Chapter 33
Cancer Biomarker Assessment Using Evolutionary Rough Multi-Objective Optimization Algorithm

Anasua Sarkar
Government College of Engineering and Leather Technology, India

Ujjwal Maulik
Jadavpur University, India

ABSTRACT

A hybrid unsupervised learning algorithm, which is termed as Evolutionary Rough Multi-Objective Optimization (ERMOO) algorithm, is proposed in this chapter. It comprises a judicious integration of the principles of the rough sets theory with the archived multi-objective simulated annealing approach. While the concept of boundary approximations of rough sets in this implementation deals with the incompleteness in the dynamic classification method with the quality of classification coefficient as the classificatory competence measurement, it enables faster convergence of the Pareto-archived evolution strategy. It incorporates both the rough set-based dynamic archive classification method in this algorithm. A measure of the amount of domination between two solutions is incorporated in this chapter to determine the acceptance probability of a new solution with an improvement in the spread of the non-dominated solutions in the Pareto-front by adopting rough sets theory. The performance is demonstrated on real-life breast cancer dataset for identification of Cancer Associated Fibroblasts (CAFs) within the tumor stroma, and the identified biomarkers are reported. Moreover, biological significance tests are carried out for the obtained markers.

INTRODUCTION

The progress of microarray technology in the field of cancer research has enabled scientists to measure the molecular signatures of cancer cells. The scientists today monitor the expression levels for differentially expressed cancer genes simultaneously over different time points under different drug treatments

DOI: 10.4018/978-1-5225-8903-7.ch033

(Tusher, 1940). In microarray analysis, the expression levels of two genes may rise and fall synchronously in response to environmental stimuli (Tusher, 1940), (Eisen, 1998). The efficient machine learning classifiers help in the diagnosis of cancer sub types for patients (Spang, 2003).

In recent times, researchers experiment for developing computational methods for analysis of RNA and gene expression profiles for oncology detection. Such computational methods are expected to promote the experimental work that needs to be carried out in the wet laboratory for analyzing biomarker RNAs. Gene expression profiling of breast tumors stratifies into breast cancer of different molecular subtypes which also co-segregate with the receptor status of the tumor cells. Therefore cancer associated fibroblasts (CAFs) within the tumor stroma may exhibit subtype specific gene expression profiles. These onco-RNA signatures may be further analyzed to find out the most significant oncological biomarkers computationally.

Clustering is one unsupervised classification method based on maximum intra-class similarity and minimum inter-class similarity. Historically Eisen et al. (Eisen, 1998) first classified groups of co-expressed genes using hierarchical clustering. Other already proposed clustering, which can be applied for cancer subtype detection are: self-organizing map (SOM) (Spang, 2003), K-Means clustering (Tavazoie, 2001), (Hoon, 2004), simulated annealing (Lukashin, 1999), graph theoretic approach (Xu, 1999), fuzzy c-means clustering (Dembele, 2003), spectral clustering (Maulik, 2013), (Sarkar, 2011), scattered object clustering (de Souto, 2008) and symmetry based clustering (Maulik, 2012), (Sarkar, 2009). Several other methods like (Maulik, 2009), (SarKar, 2009), (Bandyopadhyay, 2010) are also which may be applicable efficiently for cancer subtype detection problem.

Earlier approaches worked on detecting cancer biomarker using a hybrid method of Genetic Algorithm and all paired (AP) support vector machine (svm) approaches (Liu, 2005). Multi-category classification SVM (MC-SVM) approaches have also been explored in Gene Expression Model Selector (GEMS) system (Statnikov, 2005). Feature ranking scores for feature selection have also been experimented in SVM-RFE approach (Duan, 2005). Therefore, we can implement a method which would simultaneously optimize the features and select biomarkers considering multiple objectives.

The multi-objective optimization *(MOO)* involves the simultaneous optimization of two or more conflicting objectives, forming the Pareto-optimal *(PO)* or non-dominated set of solutions with equal importance. The perspective is different from the single-objective optimization problem with only one global optimum, which lacks significance for most real-world problems with multiple objectives.

Over the decade, a number of multi-objective techniques based on Evolutionary Algorithm *(EA)* have been suggested (Deb, 2001), (Coello, 2002), (Srinivas, 1995), (Zitzler, 1998), (Maulik, 2010), (Ganesan, 2013). The unique features behind the success of *EA* for solving *MOO* problems is their population based nature and ability to find multiple Pareto-optima simultaneously. Some of them consider soft computing approaches (Elamvazuthi, 2013), (Jimenez, 2013).

Simulated Annealing *(SA)* (Kirkpatrick, 1983) is another method utilizing principles of statistical mechanics to find minimal cost solutions for large optimization problems by minimizing the associated energy. Since each execution of the *SA* method converges to one global optimum solution, the Multi-Objective Simulated Annealing *(MOSA)* approach evolves the *PO* set of solutions in multiple *SA* runs. But the diversity of *PO* set of solutions suffers for those independent runs. Therefore, Pareto-dominance based acceptance criterion has been incorporated in *MOSA* (Suman, 2005), (Smith, 2004). To consider the amount of domination between new solution and the *PO* solutions stored in *Archive* including the current solution, a novel Archived Multi-Objective Simulated Annealing *(AMOSA)* algorithm has been proposed recently by Bandyopadhyay et.al (Bandyopadhyay, 2008) in 2008. This improved method uti-