Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 92 (2016) 442 – 449

Procedia Computer Science

2nd International Conference on Intelligent Computing, Communication & Convergence

(ICCC-2016)

Srikanta Patnaik, Editor in Chief

Conference Organized by Interscience Institute of Management and Technology

Bhubaneswar, Odisha, India

Empirical Study on Clustering Based on Modified Teaching Learning Based Optimization

Sushanta Kumar Panigrahi[a], Sabyasachi Pattnaik[b]*

[a] *Interscience Institute of Management & Technology, Bhubaneswar, Odisha, India*
[b] *Fakir Mohan University, Bhubaneswar, Odisha, India*

Abstract

In this Paper the focus is given on data clustering using Modified Teaching–Learning Based Optimization (MTLBO) a hybridization technique of TLBO. Unlike TLBO, this population based method works on the effect of influence of a teacher on learners to find the optimum solution and it has been used for clustering. The motivation behind the data clustering is to find inherent structure in the data items and grouping then on the basis of their mutual similarity. The effectiveness of the method is tested on many benchmark problems with different characteristics and the results are compared with other population based methods and finally it is implemented on clustering using neural network in data mining.

Peer-review under responsibility of the Organizing Committee of ICCC 2016

Keywords: Artificial neural network; evolutionary algorithm; particle Swarm Optimization; teaching learning based optimization; clustering.

* Corresponding author. Tel.: 91-9861261763
*E-mail address:*ctcsushanta@gmail.com

Peer-review under responsibility of the Organizing Committee of ICCC 2016
doi:10.1016/j.procs.2016.07.338

1. Introduction

Data mining is the field of research whose core exists at the intersection of statistics, machine learning, and databases. In data mining several tasks like classification, association rule mining, clustering, regression, summarization etc .are embedded with. Each of these tasks can be viewed as a type of problem to be solved by a data mining technique. In this work the focus is on data clustering.

The motivation behind the data clustering is to find inherent structure (similarity) in the data items and grouping then on the basis of their mutual similarity. A good clustering is one that achieves- High within-cluster similarity and Low inter-cluster similarity [2]. In other words Similarity among the same cluster should be high as compared to the data objects among different clusters [1]. Similarity measurement is a very important concern in data clustering. It is inversely related to distance.

Clustering technique is used to partition unlabeled scattered data set into groups of similar objects known as clusters. Usually the clusters are different from each other. Unsupervised algorithms are mainly known as clustering algorithms. Clustering techniques can be classified into types such as hierarchical and partitional. The hierarchical clustering is classified into agglomerative and divisive. In hierarchical clustering n objects will be grouped into k clusters by minimizing some measure of dissimilarity in each group and maximizing the dissimilarity of different groups [2:9].

In this paper the focus is on partitional clustering, and in particular the K-means algorithm that is one of the most efficient clustering algorithms. However, the K-means algorithm suffers from drawbacks like many local optima, because it is not convex and it heavily depends on the initial solutions [2:4].

Clustering process starts with randomly generated initial centroids and keeps reassigning the data objects various clusters based on the similarity between the data object and the cluster centroids until a termination criteria is met (e.g., the fixed number of iterations or stability in movement of data points among clusters) [4]. K-Means is the most efficient algorithm in terms of the execution time but it has a drawback that the cluster results are extremely sensitive to the selection of the initial cluster centroids and may converge to the local optimal solution [10, 11]. Bad initialization leads to bad clustering and poor convergence speed. Therefore, the initial selection of the cluster centroids decides the main processing of K-Means and the clustering result of the dataset as well. Considering these limitation, it has been proposed to use meta-optimization to improve the processing capabilities of existing clustering algorithms. Meta-optimization is an approach which allows using the combination of two or more than two algorithms to achieve a common goal. In current scenario, it will be good to utilize any global optimal searching algorithm for generating the initial cluster centroids for K-Means [2:4]. Recently many algorithms have been developed based on evolutionary algorithms like Genetic Algorithm (GA), Tabu Search (TS), Particle Swarm Optimization (PSO) and Simulated Annealing (SA) [21:24]. But the disadvantage is that most of these evolutionary algorithms are very slow to get the optimal solution.

This work presents the improvised Teaching–Learning Based Optimization (TLBO) termed as Modified Teaching–Learning Based Optimization (MTLBO**)** a hybridization technique of TLBO with evolutionary system, such as Adaptive Particle Swarm Optimization (APSO) [26, 27]. Unlike TLBO, this population based method works on the effect of influence of a teacher on learners to find the optimum solution and it has been used for clustering.

2. Cluster analysis

Clustering analysis that is an NP-complete problem to find groups in heterogeneous data by minimizing dissimilarity measures is one of the fundamental tools in data mining, machine learning and pattern classification solutions [12:20]. Clustering in N-dimensional Euclidean space R^N is the process of partitioning a given set of n points into K groups (or, clusters) based on some similarity (distance) metric that is Euclidean distance, which derived from the Minkowski metric (equations 1 and 2).

$$d(x, y) = \left(\sum_{i=1}^{m} | x_i - y_j |^r \right)^{1/r} \quad (1)$$