Partial Evaluation Genetic Algorithm Based on Fuzzy c-Means Algorithm*

Si-Ho Yoo and Sung-Bae Cho

Dept. of Computer Science, Yonsei University
134 Shinchon-dong, Sudaemoon-ku, Seoul 120-749, Korea
bonanza@sclab.yonsei.ac.kr, sbcho@cs.yonsei.ac.kr

Abstract. To find the optimal solution with genetic algorithm, it is desirable to maintain the population size as large as possible. In some cases, however, the cost to evaluate each individual is relatively high and it is difficult to maintain large population. To solve this problem we propose a partially evaluated GA based on fuzzy clustering, which considerably reduces evaluation cost without any loss of its performance by evaluating only one representative for each cluster. The fitness values of other individuals are estimated from the representative fitness values indirectly. We have used fuzzy c-means algorithm and distributed the fitness according to membership matrix. The results with nine benchmark functions are compared to six hard clustering algorithms with Euclidean distance and Pearson correlation coefficients for measuring the similarity between the representative and its members in fitness distribution.

1 Introduction

Genetic algorithm (GA) is an efficient method for machine learning, optimization and classification, based on evolution mechanisms such as biological genetics and natural selection [1]. It is required to make the population size of evolution as large as possible because GA approach evolves the population spread over the search space. However, in some domains where the cost of evaluation is relatively high, it is difficult to maintain large number of individuals in a population. Smaller population causes several negative effects such as genetic drift.

One example that cannot help utilizing only smaller population is interactive genetic algorithm (IGA) application [2]. IGA is a technique that performs optimization based on human evaluation. A human operator can obtain what he wants through repeated interaction with computer. It has a special advantage, which is to adopt user’s choice as fitness, when fitness function cannot be explicitly defined. This property allows IGA to be applied to artistic domains such as music and design, which are almost impossible to be solved with simple GA [2]. Also, there is the difficulty of high cost in solving the inverse problem with simple GA [3]. It is impossible to maintain large population size to search for the optimal solution in such domains.

* This paper was supported in part by Biometrics Engineering Research Center, KOSEF, and Brain Science and Engineering Research Program sponsored by Korean Ministry of Science and Technology in Korea.
To cope with this shortcoming an efficient GA where some parts of population are evaluated by human and remaining part by computer was proposed [4]. GA that evaluates only a few individual directly and evaluates the rest of individuals by their similarity to the selected individuals was also proposed [5]. Also, the hybrid GA based on clustering which considerably reduces the number of evaluations by evaluating only one representative of each cluster’s center after clustering individuals in population was presented [6]. However, distributing fitness values of individuals linearly with only one representative does not provide precise fitness values if the hybrid GA could not form the ideal cluster partition.

In this paper, we propose a partially evaluated GA based on fuzzy clustering algorithm to solve this problem. At first, we make $N$ clusters of individuals in a population by fuzzy c-means algorithm and calculate the fitness values of the representative in each cluster. Remaining individuals in each cluster get their fitness estimated by their membership values which indicate individuals’ degree of belongness to the clusters. The fitness value of an individual that belongs to multiple clusters is estimated by all the relevant cluster representatives. Nine benchmark functions [7] have used to test the proposed method and the results are compared to six hard clustering algorithms with Euclidean distance and Pearson correlation coefficient for fitness estimation.

2 Clustering Algorithms

Clustering refers to the process of grouping samples in order to be similar within group [8]. These groups are called clusters. There are three general categories of clustering techniques: Hierarchical clustering, partitional clustering, and overlapping clustering.

2.1 Hierarchical Clustering

Hierarchical clustering algorithm constructs a structure of clusters. In this structure a cluster can have several substructures which are composed of other clusters. There are several hierarchical clustering algorithms such as single linkage algorithm, complete linkage algorithm, and average linkage algorithm [9]. Single linkage algorithm defines inter-cluster distance as the closest distance between two samples belonging to two different clusters. Complete linkage algorithm uses the distance between the most remote samples belonging to two different clusters as inter-cluster distance. Average linkage algorithm measures the average distance between all of the samples belonging to two different clusters for inter-cluster distance.

2.2 Partitional Clustering

Different from hierarchical clustering which creates a series of nested clusters, partitional clustering usually creates one set of clusters that partition the data into groups. Samples close to one another are assumed to be similar and the goal of the partitional clustering algorithm is to group data to be close together. $k$-means algorithm and hard c-means algorithm are good examples of partitional clustering. The $k$-means algorithm, one of the most widely used ones, attempts to solve the clustering problem by optimizing a given objective function [9].