

Accepted Manuscript

Original article

An Enhanced Topologically Significant Directed Random Walk in Cancer Classification using Gene Expression Datasets

Choon Sen Seah, Shahreen Kasim, Mohd Farhan Md Fudzee, Jeffrey Mark Law Tze Ping, Mohd Saberi Mohamad, Rd Rohmat Saedudin, Mohd Arfian Ismail

PII: S1319-562X(17)30298-X
DOI: <https://doi.org/10.1016/j.sjbs.2017.11.024>
Reference: SJBS 1067

To appear in: *Saudi Journal of Biological Sciences*

Received Date: 21 September 2017
Revised Date: 8 November 2017
Accepted Date: 9 November 2017

Please cite this article as: C. Sen Seah, S. Kasim, M. Farhan Md Fudzee, J. Mark Law Tze Ping, M. Saberi Mohamad, R. Rohmat Saedudin, M. Arfian Ismail, An Enhanced Topologically Significant Directed Random Walk in Cancer Classification using Gene Expression Datasets, *Saudi Journal of Biological Sciences* (2017), doi: <https://doi.org/10.1016/j.sjbs.2017.11.024>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Topologically Significant Directed Random Walk in Cancer Classification using Gene Expression Datasets

Choon Sen Seah¹, Shahreen Kasim¹, Mohd Farhan Md Fudzee¹, Jeffrey Mark Law Tze Ping¹, Mohd Saberi Mohamad², and Rd Rohmat Saedudin³, Mohd Arfian Ismail⁴

¹*Soft Computing and Data Mining Centre, Faculty of Computer Sciences and Information Technology, Universiti Tun Hussein Onn Malaysia*

²*Faculty of Creative Technology and Heritage, Universiti Malaysia Kelantan, Karung Berkunci 01, 16300, Bachok, Kelantan, Malaysia,*

³*School of Industrial Engineering, Telkom University, 40257 Bandung, West Java, Indonesia*

⁴*Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Pahang, Malaysia*

Abstract

Microarray technology has become one of the elementary tools for researchers to study the genome of organisms. As the complexity and heterogeneity of cancer is being increasingly appreciated through genomic analysis, cancerous classification is an emerging important trend. Significant directed random walk is proposed as one of the cancerous classification approach which have higher sensitivity of risk gene prediction and higher accuracy of cancer classification. In this paper, the methodology and material used for the experiment are presented. Tuning parameter selection method and weight as parameter are applied in proposed approach. Gene expression dataset is used as the input datasets while pathway dataset is used to build a directed graph, as reference datasets, to complete the bias process in random walk approach. In addition, we demonstrate that our approach can improve sensitive predictions with higher accuracy and biological meaningful classification result. Comparison result takes place between significant directed random walk and directed random walk to show the improvement in term of sensitivity of prediction and accuracy of cancer classification.

1. Introduction

Early detection is one of the key elements in the reduction of mortality rate among disease carriers. The accurate determination type of cancer provides adequate early treatment and also to make sure that the treatment is efficient. For example, early malignant pleural mesothelioma is optimally treated by extrapleural pneumonectomy followed by radiochemotherapy, whereas metastatic lung cancer is cured by chemotherapy only [1]. Anticancer strategies are build based on tumor morphology (morphogenesis).

As technology grows, many researchers executed various investigations on the gene expression patterns and studied the gene mutation [2, 3]. Microarray has been an experimental tool to extract functional information from the genome [4]. In recent years, many researchers used microarray to profile the gene expression patterns of abnormal and normal gene in cancer [5, 6]. These kinds of studies shed light on obtaining bio-markers for cancer classification. Cancer classification enable doctors to get some insights about gene expression patterns such as gene function and interaction between genes.

Microarray has been adopted to profile gene expression datasets, and, applied in cancer classification. The success rate of cancer classification on the tools is largely dependent on data mining [7]. This is because, among gene expression datasets, only a part gives significant expression levels towards cancer. Therefore, a classification tools that can identify cancerous genes with high accuracy is needed. There are several types of cancer prediction and cancer classification approach [7,8].

In recent years, the random walk algorithm has been used by several researchers [9,10,11,12,13] to enable a more efficient process of cancer classification. In 2011, Revathy studied the usage of random walk in the improvement of cancer classification accuracy [9]. Through a multi-directed graph, the

Download English Version:

<https://daneshyari.com/en/article/8850045>

Download Persian Version:

<https://daneshyari.com/article/8850045>

[Daneshyari.com](https://daneshyari.com)