Li, Rui; Kazakovtsev, Lev A.
Comparative study of mutation operators in the genetic algorithms for the $k$-means problem.
(English) Zbl 07385110

Summary: The $k$-means problem and the algorithm of the same name are the most commonly used clustering model and algorithm. Being a local search optimization method, the $k$-means algorithm falls to a local minimum of the objective function (sum of squared errors) and depends on the initial solution which is given or selected randomly. This disadvantage of the algorithm can be avoided by combining this algorithm with more sophisticated methods such as the Variable Neighborhood Search, agglomerative or dissociative heuristic approaches, the genetic algorithms, etc. Aiming at the shortcomings of the $k$-means algorithm and combining the advantages of the $k$-means algorithm and revolutionary approach, a genetic clustering algorithm with the cross-mutation operator was designed. The efficiency of the genetic algorithms with the tournament selection, one-point crossover and various mutation operators (without any mutation operator, with the uniform mutation, DBM mutation and new cross-mutation) are compared on the data sets up to 2 millions of data vectors. We used data from the UCI repository and special data set collected during the testing of the highly reliable semiconductor components. In this paper, we do not discuss the comparative efficiency of the genetic algorithms for the $k$-means problem in comparison with the other (non-genetic) algorithms as well as the comparative adequacy of the $k$-means clustering model. Here, we focus on the influence of various mutation operators on the efficiency of the genetic algorithms only.

MSC:
90C57 Polyhedral combinatorics, branch-and-bound, branch-and-cut
90C27 Combinatorial optimization
90C09 Boolean programming
90C59 Approximation methods and heuristics in mathematical programming

Keywords:
$k$-means problem; variable neighborhood search; genetic algorithms; cross-mutation operator

Software:
UCI-ml

Full Text: DOI

References:

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.