Abstract

Clustering is one of the fundamental processes of analyzing gene expression data, basically by comparing gene expression profiles or sample expression profiles. Comparing expression profiles requires a measure apart from the actual clustering algorithm to quantify how similar or dissimilar the objects under consideration are. Various clustering algorithms have been used to analyze gene expression data. Some of these algorithms reported the incorporation of similarity measures like Euclidean Distance, Pearson Correlation and mutual information for their performance. This work considered different reported clustering algorithms for gene expression data analyses and the importance of different similarity measures for optimizing these clustering algorithms. To this end, no clustering technique in all the works investigated has been applied directly on gene expression data. It is observed that the output (distance matrix) of similarity or dissimilarity measures plays the role of input to clustering techniques, and those that did not use any of the popular proximity measures applied one or two approaches such as Constrained Coherency (CoCo), Silhouette coefficient measurement, and normalization and discretization, to refine gene expression data for improved cluster quality by speeding up the learning phase,
Optimization of Clustering Algorithms for Gene Expression Data Analysis using Distance Measures

red of computational space and handling of noise effectively.

References

gene expression data: a comparative study. BMC Bioinformatics, 9:497,
Distance-based clustering of CGH data. BIOINFORMATICS, Vol. 22 no. 16, pages 1971–1978,
via mutual information distance measure. BMC Bioinformatics, 8:111,
distance to keep. Biology Direct, 5:51.
Microarray Gene Expressions using Linear Dynamic Range of Photo Multiplier Tube.
Proceedings of the International Conference on Computing: Theory and Applications
(ICCTA’07), 0-7695-2770-1/07.
Inference and Visualization of Gene Regulatory Networks (GRN). Computational Biology and

Index Terms

Computer Science Artificial Intelligence

Keywords

Clustering Algorithms, Proximity Measure, Gene Expression Data, Distance Matrix, Microarray.