International Journal of Innovative
Computing, Information and Control
Volume 4, Number 10, October 2008

 ISSN 1349-4198
pp. 2523–2531

FUZZY CLUSTERING LEVEL ANALYSIS VIA STATISTICAL SCHEME APPLYING AKAIKE'S INFORMATION CRITERION (AIC)

Shuya Kanagawa[1] and Kimiaki Shinkai[2]

[1]Department of Systems and Information Engineering
Faculty of Engineering
Musashi Institute of Technology
Tokyo, Japan
kanagawa@si.musashi-tech.ac.jp

[2]Department of Mathematics
Graduate School of Education
Waseda University
Tokyo, Japan
building-bridges@ruri.waseda.jp

Received December 2007; revised May 2008

Abstract. *Recently Uesu and Shinkai et al. [5-7,9-11] presented a rational method to decide the number of clusters with fuzzy clustering. In other words, this is a problem to find the optimal level as to a partition tree. In this paper we would improve their methods by use of AIC (Akaike's information criterion) to show that the new statistical technique is able to find a statistically optimal level of fuzzy clustering. Furthermore we investigate advantage and disadvantage of the AIC method using some examples of fuzzy partition trees via a numerical simulation.*

Keywords: Fuzzy clustering, AIC, Fuzzy node, Fuzzy graph, Fuzzy decision, Optimal clustering level, Steepest decent method

1. **Introduction.** As for the analysis of inexact information such as human behavior, mental process, social structure and so on, we consider a fuzzy graph of some relation in these problems. Since the fuzzy graph is constructed with clusters of some levels, the clustering level analysis of fuzzy graph is required for the above applications, see e.g. Romsburg [4]. Figure 1 is a typical example of a partition tree with 10 clustering levels {**R**0.00, **R**0.13, **R**0.27, **R**0.49, **R**0.58, **R**0.59, **R**0.74, **R**0.79, **R**0.91, **R**1.00}. For example, the clustering level **R**0.91 implies that the distance between {5} and {8} is $0.09 = 1 - 0.91$ and two sample points {5} and {8} construct a cluster {5,8} and other points are separated since the distance between each pair is larger than 0.09, respectively. Furthermore the level **R**0.58 implies that the set of ten sample points {1, 2, ..., 10} is divided by 5 clusters {5, 8}, {4, 9, 1, 7, 2}, {3}, {6}, {10}. The cluster {4, 9, 1, 7, 2} consists of three clusters {4, 9}, {1, 7}, {2} and the distance between adjoining two clusters is less than $0.42 = 1 - 0.58$, respectively. However it is difficult to find the optimal level in {**R**0.00, **R**0.13, **R**0.27, **R**0.49, **R**0.58, **R**0.59, **R**0.74, **R**0.79, **R**0.91, **R**1.00}. As to the selection of the optimal level in a partition tree the steepest decent method in the multivariate analysis has been used. In 1994 Tsuda and Yamashita [8] suggested a new method to estimate the optimal level to find a unique equilibrium state for the cluster number and the cluster size which is a kind of stable classification of clusters for {1, 2, ..., 10}. Their method fits in well with actual examples. Furthermore, Uesu and Shinkai et al. [5-7,9-11] presented some rational methods to decide the optimal level of fuzzy