Abstract

Clustering is an unsupervised learning task where one seeks to identify a finite set of categories termed clusters to describe the data. The proposed system, try to exploit computational power from the multicore processors by modifying the design on existing algorithms and software. However, the existing clustering algorithms either handle different data types with inefficiency in handling large data or handle large data with limitations in considering numeric attributes. Hence, parallel clustering has come into picture to provide crucial contribution towards clustering large data. In this paper a scalable parallel clustering algorithm called Possibilistic Fuzzy C-Means (PFCM) clustering to cluster large data is introduced. In order to harvest the full power of a multi-core processor the software application must be able to execute tasks in parallel utilizing all available CPUs. To achieve this aim, it use fork/join method in java programming. It is the most effective design techniques for obtaining good parallel performance. The experimental analysis will be carried out to evaluate the feasibility of the scalable Possibilistic Fuzzy C-Means (PFCM) clustering approach. The experimental analysis showed that the proposed approach obtained upper head over existing method in terms of accuracy, classification error percentage and time.
Scalable Parallel Clustering Approach for Large Data using Possibilistic Fuzzy C-Means Algorithm

- Robert D Blumofe, The University of Texas at Austin, Scheduling Multithreaded Computations by Work stealing.
Keywords
Clustering  parallel k-means  Fuzzy C-Means  Possibilistic Fuzzy C-Means