Dynamic Data Clustering Using Stochastic Approximation Driven Multi-Dimensional Particle Swarm Optimization

Serkan Kiranyaz[1], Turker Ince[2], and Moncef Gabbouj[1,*]

[1] Tampere University of Technology, Tampere, Finland
{serkan.kiranyaz,moncef.gabbouj}@tut.fi
[2] Izmir University of Economics, Izmir, Turkey
turker.ince@ieu.edu.tr

Abstract. With an ever-growing attention Particle Swarm Optimization (PSO) has found many application areas for many challenging optimization problems. It is, however, a known fact that PSO has a severe drawback in the update of its global best (*gbest*) particle, which has a crucial role of guiding the rest of the swarm. In this paper, we propose two efficient solutions to remedy this problem using a stochastic approximation (SA) technique. For this purpose we use simultaneous perturbation stochastic approximation (SPSA), which is applied only to the *gbest* (not to the entire swarm) for a low-cost solution. Since the problem of poor *gbest* update persists in the recently proposed extension of PSO, called multi-dimensional PSO (MD-PSO), two distinct SA approaches are then integrated into MD-PSO and tested over a set of unsupervised data clustering applications. Experimental results show that the proposed approaches significantly improved the quality of the MD-PSO clustering as measured by a validity index function. Furthermore, the proposed approaches are generic as they can be used with other PSO variants and applicable to a wide range of problems.

Keywords: Particle Swarm Optimization, stochastic approximation, multi-dimensional search, gradient descent, dynamic data clustering.

1 Introduction

The particle swarm optimization (PSO) [4,10,11] exhibits certain similarities with the other evolutionary algorithms (EAs) [2]. The common point of all is that EAs are in population based nature and they can avoid being trapped in a local optimum. Thus they can find the optimum solutions; however, this is never guaranteed. In a PSO process, a swarm of particles (or agents), each of which represent a potential solution to an optimization problem, navigate through the search (or solution) space. One major drawback of PSO is the direct link of the information flow between particles and the global-best particle, *gbest*, which primarily "guides" the rest of the swarm and thus resulting in the creation of similar particles with some loss of diversity. Hence this phenomenon increases the probability of being trapped in local optima [7] and it is the main cause of the premature convergence problem especially when the search

[*] This work was supported by the Academy of Finland, project No. 213462 (Finnish Centre of Excellence Program (2006 - 2011).

C. Di Chio et al. (Eds.): EvoApplications 2010, Part I, LNCS 6024, pp. 336–343, 2010.

space is of high dimensions [10] and the problem to be optimized is multi-modal [7]. This makes it clear that at any iteration of a PSO process, *gbest* is the most important particle; however, it has the poorest update equation, i.e. when a particle becomes *gbest*, it resides on its personal best position (*pbest*) and thus both social and cognitive components are nullified in the velocity update equation. Although it guides the swarm during the following iterations, ironically it lacks the necessary guidance to do so effectively. In that, if gbest is (likely to get) trapped in a local optimum, so the rest of the swarm due to the aforementioned direct link of information flow. This deficiency has been raised in a recent work [5] where an artificial GB particle, the *aGB*, is created at each iteration as an alternative to *gbest*. However, the underlying mechanism for creating the *aGB* particle, the so-called fractional GB formation (FGBF), is not generic, rather problem dependent.

For the problem of finding a root θ^* (either minimum or maximum point) of the gradient equation: $g(\theta) \equiv \frac{\partial L(\theta)}{\partial \theta} = 0$ for some differentiable function $L : R^p \rightarrow R^1$, when g is present and L is a uni-modal function, there are powerful deterministic methods for finding the global θ^* such as traditional steepest descent and Newton-Raphson methods. However, in many real problems *g* cannot be observed directly and/or L is in multi-modal nature, which presents many deceiving local optima. This brought the era of the stochastic optimization algorithms, which can estimate the gradient and may avoid being trapped into a local optimum due to their stochastic nature. One of the most popular stochastic optimization techniques is stochastic approximation (SA), in particular the form that is called “gradient free” SA. Among many SA variants the one and somewhat different SA application is called simultaneous perturbation SA (SPSA) proposed by Spall in [8].

In this paper we shall propose two approaches, one of which drives *gbest* efficiently or simply put, *guides* it with respect to the function (or error surface). The idea behind this is quite simple: since the velocity update equation of *gbest* is quite poor, SPSA as a simple yet powerful search technique is used to *drive* it instead. The second approach has a similar motivation with the FGBF proposed in [5], i.e., an artificial Global Best (*aGB*) particle is created by SPSA this time, which is applied over the personal best (*pbest*) position of the *gbest* particle. The *aGB* particle will then guide the swarm instead of *gbest* if and only if it achieves a better fitness score than the (personal best position of) *gbest*. Note that both approaches *only* deal with the *gbest* particle and hence the internal PSO process remains as is. They are then applied to the multi-dimensional extension of PSO, the MD-PSO technique proposed in [5], which can find the optimum dimension of the solution space. SA-driven (SAD) MD-PSO is then tested and evaluated against the standalone MD-PSO application over several data clustering problems.

2 Proposed Technique: SAD MD-PSO

2.1 SPSA Overview

The goal of the deterministic optimization methods is to minimize a loss function $L : R^p \rightarrow R^1$, which is a differentiable function of θ and the minimum (or maximum) point θ^* corresponds to zero-gradient point, i.e.