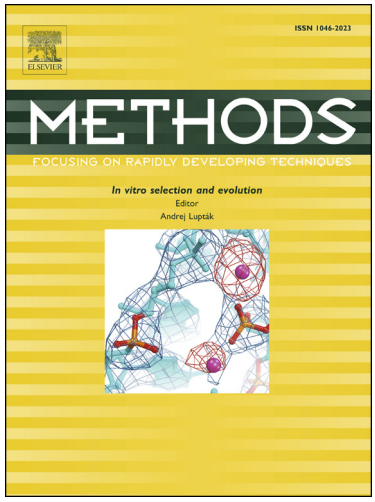

Feature Selection Methods for Big Data Bioinformatics: A Survey from the Search Perspective

Lipo Wang, Yaoli Wang, Qing Chang

PII: S1046-2023(16)30274-2
DOI: http://dx.doi.org/10.1016/j.ymeth.2016.08.014
Reference: YMETH 4063

To appear in: *Methods*

Received Date: 15 May 2016
Revised Date: 25 August 2016
Accepted Date: 30 August 2016

Please cite this article as: L. Wang, Y. Wang, Q. Chang, Feature Selection Methods for Big Data Bioinformatics: A Survey from the Search Perspective, *Methods* (2016), doi: http://dx.doi.org/10.1016/j.ymeth.2016.08.014

Feature Selection Methods for Big Data Bioinformatics: A Survey from the Search Perspective

Lipo Wang

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. Email: elpwang@ntu.edu.sg

Yaoli Wang and Qing Chang

Corresponding authors, College of Information Engineering, Taiyuan University of Technology, Taiyuan, China.

Abstract

This paper surveys main principles of feature selection and their recent applications in big data bioinformatics. Instead of the commonly used categorization into filter, wrapper, and embedded approaches to feature selection, we formulate feature selection as a combinatorial optimization or search problem and categorize feature selection methods into exhaustive search, heuristic search, and hybrid methods, where heuristic search methods may further be categorized into those with or without data-distilled feature ranking measures.

Keywords: big data; bioinformatics; biomarkers; classification; classifiers; cluster analysis; clustering; computational biology; computational intelligence; data attributes; data dimensionality reduction; data mining; drugs; evolutionary computation; evolutionary algorithms; feature extraction; feature selection; fuzzy logic; fuzzy sets; gene expression; gene regulatory networks; genetic algorithms; genetics; information analysis; kernel ridge regression; machine learning; microarray; neural nets; neural networks; neurocomputing; particle swarm optimization; pattern recognition; random forests; rough sets; soft computing; swarm intelligence; support vector machines.