Chapter 8

Application of Optimization Techniques for Gene Expression Data Analysis

Suresh Dara
DIT University, India

Arvind Kumar Tiwari
DIT University, India

ABSTRACT

The feature selection from gene expression data is the NP hard problem, few of evolutionary techniques give optimal solutions to find feature subsets. In this chapter, authors introduce some evolutionary optimization techniques and proposed a Binary Particle Swarm Optimization (BPSO) based algorithm for feature subset selection. The Feature selection is one of the important and challenging tasks for gene expression data where many traditional methods failed and evolutionary based methods were succeeded. In this study, the initial datasets are preprocessed using a quartile based fast heuristic technique to reduce the crude domain features which are less relevant in categorizing the samples of either group. The experimental results on three bench-mark datasets vis-a-vis colon cancer, defused B-cell lymphoma and leukemia data are evaluated by means of classification accuracies. Detailed comparative studies with some of popular existing algorithms like Genetic Algorithm (GA), Multi Objective GA are also made to show the superiority and effectiveness of the proposed method.

DOI: 10.4018/978-1-5225-2545-5.ch008
1. INTRODUCTION

The feature selection problem is a high-dimensional optimization problem in the nature and thus needs a solver with high exploration power. On the other hand, if alternative optimal solutions could be provided for a problem, the implementation phase may become more selective depending on the cost and limitations of domain of the problem. The high exploration power and solution conservation capability of optimization methods make them able to find multiple suitable solutions in a single run. Therefore, optimization methods can be considered as a powerful tool of finding suitable feature subsets for feature selection problem. Figure 1 shows that, various optimization techniques in hierarchal format. Here, few of traditional approaches and nontraditional approaches i.e., heuristic approaches are presented.

Feature Selection (FS) is the process of selecting optimal feature subset from a given dataset that can interpret the target concept. Any feature selection algorithm completes in four steps viz. subset generation, subset evaluation, stopping criterion and result validation. Subset generation is the process of searching subsets from the given feature space and then selected subset is evaluated to determine the goodness of feature subset under consideration. The termination of the algorithm is decided by stopping criterion. Finally, validation is performed to identify legitimate feature subsets (Tou & Gonz’alez, 1994). Feature selection has been explored extensively by researchers of data mining and machine learning since 1970s. It has exploited many domains like machine learning, data mining, pattern recognition and other related domains such as software engineering, text categorization, bioinformatics, image retrieval, intrusion detection, information and music retrieval (Mitra & Acharya, 2003).

The microarray experiments produce gene expression patterns that provide dynamic information about cell function. In a single experiment, the DNA microarray technologies can simultaneously monitor and analysis of thousands of different genes in histological or cytological specimens which helps to find diseased samples according to the different levels of expression profiles. Gene expression profiles usually contain a large number of genes but a small number of samples. An important need to analyze and interpret the huge amount of data, involving the decoding of around 24000-30000 human genes (Special Issue on Bioinformatics, 2002) is found to be an NP-Hard problem (Skowron & Rauszer, 1992).

High dimensional feature selection technique may help us to identify important features by applying certain selection criteria which reduces the computation cost and increases the classification accuracy. DNA microarray technologies have been
On the System Algebra Foundations for Granular Computing
www.igi-global.com/article/system-algebra-foundations-granular-computing/2786?camid=4v1a

Quotient Space-Based Boundary Condition for Particle Swarm Optimization Algorithm
www.igi-global.com/article/quotient-space-based-boundary-condition/53164?camid=4v1a