Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 125 (2018) 780–789

Procedia Computer Science

www.elsevier.com/locate/procedia

6th International Conference on Smart Computing and Communications, ICSCC 2017, 7-8 December 2017, Kurukshetra, India

A Novel Heuristic for Evolutionary Clustering

Pranav Nerurkar*,[a], Archana Shirke[b], Madhav Chandane[c], Sunil Bhirud[d]

[a]*Dept. of Computer Engineering & IT, VJTI, Mumbai - 400019, India*
[b]*Dept. of Computer Engineering & IT, VJTI, Mumbai - 400019, India*
[c]*Dept. of Computer Engineering & IT, VJTI, Mumbai - 400019, India*
[d]*Dept. of Computer Engineering & IT, VJTI, Mumbai - 400019, India*

Abstract

Clustering is considered a challenging problem of data mining due to its unsupervised nature. The literature is inundated with algorithms and concepts related to determining the most suitable clustering structure in data. These techniques have a mathematical model of a cluster and attempt to obtain a result that shall represent this model as closely as possible. However as the problem of clustering is NP hard such strategies have disadvantages such as converging to local optima or suffering from the curse of dimensionality. In such scenario, meta heuristics could be more suitable strategies. Such techniques utilizes biologically inspired techniques such as swarm intelligence, evolution etc. to traverse the search space. Due to their inherent parallel nature, they are most robust towards converging to a local optima. The objective (cost) function used by such meta heuristics is responsible for guiding the agents of the swarm towards the best solution. Hence it should be designed to achieve trade-off between multiple objectives and constraints and at the same time produce relevant clustering. In this paper, a cost function is proposed (PSO-2) to produce compact well separated clusters by using the concept of intra-cluster and inter-cluster distances. Experiments have been performed on artificial benchmark data-sets where performance of the particle swarm optimizer using the proposed cost function is evaluated against other evolutionary and non evolutionary algorithms. The clustering structures produced by the methods have been evaluated using distance based and internal cluster validation metrics to demonstrate that the performance of PSO-2 is comparable to other techniques.

Peer-review under responsibility of the scientific committee of the 6th International Conference on Smart Computing and Communications

Keywords: Data clustering; Applications of evolutionary algorithms; Genetic algorithm

* Corresponding author. Tel.: +91-961-999-7797.
E-mail address: pranavn91@gmail.com

1877-0509
Peer-review under responsibility of the scientific committee of the 6th International Conference on Smart Computing and Communications
10.1016/j.procs.2017.12.100