

Hybrid Gene Selection with Information Gain and Multi-Objective Evolutionary Algorithm for Leukemia Classification

M. N. F. Fajila

Department of Mathematical Sciences, Faculty of Applied Sciences, South Eastern University of Sri Lanka, Sri Lanka

Leukemia is a bone marrow cancer with various subtypes such as Acute Myeloid Leukemia and Acute Lymphoblastic Leukemia which require expertise to be identified. Morphological and histological appearances can be used to identify diseases. Yet, precise identification of subtypes is a difficult task. Therefore, subtype detection is a crucial part in prognosis. In this study, a hybrid gene selection approach Information Gain-Multi-Objective Evolutionary Algorithm (IG-MOEA) is proposed to identify Leukemia subtypes. Microarray data consists of thousands of genes where all are not corresponding to disease. Irrelevant and redundant genes have high impact on worst classification performance. Hence, IG is initially applied to preprocess the original datasets to remove irrelevant and redundant genes. Then, further MOEA is used to select a smaller subset of genes for perfect classification of new instances. Gene subset selection highly influences the classification. Further, the subsets selected intern is influenced by the algorithm used for gene selection. Moreover, informative subset of genes can be used efficiently for perfect prediction. Thus, selecting the appropriate algorithm for subset selection is important. Hence, MOEA is used in the proposed study for subset selection. The performance of proposed IG-MOEA is compared against the Information Gain-Genetic Algorithm (IG-GA) and Information Gain-Evolutionary Algorithm (IG-EA). Three Leukemia microarray datasets were used to evaluate the performance of the denoted approach. Remarkably, 100% classification was achieved for all the three datasets only with few informative genes using the proposed approach.

Keywords: Classification; Gene Selection; Microarray and Multi-Objective Evolutionary Algorithm