



Review

A survey on nature inspired metaheuristic algorithms for partitional clustering



Satyasai Jagannath Nanda^{a,*}, Ganapati Panda^b

^a Department of Electronics and Communication Engineering, Malaviya National Institute of Technology Jaipur, Rajasthan 302017, India

^b School of Electrical Sciences, Indian Institute of Technology Bhubaneswar, Odisha 751013, India

ARTICLE INFO

Article history:

Received 10 October 2012

Received in revised form

23 August 2013

Accepted 20 November 2013

Available online 18 January 2014

Keywords:

Partitional clustering

Nature inspired metaheuristics

Evolutionary algorithms

Swarm intelligence

Multi-objective Clustering

ABSTRACT

The partitional clustering concept started with K-means algorithm which was published in 1957. Since then many classical partitional clustering algorithms have been reported based on gradient descent approach. The 1990 kick started a new era in cluster analysis with the application of nature inspired metaheuristics. After initial formulation nearly two decades have passed and researchers have developed numerous new algorithms in this field. This paper embodies an up-to-date review of all major nature inspired metaheuristic algorithms employed till date for partitional clustering. Further, key issues involved during formulation of various metaheuristics as a clustering problem and major application areas are discussed.

© 2014 Published by Elsevier B.V.

Contents

| | |
|---|----|
| 1. Introduction | 2 |
| 2. Single objective nature inspired metaheuristics in partitional clustering | 3 |
| 2.1. Problem formulation | 3 |
| 2.2. Historical developments in nature inspired metaheuristics for partitional clustering | 3 |
| 2.2.1. Evolutionary algorithms in partitional clustering | 3 |
| 2.2.2. Physical algorithms in partitional clustering | 5 |
| 2.2.3. Swarm Intelligence algorithms in partitional clustering | 7 |
| 2.2.4. Bio-inspired algorithms in partitional clustering | 9 |
| 2.2.5. Other nature inspired metaheuristics for partitional clustering | 9 |
| 2.3. Fitness functions for partitional clustering | 10 |
| 2.4. Cluster validity indices | 10 |
| 3. Multi-objective algorithms for flexible clustering | 10 |
| 3.1. Problem formulation | 10 |
| 3.2. Historical development in multi-objective algorithms for partitional clustering | 11 |
| 3.3. Evaluation methods | 11 |
| 4. Real life application areas of nature inspired metaheuristics based partitional clustering | 11 |
| 5. Conclusion | 11 |
| 6. Future research issues | 12 |
| References | 13 |

* Corresponding author.

E-mail addresses: nanda.satyasai@gmail.com (S.J. Nanda), ganapati.panda@gmail.com (G. Panda).

1. Introduction

Data clustering determines a group of patterns in a dataset which are homogeneous in nature. The objective is to develop an automatic algorithm which can accurately classify an unlevelled dataset into groups. Recent literature [1–5] broadly classifies clustering algorithms into three categories: hierarchical, partitional and overlapping. The hierarchical algorithms provide a tree structure output (dendrogram plot) which represent the nested grouping of the elements of a dataset [6,7]. They do not require a priori knowledge about the number of clusters present in the dataset [8,9]. However the process involved in the algorithm is assumed to be static and elements assigned to a given cluster cannot move to other clusters [10]. Therefore they exhibit poor performance when the separation of overlapping clusters is carried out.

The overlapping nature of clusters is better expressed in fuzzy clustering [11–13]. The popular algorithms include fuzzy c-means (FCM) [14] and fuzzy c-shells algorithm (FCS) [15]. In this approach each element of a dataset belongs to all the clusters with a fuzzy membership grade. The fuzzy clustering can be converted to a crisp clustering (any element belongs to one cluster only) by assigning each element to the cluster with highest measure of membership value.

The partitional clustering divides a dataset into a number of groups based upon certain criterion known as fitness measure. The fitness measure directly affects the nature of formation of clusters. Once an appropriate fitness measure is selected the partitioning task is converted into an optimization problem (example: grouping based on minimization of distance or maximization of correlation between patterns, otherwise optimizing their density in the N dimensional space etc.). These partitional techniques are popular in various research fields due to their capability to cluster large datasets (example: in signal and image processing for image segmentation [16], in wireless sensor network for classifying the sensors to enhance lifetime and coverage [17–20], in communication to design accurate blind equalizers [21], in robotics to efficiently classify the humans based upon their activities [22], in computer science for web mining and pattern recognition [23], in economics research to identify the group of homogeneous consumers [24], in management studies to determine the portfolio [27], in seismology to classify the aftershocks from the regular background events [28], to perform high dimensional data analysis [29], in medical sciences to identify diseases from a group of patient reports and genomic studies [30], in library sciences for grouping books based upon the content [32], etc.). In all these applications the nature of patterns associated with the datasets is different from each other. Therefore a single partitional algorithm cannot universally solve all problems. Thus given a problem in hand an user has to carefully investigate the nature of the patterns associated with the dataset and select the appropriate clustering strategy.

