

DETERMINISTIC AND SIMULATED ANNEALING APPROACH TO FUZZY C-MEANS CLUSTERING

MAKOTO YASUDA

Department of Electrical and Computer Engineering
Gifu National College of Technology
Kamimakuwa 2236-2, Motosu, Gifu, 501-0495, Japan
yasuda@gifu-nct.ac.jp

Received July 2008; revised December 2008

ABSTRACT. *This paper explains the approximation of a membership function obtained by entropy regularization of the fuzzy c-means (FCM) method. By regularizing FCM with fuzzy entropy, a membership function similar to the Fermi-Dirac distribution function is obtained. We propose a new clustering method, in which the minimum of the Helmholtz free energy for FCM is searched by deterministic annealing (DA), while optimizing the parameters of the membership function by simulated annealing (SA). However, it takes a long time to execute SA repeatedly because the membership function contains an exponential function. Thus, the membership function is approximated by linear functions. Numerical experiments are performed and the obtained results indicate that this method can cluster data properly and shorten the computational time.*

Keywords: Fuzzy C-means clustering, Deterministic annealing, Simulated annealing, Fuzzy entropy, Fermi-Dirac distribution

1. **Introduction.** Statistical mechanics investigate the macroscopic properties of a physical system consisting of several elements. Recently, there has been great research interested in applying statistical mechanical models or tools to information science. Many engineering problems can be formulated as optimization problems, and the simulated annealing (SA) method [1, 6] is an efficient optimization technique for such problems [3]. SA is a stochastic relaxation method which, by analogy with the annealing process of solids, considers a cost function as the energy of a system, and as the (pseudo) temperature is decreased, randomly searches for a minimum at a high temperature and more deterministically at a low temperature. SA is a global optimization technique for solving combinatorial optimization problems as long as the cost function is definable and cooling is performed sufficiently slowly. Finding the optimal solution by SA, however, requires a very long time because SA searches stochastically at each temperature and practically functions as an approximation algorithm.

The deterministic annealing (DA) method is a deterministic variant of SA [12]. DA characterizes the minimization problem of the cost function as the minimization of the Helmholtz free energy, which depends on the temperature, and tracks its minimum while decreasing the temperature; thus, it can deterministically optimize the cost function at each temperature. Hence, DA is more efficient than SA, but it does not guarantee the optimal solution.

On the other hand, clustering is one of the major tools for data analysis. Bezdek [2] proposed the fuzzy c-means (FCM) which is now well known as the standard technique for fuzzy clustering [10]. In fuzzy theory, there exists a strong relationship between the membership functions of FCM clustering with maximum entropy or entropy regularization