

A GENETIC HARD C-MEANS CLUSTERING ALGORITHM

L. Meng¹, Q. H. Wu² and Z. Z. Yong³

^{1,2}Department of Electrical Engineering and Electronics
University of Liverpool, Liverpool, L69 3GJ, U.K.

³Department of Electronic Engineering
Shenzhen University, Shenzhen, 518060, P. R. China

Abstract. As a calculus-based method, the traditional clustering algorithm is found to be easily trapped in local optima and is very sensitive to initial conditions. A novel algorithm, the genetic hard c-means clustering algorithm (GHCMCA), has been developed to overcome these limitations. GHCMCA adopts the basic scheme of a genetic algorithm (GA) and performs a genetically guided search for the optimization of a hard c-means clustering problem. Unlike previous clustering algorithms with GAs, the population members maintained in the GHCMCA approach represent partition matrices instead of sets of cluster centers. A new genetic crossover operator has been introduced which effectively recombines important similarities between a pair of partitions. GHCMCA has been evaluated and compared with the traditional clustering algorithm and a previous genetic clustering algorithm that adopts the standard two-point crossover. Results show that a genetic approach is able to overcome the inevitable drawbacks of a hill-climbing technique. In experiments with an image data set, GHCMCA is superior to the previous genetic clustering algorithm in the sense that it converges much more quickly to the desired region in which the global optimum resides. In the cases where speed as well as performance is required, GHCMCA may thus provide a solution to the dilemma where the classical clustering algorithm can be easily trapped in different local extrema and the conventional genetic approach is time consuming.

Keywords. clustering, genetic algorithms, hard c-means, image quantization, evolutionary computation.

1 Introduction

The nature of many practical problems consists of clustering a set of objects into a collection of mutually disjoint subsets. These problems have been classified as *hard clustering problems* [15]. In general, hard clustering acts on the basis of similarities or distance (or dissimilarities) among objects, the optimal solution is the one that yields a minimal cost. The traditional *hard c-means clustering algorithm* (HCMCA) [6] has been a popular method for finding the optimal solution to hard clustering problems. Applying an iterative scheme, HCMCA starts with an initial distribution of the cluster