International Journal of Innovative
Computing, Information and Control
Volume 4, Number 6, June 2008
 ISSN 1349-4198
pp. 1465–1476

ON USING NOISING METHOD FOR CLUSTERING ANALYSIS

Yongguo Liu[1,2], Xiaorong Pu[1], Zhang Yi[1], Xiaofeng Liao[3]
and Kefei Chen[4]

[1]School of Computer Science and Engineering
University of Electronic Science and Technology of China
Chengdu 610054, P. R. China
liuyg@uestc.edu.cn

[2]State Key Laboratory of Computer Science
Institute of Software
Chinese Academy of Sciences
Beijing 100080, P. R. China

[3]Department of Computer Science and Engineering
Chongqing University
Chongqing 400044, P. R. China

[4]Department of Computer Science and Engineering
Shanghai Jiaotong University
Shanghai 200030, P. R. China

Received April 2007; revised November 2007

Abstract. *In this article, the clustering problem under the criterion of minimum sum of squares clustering is considered. This problem is a nonconvex program which possesses many local optima. Its solution often falls into these traps. In this paper, we propose a new noising method based clustering algorithm called Improved Noising Method based Clustering (INMC). In the INMC algorithm, local iteration methods are integrated to improve the solution obtained during the iteration process. In addition, a mergence and partition mode is designed to establish the neighboring solution. Experimental results on artificial and real-life data sets are given to illustrate the superiority of the INMC algorithm over some known clustering methods.*
Keywords: Clustering, Noising method, Metaheuristics

1. **Introduction.** The clustering problem is a fundamental problem that frequently arises in a great variety of application fields such as pattern recognition, machine learning and statistics. Clustering analysis is a formal study of algorithms and methods for grouping or classifying objects without category labels. A cluster is a set of entities that are alike, and entities from different clusters are not alike. In this article, we consider the minimum sum of squares clustering problem stated as follows: Given N objects in R^m, allocate each object to one of K clusters such that the sum of squared Euclidean distances between each object and the center of its belonging cluster for every such allocated object is minimized. This problem is mathematically defined as

$$\min_{W,C} J(W,C) = \sum_{i=1}^{N}\sum_{j=1}^{K} w_{ij}\|\mathbf{x}_i - \mathbf{c}_j\|^2 \tag{1}$$

where $\sum_{j=1}^{K} w_{ij} = 1$, $i = 1, \ldots, N$. If object $\mathbf{x}_i$ is allocated to cluster C_j, then w_{ij} is equal to 1; otherwise w_{ij} is equal to 0. Here, N denotes the number of objects, m denotes the