Chapter 34
Analysis of Microarray Data Using Artificial Intelligence Based Techniques

Khalid Raza
Jamia Millia Islamia, India

ABSTRACT

Microarray is one of the essential technologies used by the biologists to measure genome-wide expression levels of genes in a particular organism under some particular conditions or stimuli. As microarrays technologies have become more prevalent, the challenges of analyzing these data for getting better insight about biological processes have essentially increased. Due to availability of artificial intelligence based sophisticated computational techniques, such as artificial neural networks, fuzzy logic, genetic algorithms, and many other nature-inspired algorithms, it is possible to analyse microarray gene expression data in a better way. In this chapter, we present artificial intelligence based techniques for the analysis of microarray gene expression data. Further, challenges in the field and future work direction have also been suggested.

INTRODUCTION

The bioinformatics is an interdisciplinary area of study where one of the objectives is to deal with the analysis and interpretation of large sets of data generated from various large-scale biological experiments. The example of one such large-scale biological experiment is measuring the expression levels of tens of thousands of genes simultaneously under some environmental condition. Microarray is one of the essential technologies used by the biologists to measure genome-wide expression levels of genes in a particular organism. As microarrays technologies have become more prevalent, the challenges associated with collecting, managing, and analyzing the data from each experiment have essentially increased. Robust laboratory protocols, improved understanding of the complex experimental design and falling prices of commercial platforms, all these have combined to drive the field to more complex experiments, generating huge amounts of data (Brazma & Vilo, 2000).

DOI: 10.4018/978-1-5225-8903-7.ch034

With the help of measured transcription levels of genes under different biological conditions (e.g. at various developmental stages and in different tissues), biologists are able to develop gene expression profiles that differentiate the functionality of each gene in the genome. The gene expression profiles are organized in the form of a matrix, where rows represent genes, columns represent samples/replicas, and each cell of the matrix contains a numeric value representing the expression level of a gene in a particular sample. Generally, such a table is called gene expression matrix. These gene expression matrices can be used to correlate the over expression of certain genes with a certain disease and can help researchers to discover other conditions affecting the expression level of these genes. The gene expression matrix can also be used to identify other set of genes having co-expression profile patterns. Hence, suitable compounds (potential drugs) can be investigated that can lower the expression level of these overexpressed genes (Babu, 2004).

Many sophisticated statistical and computational tools have been developed to help biologists analyse gene expression data and to identify novel targets from their experimental data (Deng et al., 2009; Debouck et al., 1999). Among these techniques, clustering and statistical methods are most commonly used data analysis methods. Clustering generally groups the gene expression data with similar expression pattern, i.e. co-expressed genes. However, clustering approach suffers from several drawbacks (Bassett et al., 1999). The statistical methods help to analyse gene expression data and infer relationships between genes. However, it fails to provide complex regulatory relationships among genes.

The chapter is organized as follows. Section 2 describes the background of Microarray experiments and data generation. Section 3 covers the applications of Microarrays and Section 4 describes artificial intelligence based techniques, and reviews its application in the analysis of Microarray data. Section 5 summarizes the chapter and presents research challenges and future work directions.

Microarray Technology

With the help of Microarray technology, one can measure the expression level of all genes in a genome simultaneously. By measuring and comparing the expression level of genes in an normal and diseased cell, it would be possible to identify genes which are responsible for various diseases. Due to unprecedented amount of large biological data generated out of microarray experiments, the focus of research has shifted from the generation of data to analysis, interpretation, and presentation of data in the most efficient manner (Hood, 2003; Kitano, 2002; Kitano, 2002a). With the help of these technologies, researchers can find answers to some of the challenging questions like;

1. What are the functions of different genes?
2. In what cellular processes do these genes participate?
3. How genes are regulated?
4. How genes and their products (proteins) do interact, and what are these interaction networks?
5. How expression level of genes differs in different cell types and states?
6. How expressions of genes are affected by various disease or drug treatments?

Microarrays are frequently used in biomedical research to tackle a number of problems, including classification of tumors, and gene expression response to different stress conditions. A central and frequently asked question in microarray is the identification of differentially expressed genes (DEGs). The DEGs are those genes whose expression levels are associated with a response or covariate of interest