The K-means algorithm is the most fundamental partitional clustering concept which was published by Lloyd of Bell Telephone laboratories in 1957 [373–375]. After 50 years of its existence, till date this algorithm is still popular and widely used for high dimensional datasets due to its simplicity and lower computational complexity [34,376,377]. In this case the minimization of Euclidean distance between elements and cluster center is considered as optimization criterion. Inspired by K-means a number of gradient algorithms for partitional clustering are developed by researchers which include bisecting K-means [35] (recursively dividing the dataset into two clusters in each step), sort-means [36] (means are shortened in the order of increasing distance from each mean to speed up the traditional process), kd-tree [37] (determines the closest cluster centers for all the data points),

X-means [38] (determines the best number of clusters K by optimizing a criterion such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC)), k-harmonic means [39] (instead of minimum of Euclidean distance the harmonic mean is taken), k-modes algorithm [40,41] (selects k initial modes instead of centers, followed by allocating every object to the nearest mode), Kernel K-means [42] (to detect arbitrary shaped clusters, with proper choice of the kernel function), K-medoid [43] (cluster center is represented by median of the data instead of the mean). These algorithms are computationally simpler, but are often trapped into local optimums due to hill-climbing approach (of cluster center moment in case of K-means). On the other hand, the nature inspired metaheuristics employ a population to explore the search space and thus ensure greater probability to achieve optimal cluster partitions.

Literature review [44,45] reveals the recent trend to name all stochastic algorithms with randomization and local search as 'metaheuristic'. The randomization process generates arbitrary solutions, which explore the search space and are responsible to achieve global solution. The local search is responsible to determine convergence and focus on achieving good solutions in a specific region. The first nature inspired metaheuristic is genetic algorithm (GA) developed by Holland and his colleagues in 1975 [46,47]. It is followed by development of simulated annealing (SA) by Kirkpatrick in 1983 [48]. Recent literature reports many established nature inspired metaheuristics which are enlisted in Table 1. These algorithms are broadly classified into Evolutionary Algorithms, Physical Algorithms, Swarm Intelligence, Bio-inspired Algorithms and others. Table 1 lists these algorithms which are further divided into single objective and multi-objectives depending on the number of objective functions that they simultaneously optimize to achieve the solution.

The fundamental approach to develop nature inspired metaheuristics based clustering algorithm using simulated annealing was proposed by Selim and Alsultan [159] in 1991. Then Bezdek et al. [100] proposed the evolutionary approach to develop clustering using genetic algorithm in 1994. The research article by Sarkar and Yegnarayana [101] highlights the core issues involved in evolutionary programming for development of clustering algorithm. Lumer and Faieta first explored the use of swarm nature of clustering ants [191]. Subsequently the swarm intelligence algorithms like ant colony optimization [183] and particle swarm optimization [219] have been applied for cluster analysis.

This paper presents an in-depth survey of nature inspired metaheuristic algorithms used for partitional clustering. The paper focuses on the nature inspired metaheuristics that have been used for cluster analysis in the last two decades. Few interesting review articles on cluster analysis with overwhelming citation by researchers have been published by Jain et al. [34], Hruschka et al. [3], Xu and Wunsch [91], Freitas [92], Paterlini and Minerva [93], Jafar and Sivakumar [94]. To the best of our knowledge a review paper employing recently developed nature inspired metaheuristics for partitional clustering has not been reported. In 2009 Hruschka et al. [3] have focused on the initialization procedures, crossover, mutation, fitness evaluation and reselection associated with genetic type evolutionary algorithms for single and multiobjective cases. Jain et al. [34] have dealt with key issues of clustering, users dilemma and have suggested corresponding solutions. Jafar and Sivakumar [94] highlighted the developments in ant algorithm for cluster analysis. The book chapter by Abraham et al. [4] focuses on the use of PSO and ant algorithm for clustering task. The basic principles and methods of clustering are embodied in the books [1,95–99].

Keeping the current research trends in mind the present paper contributes in the survey of partitional clustering in terms of four aspects: (1) systematic review on all the single objective nature

Download English Version:

<https://daneshyari.com/en/article/493978>

Download Persian Version:

<https://daneshyari.com/article/493978>

[Daneshyari.com](https://daneshyari.com)