

Chapter 10

A Hybrid EM–Based Boosting Classification Model for Microarray Somatic Disease Prediction

Shaik Mahaboob Basha

Acharya Nagarjuna University, India

Nagaraju Devarakonda

Vellore Institute of Technology, India

ABSTRACT

As the size of the micro-array disease databases increase, finding an essential feature set for the classification problem is complex due to the large data size and sparsity problems. Traditional feature subset models are based on static clustering and classification models due to the fixed sized dimensions cluster-based disease prediction process. Sparsity, missing values, and imbalance are the major issues that affect the selection of essential feature clusters for data classification process. In this chapter, a hybrid cluster-based Bayesian probability estimation model is proposed in order to predict the disease class label on high dimensional databases. The proposed cluster-based classification model selects optimal clusters for feature ranking and classification problems to improve the true positive rate and accuracy. Experimental results are simulated on different training datasets for accuracy prediction. The results proved that the gene-disease-based patterns have better optimization than the conventional methods in terms of statistical metrics and classification models.

DOI: 10.4018/978-1-7998-9426-1.ch010

INTRODUCTION

As the size of the microarray datasets is growing day by day, finding a significant feature in the large feature space have become highly complicated because of data size and sparsity issues (Nagpal, 2018). Feature ranking of Microarray and classification is the main difficulties to technical and biomedical researchers because of its high dimensional feature space and restricted samples. Each microarray contains many identical DNA molecules that are used to identify a specific gene-related disease (X. H. Han, 2019). Microarrays are available in a wide variety of technologies. Microarrays are most commonly used to quantify mRNAs transcribed by various genes and different encoding proteins. Many cell types extract RNA and convert it into cDNA or cRNA, preferably one cell type.

RtPCR should upgrade copies. Fluorescent tags are added enzymatically within a cDNA/cRNA sequence or can be affixed to a further DNA strand, chemically or in a second. DNA's knowledge microarray processes give thousands of genes under several experimental gene expression levels. (S. Sayed, 2018) Proposed analysing the data on DNA microarray expression as a powerful tool for biological mechanisms study and developing predictive and prognostic categorisers to identify the patients who need treatment and the best treatment applicants. In (M. Sun, 2019), examining the data obtained through the microscope technology was very practical to understand how the genetic information turns into practical genetic products. Such a biclustering examination can determine a collection of genes under a set of provisional conditions. (M. Daoud, 2019) Proposed a method for classifying trajectories in road networks for discriminative patterns. By analysing the conduct of road trajectories, they found that the order of these visited locations was essential to improving classification accuracy, apart from the locations they had visited. This method challenged sequential patterns on the analysis feature as they retained order information to be good applicants.

The successful diagnosis and treatment of cancer needed to be properly identified and classified as cancer types. Certain gene expression analysis using a built-in CMOS microarray showed a successful diagnosis of time-resolved fluorescence. (Halder, 2019) Have likewise submitted OpenFlyData to the *Drosophila Melanogaster* exemplary web data that integrates gene expression data. Combining heterogeneous data across distributed sources is an important requirement for silicon-bioinformatics to support translational research. One of the major drawbacks in cancer data sets class discovery is that the cancer gene expression profiles contained many genes and lots of noisy genes. Reduce the effect of noisy genes on the expression profile of cancer genes. The two new consensus frameworks for gene-expression profile cancer discovery have been suggested by Zhiwen Y and others, namely triple spectral clusters (SC3) and dual spectral clusters (SC2 N cut). Although Mining Discriminatory paths

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/a-hybrid-em-based-boosting-classification-model-for-microarray-somatic-disease-prediction/300220

Related Content

FSR Evaluation Using the Suboptimal Operational Values

Osama H.S. Khader (2009). *Integrated Approaches in Information Technology and Web Engineering: Advancing Organizational Knowledge Sharing* (pp. 203-211).

www.irma-international.org/chapter/fsr-evaluation-using-suboptimal-operational/23995

Cloud Computing and Frameworks for Organisational Cloud Adoption

Victor Chang, Robert John Walters and Gary B. Wills (2016). *Web-Based Services: Concepts, Methodologies, Tools, and Applications* (pp. 683-707).

www.irma-international.org/chapter/cloud-computing-and-frameworks-for-organisational-cloud-adoption/140823

Horizontal Web Searching and Navigational Resource Identification

(2021). *Result Page Generation for Web Searching: Emerging Research and Opportunities* (pp. 16-27).

www.irma-international.org/chapter/horizontal-web-searching-and-navigational-resource-identification/268294

On Using Multiple Disabilities Profiles to Adapt Multimedia Documents: A Novel Graph-Based Method

Asma Saighi, Zakaria Laboudi, Philippe Roose, Sébastien Laborie and Nassira Ghoulmi-Zine (2020). *International Journal of Information Technology and Web Engineering* (pp. 34-60).

www.irma-international.org/article/on-using-multiple-disabilities-profiles-to-adapt-multimedia-documents/258738

PECA: Power Efficient Clustering Algorithm for Wireless Sensor Networks

Maytham Safar, Hasan Al-Hamadi and Dariush Ebrahimi (2011). *International Journal of Information Technology and Web Engineering* (pp. 49-58).

www.irma-international.org/article/peca-power-efficient-clustering-algorithm/52805