

Mencari korelasi bicluster pada data ekspresi gen penyakit diabetes melitus dengan menggunakan metode three-phase biclustering = Finding correlated bicluster from gene expression data of diabetes melitus using three-phase biclustering

Soeganda Formalidin, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20467557&lokasi=lokal>

Abstrak

Penelitian ini bertujuan untuk mencari korelasi yang kuat antar gen dan kondisi dari data ekspresi gen penyakit Diabetes Melitus (DM) pada sampel obesitas dan sampel kurus dengan menggunakan metode three phase biclustering. Tahap pertama pada metode ini adalah dengan menggunakan matriks dekomposisi Singular Value Decomposition (SVD) yang mentransformasikan data menjadi dua matriks berbasis gen dan kondisi. Selanjutnya pada tahap kedua menggunakan metode partisi Partition Around Medoids (PAM) pada dua matriks gen dan kondisi menggunakan jarak Euclidean sehingga jika digabung akan membentuk bicluster yang pada tahap tiga akan dievaluasi dengan menggunakan modifikasi lift algorithm berdasarkan korelasi Pearson yang cocok untuk mendeteksi bicluster model additive-multiplicative. Hasil dari implementasi algoritma yang digunakan pada dataset microarray dinamakan -corbicluster yang memiliki korelasi yang tinggi antar gen dan sampel. Implementasi dari tahap pertama dan kedua (SVD-PAM) pada dataset DM dengan 1331 gen terseleksi menghasilkan 8 bicluster. Sedangkan hasil tahap ketiga yaitu modifikasi algoritma lift pada kedelapan bicluster ini menghasilkan 3 -corbicluster dengan masing-masing nilai korelasi yang tinggi yaitu 0,097, 0,095, 0,085, sehingga metode yang diusulkan dan hasil analisisnya pada gen dan sampel penyakit DM memiliki potensi besar ke depannya dalam penelitian pada bidang medis.

.....

The purpose of this research is to find strong correlation among genes and conditions of Diabetes Melitus genes expression data which samples are obese and lean people using three phase biclustering. First step is to use matrix decomposition Singular Value Decomposition (SVD) to decompose matrix gene expression data into two global based gene and condition matrices. Second step is to use partition method Partition Around Medoid (PAM) to cluster gene and condition based matrices using Euclidean distance, forming several biclusters which further evaluated using modified lift algorithm based on Pearson correlation which is very appropriate method to detect additive-multiplicative bicluster type. The resulting bicluster of the proposed algorithm having strong correlation among genes and samples to microarray dataset are called -corbicluster. Implementation of the first and second step (SVD-PAM) to dataset DM with 1331 selected genes produces 8 biclusters. For the third step using modified lift algorithm to these 8 biclusters produces 3 -corbiclusters having strong correlation values: 0,097, 0,0095, 0,085, so that the proposed method and the result of analysis to genes and samples of DM have high potential in future medical researches.