

Membangun THD-Tricluster Berbasis Biclustering dengan Menggunakan Transposed Virtual Error pada Data Ekspresi Gen = Building Biclustering Based THD-Tricluster Using Transposed Virtual Error in Gene Expression Data

Siregar, Angrainy Togi Marito, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20509937&lokasi=lokal>

Abstrak

Data tiga dimensi banyak ditemukan terutama dalam bidang biomedical dan sosial. Contoh data tiga dimensi adalah data gen-sampel-waktu, individual-fitur-waktu, atau node-node-time, yang umumnya disebut sebagai data konteks-atribut-observasi. Salah satu teknik dalam mengelola data tiga dimensi yaitu triclustering. Penelitian ini menggunakan metode THD-Tricluster untuk mendapatkan hasil triclustering pada data 3 dimensi penyakit AIDS yang disebabkan oleh virus Human Immunodeficiency Virus Tipe 1 (HIV-1). Pencarian triclustering tersebut dilakukan melalui tahap biclustering yang dilakukan pada setiap kondisi. Penelitian ini dengan memodifikasi algoritma Cheng & Church (CC) untuk menghasilkan biclustering dengan ukuran yang digunakan yaitu transposed virtual error) yang mampu mengelompokkan data berdasarkan pola pergeseran (shifting) dan penskalaan (scaling). Hasil bicluster tersebut yang kemudian diiris pada masing-masing kondisi sehingga menghasilkan triclustering. Keseluruhan penelitian ini dilakukan pada program RStudio. Setiap tricluster memiliki suatu kedalaman yang dapat ditentukan bergantung pada banyak kondisi yang diiris. Implementasi ukuran transposed virtual error pada triclustering ini menghasilkan 4 tricluster pada kedalaman empat. Evaluasi tricluster tersebut dilakukan dengan menggunakan inter temporal homogeneity dan diperoleh nilai korelasi tricluster antar kondisi di atas 0.9. Dari hasil triclustering tersebut dianalisis probe id gen yang berpengaruh pada penyakit AIDS. Dari analisis tersebut ditemukan 2 simbol gen yang berhubungan dengan penyakit AIDS yang disebabkan oleh HIV-1 yang ada pada setiap kondisi penderita HIV-1 normal, akut, kronis, dan nonprogressor yaitu HLA-C dan ELF-1.

Three-dimensional data are mainly found in biomedical and social fields. Examples of three-dimensional data are gene-sample-time, individual-feature-time, or node-node-time Data, commonly referred to as context-attribute-observation data. One of the techniques in managing three-dimensional data is triclustering. This study uses the THD Tricluster method to obtain triclustering results in 3-dimensional data on AIDS caused by the HIV-1 virus. The Triclustering search is carried out through biclustering stages carried out in each condition. This study modifies the Cheng & Church (CC) algorithm to produce a bicluster using the measure called transposed virtual error (which is able to group data based on shifting and scaling patterns. The biclustering results are then sliced under each condition to produce a tricluster. This research was conducted in the RStudio program. Each tricluster has a depth that can be determined depending on the many overlapping conditions. The implementation of the transposed virtual error size on biclustering results in 4 triclusters with a depth of four. The tricluster evaluation was carried out using inter temporal homogeneity and obtained a tricluster correlation value between conditions that was above 0.9. Then, the probe id genes that affect AIDS were analyzed from the results of the triclustering. From this analysis, 2 symbol genes associated with AIDS caused by HIV-1 were found in every condition of HIV-1

sufferers, which are normal, acute, chronic, and non-progressor HIV-1 sufferers, namely HLA-C and ELF-1.