

Algoritma biclustering untuk matriks biner berdasarkan jarak hamming dan penerapannya dalam memprediksi interaksi protein antara protein HIV-1 dan protein manusia = Biclustering algorithm for binary matrix based on hamming distance and its application in predicting protein interactions between HIV-1 protein and human protein

Tampubolon, Patuan Pangihutan, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20493504&lokasi=lokal>

Abstrak

Kebutuhan untuk mendapatkan pola yang terbentuk dari matriks biner pada masa ini dan mendatang, meningkat dengan pesat. Data dari 'clickstream' pengguna internet, 'face-recognition', matriks setelah dilakukan prapengolahan dari data kategorik, interaksi protein-protein dan masih banyak daftar lainnya yang menghasilkan matriks biner. Salah satu pola yang dapat dibentuk dari matriks biner merupakan satu himpunan submatriks yang semua entrinya bernilai 1. Submatrik tersebut disebut dengan 'bicluster' dengan jenis nilai konstan. Permasalahan dari pembentukan 'bicluster' disebut dengan 'biclustering'. Permasalahan tersebut tergolong dalam permasalahan 'NP-complete'. Meskipun demikian, hasil yang suboptimal mampu didapatkan dengan membuat algoritma 'biclustering'.

Penelitian ini mengusulkan suatu algoritma 'biclustering' baru dengan menggunakan jarak 'Hamming' antara satu kolom dengan kolom yang lainnya pada matriks biner. Algoritma yang diberi nama 'bicHPT' ('biclustering based on Hamming distance Pattern Table') ini, mampu membuat satu himpunan 'bicluster' dengan lima langkah, yaitu mereduksi kolom matriks, membuat tabel jarak 'Hamming', mencari kandidat 'bicluster', menyaring kandidat 'bicluster', dan membentuk 'bicluster'. Setelah uji coba performa, algoritma 'bicHPT' mampu menghasilkan satu himpunan 'bicluster', bahkan mampu mengungguli algoritma lain dalam hal jumlah 'bicluster' yang dibentuk. Algoritma ini juga mampu untuk diaplikasikan sebagai salah satu unsur yang digunakan untuk memprediksi interaksi protein-protein baru, antara protein 'Human Immunodeficiency Virus type' 1 (HIV-1) dan protein manusia. Total interaksi baru yang didapatkan dengan menggunakan algoritma ini ada sebanyak 482 interaksi.

.....

The demand to obtain patterns from a binary matrix today and in the future is rapidly increasing. Data from internet users clickstreams, face-recognition, the matrix after preprocessing categorical data, protein-protein interactions, and so on that will produce a binary matrix. One kind of pattern that might be obtained from a binary matrix is a set of submatrices which all their entries have the value of 1. A submatrix is called with bicluster with constant values. The problem to make biclusters is called with biclustering. This problem is NP-complete. Although, the suboptimal solution might be obtained with constructing a biclustering algorithm.

This research proposes a novel biclustering algorithm based on Hamming distance among each column in a binary matrix. The algorithm which called with \pt (biclustering based on Hamming distance Pattern Table) can produce biclusters in 5 steps, which are, the column reduction of the matrix, constructing Hamming distance table, finding bicluster candidate, filtering bicluster candidate and forming the biclusters. After testing the performance, this algorithm can produce biclusters. Moreover, it can outperform another algorithm in numbers of biclusters. This algorithm is also succeeded to be applied as one of the elements to

predict protein-protein interaction between Human Immunodeficiency Virus type 1 protein (HIV-1) and human protein. The total new interactions which using this algorithm are 482 interaction.