combination of K-Medoids clustering algorithm and Bat Algorithm. In this paper we show the difference between K-Medoid Clustering Technique with Bat Algorithm & K-medoid itself.

Keywords

Swarm Intelligence, Bat Algorithm, Clustering, K-Medoids Clustering Technique

1. INTRODUCTION

Swarm Intelligence is an innovative distributed intelligent paradigm or the collective behaviour in a decentralized system which is done by the groupings of individuals which are interacting locally with one another. This system often found in nature, such examples are ant colonies, population of bats, fish schooling, honey bees, spiders, and particles present in cloud, herding and so many more. These insects have intelligent group behaviour. We can merge this method with clustering due to its inconvenient implementation. Bees, ants, bats are the insects that do big tasks although they are not of high intelligence. So we can able to manage the large data sets of individuals who interact with other through communication to produce a global behaviour. SI derives from the nature inspired form nature. Bat Algorithm is a swarm based method that is used in this paper for optimal solution combining with the Clustering method for large datasets. Clustering is one of the unsupervised learning method in which we partitioned the data objects into subsets which is known as cluster. A greedy approach for partitioning clustering, which is used in the most of application, is K-Medoids algorithm for improving the quality of cluster. In this method, the prototype is often medoid, we can say as a representative object or point of a cluster. Therefore, identifying the initial representative object selection is very important topic in clustering. The bat inspired algorithm (BA) was used for overcome such a problem of selection of representative object.

2. CONCEPTUAL DEFINITION AND DETAILS

2.1 BAT ALGORITHM

Bat algorithm is swarm intelligence based algorithm which is worked on the echolocation of bats. This algorithm was developed by Xin-She Yang in 2010. It is a new metaheuristic algorithm for solving the many optimization problems. Bats are based on the echolocation behaviour of bats. They can find their prey o food and also they can know the different type of insects even in a complete darkness. Since we know that microbats are the insectivore who have the quality of fascinating. These bats use a type of sonar namely as echolocation. They emit a loud sound pulse and detect a echo that is comes back from their surrounding objects. Their pulse varying in properties and will be depend on the species. Their loudness is also varying. When they are searching for their prey, their loudness is loudest if they are far away from the prey and they will become slow when they are nearer to the prey. Now for emission and detection of echo which are generated by them, they use time delay. And this time delay is between their two ears and the loudness variation of echoes.

The following formulae are used for their position xi and velocities vi when they are updated:

$$f_i = f_{min} + (f_{max} - f_{min})\beta, \qquad ...(2.1.1)$$
$$v_{ti} = v_{t-1i} + (x_{ti} - x*)f_i, \qquad ...(2.1.2)$$
$$x_{ti} = x_{t-1i} + v_{ti}, \qquad ...(2.1.3)$$

We know that loudness is decreases when bat found its prey or food but the rate of pulse emission increases. For simplicity we use loudness A0 = 1 and minimum Amin = 0 that means a bat found their prey and they stop making sound.

where $\beta \in [0, 1]$ is a random vector drawn from a uniform distribution. Here **x**∗ is the current global best location (solution) which is located after comparing all the solutions among all the *n* bats. Generally speaking, depending on the domain size of the

problem of interest, the frequency *f* is assigned to *f*min = 0 and *f*max = 100 in practical implementat ion. Initially, each bat is randomly given a frequency which is drawn uniformly from [*f*min, *f*max]. For the local search part, once a solution is selected among the current best solutions, a new solution for each bat is generated locally using random walk.

$$X_{new} = x_{old} + \varepsilon A_t \qquad ...(2.1.4)$$

The update of the velocities and positions of bats have some similarity to the procedure in the standard particle swarm optimization as *f*i in essence controls the pace and range of the movement of the swarming particles. To some degree, BA can be considered as a balanced combination of the standard

particle swarm optimization and the intensive local search controlled by the loudness and pulse rate. Furthermore, the loudness *Ai* and the rate *ri* of pulse emission update accordingly as the iterations proceed

2.2 K-MEDOIDS CLUSTERING ALGORITHM

K-Medoids is the most popular clustering algorithm in which a data set of n object is clustered with k number of cluster, which is given by the user. This algorithm works on the principle of minimizing the dissimilarities between each & every object

in a cluster and its representative object. Initially, we have to choose the representative object arbitrarily and this object is known as Medoid. Find a single partition of the data into K clusters such that each cluster has a most representative point, i.e., a point that is the most "centrally" located point in the cluster with respect to some measure, e.g., distance. These representative points are called Medoids.

