

Implementasi spectral clustering pada data microarray gen karsinoma menggunakan algoritma K-means = The implementation of spectral clustering on microarray data on carcinoma genes using K-means algorithm

Frisca, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20446400&lokasi=lokal>

Abstrak

Spectral clustering adalah salah satu algoritma clustering modern yang paling terkenal. Sebagai teknik clustering yang efektif, metode spectral clustering muncul dari konsep teori graf spektral. Metode spectral clustering membutuhkan algoritma partisi. Ada beberapa metode partisi termasuk PAM, SOM, Fuzzy c-means, dan k-means. Berdasarkan penelitian yang telah dilakukan oleh Capital dan Choudhury pada 2013, ketika menggunakan Euclidian distance, k-means memberikan akurasi yang lebih baik dibandingkan dengan algoritma PAM. sehingga, makalah ini menggunakan algoritma k-means. Keuntungan utama dari spectral clustering adalah mengurangi dimensi data, terutama dalam hal ini untuk mengurangi dimensi yang besar dari data microarray.

Microarray data adalah chip berukuran kecil yang terbuat dari slide kaca yang berisi ribuan bahkan puluhan ribu jenis gen dalam fragmen DNA yang berasal dari cDNA. Aplikasi data microarray secara luas digunakan untuk mendeteksi kanker, misalnya adalah karsinoma, di mana sel-sel kanker mengekspresikan kelainan pada gen-nya. Proses spectral clustering dimulai dengan pengumpulan data microarray gen karsinoma, preprocessing, menghitung similaritas, menghitung , menghitung nilai eigen dari , membentuk matriks , dan clustering dengan menggunakan k-means. Dari hasil pengelompokan gen karsinoma pada penelitian ini diperoleh dua kelompok dengan nilai rata-rata Silhouette maksimal adalah 0.6336247. Proses clustering pada penelitian ini menggunakan program open source R.

.....Spectral clustering is one of the most famous modern clustering algorithms. As an effective clustering technique, spectral clustering method emerged from the concepts of spectral graph theory. Spectral clustering method needs partitioning algorithm. There are some partitioning methods including PAM, SOM, Fuzzy c means, and k means. Based on the research that has been done by Capital and Choudhury in 2013, when using Euclidian distance k means algorithm provide better accuracy than PAM algorithm. So in this paper we use k means as our partition algorithm. The major advantage of spectral clustering is in reducing data dimension, especially in this case to reduce the dimension of large microarray dataset.

Microarray data is a small sized chip made of a glass plate containing thousands and even tens of thousands kinds of genes in the DNA fragments derived from doubling cDNA. Application of microarray data is widely used to detect cancer, for the example is carcinoma, in which cancer cells express the abnormalities in his genes. The spectral clustering process is started with collecting microarray data of carcinoma genes, preprocessing, compute similarity matrix, compute , compute eigen value of , compute , clustering using k means algorithm. In this research, Carcinoma microarray data using 7457 genes. The result of partitioning using k means algorithm is two clusters clusters with maximum Silhouette value 0.6336247.