International Journal of Innovative
Computing, Information and Control
Volume **8**, Number **6**, June **2012**

ICIC International ©2012 ISSN 1349-4198
pp. **4285–4297**

SELECTING A SMALL SUBSET OF INFORMATIVE GENES FROM GENE EXPRESSION DATA BY USING A MODIFIED BINARY PARTICLE SWARM OPTIMISATION

Mohd Saberi Mohamad[1,*], Sigeru Omatu[2], Safaai Deris[1]
Michifumi Yoshioka[3] and Zuwairie Ibrahim[4]

[1]Artificial Intelligence and Bioinformatics Research Group
Faculty of Computer Science and Information Systems
Universiti Teknologi Malaysia
81310 Skudai, Johor, Malaysia
[*]Corresponding author: saberi@utm.my

[2]Department of Electronics, Information and Communication Engineering
Osaka Institute of Technology
5-16-1 Omiya, Asahi-ku, Osaka 535-8585, Japan

[3]Department of Computer Science and Intelligent Systems
Graduate School of Engineering
Osaka Prefecture University
1-1 Gakuen-cho, Nakaku, Sakai, Osaka 599-8531, Japan

[4]Faculty of Electrical and Electronic Engineering
Universiti Malaysia Pahang
26600 Pekan, Pahang, Malaysia

Received February 2011; revised September 2011

Abstract. *Gene expression technology, especially microarrays, can be used to measure the expression levels of thousands of genes simultaneously in biological organisms. Gene expression data produced by microarrays are expected to be useful for cancer classification. To select a small subset of informative genes for cancer classification, many researchers have analysed the gene expression data using various computational intelligence methods. However, due to the small number of samples compared with the huge number of genes (high-dimensional data), irrelevant genes, and noisy genes, many of the computational methods face difficulties in selecting the small subset. Thus, we propose a modified binary particle swarm optimisation to select a small subset of informative genes that are relevant for the cancer classification. In the proposed method, we introduce the particle speed and a rule for increasing the probability of bits in a particle's position to be zero. The method was empirically applied to a suite of four well-known benchmark gene expression data sets. The experimental results demonstrate that the proposed method outperforms the conventional version of binary particle swarm optimisation (BPSO) and other related works in terms of classification accuracy and the number of selected genes. In addition, this method also produces lower running times compared to BPSO.*
Keywords: Binary particle swarm optimisation, Gene selection, Gene expression data, Cancer classification

1. **Introduction.** Advances in the area of microarray-based gene expression analysis have led to a promising future for the diagnosis using new molecular-based approaches [1]. Microarray technology allows scientists to measure the expression levels of thousands of genes simultaneously, producing gene expression data that contain useful genomic, diagnostic, and prognostic information for researchers [2]. Comparisons between the gene expression levels of cancerous and normal tissues can be performed, and these comparisons