International Conference on Advances in Bioprocess Engineering and Technology 2016 (ICABET 2016)

Selection of Relevant Genes for Pediatric Leukemia using Co-Operative Multiswarm

Prativa Agarwalla[a]*, Sumitra Mukhopadhyay[b]

aHeritage Institue of Technology, Anandapur, Kolkata-107, India
bInstitute of Radiophysics & Electronics, Kolkata-009, India

Abstract

DNA microarray has gained an increasing popularity for the experimental study of gene-expression for cancer detection. The selection of relevant genes is very important for medical diagnosis as well as to understand how a genome as a whole works for a particular disease. We have proposed a Co-operative Multiswarm Particle Swam Optimization (CMPSO) based technique to find co-regulated gene groups associated with cancer sample categories. The proposed CMPSO is applied to pediatric leukemia dataset and the result is compared with conventional techniques. Proposed multiswarm based technique generates more precise groups of genes which provide important contribution for pediatric leukemia detection.

1. Introduction

Pediatric Leukaemia, cancer of the white blood cells, is the most frequent type of cancer among the children. Worldwide, per year the incidence of childhood cancer is more than 175,000, and the mortality rate is approximately 96,000 per year [1]. To diagnose a new case of pediatric cancer, selection of biologically relevant genes which are

* Corresponding author. Tel.: +91-8017274450;
E-mail address:prativa.agarwalla@heritageit.edu.

contributing to the cancer, is one of the major task. For a functional genomics study of cancer, one of the most popular and powerful tool is DNA microarray[2],which arranges thousands of gene sequence of different samples in a single microscopic slide.The huge number of relevant, irrelevant and noisy genes presented in microarray data makes a major challenge for the relevant gene selection for cancer detection. Several methods for informative gene selection have been proposed such as statistical techniques [8], Fuzzy clustering based approach [9], Pareto front based technique [10] etc. Generally the resulting top ranked relevant genes or features which are produced by these techniques, are often redundant. Several bio-inspired optimizations are also proposed to select informative genes from microarray by partitioning the thousand of genes into biologically meaningful subsets [3]. Among various existing bio-inspired optimization techniques, PSO is one of the popular techniques which is successfully applied for the clustering of cancer genes [7], [11], [12]. Monalisa et.al [12] has applied Binary Particle swarm optimization based approach to identify the pathway markers from the gene expression data. Again, multi-objective Pareto front based PSO technique is applied for the identification of non-redundant gene markers [7]. Also, the relevant genes are selected using hybrid PSO [11] based clustering method. Generally, these techniques are used to find accurate cluster centre formulating the problem as an optimization problem.

   In case of generalized optimization problems, multi-swarm PSO [13] approach produces better result than the conventional PSO [4] as its gives better diversity to the swarm and the problem of local trapping and premature convergence can be avoided. Thus multi swarm PSO can be applied in the domain of gene selection problem to find more accurate and well separated cluster centres which can be used to select functionally active genes from the huge number of genes present in microarray data.

In this paper, we have proposed multi-swarm based co-operative Particle Swarm Optimization (CMPSO) based gene selection approach that can generate accurate gene expression centre which represents the signature genes for the particular class of cancer. Proposed CMPSO avoids the problem of local trapping and premature convergence of the original PSO and provides a good diversity to the swarm. First, the genes are pre-processed for the computation purpose and noisy genes are filtered out using a CMPSO based filtering technique. To separate the genes in well separated clusters, cluster centroids are identified and after clustering the top ranked genes are selected. The proposed CMPSO technique helps to provide more accurate cluster centres when clustering of genes is formulated as an optimization problem. We have applied the CMPSO based gene selection technique to some pediatric cancer dataset and also the performance of the proposed technique is compared with other some popular techniques [6]-[8], [12] to validate its superiority.

   The remaining section of the paper is organized as follows: Section.2 presents a general overview of the traditional clustering techniques, a PSO algorithm and the application of PSO in clustering algorithm. Section.3 describes the new proposed CMPSO variant, and the proposed CMPSO based gene selection technique. In the section.4 we have described the result of the proposed technique when applied on pediatric leukemia dataset. Next the biological relevance of the result is given. Finally, Section 6 provides the conclusion.

2. Preliminaries

2.1. Concept of PSO algorithm

   Particle Swarm Optimization algorithm (PSO), originally developed by Eberhart and Kennedy in 1995 [4], inspired by the social behaviour of a bird flock. In the PSO algorithm, swarms are considered as particles whose location in the multi-dimensional problem space represents the possible solutions for the problem. The velocity and direction of each particle moving along each dimension of the problem space will be altered with each generation of movement. The velocity and position update rule for i^{th} particle at t^{th} generation are given below where c_1, the cognitive learning factor, represents the attraction of a particle toward its own success p_{best} and c_2, social learning factor, represents the attraction of a particle toward the swarm's best position g_{best}, and w is the inertia weight.

$$v^i(t) = w * v^i(t-1) + c_1 * rand * (p_{best}^i - x^i) + c_2 * rand * (g_{best}^i - x^i) \tag{1}$$

$$x^i(t) = x^i(t-1) + v^i(t) \tag{2}$$