3. METHODOLOGY

In this paper, we proposed a new algorithm for optimized cluster analysis. We use K- Medoids clustering algorithm with Bat Algorithm. The scope of this method is that for giving a better efficiency and cluster centre. We already know that it is difficult to choosing the initial representative object in the K-Medoids clustering algorithm. But we can overcome this issue by using Bat Algorithm by initializing the location of cluster. We will introduce a new technique in this paper.

3.1 Problem Formulation

In this paper, we combine the K-Medoids and Bat Algorithm. Since we know that in K-Medoids clustering Algorithm, initially we have to assign a representative object arbitrarily but it is difficult to choose this value arbitrarily. This is issue occurred in the K-Medoids method. In Bat Algorithm, we can optimize the solution for better efficiency. So for finding the representative object in a data set in K-Medoids clustering technique, we apply Bat Algorithm to this technique. In this paper we included two methods and combing these methods for better result in clustering and efficiency. These two parameters are used in this paper. The proposed algorithm is given in as:

- Initializing the objective function f(x)
- Initialize the population of bat xi where i=1,2,3....n and velocity vi
- Set the maximum iteration max_iter
- Set time t=1
- Define pulse frequency fi for bats at each position
- Define the pulse rate that is emitted by the bat ri and the loudness Ai
- While(t < max_iter)
- Generate the new solutions by adjusting frequencies and update velocity and location using formulae (equation (1) to (3))
  - o Initialize k as initial representative object
  - o Assign each xi to its nearest representative object with the most similar based on the medoids of the objects in a cluster
  - o Select a non representative object randomly Orandom
  - o Calculate the cost S
  - o If S < 0
  - o Swap the Oi with Orandom
  - o Else
  - o Go to step b
  - o If (rand > ri)
- Select a solution from the best solutions
- Generate a local solution
- End if
- Generate a new solution by the flying of bats ramdomly
- If (rand <Ai)
- Accept the new solution
- Increment the pulse rate and decrement the loudness
- End if
- Rank the bats and find the bet population in bat x*
- Increase the t
- End while
- Result will be given

4. RESULTS

If we have small dataset then there is no issue occurs in choosing medoid in a cluster. But if there is large dataset then we face a problem in choosing a cluster point. For example if we have 3 clusters set in a small data such as set of 12 inputs data then there we easily find the medoid by using K- medoid clustering technique only. But if we have a large randomly datasets then there is issue for choosing the medoid in a cluster so to solve this problem we use k-medoid clustering technique with bat algorithm which is nature inspired algorithm that give a better optimal solution such as if we have a dataset of 50 randomly inputs and we have to make 4 clusters then we use K-medoid clustering technique with Bat Algorithm to make the segregate solution. So the proposed algorithm is implemented in MATLAB 2010b. MATLAB stands for matrix laboratory. MATLAB is a fourth generation programming language which is used to calculate scientific and engineering calculations to solve problems. MATLAB is also used for automatic control, signal processing and solving linear equations.

In this section we described what will be the use by using these two different Algorithms. They give an Optimal Solution in a large dataset. This combination makes a hybrid approach which is very useful in the data mining as well as in swarm intelligence for the optimization. They are better or we say a clear way to present the data objects present in data sets. Through these objectives, we know the use of this hybrid approach for the clustering as well as achieving better efficiency.

