

Implementasi triclustering menggunakan metode $\hat{\lambda}$ -trimax pada data ekspresi gen microarray = Triclustering implementation using $\hat{\lambda}$ -trimax metode on microarray gene expression dataset

Noval Saputra, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20509671&lokasi=lokal>

Abstrak

Analisis triclustering merupakan teknik analisis pada data 3D (observasi – atribut – konteks). Analisis triclustering dapat mengelompokkan observasi pada beberapa atribut dan konteks secara bersamaan. Analisis triclustering telah sering diterapkan untuk menganalisis data ekspresi gen microarray. Penelitian ini menggunakan metode $\hat{\lambda}$ -Trimax untuk melakukan analisis triclustering pada data ekspresi gen microarray. Metode $\hat{\lambda}$ -Trimax bertujuan untuk menemukan tricluster yang memiliki mean square residual kecil dari $\hat{\lambda}$ dan volume maksimal. Tricluster diperoleh dengan cara melakukan penghapusan node dari data 3D dengan menggunakan algoritma multiple node deletion dan single node deletion. Kandidat tricluster yang telah didapatkan, dilakukan pengecekan kembali dengan menambahkan beberapa node yang telah dihapus sebelumnya menggunakan algoritma node addition. Pada penelitian ini dilakukan perbaikan program pada metode $\hat{\lambda}$ -Trimax dan juga menambahkan penghitungan evaluasi tricluster yang dihasilkan. Implementasi metode $\hat{\lambda}$ -Trimax dilakukan pada data ekspresi gen dari proses diferensiasi human induced pluripoten stem cell (HiPSC) dari pasien penyakit jantung. Ekspresi gen diukur pada 12 titik waktu dan 3 replikasi. Dari beberapa simulasi yang dilakukan, metode $\hat{\lambda}$ -Trimax memberikan hasil terbaik ketika $\hat{\lambda} = 0,0068$ dan $\hat{\lambda} = 1,2$. Berdasarkan tricluster yang dihasilkan dari simulasi terbaik tersebut, dipilih 5 tricluster yang diduga sebagai ciri-ciri penyakit jantung. Lima tricluster ini dapat menjadi pertimbangan bagi ahli medis untuk melakukan tindakan lebih lanjut terhadap pasien.

Triclustering analysis is an analysis technique on 3D data (observation - attribute - context). Triclustering analysis can group observations on several attributes and contexts simultaneously. Triclustering analysis has been frequently applied to analyze microarray gene expression data. This study used the $\hat{\lambda}$ -Trimax method to perform triclustering analysis on microarray gene expression data. The $\hat{\lambda}$ -Trimax method aims to find a tricluster that has a mean square residual smaller than $\hat{\lambda}$ and a maximum volume. Tricluster is obtained by deleting nodes from 3D data using multiple node deletion and single node deletion algorithms. The tricluster candidates that have been obtained are checked again by adding some previously deleted nodes using the node addition algorithm. In this research, the program improvement of the $\hat{\lambda}$ -Trimax method was carried out and also the calculation of the resulting tricluster evaluation. The implementation of the $\hat{\lambda}$ -Trimax method was carried out on gene expression data from the differentiation process of human induced pluripotent stem cells (HiPSC) from patients with heart disease. Gene expression was measured at 12 time points and 3 replications. From several simulations performed, the $\hat{\lambda}$ -Trimax method gives the best results when $\hat{\lambda} = 0.0068$ and $\hat{\lambda} = 1.2$. Based on the tricluster generated from the best simulation, 5 tricluster were selected which were suspected as a characteristic of heart disease. These five tricluster can be a consideration for medical experts to take further action on patients.