

Implementasi metode spectral clustering-partitioning around medoids (PAM) dengan algoritma similaritas paralel berbasis cuda pada data microarray gen karsinoma = Implementation of spectral clustering partitioning around medoids (PAM) method with parallel similarity algorithm based on cuda in microarray data of carcinoma genes

Rosalia Deviana Cahyaningrum, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20446402&lokasi=lokal>

Abstrak

Penelitian ini bertujuan untuk mengimplementasikan spectral clustering-PAM dengan menggunakan algoritma similaritas serial dan mengimplementasikan algoritma similaritas paralel berbasis CUDA dalam metode spectral clustering pada data microarray gen karsinoma. Implementasi dibantu dengan perangkat lunak R berbasis open source yang digunakan pada algoritma spectral clustering-PAM dengan algoritma similaritas serial dan CUDA yang digunakan pada algoritma similaritas paralel. Pengelompokan data microarray gen karsinoma diawali dengan menormalisasi data menggunakan normalisasi min-max. Pada algoritma spectral clustering-PAM, pertama-tama similaritas antar gen karsinoma dihitung. Selanjutnya, membentuk matriks Laplacian ternormalisasi dari matriks diagonal dan matriks Laplacian tak ternormalisasi. Langkah berikutnya yaitu menghitung eigenvalue dari matriks Laplacian ternormalisasi dan menentukan eigenvector dari eigenvalue terkecil matriks Laplacian ternormalisasi yang disusun menjadi dataset baru untuk dipartisi setiap barisnya menggunakan metode PAM. Berdasarkan running time, waktu yang dibutuhkan untuk menghitung nilai similaritas secara paralel di CUDA 378 kali lebih cepat daripada secara serial di R. Hasil penelitian menunjukkan bahwa spectral clustering-PAM mengelompokkan data microarray gen karsinoma menjadi dua cluster dengan nilai rata-rata silhouette yaitu 0,6458276.

.....This research aims to implement the spectral clustering PAM using serial similarity algorithm and implement parallel similarity algorithm based on CUDA in spectral clustering method on microarray data of carcinoma genes. Implementation assisted with software based on open source R used in spectral clustering algorithm PAM with serial similarity algorithm and CUDA used to parallel similarity algorithm. Clustering microarray data of carcinoma genes preceded by normalizing the data using min max normalization. In the spectral clustering PAM algorithm, first of all, similarity between genes of carcinoma calculated. Furthermore, forming the normalized Laplacian matrix from diagonal matrix and unnormalized Laplacian matrix. The next step is to calculate the eigenvalues of normalized Laplacian matrix and determine the eigenvectors of k smallest eigenvalues of normalized Laplacian matrix is organized into a new dataset to be partitioned each line using PAM. Based on the running time, the time required to calculate the value of parallel similarity in CUDA is 378 times faster than a serial in R. The results showed that spectral clustering PAM classify microarray data of carcinoma genes into two clusters with an average silhouette value is 0,6458276.