Fig. 1.1 depicts that if we have small dataset as shown in figure then we easily choose the cluster point either by using K-medoid clustering Technique with Bat Algorithm or by using K-medoid only. As in figure the maroon square and green diamond both are show the same cluster point. So this

figure depicts the cluster point either using the proposed algorithm or using K-medoid only.
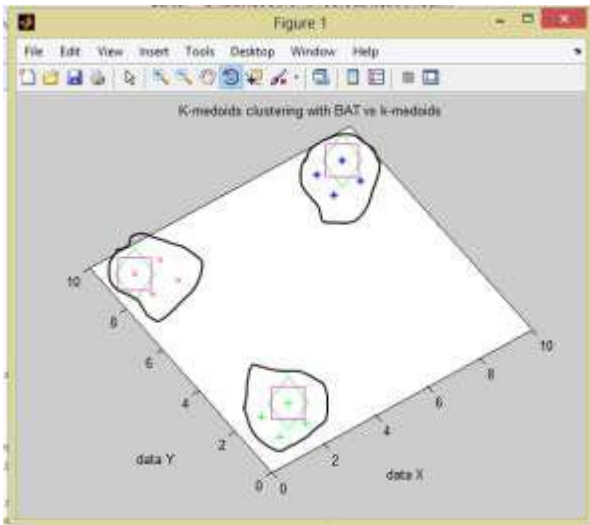

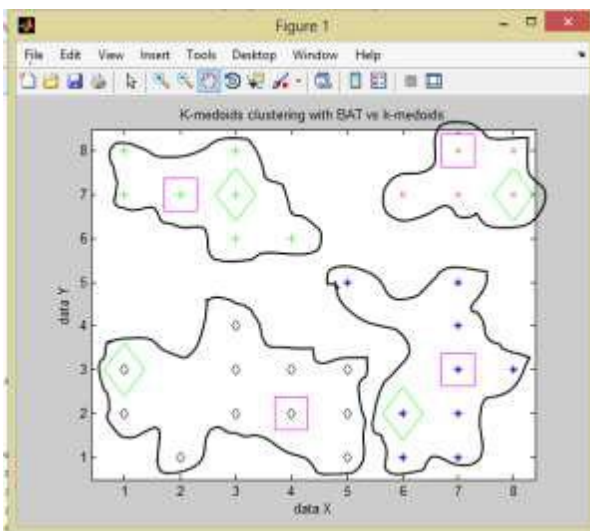**Fig 1.1 No. of cluster is 3 with small dataset**

Fig. 1.2 Number of cluster is 4 with randomly large dataset

This figure shows the proposed algorithm. In this figure we give the comparison between K-medoid clustering with BAT and K-Medoid clustering Technique only. The maroon square gives the cluster point by using K-medoid clustering

technique with Bat Algorithm and the green diamond is for K-medoid clustering only.

5. CONCLUSION

Swarm intelligence has the capability to recover path with minimum complexity. In this section we described what will be the use by using these two different Algorithms. They give an Optimal Solution in a large dataset. This combination makes a hybrid approach which is very useful in the data mining as well as in swarm intelligence for the optimization. They are better or we say a clear way to present the data objects present in data sets. Through these objectives, we know the use of this hybrid approach for the clustering as well as achieving better efficiency.

6. REFERENCES

[1] G. Komarasamy, Amitabh Vahi: A Optimized K-Means Clustering Technique using Bat Algorithm, Bannari Amman Institute of Technology (2012)

[2] Xin-She Yang: A new metaheuristic Bat inspired Algorithm. Studies in Computational Intelligence, Springer (2010) ;loki

[3] Xin-She Yang: Bat Algorithm for Multi- Objective Optimization, Department of Engineering, University of Cambridge (2011)

[4] Raghuvira Pratap A, K survarna Vani, J Rama Devi, Dr. K Nageswara Rao: An Efficient Density based Improved KMedoids Clustering Algorithm, IJACSA (International Journal of Advanced Computer Science and Application), (2011)(Raghuvira Pratap A, 2011)

[5] Shalini S Singh, NC Chauhan: K-Means v/s K-Medoids: A Comparative Study, National Conference on recent trends in Engineering And Technology (Shalini Singh,2011)

[6] Peng Jin, Yun-Long-Zhu, Kun-Yuan-Hu: A Clustering Algorithm for Data Mining Based on Swarm Intelligence, Shenyang Institute of Automation of the Chinese Academy of Sciences, International Conference, china (2007)

[7] Yue-jiao Gong, Rui-tian XU and Jun Zhang: A Clustering-based Adaptive Parameter Control Method for Continuous Ant Colony Optimization, Department of Computer Science, China (2009)