

Metode timesvector triclustering dan penerapannya pada data ekspresi gen = Timesvector triclustering method and its application on gene expression data

Ika Marta Sari, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20509669&lokasi=lokal>

Abstrak

Analisis tri-clustering merupakan pengembangan dari analisis clustering dan analisis bi-clustering. Tujuan dari analisis tri-clustering yaitu mengelompokkan data tiga dimensi secara simultan atau bersamaan. Data tiga dimensi tersebut dapat berupa observasi, atribut, dan konteks. Salah satu pendekatan yang digunakan dalam analisis triclustering, yaitu pendekatan berdasarkan pattern contohnya, adalah metode Timesvector. Metode Timesvector bertujuan untuk mengelompokkan matriks data yang menunjukkan pola yang sama atau berbeda pada data tiga dimensi. Metode Timesvector memiliki langkah kerja yang dimulai dengan mereduksi matriks data tiga dimensi menjadi matriks data dua dimensi untuk mengurangi kompleksitas dalam pengelompokkan. Pada metode ini akan digunakan algoritma Spherical K-means dalam pengelompokkannya. Tahap selanjutnya, yaitu mengidentifikasi pola dari cluster yang dihasilkan pada Spherical K-means. Pola yang dimaksud terdiri dari tiga jenis, yaitu DEP (Differentially Expressed Pattern), ODEP (One Differentially Expressed Pattern), dan SEP (Similarly Expressed Pattern). Penerapan dari metode Timesvector dilakukan pada data ekspresi gen yaitu data tumor otak yang dilakukan dalam 6 skenario. Masing-masing skenario menggunakan banyak cluster yang sama tetapi nilai threshold yang berbeda-beda. Hasil dari ke enam skenario akan divalidasi menggunakan nilai coverage dan nilai tricluster diffusion (TD). Hasil penerapan metode timesvector menunjukkan bahwa dengan menggunakan threshold sebesar 1,5 memberikan hasil yang paling optimal karena memiliki nilai coverage yang tinggi sebesar 57% dan nilai TD yang rendah sebesar $2,95594E-06$. Nilai coverage yang tinggi menunjukkan kemampuan metode dalam mengekstrak data dan nilai TD yang rendah menunjukkan bahwa tricluster yang dihasilkan memiliki volume yang besar dan koherensi yang tinggi. Berdasarkan pola yang dihasilkan menggunakan skenario yang optimal diperoleh sebanyak 49 ODEP cluster dengan pasien ke-empat selalu memiliki pola ekspresi yang berbeda dibandingkan dengan pasien lainnya. Hal ini dapat digunakan oleh ahli medis untuk melakukan tindakan selanjutnya terhadap pasien tumor otak.

Triclustering analysis is the development of clustering analysis and biclustering analysis. The purpose of triclustering analysis is to group three-dimensional data simultaneously or simultaneously. The three-dimensional data can be in the form of observations, attributes, and context. One of the approaches used in triclustering analysis, namely an approach based on a pattern, for example, is the Timesvector method. Timesvector method aims to group data matrices that show the same or different patterns in three-dimensional data. The Timesvector method has a work step that starts with reducing the three-dimensional data matrix to a two-dimensional data matrix to reduce complexity in a grouping. In this method, the Spherical K-means algorithm will be used in grouping it. The next step is to identify the pattern of the clusters generated in the Spherical K-means. The pattern referred to consists of three types, namely

DEP (Differentially Expressed Pattern), ODEP (One Differentially Expressed Pattern), and SEP (Similar Expressed Pattern). The application of the Timesvector method was carried out on gene expression data, namely brain tumor data carried out in 6 scenarios. Each scenario uses the same many clusters but different threshold values. The results of the six scenarios will be validated using the coverage value and the tricluster diffusion (TD) value. The results of applying the timesvector method show that using a threshold of 1.5 gives the most optimal results because it has a high coverage value of 57% and a low TD value of 2.95594E-06. A high coverage value indicates the method's ability to extract data and a low TD value indicates that the resulting tricluster has a large volume and high coherence. Based on the pattern generated using the optimal scenario, there were 49 ODEP clusters with the fourth patient always having a different expression pattern compared to other patients. This can be used by medical experts to perform further action on brain tumor patients.