

Metode Delta Trimax dengan Silhouette Coefficient dan Genetic Algorithm Optimization Menggunakan Analisis Gene Ontology pada Data Ekspresi Gen Tiga Dimensi = Delta Trimax Method with Silhouette Coefficient and Genetic Algorithm Optimization Using Gene Ontology Analysis on Three-Dimensional Gene Expression Data

Muhammad Fabian Alkautsar, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920551784&lokasi=lokal>

Abstrak

Analisis triclustering adalah metode data mining yang memiliki tujuan untuk mengelompokkan data tiga dimensi. Metode ini kerap kali digunakan untuk bidang bioinformatika. Pada penelitian ini digunakan metode analisis triclustering delta trimax. Delta Trimax pada intinya adalah metode analisis triclustering yang bertujuan untuk menemukan tricluster yang memiliki nilai MSR yang lebih kecil dari nilai threshold (θ) yang telah ditentukan. Penggunaan silhouette coefficient pada penelitian ini adalah untuk membantu menentukan nilai threshold (θ) tersebut. Hasil triclustering delta trimax nantinya dievaluasi dengan menggunakan Triclustering Quality Index (TQI). Genetic algorithm (GA) adalah sebuah algoritma pencarian yang efisien yang didasari oleh evolusi biologis dan genetika alam. Algoritma GA digunakan untuk menemukan solusi terbaik. Terdapat tiga operator genetika yang digunakan di dalam GA, yaitu seleksi, crossover, dan mutasi. Pada penelitian ini, digunakan data ekspresi gen tiga dimensi dari sel kanker paru-paru fase stabil (A549) yang diberi perlakuan obat kemoterapi Motexafin Gadolinium (MGd) dan mannitol sebagai grup kontrol, dimana ekspresi gen diamati dalam 6 kondisi dan 3 titik waktu. Tujuan dari penelitian ini adalah untuk mengetahui apa kumpulan gen yang memiliki respon baik terhadap pemberian obat kemoterapi MGd dan kondisi apa yang mempengaruhinya. Pada penelitian ini, himpunan tricluster yang memiliki kualitas terbaik berdasarkan Triclustering Quality Index (TQI) adalah himpunan tricluster yang dihasilkan dengan nilai $\theta = 0,004$. Berdasarkan himpunan tricluster tersebut, didapatkan informasi penting mengenai kumpulan gen yang memiliki respon baik terhadap pemberian MGd tapi efek obat MGd tidak bertahan di setiap titik waktu. Terdapat juga gen yang menunjukkan respon baik pemberian obat kemoterapi MGd, tetapi efektivitasnya tidak terlalu maksimal karena responnya beririsan dengan subjek yang hanya diberikan mannitol. Setelah itu, dilihat bagaimana hubungan gen yang berasal dari keseluruhan dataset dengan penyakit melalui gene ontology sebagai informasi tambahan untuk perkembangan obat MGd. Nilai fold enrichment tertinggi pada GO biological process adalah Cytoplasmic Translation, pada GO Cellular Component adalah cytosolic ribosome, dan pada GO Molecular Function adalah structural constituent of ribosome.

.....Triclustering analysis is a data mining method aimed at grouping three-dimensional data. This method is often used in the field of bioinformatics. In this study, the delta trimax triclustering analysis method is used. Delta Trimax essentially aims to find triclusters with Mean Squared Residue (MSR) values smaller than a predetermined threshold (θ). The silhouette coefficient is used in this study to help determine the threshold (θ). The results of the delta trimax triclustering are then evaluated using the Triclustering Quality Index (TQI). The genetic algorithm (GA) is an efficient search algorithm based on biological evolution and natural genetics. GA is used to find the best solution. There are three genetic operators used in GA: selection, crossover, and mutation. In this study, three-dimensional gene expression data from stable phase lung cancer

cells (A549) treated with the chemotherapy drug Motexafin Gadolinium (MGd) and mannitol as a control group were used, where gene expression was observed under 6 conditions and 3 time points. The aim of this study is to identify which sets of genes respond well to MGd chemotherapy and which conditions influence these responses. The set of triclusters with the highest quality based on the Triclustering Quality Index (TQI) was obtained with $\alpha=0.004$. From this set of triclusters, important information was obtained about the sets of genes that respond well to MGd, but the effect of MGd does not persist at every time point. There are also genes that show a good response to MGd chemotherapy, but its effectiveness is not maximized because the response overlaps with subjects that were only given mannitol. Subsequently, the relationship between genes from the entire dataset and the disease is observed through gene ontology as additional information for the development of MGd drugs. The highest fold enrichment value in the GO biological process is Cytoplasmic Translation, in the GO Cellular Component is cytosolic ribosome, and in the GO Molecular Function is structural constituent of ribosome.