

Particle swarm optimization with selective particle regeneration for data clustering

Chi-Yang Tsai*, I-Wei Kao

Department of Industrial Engineering and Management, Yuan Ze University, Taoyuan 320, Taiwan, ROC

ARTICLE INFO

Keywords:

Data clustering
Particle swarm optimization
K-means algorithm

ABSTRACT

This paper presents selective regeneration particle swarm optimization (SRPSO), a novel algorithm developed based on particle swarm optimization (PSO). It contains two new features, unbalanced parameter setting and particle regeneration operation. The unbalanced parameter setting enables fast convergence of the algorithm and the particle regeneration operation allows the search to escape from local optima and explore for better solutions. This algorithm is applied to data clustering problems for performance evaluation and a hybrid algorithm (KSRPSO) of K-means clustering method and SRPSO is developed. In the conducted numerical experiments, SRPSO and KSRPSO are compared to the original PSO algorithm, K-means, as well as, other methods proposed by other studies. The results demonstrate that SRPSO and KSRPSO are efficient, accurate, and robust methods for data clustering problems.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

In recent years, meta-heuristic algorithms have been applied to a variety of complex problems in order to obtain quality solutions within acceptable computation time. Proposed by Kennedy and Eberhart (1995), particle swarm optimization (PSO) has been drawing attention of many researchers. This algorithm simulates the social behavior of animals such as birds and fish in nature. Individuals in a flock of birds or a school of fish exchange previous experience and make adjustment accordingly so that they can move toward the objective. The concept is adopted by PSO in searching for optimal solutions.

PSO has been widely applied in many research areas and real-world engineering fields. Examples include task assignment and scheduling (Liu, Wang, & Jin, 2008; Sha & Hsu, 2008), data clustering (Kao, Zahara, & Kao, 2008), power flow analysis (Acharjee & Goswami, 2009), pattern recognition (Lin, Wang, & Lee, 2009), roundness measurement (Sun, 2009) demand forecast (Alper, 2008; Gao, Yang, Zhou, & Hu, 2006), financial decisions (Yannis, Magdalene, Michael, & Constantin, 2009), product plans (Wang, Che, & Wu, 2010) and layout design (Onut, Tuzkaya, & Doğac, 2008; Zeng, Zhu, Shen, & Qi, 2007). It has been demonstrated that PSO performs well in many optimization problems. However, it was observed that the algorithm did not perform well at times. The conversion may be slow when solving complex problems and the search can be occasionally trapped in local optima. Many attempts have been made to improve the algorithm's efficiency and robustness.

One of the recent efforts to improve PSO is the selective regeneration particle swarm optimization (SRPSO) proposed by Tsai and Kao (2009), where the basic concept of the algorithm was introduced and the algorithm was applied to multimodal functions for preliminary evaluation of efficiency. This paper follows their work and the goal is two-fold. Firstly, how the operation of selective particle regeneration is designed and incorporated into PSO is illustrated. The intuition and detailed procedure of the algorithm is shown. Examples are provided to demonstrate the effect of the designed operation. Secondly, SRPSO is applied to data clustering. Its performance is evaluated and compared to other methods.

Data clustering is the classification of similar objects into different groups, or more precisely, the partitioning of a dataset into subsets, so that the data in each subset share some common traits. It was proven that clustering problem is NP problem (Adib, 2005). By clustering, sparse and dense regions can be identified and one can therefore discover overall distribution patterns and interesting correlations among data attributes. There are two main types of clustering algorithm, hierarchical and partitional clustering approaches. Hierarchical clustering approach aims at grouping data through repeated cluster separation or agglomeration. Partitional clustering approach attempts to directly decompose data into disjoint clusters based on an objective function such as minimizing the distance between data points and cluster centers. This paper shows how to apply SRPSO to data clustering problems. K-means, a common method for data clustering problems, is also incorporated into SRPSO for performance improvement.

There were many improved or hybridized PSO proposed in past several years. Fan, Liang, and Zahara (2004) developed the hybrid Nelder–Mead (NM)-particle swarm optimization algorithm based on the NM simplex search method and PSO. Wang, Qiu, and Bai (2005) developed a hybrid technique based on particle swarm

* Corresponding author. Tel.: +886 3 463 8800x2504; fax: +886 3 463 8907.

E-mail address: iecytsai@saturn.yzu.edu.tw (C.-Y. Tsai).

optimization algorithm combined with the nonlinear simplex search method (HNM-PSO) for multimodal function optimization. A hybrid algorithm of the genetic algorithm and particle swarm optimization (GA-PSO) was developed by Kao and Zahara (2007). It was applied to solve continuous multimodal function. Coelho (2008) presented a quantum-behaved PSO (QPSO) using chaotic mutation operator and applied QPSO to solve a well-studied continuous optimization problem of mechanical engineering design. Ling et al. (2008) proposed a new hybrid particle swarm optimization method which incorporates a wavelet-theory-based mutation operation. It applied the wavelet theory to enhance the PSO in exploring the solution space more effectively for a better solution.

