ILRC : A Hybrid Biomarker Discovering Algorithm Based on Improved L1 Regularization and Clustering in Microarray Data

Kun Yu

Northeastern University

Weidong Xie (✉ 1910638@stu.neu.edu.cn)

Northeastern University <https://orcid.org/0000-0003-1930-4509>

Linjie Wang

Northeastern University

Wei Li

Northeastern University

Research

Keywords: Microarray, Feature selection, Biomarker, Machine Learning

Posted Date: June 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-572610/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Finding significant genes or proteins from gene chip data for disease diagnosis and drug development is an important task, and the challenge comes from the curse of the data dimension. It is of great significance to use machine learning methods to find important features from the data and build an accurate classification model.

Results: The proposed Method has proved superior to the published advanced hybrid feature selection method and traditional feature selection method on different public microarray data sets. In addition, the results on the cleft lip and palate data set with known biomarkers provided by the cooperative hospital show that compared with other methods, our method can preferentially select these biomarkers.

Method: In this paper, a feature selection algorithm ILRC based on clustering and improved L1 regularization is proposed. In this method, the features are first clustered, and the redundant features in the sub-clusters are deleted. Then all the remaining features are iteratively evaluated using ILR, and the final result is output according to the cumulative weight reordering.

Conclusion: The proposed method can effectively remove redundant features. The algorithm's output has high stability and classification accuracy and can potentially select potential biomarkers.

Full Text

This preprint is available for [download as a PDF](#).