# FUZZY C-MEANS ALGORITHM BASED ON PSO AND MAHALANOBIS DISTANCE

HSIANG-CHUAN LIU[1], JENG-MING YIH[2], WEN-CHIH LIN[3]
AND TUNG-SHENG LIU[4]

[1]Department of Bioinformatics
[3]Department of Computer Science and Information Engineering
Asia University
500, Lioufeng Rd., Wufeng, Taichung, 41354, Taiwan
lhc@asia.edu.tw; kbc@mail.ntcu.edu.tw

[2]Graduate Institute of Educational Measurement and Statistics
Department of Mathematics Education
National Taichung University
No. 140, Ming-sheng Road, Taichung, 40306, Taiwan
yih@mail.ntcu.edu.tw

[4]Forensic Science Center of Military Police Command, Taiwan
tommy0366@yahoo.com.tw

ABSTRACT. *Some of the well-known fuzzy clustering algorithms are based on Euclidean distance function, which can only be used to detect spherical structural clusters. Gustafson-Kessel (GK) clustering algorithm and Gath-Geva (GG) clustering algorithm were developed to detect non-spherical structural clusters. Both of GG and GK algorithms suffer from the singularity problem of covariance matrix and the effect of initial status. In this paper, a new Fuzzy C-Means algorithm based on Particle Swarm Optimization and Mahalanobis distance without prior information (PSO-FCM-M) is proposed to improve those limitations of GG and GK algorithms. And we point out that the PSO-FCM algorithm is a special case of PSO-FCM-M algorithm. The experimental results of two real data sets show that the performance of our proposed PSO-FCM-M algorithm is better than those of the FCM, GG, GK algorithms.*
**Keywords:** Fuzzy C-Means algorithm, Mahalanobis distance, PSO-FCM algorithm, PSO-FCM algorithm

1. **Introduction.** Clustering technique plays an important role in data analysis and interpretation. It groups data into clusters so that the data objects within a cluster have high similarity in comparison to one another, but are very dissimilar to those data objects in other clusters. Fuzzy clustering is a branch in clustering analysis and it is widely used in the pattern recognition field. The well-known ones, such as Bezdek's Fuzzy C-Means (FCM) [1], Pal, Pal and Bezdek's "Possibility C-Means (PCM)" [2], and "Fuzzy Possibility C-Means (FPCM)" [3], are all based on Euclidean distance. These fuzzy clustering algorithms can only be used to detect the data classes with the same super spherical shapes.

By extending Euclidean distance to Mahalanobis distance, Gustafson-Kessel (GK) clustering algorithm [4] and Gath-Geva (GG) clustering algorithm [5] were developed to detect non-spherical structural clusters. There are some limitations of GG and GK algorithms. In GK algorithm, the Mahalanobis distance with the fuzzy covariance matrix was applied