

Accepted Manuscript

Fusion of expression values and protein interaction information using multi-objective optimization for improving gene clustering

Pratik Dutta, Sriparna Saha



PII: S0010-4825(17)30239-1

DOI: [10.1016/j.combiomed.2017.07.015](https://doi.org/10.1016/j.combiomed.2017.07.015)

Reference: CBM 2726

To appear in: *Computers in Biology and Medicine*

Received Date: 5 May 2017

Revised Date: 28 July 2017

Accepted Date: 28 July 2017

Please cite this article as: P. Dutta, S. Saha, Fusion of expression values and protein interaction information using multi-objective optimization for improving gene clustering, *Computers in Biology and Medicine* (2017), doi: [10.1016/j.combiomed.2017.07.015](https://doi.org/10.1016/j.combiomed.2017.07.015).

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.

Fusion of Expression Values and Protein Interaction Information using Multi-objective Optimization for Improving Gene Clustering

Pratik Dutta, Sriparna Saha

*Department of Computer Science and Engineering
Indian Institute of Technology, Patna
Bihar, India*

Abstract

One of the crucial problems in the field of functional genomics is to identify a set of genes which are responsible for a particular cellular mechanism. The current work explores the usage of a multi-objective optimization based genetic clustering technique to classify genes into groups with respect to their functional similarities and biological relevance. Our contribution is two-fold: firstly a new quality measure to compute the goodness of gene-clusters namely protein-protein interaction confidence score is developed. This utilizes the confidence scores of the protein-protein interaction networks to measure the similarity between genes of a particular cluster with respect to their biochemical protein products. Secondly, a multi-objective based clustering approach is developed which intelligently uses integrated information of expression values of microarray dataset and protein-protein interaction confidence scores to select both statistically and biologically relevant genes. For that very purpose, some biological cluster validity indices, viz. biological homogeneity index and protein-protein interaction confidence score, along with two traditional internal cluster validity indices, viz. fuzzy partition coefficient and Pakhira-Bandyopadhyay-Maulik-index, are simultaneously optimized during the clustering process. Experimental results on three real-life gene expression datasets show that the addition of new objective capturing protein-protein interaction information aids in clustering the

Email address: pratik.pcs16@iitp.ac.in , sriparna@iitp.ac.in (Pratik Dutta, Sriparna Saha)

Download English Version:

<https://daneshyari.com/en/article/4964864>

Download Persian Version:

<https://daneshyari.com/article/4964864>

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