

Implementasi metode similarity based biclustering (SBB) pada microarray data ekspresi gen dengan menggunakan algoritma partisi k-means = Implementation of similarity based biclustering (SBB) on microarray gene expression data by using k means partition algorithm

Sofia Debi Puspa, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20467563&lokasi=lokal>

Abstrak

ABSTRAK

Penelitian ini bertujuan untuk mengimplementasikan similarity based biclustering SBB dalam memperoleh bicluster sekumpulan gen dengan ekspresi yang similar di bawah kondisi tertentu yang signifikan pada data microarray. Secara teoritis similarity based biclustering terdiri atas tiga tahap utama, yaitu: membangun matriks similaritas baris gen dan matriks similaritas kolom kondisi, mempartisi masing-masing matriks similaritas dengan hard clustering khususnya dalam penelitian ini menggunakan partisi k-means, dan ekstrak bicluster. Sebelum mengimplementasikan metode SBB, strategi seleksi gen diterapkan dan selanjutnya dilakukan normalisasi. Perolehan evaluasi indeks silhouette pada dataset diabetic nephropathy, diabetic retinopathy dan lymphoma berturut-turut pada cluster kondisi yaitu 0,8304; 0,7853 dan 0,7382, sedangkan indeks silhouette untuk cluster gen yaitu 0,5382; 0,5408 dan 0,5464. Dan dari hasil analisis cluster kondisi, akurasi dari dataset diabetic nephropathy dan diabetic retinopathy yaitu 100, sedangkan dataset lymphoma yaitu 98. Selanjutnya dapat diketahui regulasi proses seluler yang terjadi pada bicluster dari ketiga dataset. Hasil analisis menunjukkan bahwa gen-gen yang diperoleh dari bicluster sesuai dengan fungsi gen dan proses biologis didukung oleh GO enrichment sehingga menjadi potensi yang besar bagi praktisi medis dalam tindak lanjut suatu penyakit yang diderita oleh pasien.

ABSTRACT

This study aims to implement similarity based biclustering SBB in obtaining a bicluster a set of genes that exhibit similar levels of gene expression under certain conditions that is significant in microarray data. Theoretically, similarity based biclustering consists of three main phase constructing the row gene similarity matrix and the column condition similarity matrix, partitioning each matrix similarity with hard clustering especially in this research using k means partition, and extracting bicluster. Before implementing the SBB method, the gene selection strategy is applied and subsequently normalized. The acquisition of silhouette index evaluation in diabetic nephropathy, diabetic retinopathy, and lymphoma on cluster condition respectively is 0.8304, 0.7853 and 0.7382, while the silhouette index for the gene cluster is 0.5382, 0.5408 and 0.5464. In addition, according to the cluster condition analysis, accuracy of dataset diabetic nephropathy and diabetic retinopathy is 100, whereas dataset lymphoma is 98. Furthermore, it can be known cellular regulation that occurs on the bicluster of the three datasets. The results of the analysis show that the genes obtained from bicluster are relevant to the function of genes and biological processes supported by GO enrichment, therefore it becomes a great potential for medical practitioners in the follow up of a disease suffered by the patient.