

Triclustering dengan Coarse-Grain Parallel Genetic Algorithm Menggunakan Identifikasi Gene Ontology dan KEGG Pathway pada Data Ekspresi Gen Microarray = Triclustering with Coarse-Grain Parallel Genetic Algorithm Using Gene Ontology and KEGG Pathway Identification on Microarray Gene Expression Data

Aisyah Gefira, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920551669&lokasi=lokal>

Abstrak

Teknologi microarray adalah teknologi di bidang bioinformatika yang digunakan untuk mengukur ekspresi gen dalam berbagai kondisi eksperimental dan menghasilkan data ekspresi gen. Salah satu metode yang dapat digunakan untuk menganalisis data ekspresi gen adalah triclustering, Triclustering adalah metode pengelompokan data tiga dimensi berdasarkan karakteristik yang serupa. Berbagai algoritma seperti TriGen, -Trimax dan EMOA--Trimax dikembangkan untuk melakukan triclustering. Namun, seluruh algoritma tersebut menghadapi masalah yang sama, yaitu waktu komputasi yang lama. Penelitian ini menggunakan Coarse-grain Parallel Genetic Algorithm (CgPGA) untuk mengatasi tantangan waktu komputasi pada triclustering. Algoritma ini membagi data menjadi beberapa subpopulasi dan menjalankan proses evolusi genetik secara paralel menggunakan enam core. Penelitian ini mengusulkan penggunaan CgPGA untuk mempercepat proses triclustering pada data ekspresi gen darah microarray tiga dimensi yang dipengaruhi empat jenis minuman dan diukur pada lima titik waktu. Kualitas tricluster dievaluasi menggunakan fitness function yang diadaptasi dari Mean Square Residue (MSR), weights, dan distinction. Hasil penelitian ini menunjukkan bahwa CgPGA secara signifikan mengurangi waktu komputasi dengan bekerja 18,24 kali lebih cepat dibandingkan dengan Genetic Algorithm secara serial yang diukur dengan speedup. CgPGA berhasil diterapkan untuk melakukan triclustering pada data ekspresi gen tiga dimensi microarray berdasarkan kemiripan pola ekspresi gen. 10 tricluster yang dihasilkan memiliki kombinasi gen, kondisi, dan waktu yang beragam, serta memiliki fitness score tinggi yang berkisar antara 2514,542745 hingga 2568,106026. Tricluster 5 memiliki fitness score tertinggi yaitu sebesar 2568,106026. Selanjutnya, hasil triclustering dianalisis menggunakan Gene Ontology (GO) dan KEGG Pathway untuk mengidentifikasi informasi gen dan interaksi antar gen di dalam tricluster. Hasil penelitian menunjukkan bahwa GO berhasil mengidentifikasi gen-gen di dalam tricluster dalam aspek proses biologis, fungsi molekuler, dan komponen seluler. Analisis KEGG Pathway menunjukkan bahwa gen-gen di dalam tricluster berpartisipasi dalam berbagai jalur biokimia, salah satunya adalah jalur diabetic cardiomyopathy. Secara keseluruhan, penelitian ini memberikan kontribusi signifikan dalam analisis data genomik dengan mengimplementasikan teknik triclustering menggunakan CgPGA yang efektif dan efisien, serta memperluas pengetahuan tricluster melalui identifikasi informasi genetik yang relevan dan meningkatkan pemahaman tentang interaksi biologis yang terjadi pada suatu kelompok gen berpola ekspresi serupa menggunakan GO dan KEGG Pathway.

.....Microarray technology is a bioinformatics tool utilized to measure gene expression across various experimental conditions, generating comprehensive gene expression data. Triclustering, a method for clustering three-dimensional data based on similar characteristics, is one approach to analyzing this data. Despite the development of several algorithms for triclustering, such as TriGen, Î´-Trimax, and EMOA-Î´-Trimax, they all encounter the challenge of lengthy computation times. This study addresses this issue by

employing the Coarse-grain Parallel Genetic Algorithm (CgPGA). The algorithm mitigates computational time by dividing the data into several subpopulations and executing the genetic evolution process in parallel across six cores. The study demonstrates the application of CgPGA to expedite the triclustering process on three dimensional microarray blood gene expression data, influenced by four types of beverages and measured at five different time points. The triclusters' quality is assessed using a fitness function adapted from Mean Square Residue (MSR), weights, and distinction. Results indicate that CgPGA significantly reduces computation time, operating 18,24 times faster than the serial Genetic Algorithm as measured by speedup. CgPGA effectively performs triclustering on three-dimensional microarray gene expression data. The 10 resulting triclusters exhibit diverse combinations of genes, conditions, and time points, and have high fitness scores ranging from 2514.542745 to 2568.106026. Tricluster 5 has the highest fitness score of 2568.106026. Further analysis of the triclustering results using Gene Ontology (GO) and KEGG Pathway reveals gene information and interactions within the triclusters. GO analysis successfully identifies genes within the triclusters in terms of biological processes, molecular functions, and cellular components, while KEGG Pathway analysis shows that the genes participate in various biochemical pathways, including the diabetic cardiomyopathy pathway. Overall, this study significantly contributes to genomic data analysis by implementing an efficient and effective triclustering technique using CgPGA, expanding the understanding of triclusters by identifying relevant genetic information, and enhancing the comprehension of biological interactions within gene groups exhibiting similar expression patterns using GO and KEGG Pathway.