Chapter 1
Metaheuristic-Based Hybrid Feature Selection Models

Sujata Dash
North Orissa University, India

ABSTRACT

This chapter focuses on key applications of metaheuristic techniques in the field of gene selection and classification of microarray data. The metaheuristic techniques are efficient in handling combinatorial optimization problems. In this article, two different types of metaheuristics such as Genetic algorithm (GA) and Particle Swarm Optimization (PSO) are hybridized with fuzzy-rough (FR) method for optimizing the subset selection process of microarray data. The FR method applied here deals with impreciseness and uncertainty of microarray data. The predictive accuracy of the models is evaluated by an adaptive neural net ensemble and by a rule based classifier MODLEM respectively. Moreover, the learning efficiency of the ensemble is compared with base learners and with two classical ensembles. The rule based classifier generates a set of rules for disease diagnosis and prognosis and enables to study the function of genes from gene ontology website. The experimental results of both the models prove that, hybrid metaheuristic techniques are highly effective for finding potential genes.

1. INTRODUCTION

Microarray technology produces high dimensional datasets by measuring the expression levels of tens of thousands of genes in a single experiment under varying conditions. It has become an indispensable tool for biological, medical and pharmaceutical researchers to get a better understanding of the diseases at genomic level. On the other hand, the inherent problem i.e., large number of features and small sample size of microarray dataset makes the analysis process difficult for the problem. Typically, are latively small numbers of features are found to be strongly correlated with the phenotypes in question? Therefore, to identify these discriminative features from gene expression dataset, a data mining tool (Witten, Frank & Hall, 2011) known as feature selection technique plays an important role. The methods specifically used for feature selection can be categorized into two major groups namely, filter and wrapper methods (Saeys, Inza&Larranaga, 2007; Guyon, Nikravesh, Zadeh, 2006; Dash& Patra, 2016a).

DOI: 10.4018/978-1-5225-2857-9.ch001

Filter methods select features considering individual characteristic of each feature without taking into account the mutual dependencies among features. Then the features are sorted by their assigned ranks. The top ranked features are kept for further analysis by removing low ranked features. Actually, these selected features are used to develop the diagnostic model to efficiently predict the diseases. On the contrary, in wrapper methods, a search algorithm is wrapped around a learning algorithm: so that an estimated learning accuracy for all subsets can be calculated to derive an optimal one. This method is computationally intensive in comparison to filter methods because to obtain an optimal subset of features all possible subsets need to be examined which is practically a difficult task. Thus, to alleviate this difficulty in wrapper methods a metaheuristic search strategy (Ghosh & Jain (Eds),2005; Akadi, Amine, Ouardighi & Aboutajdine, 2009; Dash, 2016) may be adopted from a set of heuristic or stochastic algorithms which can be able to generate an optimal subset effectively.

In the gene selection process, an optimal gene subset is always relative to a certain criterion. Several information measures such as entropy, mutual information (Ding& Peng, 2005)and *f*-information (Maji, 2009) have successfully been used in selecting a set of relevant and non redundant genes from a microarray data set. An efficient and effective reduction method is necessary to cope with large amount of data by most techniques. Growing interest in developing methodologies that are capable of dealing with imprecision and uncertainty is apparent from the large scale research that are currently being done in the areas related to fuzzy (Zadeh, 1965) and rough sets. Rough set theory (RST) was introduced by Pawlak (1982) and has been used widely by researchers as a classifier and selection technique. The success of rough set theory is due to three aspects of the theory. First, only hidden facts in the data are analyzed. Second, additional information about the data is not required for data analysis. Third, it finds a minimal knowledge representation for data. Due to this Rough set theory is used in complement with other concepts such as, fuzzy set theory. The two fields may be considered similar in the sense that both can tolerate inconsistency and uncertainty. The only difference among these two fields is the type of uncertainty and their approach to it. While fuzzy sets are concerned with vagueness, rough sets are concerned with indiscernibility. The fuzzy- rough set-based approach considers the extent to which fuzzified values are similar.

The ensemble learning approach constructs several classifier models for the original dataset and then combines the predictive outputs to identify an unknown sample. The motivation of combinings several classifiers is to improve the classification efficiency which in turn depends on the accuracy and diversity (Yang P., Yang H., Bing, Zomaya, 2010) of the base classifiers. The ensemble techniques very popular in the field of classification and pattern recognition as it increases the generalization and percentage of classification by aggregating (Chen, Hong, Deng, Yang, Wei & Cui, 2015) the outcome of finite number of neural network classifiers (Lee, Hong & Kim, 2009a). However, neural network ensemble learning has been used in many problems, such as, face recognition (Lee, Hong &Kim, 2009 b), digital image processing (Liu, Cui, Jiang & Ma, 2004) and medical diagnosis (Huang,Zhou, Zhang & Chen, 2000) and has given outstanding performance in terms of classification accuracy.

In the last 30 years, a new kind of approximate algorithm has emerged which basically tries to combine basic heuristic methods in higher level frameworks aimed at efficiently and effectively exploring asearch space. These methods are nowadays commonly called *metaheuristics*. The term *metaheuristic*, first introduced in (Glover, 1986), derives from the composition of two Greek words. *Heuristic* derives from the verb *heuriskein,* which means “to find”, while the suffix *meta* means “beyond, in an upper level”. Before this term was widely adopted, metaheuristics were often called *modern heuristics* (Reeves, 1993).This class of algorithms includes but is not restricted to, Ant Colony Optimization (ACO), Evolu-