

Analisis Tricluster Menggunakan Metode Hybrid ?-TRIMAX Discrete Firefly Optimization pada Data Ekspresi Gen Tiga Dimensi = Tricluster Analysis with The Hybrid I'-TRIMAX Discrete Firefly Optimization Method on Three-Dimensional Gene Expression Data

Rana Amalia Putri, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920531222&lokasi=lokal>

Abstrak

Triclustering merupakan teknik data mining untuk mengelompokkan data tiga dimensi (observasi – atribut – konteks). Metode triclustering mampu menemukan kelompok observasi dengan karakteristik yang mirip pada beberapa atribut dan konteks sekaligus. Analisis tricluster kerap diimplementasikan pada data ekspresi gen 3D (gen – kondisi – waktu). Pada penelitian ini dilakukan analisis tricluster dengan metode Hybrid Discrete Firefly Optimization (DFO) pada data ekspresi gen 3D. Metode ini merupakan kombinasi dari metode dan Discrete Firefly Optimization (DFO). Metode DFO merupakan metode optimasi yang terinspirasi dari perilaku firefly (kunang-kunang) yang selalu bergerak ke arah firefly lain dengan cahaya yang lebih terang. Metode DFO bertujuan untuk memperoleh tricluster terbaik dengan volume maksimum dari suatu populasi tricluster. Pada metode DFO, suatu tricluster dengan volume lebih kecil akan bergerak ke arah tricluster lain dengan volume lebih besar dengan bantuan crossover operator. Setiap iterasi pada metode DFO dilakukan sampai semua tricluster dibandingkan volumenya dengan satu sama lain sehingga tricluster dengan volume lebih kecil dapat diperbaiki posisinya dan meningkat volumenya. Kebaruan dari penelitian ini merupakan pembentukan populasi tricluster awal dilakukan dengan metode . Metode bertujuan untuk menghasilkan tricluster dengan Mean Squared Residue (MSR) minimum, yaitu di bawah threshold tertentu. Populasi tricluster dengan karakteristik demikian meringankan beban dan waktu komputasi pada metode DFO. Data yang digunakan pada analisis tricluster ini merupakan data ekspresi gen dari sel kanker payudara MCF-7 ketika disuntikkan 100 nM hormon estrogen untuk empat waktu berbeda, yaitu selama 0 jam, 3 jam, 6 jam, dan 12 jam. Masing-masing perlakuan direplikasi sebanyak tiga kali. Pada penelitian ini diperoleh bahwa simulasi yang menghasilkan tricluster akhir terbaik dengan TQI terkecil adalah simulasi dengan nilai sebesar 3E-05. Gen-gen yang termasuk ke dalam tricluster ini memiliki profil ekspresi yang mirip ketika disuntikkan 100nM estrogen untuk tiga replikasi dan empat waktu. Estrogen merupakan salah satu hormon yang mempengaruhi pertumbuhan kanker payudara. Oleh karena itu, kumpulan gen dari tricluster yang diperoleh dapat dijadikan acuan pada penelitian selanjutnya oleh ahli medis terkait pengaruh estrogen terhadap kanker payudara dan untuk menentukan target terapi gen kanker payudara.

.....Triclustering is a data mining technique that is used to group observations on three dimensional data which consists of observation, attribute, and context dimension. Triclustering is used to find a group of observations with high similarity on several attributes and several contexts. Triclustering analysis is often implemented on three dimensional gene expression dataset. The 3D gene expression dataset dimensions consist of gene, condition, and time. The triclustering method that is used on this research is the Hybrid Discrete Firefly Optimization (DFO). This method is a combination of and DFO. Discrete Firefly Optimization (DFO) is an optimization method that was inspired by firefly movement in the wild in which a less bright firefly always moves to the one with brighter light. In triclustering analysis, every tricluster is called a firefly. Every firefly has a fitness value which is measured by its volume and a position which is

represented by a binary encoding of its elements. The DFO method is used to find the best tricluster with maximum volume from a population of triclusters. In the DFO method, a tricluster with smaller volume will move towards a tricluster with larger volume. The movement is done by updating the smaller volume tricluster position. The new position is obtained through crossover operator. Every iteration of DFO is completed once every tricluster fitness value is compared to each other. Repeating the DFO iteratively will give the end result of the best tricluster with maximum volume. The novelty of this research is the addition of method in constructing the initial tricluster population. The population created from will have MSR lower than the threshold so the computation time in the optimization step can be reduced. The Hybrid DFO method is implemented on a 3D gene expression dataset related to the MCF-7 breast cancer cell. The cell is injected with 100nM estrogen for four different times, i.e. 0, 3, 6, and 12 hours. Every experiment is repeated three times. This research finds that the simulation with the best tricluster based on the TQI score is the simulation with equals to $3E-05$. The genes that are included in the tricluster are found to have exhibit similar expression when injected by 100nM for three conditions and four time periods. Estrogen is one of the hormones that can affect breast cancer growth. Therefore, the genes that are included in the tricluster can be used as a reference for medical researchers in future research related to estrogen-induced breast cancer and to assist the selection for genes target in breast cancer therapy.