World Academy of Science, Engineering and Technology
International Journal of Biotechnology and Bioengineering
Vol:16, No:10, 2022

A Local Tensor Clustering Algorithm to Annotate Uncharacterized Genes with Many Biological Networks

Authors : Paul Shize Li, Frank Alber

Abstract : A fundamental task of clinical genomics is to unravel the functions of genes and their associations with disorders. Although experimental biology has made efforts to discover and elucidate the molecular mechanisms of individual genes in the past decades, still about 40% of human genes have unknown functions, not to mention the diseases they may be related to. For those biologists who are interested in a particular gene with unknown functions, a powerful computational method tailored for inferring the functions and disease relevance of uncharacterized genes is strongly needed. Studies have shown that genes strongly linked to each other in multiple biological networks are more likely to have similar functions. This indicates that the densely connected subgraphs in multiple biological networks are useful in the functional and phenotypic annotation of uncharacterized genes. Therefore, in this work, we have developed an integrative network approach to identify the frequent local clusters, which are defined as those densely connected subgraphs that frequently occur in multiple biological networks and consist of the query gene that has few or no disease or function annotations. This is a local clustering algorithm that models multiple biological networks sharing the same gene set as a three-dimensional matrix, the so-called tensor, and employs the tensor-based optimization method to efficiently find the frequent local clusters. Specifically, massive public gene expression data sets that comprehensively cover dynamic, physiological, and environmental conditions are used to generate hundreds of gene co-expression networks. By integrating these gene co-expression networks, for a given uncharacterized gene that is of biologist's interest, the proposed method can be applied to identify the frequent local clusters that consist of this uncharacterized gene. Finally, those frequent local clusters are used for function and disease annotation of this uncharacterized gene. This local tensor clustering algorithm outperformed the competing tensor-based algorithm in both module discovery and running time. We also demonstrated the use of the proposed method on real data of hundreds of gene co-expression data and showed that it can comprehensively characterize the query gene. Therefore, this study provides a new tool for annotating the uncharacterized genes and has great potential to assist clinical genomic diagnostics.

Keywords : local tensor clustering, query gene, gene co-expression network, gene annotation
Conference Title : ICBCMB 2022 : International Conference on Bioinformatics and Computational Molecular Biology
Conference Location : Barcelona, Spain
Conference Dates : October 20-21, 2022