國立金門技術學院資訊管理系
National Kinmen Institute of Technology
INFORMATION MANAGEMENT

Automatic Particle Swarm Optimization Clustering Algorithm

Ching-Yi Chen, Hsuan-Ming Feng, and Fun Ye
International Journal of Electrical Engineering. vol. 13, no. 4, pp.379-387, 2006

Abstract

•Automatic particle swarm optimization (AUTO-PSO) clustering algorithm with cluster validity measure is developed.,
•To efficiently decide the number of clusters, extract its associated cluster centers, and classify the correct catalog from various data sets.
•Simulation results for two artificial data sets and one real-life IRIS data are proposed to show the efficiency of the AUTO-PSO algorithms.

1. Introduction

Three important items in the design of clustering process, to choose the optimal number of clusters, to measure the similarity between training data points and desired cluster centers, and to quickly approach the desired cluster centers.

2. AUTO-PSO CLUSTERING ALGORITHM

•The proposed AUTO-PSO clustering algorithm is not only automatically determines the true number of the cluster centers but also extract real cluster centers and make a good classification..

•**Steps :**

(1) Randomly Initialize and its associated velocity for all particles in the form

$$X_p=[z_p,\gamma_p]=[z_{1,i}^p,z_{2,i}^p,\ldots z_{Ti}^p,\gamma_1^p,\gamma_2^p,\ldots\gamma_T^p] \qquad z_{j,i}^p = U_i^{\min} + rand(\ )\cdot(U_i^{\max}-U_i^{\min})$$

$$v_{j,i}^p = rand\ (\ )\cdot\frac{(U_i^{\max}-U_i^{\min})}{\lambda} \qquad (1)$$

, where

(2) Select the active solution (i.e. active cluster centers) of every particle from initial population with the IF-THEN rules:

IF THEN the *j*th candidate cluster center is ACTIVE (2)

ELSE IF THEN the *j*th candidate cluster center is INACTIVE.

(3) Determine unreasonable solutions

(a) First calculate the counting number

(b) Check if the counting number is smaller than 2

If Yes then

(c) Refine cluster centers by the simple average computation with N/K (3)

(4)Calculate fitness function of each particle with $F=\frac{1}{CS_i+eps}$ (4)

•Where $$CS(K)=\frac{\frac{1}{K}\sum_{i=1}^{K}\left\{\frac{1}{N_i}\sum_{x_j\in C_i}\max_{x_k\in C_i}\{d(x_j,x_k)\}\right\}}{\frac{1}{K}\sum_{i=1}^{K}\left\{\min_{j\in K,j\neq i}\{d(z_i,z_j)\}\right\}}=\frac{\sum_{i=1}^{K}\left\{\frac{1}{N_i}\sum_{x_j\in C_i}\max_{x_k\in C_i}\{d(x_j,x_k)\}\right\}}{\sum_{i=1}^{K}\left\{\min_{j\in K,j\neq i}\{d(z_i,z_j)\}\right\}} \qquad (5)$$

• The objective of the PSO is to minimize the CS measure for achieving proper clustering results.

(5) Applied the PSO learning algorithm to determine optimizations from the initializations.

3. Simulation Results

•(1) Simulation result for mixture of spherical and ellipsoidal data set is shown in Fig.1, Data set with five spherical clusters is shown in Fig.2, and IRIS data set with 150 data points separated into three subclasses is shown in Fig.3.

4 Conclusion

An AUTO-PSO algorithm has been developed to solve several types of clustering problems, which automatically determines the real cluster number, selects desired cluster centers and finds good classification result at the same time. The evolutional PSO learning method with CS validity measure yields a very efficient and simple clustering analysis.

