عنوان مقاله:

The Application of Bi-clustering and Bayesian Network for Gene Sets Network Construction in Breast Cancer Microarray Data

محل انتشار:

مجله سرطان خاورمیانه, دوره 13, شماره 4 (سال: 1401)

تعداد صفحات اصل مقاله: 17

نویسندگان:

Ahmad Sohrabi - *Department of Biostatistics, School of Public Health, Iran University of Medical Sciences, Tehran, Iran*

Neda Saraygord-Afshari - *Department of Medical Biotechnology, Faculty of Allied Medical Sciences, Iran University of Medical Sciences, Tehran, Iran*

Masoud Roudbari - *Department of Biostatistics, School of Public Health, Iran University of Medical Sciences, Tehran, Iran*

خلاصه مقاله:

Background: Breast cancer is one of the most prevalent types of cancer in Iranian women and the second cause of death in women worldwide. Gene mutations are the key determinants of the disease; therefore, the genetic study of this disease is of paramount importance. One of the genetic evaluation methods of this disease is microarray technology, which allows the examination of the simultaneous expression of thousands of genes. Clustering is the method for analyzing high-dimension data, which we used in the present research for collecting similar genes in separated clusters.Method: A descriptive and inferential statistical analysis was carried out to evaluate unsupervised learning models of gene expression analysis and five bi-clustering methods (including PLAID (PL), Fabia, Bimax, Cheng & Church (CC), and Xmotif) were compared. For this purpose, we obtained the microarray gene expression data for lapatinib-resistant breast cancer cell lines from previously published research. The enrichment efficacy of the clusters was evaluated with gene ontology, and the results of these five models were compared with the Jaccard index, variance stability, least-square error, and goodness of fit indices. Furthermore, the results of the best model were assessed for building a genes sets network with Bayesian networks.Results: After preprocessing, clustering was performed on the data with the dimension (۴۷۱۰ × ۱۸) of the genes. Four models, except for CC, successfully found bi-clusters in the data set. The data evaluation revealed that the results of the models were almost the same, but the PL model performed better than the others, finding ۱۱ bi-clusters; this model was used to build the network of gene sets.Conclusion: According to the results, the PL method was suitable for clustering the data. Accordingly, it could be recommended for data analysis. In addition, the gene sets network formed on gene expression data was incompetent.

کلمات کلیدی:

Breast cancer, Bi-clustering, Cluster Analysis, Microarray data, Gene expression, Neoplasms, Bayesian network

لینک ثابت مقاله در پایگاه سیویلیکا:

https://civilica.com/doc/1819109