International Journal of Innovative
Computing, Information and Control
Volume **7**, Number **5(A)**, May **2011**

ICIC International ©2011 ISSN 1349-4198
pp. **2235–2246**

A SIMPLIFIED SWARM OPTIMIZATION FOR DISCOVERING THE CLASSIFICATION RULE USING MICROARRAY DATA OF BREAST CANCER

Wei-Chang Yeh, Wei-Wen Chang and Chuan-Wei Chiu

Department of Industrial Engineering and Engineering Management
National Tsing Hua University
No. 101, Sec. 2, Kuang-Fu Rd., Hsinchu 30013, Taiwan
changwerwen@gmail.com

Received December 2009; revised March 2010

Abstract. *Microarray data analysis is a major line of research in bioinformatics. A significant trend in bioinformatics is identifying genes or gene groups that differentiate diseased tissues. Classification is necessary to make microarray data useful for application in medicine, and in related research such as disease diagnosis. Classification models have been developed using statistical methods such as logistic and multi-normal regression for data mining. However, the complexities of real-world classification problems, such as those in the medical domain, are highly dimensional. General statistical methods are inadequate for these complex problems. This study proposes simplified swarm optimization (SSO), an efficient methodology for discovering breast cancer classification rules. The data set was derived from the Stanford microarray database. The proposed approach enables simultaneous feature selection and pattern recognition. Experimental results indicate that SSO outperforms general data mining methods such as decision tree, neural network, support vector machine, etc. The proposed approach has potential applications in hospital decision-making and research such as predictive medicine.*
Keywords: Microarray data, Simplified swarm optimization, Pattern recognition

1. **Introduction.** Recent studies of tumor cell-specific gene expression patterns use the molecular characteristics of tumor tissue for diagnosis. Since microarray technology can screen thousands of genes simultaneously, microarray data analysis is considered a major advance in tumor diagnosis. Microarray data analysis is a major research area in bioinformatics research. An important bioinformatics trend is the identification of genes or groups of genes to differentiate diseased from normal tissues. However, a major problem in bioinformatics is using the identified genes to classify tissues as cancerous or normal. Advances in microarray technology are steadily increasing the body of available data. It is essential to identify the biologically relevant groups of genes and/or samples using some data mining techniques [1]. For practical application of microarray data in medicine and in related research such as predictive medicine, efficient classification is essential. Classification involves the partitioning of a target variable into predefined groups or classes. A classification system uses labeled data instances to determine the target variables of new data instances. The discovered knowledge is generally presented as if-then prediction rules, which have the advantage of being high level, symbolic knowledge representations, that improve comprehensibility of the discovered knowledge [2]. Microarray analysis also provides quantitative information about cell transcription profiles. To analyze microarray data sets, machine learning methodologies have been applied by bioinformatics researchers. Some machine learning approaches are widely used to classify and mine biological data sets