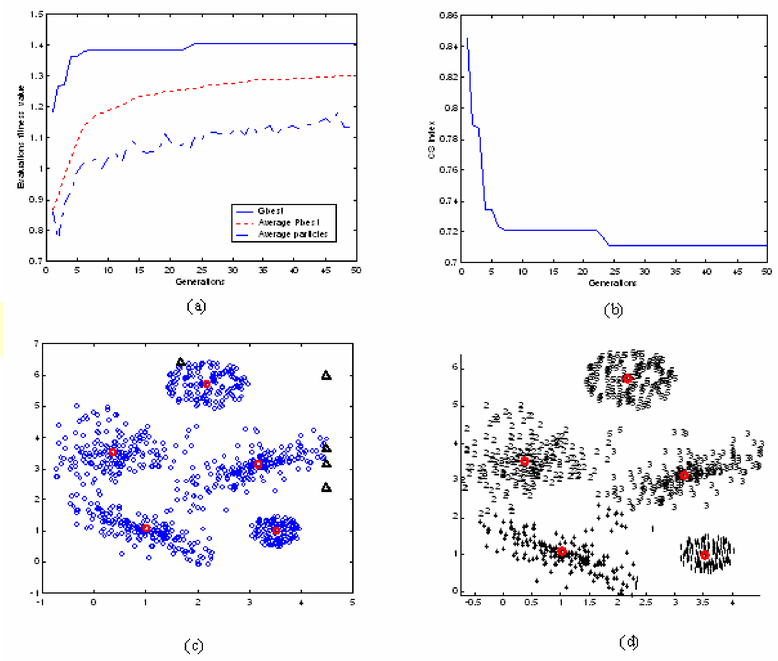

Fig. 1Response of AUTO-PSO algorithm in Example 1. (a) Fitness value against generation with (solid), average (dash) and average particles (dash-dotted), (b) CS measure against generation for gbest, (c) Final selected , where ‘Δ’ is disabled and ‘◯’ is active, (d)The optimal classification result by the selected cluster centers.

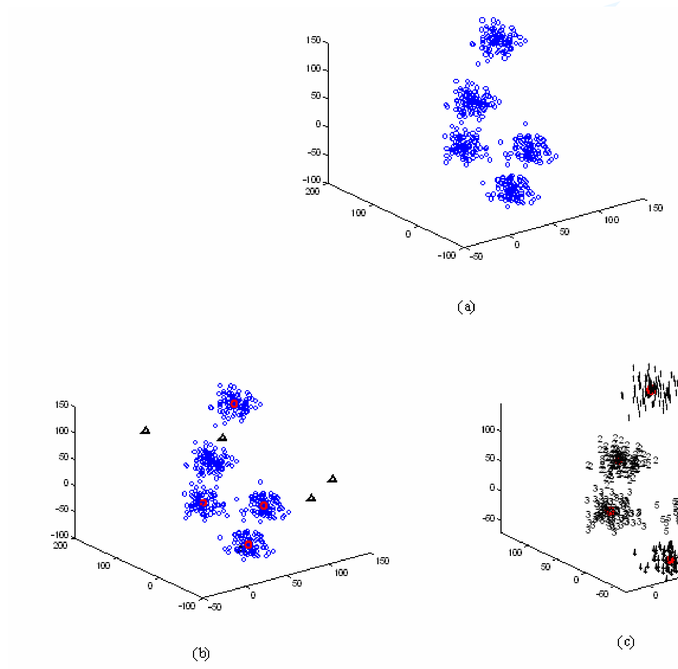

Fig 2. Response of AUTO-PSO algorithm in example 2. (a) The data set used in example 3. (b) Final , where ‘Δ’ is disable and ‘◯’ is active. (c) The optimal classification result by the selected cluster centers.

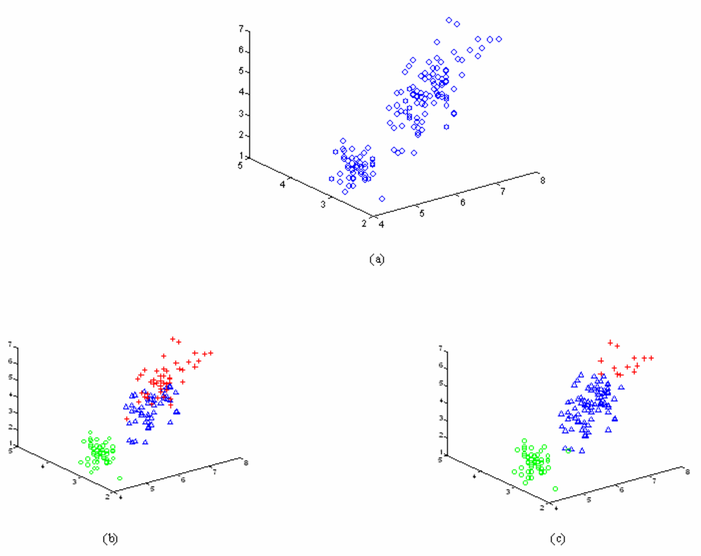

Fig. 3 Response of AUTO-PSO algorithm in example 3. (a) The three-dimensional plot for the four-dimensional IRIS data . (b) The three-dimensional plot for the four-dimensional IRIS data: IRIS setosa (◯), IRIS versicolor (△), and IRIS virginica (+). (c) The clustering results achieved by the proposed method.