

Analisis triclustering menggunakan algoritma order preserving triclustering (OPTricluster) dan implementasinya pada data ekspresi gen = Analysis of triclustering using order preserving triclustering (OPTricluster) algorithm and its implementation on gene expression data

Dea Siska, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20509744&lokasi=lokal>

Abstrak

Metode *triclustering* merupakan pengembangan dari metode *clustering* dan *biclustering*. Berbeda dengan metode *clustering* dan *biclustering* yang bekerja pada data dua dimensi, *triclustering* bekerja pada data tiga dimensi yang disusun dalam bentuk matriks. Matriks ini terdiri dari dimensi observasi, atribut, dan konteks. *Triclustering* mampu mengelompokkan ketiga dimensi tersebut secara simultan dan membentuk kelompok berupa subruang yang disebut *tricluster*. Metode ini umumnya diimplementasikan dalam bidang bioinformatika, terkhususnya dalam analisis data ekspresi gen tiga dimensi untuk menemukan profil ekspresi gen. Data atau matriks ini terdiri dari dimensi gen, kondisi eksperimen, dan waktu eksperimen (*time point*). Salah satu algoritma *triclustering*, yaitu *Order Preserving Triclustering* (OP*Tricluster*), adalah algoritma yang menggunakan pendekatan *pattern based* dan digunakan untuk menganalisis data ekspresi gen tiga dimensi yang merupakan *short time series* 3-8 *time point*. OP*Tricluster* membentuk *tricluster* dengan mengidentifikasi gen-gen yang memiliki perubahan ekspresi yang sama di sepanjang *time points* pada sejumlah kondisi eksperimen.

Dalam penelitian ini, OP*Tricluster* diimplementasikan pada data ekspresi gen sejumlah pasien *yellow fever* pasca vaksinasi dengan beberapa skenario yang menggunakan *threshold* yang berbeda-beda. Skenario dengan *threshold* yang optimum ditunjukkan oleh rata-rata skor *Tricluster Diffusion* terendah. *Tricluster-tricluster* yang dihasilkan berhasil menunjukkan hubungan biologis di antara pasien-pasien tersebut, di mana vaksin cenderung memberikan reaksi yang lebih signifikan pada pasien pria dibandingkan pasien wanita. Selain itu, ditemukan anomali pada pasien-pasien tersebut.

Triclustering method is the development of *clustering method* and *biclustering method*. Unlike *clustering* and *biclustering* that works on two-dimensional data, *triclustering* works on three-dimensional data that arranged in the form of a matrix consisting of observations, attributes, and contexts dimensions. *Triclustering* is able to group these dimensions simultaneously and form a subspace called a *tricluster*. This method is generally implemented in analysis of three-dimensional gene expression data to find profiles of gene expression. This data or matrix consists of genes, experimental conditions and time points dimensions. One of the *triclustering* algorithms, *Order Preserving Triclustering* (OP*Tricluster*), is an algorithm that uses a *pattern-based approach* and used to analyze short time series data (3-8 time points). The OP*Tricluster* forms the *tricluster* by identifying genes that have the same expression change across time points under a number of experimental conditions. The change in expression is expressed in a rank pattern which is divided based on three types of patterns, namely constant, conserved and divergent patterns.

In this study, OPTricluster was implemented in gene expression data of yellow fever patients after vaccination using several scenarios with different thresholds. The scenario with the optimum threshold is indicated by the lowest average Tricluster Diffusion score. The resulting triclusters were successful in showing biological relationships among these patients, where the vaccine tending to have a more significant reaction in male patients than in female patients. In addition, anomalies were found in these patients.