To improve heuristic algorithm performance for clustering, several methods have been proposed. Among them, K-Means is one of the widely used clustering techniques. The term K-Means was first used by MacQueen (1967). This technique groups data vectors into a predefined number of clusters on the basis of the Euclidean distance as the similarity measure. In order to improve the efficiency, many researchers combined heuristic algorithms with K-means to solve data clustering problem. Bandyopadhyay, Maulik, and Malay (2001) presented an efficient partitional clustering technique, called SAKM-clustering, that integrates the power of simulated annealing for obtaining minimum energy configuration, and the searching capability of K-means algorithm was examined in the research. Bandyopadhyay and Maulik (2002) developed a genetic algorithm-based efficient clustering technique which is called KGA-clustering. It is superior over the K-means algorithm and another genetic algorithm-based clustering method was extensively demonstrated for several artificial and real life data sets. A real life application of the KGA-clustering in classifying the pixels of a satellite image of a part of the city of Mumbai was provided. Kao, Tsai, and Wang (2007) developed an improved particle swarm optimization algorithm for data clustering. A bouncing mechanism was designed such that when particles reach the boundary of the search space, they will be bounced back and given proper directions. Two reflex schemes were implemented on PSO to improve the efficiency. Kao et al. (2008) applied hybrid NM-PSO and K-means (K-NM-PSO) to solve data clustering problem. K-NM-PSO algorithm is tested on nine data sets, and its performance is compared with those of PSO, NM-PSO, K-PSO, GA, KGA and K-means clustering. Results show that K-NM-PSO is both robust and suitable for handling data clustering.

The rest of the paper is organized as follows: the original particle swarm optimization and SRPSO are introduced in Sections 2 and 3, respectively. Section 4 presents the considered methods for data clustering, includes K-means, PSO SRPSO and KSRPSO. Experiment setting and results are provided in Section 5. Finally, the conclusion is presented in the last section.

2. Particle swarm optimization

Particle swarm optimization (PSO) is inspired by the social behavior observed in flocks of birds and schools of fish. In nature, there is a leader who leads the bird or fish group to move, as illustrated in Fig. 1. Most members of the group follow the leader. In PSO, a potential solution to the considered problem is represented by a particle, similar to the individuals in the bird and fish group. Each particle travels in the solution space and attempts to move toward a better solution by changing its direction and speed based on its own past experience and the information from the current best particle of the swarm.

The procedure of PSO is described as follows:

- (1) **Particle initialization:** An initial swarm of particles is generated in search space. Usually, the population size is decided by the dimension of problems.

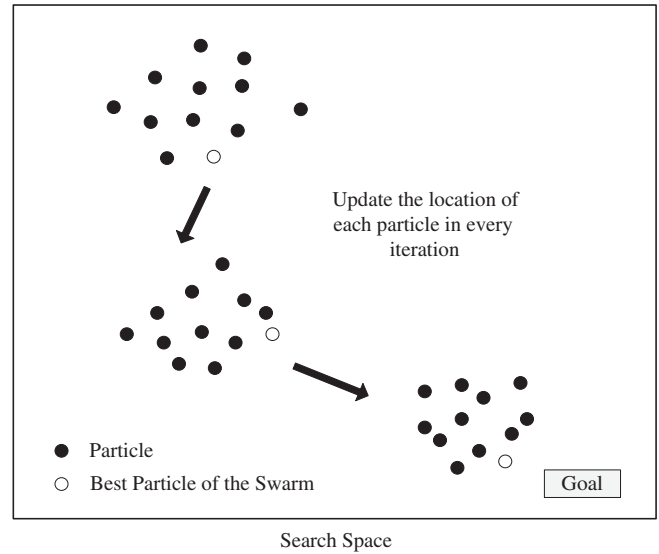


Fig. 1. Swarm following the best particle to moves to the goal.

- (2) **Velocity and position update:** In each iteration, a new velocity value for each particle is calculated based on its current velocity, the distance from its previous best position, and the distance from the global best position. The new velocity value is then used to calculate the next position of the particle in the search space. The particle's velocity and position are dynamically updated as follows:

$$V_{id}^{new} = w \times V_{id}^{old} + c_1 \times rand \times (P_{id} - x_{id}^{old}) + c_2 \times rand \times (P_{gd} - x_{id}^{old}), \quad (1)$$

$$x_{id}^{new} = x_{id}^{old} + V_{id}^{new}. \quad (2)$$

The new velocity of a particle, V_{id}^{new} , is updated by Eq. (1) taking into consideration of the particle's previous velocity, V_{id}^{old} , and previous position, x_{id}^{old} . $w = [0.5 + rand/2]$ is an inertia weight and $rand$ is a uniformly generated random number between 0 and 1. The cognition parameter, c_1 , and social parameter, c_2 , are acceleration coefficients that are conventionally set to a fixed value between 0 and 2 (Kennedy & Eberhart, 1995). P_{id} is the previous individual best position of this particle and P_{gd} is the current global best position. Eq. (2) then calculates the new position of the particle, x_{id}^{new} .

- (3) **Evaluation and update of best locations:** The fitness value of each particle is calculated by the objective function. The values of P_{id} and P_{gd} are then evaluated and replaced if better particle best position or global best position is obtained.
- (4) **Termination:** Steps (2) and (3) are repeated iteratively until the termination condition is met.

3. Selective regenerated particle swarm optimization

The proposed selective regenerated particle swarm optimization (SRPSO) aims at improving the original PSO. Two new features are designed. A suggestion on the setting of cognition and social parameters, c_1 and c_2 , is proposed to accelerate convergence. In addition, a selective particle regeneration mechanism for avoiding the search trapped in local optima.

3.1. Cognitive and social parameter setting

As shown in Eq. (1), the new velocity of a particle is determined based on the best individual position and the knowledge of the

Download English Version:

<https://daneshyari.com/en/article/10322865>

Download Persian Version:

<https://daneshyari.com/article/10322865>

[Daneshyari.com](https://daneshyari.com)