Abstract

In data mining, K-means clustering is well known for its efficiency in clustering large data sets. The main aim in grouping data points into clusters is to lump similar items together in the same cluster such that objects lying in one cluster should be as close as possible to each other (homogeneity) and objects lying in different clusters are further apart from each other.

However, there exist some flaws in classical K-means clustering algorithm. First, the algorithm is sensitive in selecting initial centroids and can be easily trapped at a local minimum with regards to the measurement (the sum of squared errors). Secondly, the KM problem in terms of finding a global minimal sum of the squared errors is NP-hard even when the number of the clusters is equal to 2 or the number of attributes for data point is 2, so finding the optimal clustering is believed to be computationally intractable.

In this dissertation, KM clustering problem is solved by optimized KM. The proposed algorithm is named as BPSO in which the issue of how to derive an optimization model for the minimum
sum of squared errors for a given data set is considered. Two evolutionary optimization
algorithms BFO and PSO are combined to optimize KM algorithm to guarantee that the result of
clustering is more accurate than clustering by basic KM algorithm. F-measure is used to do
comparison of both basic K-means and BPSO algorithm.

References

1. Nikhil Kushwaha, Vimal Singh Bisht, Gautam Shah, “Genetic Algorithm based Bacterial
Foraging Approach for Optimization”, International Journal of Computer Applications (IJCA),
2012.

2. Tarun Kumar Sharma, Millie Pant “Improved Swarm Bee Algorithm for Global
Optimization”, International Journal of Computer Applications (IJCA), International Conference
on Recent Advances and Future Trends in Information Technology (iRAFIT2012).


4. Yang Yong, “The Research of Imbalanced Data Set of Sample Sampling Method Based
on K-Means Cluster and Genetic Algorithm”, International Conference on Future Electrical

5. Youguo Li, Haiyan Wu, “A Clustering Method Based on K-Means Algorithm”, International

6. Sunita Sarkar, Arindam Roy and Bipul Shyam Purkayastha, “Application of Particle
Swarm Optimization in Data Clustering: A Survey”, International Journal of Computer

Algorithm and its Applications to Solve Simultaneous Equations”, International Journal of

8. Hlaudi Daniel Masethe, Mosima Anna Masethe,” Prediction of Heart Disease using
Classification Algorithms”, Proceedings of the World Congress on Engineering and Computer
Science, 2014 Vol. II.


and Pharmaceutical Research, 2014.

11. Poonam Sehrawat, Manju,” Association Rule Mining Using Particle Swarm
Optimization”, International Journal of Innovations & Advancement in Computer Science,

12. Sanjay Tiwari, Mahinder Kumar Rao,” Optimization In Association Rule Mining Using
Distance Weight Vector And Genetic Algorithm” International Journal of Advanced Technology
& Engineering Research (IJATER), Volume 4, Issue 1, Jan. 2014.

13. P.Kalyani,” Medical Data Set Analysis Ñ A Enhanced Clustering Approach” International
January-February 2014.

Mining Techniques”, International Journal of Computer Applications, Volume 97– No.13, July
2014.

15. Sandeep U. Mane, Pankaj G. Gaikwad,” Hybrid Particle Swarm Optimization (HPSO) for


**Index Terms**

Computer Science  
Artificial Intelligence

**Keywords**

PSO (Particle Swarm Optimization), BFO (Bacterial Foraging Optimization), KDD (Knowledge Discovery in Databases), BPSO (Bacterial Particle Swarm Optimization), KM (K-Means) etc.