Parallelization of a Modified Firefly Algorithm using GPU for Variable Selection in a Multivariate Calibration Problem

Lauro C. M. de Paula, Instituto de Informática, Universidade Federal de Goiás, Goiânia, Goiás, Brazil

Anderson S. Soares, Instituto de Informática, Universidade Federal de Goiás, Goiânia, Goiás, Brazil

Telma W. L. Soares, Instituto de Informática, Universidade Federal de Goiás, Goiânia, Goiás, Brazil

Alexandre C. B. Delbem, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, São Paulo, Brazil

Clarimar J. Coelho, Departamento de Computação, Pontifícia Universidade Católica de Goiás, Goiânia, Goiás, Brazil

Arlindo R. G. Filho, Departamento de Sistemas e Controle, Instituto Tecnológico de Aeronáutica, São José dos Campos, São Paulo, Brazil

ABSTRACT

The recent improvements of Graphics Processing Units (GPU) have provided to the bio-inspired algorithms a powerful processing platform. Indeed, a lot of highly parallelizable problems can be significantly accelerated using GPU architecture. Among these algorithms, the Firefly Algorithm (FA) is a newly proposed method with potential application in several real world problems such as variable selection problem in multivariate calibration. The main drawback of this task lies in its computation burden, as it grows polynomially with the number of variables available. In this context, this paper proposes a GPU-based FA for variable selection in a multivariate calibration problem. Such implementation is aimed at improving the computational efficiency of the algorithm. For this purpose, a new strategy of regression coefficients calculation is employed. The advantage of the proposed implementation is demonstrated in an example involving a large number of variables. In such example, gains of speedup were obtained. Additionally the authors also demonstrate that the FA, in comparison with traditional algorithms, can be a relevant contribution for the variable selection problem.

Keywords: Firefly Algorithm (FA), Graphics Processing Units (GPU), Multiple Linear Regression, Parallelization, Variable Selection

DOI: 10.4018/ijncr.2014010103
1. INTRODUCTION

Multivariate calibration refers to construction procedure of a mathematical model that establishes the relationship between the properties measured by an instrument and the concentration of a sample to be determined (Brown et al., 1994). The building of a model from a subset of explanatory variables usually involves two conflicting objectives:

- Extracting as much information from a measured data with many possible independent variables;
- Decreasing the cost of obtaining data by using the smallest set of independent variables that results in a model with high accuracy and low variance.

The balance between these two commitments is achieved using variable selection techniques. The application of multivariate calibration had a breakthrough and nowadays are very popular (Ferreira et al., 1999). One of the most interesting features of modern instrumental methods is the number of variables that can be measured in a single sample. Recently, devices as spectrophotometers have generated large amount of data with thousands of variables. As a consequence, the development of efficient algorithms for variable selection is important in order to deal with data even larger (George, 2000; Coifman & Wickerhauser, 1992; Paula et al., 2013). Furthermore, a high-performance computing framework can significantly contribute to efficiently construct an accurate model (Chau et al., 2004). In this context, this paper proposes an implementation of a modified Firefly Algorithm (FA) for variable selection in a multivariate calibration problem (George, 2000; Westad, 2000; Coifman & Wickerhauser, 1992; Filho et al., 2011).

The FA is a metaheuristic inspired by the flashing behaviour of fireflies (Yang, 2008). Recent works have used FA to solve various types of problems. For instance, Yang (2009) provides a detail description of a new FA for multimodal optimization applications. Lukazik and Zak (2009) provide an implementation of the FA for constrained continuous optimization. Yang (2010) shows the use of the FA for nonlinear design problems. Senthilnath et al. (2011) use the FA for clustering on benchmark problems and compares its performance with other nature-inspired techniques. Gandomi et al. (2011) use the FA for mixed-continuous and discrete-structural optimization problems. Jati et al. (2011) apply the FA for travelling salesman problem. Banati and Monika (2011) propose a new feature selection approach that combines the Rough Set Theory with the nature-inspired FA to reduce the dimensionality of data containing large number of features. Horng (2012) proposes a new method based on the FA to construct the codebook of vector quantization for image compression. Finally, Abdullah et al. (2013) introduce a new hybrid optimization method incorporating the FA and the evolutionary operation of the differential evolution method.

In all these works, the experimental results show that the FA scores over other algorithms in terms of computing time and optimality. As far as we know, there is no work in the literature that uses the FA to select variables in multivariate calibration problems.

This paper also investigates the use of a Graphics Processing Unit (GPU) to parallelize the computation of the vector of regression coefficients in the problem of Multiple Linear Regression (MLR). GPUs have been employed to improve performance of computing applications usually handled by a Central Processing Unit (CPU). Husselman and Hawick (2012 and 2014) are the only works we have found so far that present a GPU-based implementation of a FA. Furthermore, estimates from the proposed FA (FA-MLR) are compared with predictions from the following traditional algorithms: Successive Projections Algorithm for MLR (SPA-MLR) (Filho et al., 2011), Genetic Algorithm for MLR (GA-MLR) (Soares et al., 2013) and Partial Least Squares (PLS) (Haaland & Thomas, 1988).

The remaining of this paper is organized as follows. Section 2 describes the multivariate
Anti-Synchronization of Pan and Lorenz-Lu-Liu-Cai Chaotic Systems by Active Nonlinear Control
www.igi-global.com/article/anti-synchronization-pan-lorenz-liu/74333?camid=4v1a