

Analisis biclustering menggunakan metode plaid model dan implementasinya pada data ekspresi gen = Biclustering analysis uses the plaid model method and its implementation on gene expression data

Achmad Eriza Aminanto, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20501598&lokasi=lokal>

Abstrak

Analisis *biclustering* merupakan pengembangan analisis *clustering*, dimana analisis *biclustering* merupakan proses partisi data matriks menjadi sub-matriks berdasarkan baris dan kolom secara simultan. Salah satu metode analisis *bicluster* yaitu dengan menggunakan model probabilistik, contohnya adalah *Plaid model* yang dapat memberikan hasil *bicluster* yang bersifat *overlapping*. *Plaid model*, memperhitungkan nilai elemen yang diberikan dari suatu sub-matriks tertentu, sehingga pada analisis *biclustering* dapat dilihat sebagai jumlah kontribusi atau efek dari *bicluster* tertentu. Tahapan analisis *biclustering* dengan *plaid model* diawali dengan input data berbentuk matriks, kemudian dilakukan penaksiran model awal dan membuat matriks residual dari model tersebut. Kemudian penentuan kandidat *bicluster*. Kandidat tersebut ditaksir parameter efeknya dan parameter keanggotaan *bicluster*. Terakhir dilakukan pemangkasan kandidat *bicluster* tersebut. Implementasi dilakukan pada data matriks ekspresi gen berupa data numerik yaitu data penyakit kanker usus, dimana baris berisikan observasi atau pasien sedangkan kolom berisikan jenis dari gen yang dilakukan dalam 6 skenario. Masing-masing skenario menggunakan parameter model dan nilai *threshold* berbeda. Validasi hasil implementasi menggunakan indeks *Jaccard* yaitu kedektahan hasil anggota *bicluster* dan variansi koherensi. Hasil implementasi menunjukkan penggunaan model yang lebih sederhana yang hanya menggunakan efek *mean* memberikan variansi koherensi yang lebih tinggi dibandingkan penggunaan model yang berisi *mean*, efek baris, dan efek kolom dari *bicluster*.

Biclustering analysis is the development of clustering analysis, which is the process of partitioning matrix data into sub-matrices based on rows and columns simultaneously. One method of bicluster analysis is using probabilistic model, for example the Plaid model that provide overlapping bicluster. Plaid model, calculates the value of an element given from a particular sub-matrix, thus can be seen as number of contributions of particular bicluster. The process begins with matrix data input, then an initial model is assessed and makes a residual matrix from the model. Then determining bicluster candidates. The candidate assessed for its effect parameters and bicluster membership parameters. Finally, the bicluster candidate was pruned. The implementation is carried out on the gene expression matrix data in form of numerical data, namely colon cancer data, where the rows contain observations while the columns contain the types of genes carried out in 6 scenarios. Each scenario uses different model parameters and threshold values. Validation of the implementation results using Jaccard index and coherence variance. Implementation results show that simpler model which only uses mean effect gives higher coherence variance than using model that contains mean, row, and column effect of the bicluster.