International Journal of Innovative Computing, Information and Control
Volume **5**, Number **8**, August **2009**

ICIC International ©2009 ISSN 1349-4198
pp. **2189–2202**

A CYCLIC HYBRID METHOD TO SELECT A SMALLER SUBSET OF INFORMATIVE GENES FOR CANCER CLASSIFICATION

Mohd Saberi Mohamad, Sigeru Omatu, Michifumi Yoshioka

Department of Computer Science and Intelligent Systems
Graduate School of Engineering
Osaka Prefecture University
Sakai, Osaka 599-8531, Japan
mohd.saberi@sig.cs.osakafu-u.ac.jp; { sigeru; yoshioka }@cs.osakafu-u.ac.jp

Safaai Deris

Department of Software Engineering
Faculty of Computer Science and Information Systems
University Teknologi Malaysia
81310 Skudai, Johore, Malaysia
safaai@utm.my

Received July 2008; revised November 2008

Abstract. *Microarray data are expected to be useful for cancer classification. The main problem that needs to be addressed is the selection of a smaller subset of genes from the thousands of genes in the data that contributes to a cancer disease. This selection process is difficult due to many irrelevant genes, noisy data, and the availability of the small number of samples compared to the huge number of genes (higher-dimensional data). Hence, this paper aims to select a smaller subset of informative genes that is the most relevant for the cancer classification. To achieve the aim, a cyclic hybrid method has been proposed. Five real microarray data sets are used to test the effectiveness of the method. Experimental results show that the performance of the proposed method is superior to other experimental methods and related previous works in terms of classification accuracy and the number of selected genes. In addition, a scatter gene graph and a list of informative genes in the best gene subsets are also presented for biological usage.*
Keywords: Cancer classification, Gene selection, Genetic algorithm, Cyclic hybrid method, Microarray data

1. **Introduction.** Advances in the area of microarray-based gene expression analysis have led to a promising future of cancer diagnosis using new molecular-based approaches. Microarray technology is used to measure the expression levels of thousands of genes simultaneously, and finally produce microarray data. A comparison between the gene expression levels of cancerous and normal tissues can also be done. This comparison is useful to select those genes that might anticipate the clinical behaviour of cancers. Thus, there is a need to select informative genes that contribute to a cancerous state. However, the gene selection process poses a major challenge because of the following characteristics of microarray data: the huge number of genes compared to the small number of samples (higher-dimensional data), irrelevant genes, and noisy data.

To overcome the challenge, a gene selection method is used to select a subset of genes that maximises the classifier's ability to classify samples more accurately. The gene selection method has several advantages such as improving classification accuracy, reducing the dimensionality of data, and removing irrelevant and noisy genes.