

Implementasi Triclustering Menggunakan Metode Fuzzy Cuckoo Search berdasarkan Distribusi Gaussian pada Data Ekspresi Gen = The Implementation of Triclustering using Fuzzy Cuckoo Search Based on Gaussian Distribution on Gene Expression Data

Akira Andriani, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20525381&lokasi=lokal>

Abstrak

Analisis clustering merupakan proses pengelompokan yang bertujuan untuk menemukan kelompok atau cluster yang didalamnya memiliki karakteristik yang serupa. Seiring berjalannya waktu, teknik clustering berkembang menjadi biclustering dan triclustering, di mana dalam triclustering data yang digunakan adalah data tiga dimensi. Triclustering mampu mengelompokkan ketiga dimensi tersebut secara bersamaan yang nantinya kelompok yang dihasilkan disebut dengan tricluster. Pada penelitian ini, digunakan metode Fuzzy Cuckoo Search (FCS) untuk mengimplementasikan triclustering pada data ekspresi gen tiga dimensi. FCS mengaplikasikan konsep Fuzzy C-Means (FCM) ke dalam algoritma cuckoo search. Penggunaan fungsi objektif FCM dalam FCS dapat mengatasi ketidakjelasan (uncertainty) dalam data, khususnya pada data ekspresi gen. Dalam metode cuckoo search, pencarian 'solusi' tricluster digambarkan dengan spesies cuckoo yang meletakkan telur di sarang burung lain. Berbeda dengan cuckoo search pada umumnya yang menggunakan metode random walk levy flight untuk pencarian solusi, pada penelitian ini, digunakan metode lain, yaitu metode random walk distribusi gaussian, di mana hal tersebut merupakan sebuah kebaruan dalam penelitian ini. Cuckoo search dalam metode FCS merupakan metode metaheuristik, sehingga dapat digunakan dalam berbagai masalah analisis data, termasuk data ekspresi gen. Metode FCS berdasarkan distribusi gaussian diimplementasikan pada data ekspresi gen tiga dimensi dari gen otot rangka yang diberi infus IL-6, di mana ekspresi gen diamati pada 3 subjek dan 3 titik waktu yang berbeda. Metode ini dievaluasi menggunakan ukuran evaluasi Triclustering Quality Index (TQI). Dari skenario yang dilakukan, metode FCS memberikan hasil terbaik dengan rata-rata TQI terendah ketika menggunakan nilai gaussian dan probabilitas. Hasil implementasi metode FCS menunjukkan 4 tricluster yang diduga sebagai kumpulan gen yang berekspresi atas respon dari IL-6. Kelompok gen yang diperoleh dari tricluster dapat digunakan sebagai target oleh ahli medis dalam pengembangan di bidang pengobatan penyakit seperti kanker, diabetes, paru-paru, atau gagal jantung yang menargetkan gen-gen dalam kelompok tricluster tersebut.

.....Clustering analysis is a grouping process that aims to find clusters such that objects in the same clusters have similar characteristics. Over time, clustering developed into biclustering and triclustering, wherein triclustering use three-dimensional dataset. Triclustering is able to group these three dimensions simultaneously and form groups called tricluster. This study used the Fuzzy Cuckoo Search (FCS) method to implement triclustering on three-dimensional gene expression data. FCS applies the Fuzzy C-means (FCM) concept to the cuckoo search algorithm. The use of the objective function of FCM in FCS can overcome the uncertainty in the data, especially in gene expression data. In the cuckoo search, finding the tricluster is described with cuckoo species laying their egg in the nests of other birds. The egg laid on the nest represents a 'solution' which is an update of the tricluster from the previous tricluster. Unlike cuckoo search in general, in this study, to find the tricluster solutions, it use gaussian random walk instead of levy

flight random walk. Cuckoo search in the FCS method is a metaheuristic method, so it can be used in various data analysis problems, including gene expression data. FCS based on Gaussian distribution was implemented on three-dimensional gene expression data of skeletal muscle genes given IL-6 infusion, where the gene expression was observed in 3 subjects and 3 different time points. Of the 36 simulations performed, the FCS method gives the best results with the lowest average TQI when using gaussian values and probability. This method was evaluated using the Triclustering Quality Index (TQI) evaluation measure. The result of the implementation of FCS shows 4 triclusters which were suspected to be a collection of genes that change in response to IL-6. The gene groups obtained from the tricluster can be used as a consideration by medical professionals in the development of the treatment of diseases such as cancer, diabetes, pulmonary disease, or heart failure that target the genes in the tricluster group.