

Analisis Triclustering menggunakan Metode Gabungan Fuzzy Cuckoo Search Berdasarkan Gaussian Distribution dengan -Trimax pada Data Ekspresi Gen = Analysis of Triclustering using Combined Method of Fuzzy Cuckoo Search Based on Gaussian Distribution and δ -Trimax on Gene Expression Data

Muhamad Ido Raskapati, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920530947&lokasi=lokal>

Abstrak

Analisis triclustering adalah salah satu metode data mining yang memiliki tujuan mengelompokkan data berbentuk tiga dimensi. Triclustering umumnya digunakan pada bidang bioinformatika untuk menganalisis kesamaan ekspresi gen suatu eksperimen pada titik waktu tertentu. Analisis triclustering yang dilakukan pada penelitian ini menggunakan metode gabungan Fuzzy Cuckoo Search berdasarkan Gaussian Distribution dengan -Trimax. Metode ini merupakan penggabungan algoritma nodes deletion pada Trimax dengan algoritma optimasi Fuzzy Cuckoo Search. Algoritma nodes deletion pada -Trimax digunakan pada fase pembentukan populasi awal tricluster. Konsep algoritma nodes deletion yaitu dapat menghasilkan himpunan tricluster dengan Mean Square Residue (MSR) di bawah threshold dan mendekati 0. Algoritma optimasi Cuckoo Search adalah algoritma pencarian solusi tricluster, digambarkan dengan konsep parasitisme spesies burung cuckoo. Pada penelitian ini, Cuckoo Search menggunakan random walk Gaussian Distribution untuk pencarian solusi tricluster. Berdasarkan hal ini komputasi algoritma Cuckoo Search menjadi lebih efisien dan efektif dalam menghasilkan himpunan tricluster yang lebih optimal dan mempercepat waktu komputasi. Fuzzy Cuckoo Search adalah pengembangan dari Cuckoo Search yang menggunakan fungsi objektif Fuzzy C-Means untuk mengatasi ketidakjelasan (uncertainty) dalam data ekspresi gen. Analisis triclustering menggunakan metode gabungan Fuzzy Cuckoo Search berdasarkan Gaussian Distribution dengan -Trimax digunakan pada data ekspresi gen tiga dimensi sel fibroblas yang diberikan perlakuan dengan Egr-1 dan Tgf-, di mana ekspresi gen diamati pada 6 kondisi dan 2 titik waktu. Pada penelitian ini, himpunan tricluster yang memiliki kualitas terbaik berdasarkan Triclustering Quality Index adalah himpunan tricluster yang dihasilkan dengan nilai $\alpha = 0,015$ dan $\beta = 0,50$. Berdasarkan himpunan tricluster tersebut, didapatkan informasi penting mengenai kumpulan gen yang memiliki respon baik terhadap pemberian perlakuan dengan Egr-1, Tgf- dan bertahan setiap titik waktu. Kumpulan gen tersebut dilakukan Gene Ontology (GO) yang diuji menggunakan Fisher's exact dengan tingkat signifikansi 0,05 dan dikoreksi dengan False Discovery Rate. Hasil GO tersebut terdiri dari 219 GO Terms Biological Process, 28 GO Terms Molecular Function, dan 52 GO Terms Cellular Component. GO Terms dari masing-masing aspek GO tersebut dapat dijadikan bahan untuk penelitian di bidang bioinformatika untuk menganalisis hubungan GO Terms terhadap penyakit Systemic Sclerosis (SSc).

.....Triclustering analysis is one of the data mining methods aimed at clustering three dimensional data. Triclustering is commonly used in the field of bioinformatics to analyze the similarity of gene expression in an experiment at specific time points. The triclustering analysis in this research uses a combined method of Fuzzy Cuckoo Search based on Gaussian Distribution with -Trimax. This method combines the nodes deletion algorithm of -Trimax with the optimization algorithm of Fuzzy Cuckoo Search. The nodes deletion algorithm of -Trimax is used in the initial population formation phase of the tricluster. The concept of the

nodes deletion algorithm is to produce tricluster sets with Mean Square Residue (MSR) below the threshold and close to 0. The optimization algorithm of Cuckoo Search is a search algorithm for tricluster solutions, depicted with the parasitism concept of cuckoo bird species. In this research, Cuckoo Search uses random walk Gaussian Distribution for tricluster solution search. This enhances the efficiency and effectiveness of the Cuckoo Search algorithm in producing more optimal tricluster sets and accelerating the computation time. Fuzzy Cuckoo Search is an extension of Cuckoo Search that employs Fuzzy C-Means objective function to handle uncertainty in gene expression data. The triclustering analysis using the combined method of Fuzzy Cuckoo Search based on Gaussian Distribution with -Trimax is applied to the three-dimensional gene expression data of fibroblast cells treated with Egr-1 and Tgf-1, where gene expressions are observed under 6 conditions and 2 time points. In this research, the tricluster set with the best quality based on the Triclustering Quality Index (TQI) is obtained with $\alpha = 0.015$ and $\beta = 0.50$. Based on this tricluster set, important information is derived regarding groups of genes that respond well to treatment with Egr 1, Tgf, and persist at each time point. These gene groups are subjected to Gene Ontology (GO) analysis, which is tested using Fisher's exact test with a significance level of 0.05 and corrected with False Discovery Rate. The GO results consist of 219 GO Terms Biological Process, 28 GO Terms Molecular Function, and 52 GO Terms Cellular Component. The GO Terms from each aspect can be utilized for further research in the field of bioinformatics to analyze the relationship of GO Terms with Systemic Sclerosis (SSc) disease.