

Metode clustering tiga dimensi pada data ekspresi gen menggunakan pendekatan kubus gen = Three-dimensional clustering method on gene expression dataset using the gene cube approach.

Almaira Nabila Ayudhiya, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20513542&lokasi=lokal>

Abstrak

Saat ini, minat terhadap analisis data ekspresi gen telah meningkat pesat. Hal ini dikarenakan analisis informasi yang dihasilkan dari analisis data ekspresi gen dapat memungkinkan untuk menemukan fenomena biologis tertentu. Teknik clustering merupakan teknik yang telah banyak diaplikasikan sebagai alat untuk memahami informasi dalam data ekspresi gen. Seiring berkembangnya teknologi, digunakan teknik triclustering dalam banyak analisis biologis lainnya, tetapi belum ada satupun dari analisis ini yang memperhitungkan peran kromosom. Terdapat sumber yang menyatakan bahwa data ekspresi gen dikendalikan oleh elemen pengatur yang dapat berlokasi di sepanjang kromosom, bahkan dalam beberapa kasus juga dapat berlokasi di kromosom lain. Elemen pengatur yang dimaksud adalah suatu protein yang dihasilkan oleh gen regulator, yaitu gen yang produk ekspresinya berperan mengatur ekspresi gen lain. Pada penelitian ini diperhitungkan identitas kromosom dari setiap gen sebelum dilakukannya clustering, dengan cara membentuk struktur tiga dimensi, yaitu kromosom-gen-observasi. Struktur ini disebut sebagai kubus gen. Pendekatan kubus gen ini diimplementasikan pada data ekspresi gen penyakit kanker kandung kemih. Untuk menguji apakah struktur kubus gen ini dapat memberikan informasi yang diinginkan dengan cara menemukan pola pada ekspresi gen, diterapkan metode K-means clustering yang langkah awalnya dioptimasi menggunakan algoritma K-means++ dan metode Delta-Trimax triclustering. Metode K-means clustering menghasilkan cluster-cluster yang optimal dengan indeks Davies Bouldin yang rendah pada sumbu gen, sumbu observasi, dan sumbu kromosom. Sedangkan, metode Delta-Trimax triclustering menghasilkan tricluster-tricluster optimal dengan menggunakan threshold terbaik yang diperoleh berdasarkan nilai tricluster diffusion terkecil. Tricluster ini dapat menjadi pertimbangan bagi ahli medis untuk digunakan sebagai target terapi. Struktur kubus gen terbukti dapat memberikan informasi yang diinginkan menggunakan sudut pandang baru, yaitu berupa dimensi kromosom.

.....Nowadays, the interest in gene expression data analysis has grown rapidly. This is because the analysis of information generated from gene expression data analysis has enabled the founding of biological phenomenon. As how the data itself is reasoned, the clustering approach has become a technique that has been widely applied in understanding gene expression data. As technology advances, the triclustering technique has been used for many biological analysis, but none has yet taken the role of chromosomes into account. There are sources claiming that the expression of gene data is controlled by regulatory elements which can be located alongside a chromosome, in some cases, even located on another chromosome. Said regulatory elements are proteins produced by a gene regulator, namely genes whose expression products play a role in regulating the expression of other genes. In this study, the chromosome identity of each gene was accounted before doing clustering, which was done by forming a three-dimensional structure, i.e. chromosome-gene-observation. The structure is known as gene cube. This gene cube approach is implemented in the expression of bladder cancer gene data. To find out whether the cube structure of this gene can provide the desired information through finding the pattern of gene expression, K-means clustering

which initial step is optimized using the K-means++ algorithm and Delta-Trimax triclustering method are applied. The K-means clustering method produces optimal clusters with small value of Davies Bouldin index on the gene axis, observation axis, and chromosome axis. Meanwhile, the Delta-Trimax triclustering method produces optimal triclusters using the best threshold obtained based on the smallest value of tricluster diffusion. This tricluster can be a consideration for medical experts to be used as therapy target. The gene cube structure has been shown to provide the desired information from a new perspective, namely the dimensions of chromosome.