Chapter 33
GeneticTKM:
A Hybrid Clustering Method Based on Genetic Algorithm, Tabu Search and K-Means

Masoud Yaghini
Iran University of Science and Technology, Iran

Nasim Gereilinia
Iran University of Science and Technology, Iran

ABSTRACT

The clustering problem under the criterion of minimum sum square of errors is a non-convex and non-linear problem, which possesses many locally optimal values, resulting that its solution often being stuck at locally optimal solution. In this paper, a hybrid genetic, tabu search and k-means algorithm, called GeneticTKM, is proposed for the clustering problem. A new mutation operator is presented based on tabu search algorithm for the proposed hybrid genetic method. The key idea of the new operator is to produce tabu space for escaping from trap of local optimal and finding better solution. The results of the proposed algorithm are compared with other clustering algorithms such as genetic algorithm; tabu search and particle swarm optimization by implementing them and using standard and simulated data sets. The authors also compare the results of the proposed algorithm with other researchers' results in clustering the standard data sets. The results show that the proposed algorithm can be considered as an effective and efficient algorithm to find better solution for the clustering problem.

1. INTRODUCTION

The clustering is the process of grouping a set of objects into classes (Han & Kamber, 2006). The objects in the same cluster are similar to one another and are dissimilar to the objects in other clusters. k-means is one of the popular clustering methods. This algorithm divides a set of n objects in to k clusters so that the resulting intracluster similarity is high but the intercluster similarity is low. The k-means algorithm proceeds as follows. First, it randomly selects k of the objects as cluster centers. For each of the remaining objects, an object is assigned to the cluster to which it is the

DOI: 10.4018/978-1-4666-8473-7.ch033

most similar, based on the distance between the object and the cluster center. It then computes the new cluster centers as mean value of the objects in each cluster. The above process continues until the criterion function converges. Usually, the square-error criterion is used. It is calculated as follows:

$$F = \sum_{i=1}^{k} \sum_{p \in c_i} \left| p - m_i \right|^2 \tag{1}$$

where F is the sum of the square error for all objects in the data set; p is the point in space representing a given object; and m_i is the mean of cluster C_i. In other words, the distances between objects and their cluster centers are squared and then summed (Han & Kamber, 2006). The major drawback of k-means algorithm is that the clustering quality is sensitive to the initial election of cluster centers and may converge to the local optimal (Abdel-Kader, 2010).

In recent years, the various metaheuristic algorithms such as genetic algorithm and tabu search are used to avoid getting trapped in local optimal solutions in solving the clustering problem. In domain of genetic algorithm, a genetically based algorithm, which is composed of split and merge algorithm proposed in Garai and Chaunhuri (2004). Laszlo and Mukherjee (2006) presented a genetic algorithm to select a good set of centers for k-means to produce near optimal partitions. Their approach uses hyper-quad trees. The main drawback of this algorithm is that it is only suitable for low-dimensional data, so Laszlo and Mukherjee (2007) presented another method that exploits region-based crossover, but unlike the hyper-quad tree approach, it scales well to higher-dimensional data sets. Hong & Kwong (2008) proposed a novel clustering algorithm that combines the steady-state genetic algorithm and the ensemble learning method. Chang et al. (2009) suggested a Genetic Algorithm with Gene Rearrangement (GAGR). In their algorithm, in order to reduce the degeneracy, a gene rearrangement

of the chromosome has been used. Degeneracy occurs when different chromosomes showing the same cluster result. Moreover, a new path-based crossover operator has been presented. This operator builds a path from one chromosome to another chromosome.

In domain of tabu search algorithm Sung and Jin (2000) proposed a clustering algorithm by using the tabu search method which is combined with two newly exploited procedures. These procedures are called packing and releasing. The proposed algorithm of this paper consists of two parts. First part is to find the initial solution and the second part is to improve the initial solution. In the first step, the elements that having a high possibility to be grouped into a cluster are packed, then in the improvement step, any packed elements are separated from each other. A fuzzy tabu search algorithm, which uses a fuzzy method in the determination of two main parameters, is presented in Xu et al. (2002). These parameters are tabu list size and the selection of neighborhood. Experimental results show the fuzzy tabu search works better than standard TS. Pan and Senior (2007) proposed a framework of automatic clustering algorithms called ETSAs. The ETSAs can find the optimal number of clusters and can also create the clusters with good quality at the same time. A new algorithm that is named TabuKHM was presented by Gungor and Unler (2008). This algorithm is based on tabu search technique and k-Harmonicmeans. The experiment results shows robustness of the TabuKHM algorithm. Liu et al. (2008) designed a tabu search based clustering method called TS-Clustering. In this algorithm, five improvement operations are given. The improvement operation is used to improve clustering solution and to speed up the convergence of the clustering algorithm. In this algorithm also three neighborhood modes are given. The neighborhood mode is used to create the neighboring solutions. Experimental results show the superiority of the TS-Clustering algorithm over some known clustering techniques.

Related Content

MARS and Neural Network Models for Shear Strength Prediction of Squat Reinforced Concrete Walls: Shear Strength of Squat Reinforced Concrete Walls
Anthony T.C. Gohand Wengang Zhang (2017). *Modeling and Simulation Techniques in Structural Engineering (pp. 294-327).*
www.irma-international.org/chapter/mars-and-neural-network-models-for-shear-strength-prediction-of-squat-reinforced-concrete-walls/162923

Maturity and Process Capability Models and Their Use in Measuring Resilience in Critical Infrastructure Protection Sectors
Clemith J. Houston Jr.and Douglas C. Sicker (2015). *Transportation Systems and Engineering: Concepts, Methodologies, Tools, and Applications (pp. 506-526).*
www.irma-international.org/chapter/maturity-and-process-capability-models-and-their-use-in-measuring-resilience-in-critical-infrastructure-protection-sectors/128682

Engaging Students' Learning in the Built Environment Through Active Learning
Lloyd Martin Scott (2020). *Claiming Identity Through Redefined Teaching in Construction Programs (pp. 1-25).*
www.irma-international.org/chapter/engaging-students-learning-in-the-built-environment-through-active-learning/234857

Ancient Materials and Singular Constructions: Numerical, Experimental, and Heritage Strategies to Preserve Masonry Structures in Seismic Areas
Paloma Pineda (2016). *Civil and Environmental Engineering: Concepts, Methodologies, Tools, and Applications (pp. 340-359).*
www.irma-international.org/chapter/ancient-materials-and-singular-constructions/144503

The Potential for Rail Transit as a Way to Mitigate Accident Risk: A Case Study in Chennai
Sumeeta Srinivasan (2016). *Handbook of Research on Emerging Innovations in Rail Transportation Engineering (pp. 23-39).*
www.irma-international.org/chapter/the-potential-for-rail-transit-as-a-way-to-mitigate-accident-risk/154407