

Evaluasi Kualitas Grafis Tricluster pada Algoritma Order Preserving Tricluster (OPTricluster) dengan Menggunakan Multi Slope Measure (MSL) pada Data Gene-Sample-Time (GST) = Tricluster Graphical Quality Evaluation on the Order Preserving Tricluster (OPTricluster) Algorithm using Multi Slope Measure (MSL) on Gene-Sample-Time (GST) Data

Luthfi Zahiya, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920541677&lokasi=lokal>

Abstrak

Triclustering merupakan metode unsupervised learning yang bekerja pada data tiga dimensi dengan dimensi-dimensi berupa observasi, atribut, dan konteks. Tujuan dari triclustering adalah untuk membentuk himpunan submatriks yang disebut sebagai tricluster berdasarkan ketiga dimensi pada data yang diberikan. Data tiga dimensi banyak ditemukan dalam bidang biomedis, di mana hal tersebut turut mendorong penggunaan dan pengembangan triclustering untuk keperluan analisis data ekspresi gen di mana salah satu alternatif dimensi-dimensinya adalah gen, sampel, dan waktu (gene, sample, time) dan biasa juga disebut sebagai data GST. Salah satu metode triclustering yang dikembangkan untuk menganalisis data tiga dimensi short time-series adalah Order Preserving Tricluster (OPTricluster).

OPTricluster mempertahankan sifat natural dimensi waktu yang berurutan dan mengelompokkan data ke dalam tricluster berdasarkan pola. Penelitian ini mengimplementasikan metode OPTricluster pada data GST biopsi otot rangka pria sehat yang berpuasa selama 24 jam. Tricluster-tricluster yang terbentuk dievaluasi menggunakan nilai Multi Slope Measure (MSL) untuk mengetahui kualitas grafis tricluster. MSL dikembangkan untuk mengevaluasi tricluster berisi data GST dan nilai MSL yang rendah menandakan kualitas tricluster yang baik. Berdasarkan analisis data GST, ditemukan bahwa penggunaan $\hat{\Gamma}=1,7$ menghasilkan tricluster pola constant dan divergent terbaik dan penggunaan $\hat{\Gamma}=1,5$ menghasilkan tricluster pola conserved terbaik. Namun, tricluster yang terbentuk mengindikasikan bahwa puasa selama 24 jam tidak menyebabkan banyak perubahan nilai ekspresi gen pada otot rangka manusia.

Triclustering is a method of unsupervised learning that runs on three-dimensional data where the dimensions are observation, attribute, and context. The objective of triclustering is to create a set of triclusters based on the three dimensions within the provided data. The abundance of three-dimensional data in the biomedical field is a big factor on the utilization and improvement of triclustering, particularly in the analysis of gene expression data which in it are the dimensions gene, sample, and time or abbreviated as GST data. One of the triclustering method developed to analyze three-dimensional short time-series data is Order Preserving Tricluster (OPTricluster). OPTricluster preserves the natural sequential properties of the time dimension and organizes data into triclusters based on patterns. This study applies the OPTricluster method on the GST data from muscle biopsies from fasting healthy men. The formed triclusters are evaluated graphically using the Multi Slope Measure (MSL). A small MSL score indicates a good tricluster. Based on the analysis of GST data, $\hat{\Gamma}=1,7$ produces the best constant and

divergent triclusters and $\hat{I} \leq 1,5$ produces the best conserved triclusters. However, the triclusters formed suggest that fasting for 24 hours doesn't have a lot of effect on gene expressions in human muscle.