

Metode Imputasi Sequential Biclustering Berbasis Shifting-and-Scaling Similarity dan Mean Squared Residue (SSSim-MSR) pada Data Ekspresi Gen Dua Dimensi = Sequential Biclustering Imputation Method Based on Shifting-and-Scaling Similarity and Mean Squared Residue on Two-Dimensional Gene Expression Data

Adib Hanafi, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20529058&lokasi=lokal>

Abstrak

Teknologi microarray merupakan salah satu teknologi yang berkembang dalam bidang bioinformatika. Salah satu teknologi microarray dalam bidang kesehatan, yaitu untuk mendeteksi adanya gen pada DNA individu yang menghasilkan data ekspresi gen. Pada data ekspresi gen, sering kali ditemukan informasi yang hilang sehingga membuat terhambatnya analisis lebih lanjut pada data ekspresi gen. Pada penelitian ini, diusulkan metode imputasi missing values Sequential Biclustering berbasis Shifting-and-Scaling Similarity dan Mean Square Residue (SSSim-MSR). Penentuan anggota bicluster dengan kesamaan sifat co-expressed dan pendeteksian pola shifting-and-scaling dilakukan berdasarkan pada skor Mean Squared Residue (MSR) dan skor Shifting-and-Scaling Similarity (SSSim) antara masing-masing gen dengan gen yang mengandung missing values. Performa metode diukur dengan skor korelasi Pearson dan skor NRMSE, lalu dibandingkan dengan metode Chronological Biclustering berbasis PCor-MSRE. Berdasarkan pada skor korelasi Pearson, metode Sequential Biclustering dengan basis SSSim-MSR merupakan metode yang cukup baik dibandingkan metode Chronological Biclustering berbasis PCor-MSRE pada missing rate sebesar 20% dan 50% untuk setiap nilai k. Untuk setiap missing rate pada nilai k lainnya, skor korelasi Pearson yang dihasilkan belum tentu bernilai lebih besar untuk nilai k yang lebih besar. Hal ini dapat terjadi karena persebaran proporsi pola shifting-and-scaling dan yang tidak berpola shifting-and-scaling pada data yang digunakan cenderung sama, sehingga pada tahap pembentukan bicluster yang didasarkan pada keserupaan pola dan pendeteksian pola shifting-and-scaling dapat memengaruhi keserupaan pola yang dibentuk.

.....Microarray technology is one of the emerging technologies in the field of bioinformatics. One of the microarray technologies in the health sector is to detect the presence of genes in individual DNA that produce gene expression data. In gene expression data, missing information is often found, which hinders further analysis of gene expression data. In this study, a method of imputing missing values Sequential Biclustering based on Shifting-and-Scaling Similarity and Mean Square Residue (SSSim - MSR) is proposed. Determination of bicluster members with similar co-expressed characteristics and detection of shifting-and-scaling patterns is carried out based on the score. Mean Squared Residue (MSR) and Shifting-and-Scaling Similarity (SSSim) scores between each gene and genes containing missing values. The performance of the method was measured by the Pearson correlation score and the NRMSE score, then compared with the Chronological Biclustering method on the basis of PCor – MSRE. Based on the Pearson correlation score, the Sequential Biclustering method on the basis of SSSim – MSR is a fairly good method compared to the Chronological Biclustering method at a missing rate of 20% and 50% for each value of k. For each other missing rate for k values, the resulting Pearson correlation score is not necessarily greater for larger k values. This can happen because the proportions of shifting-and-scaling and non-shifting-and-scaling patterns in the data used tend to be the same, so that at the stage of bicluster formation based on

pattern similarity and detection of shifting-and-scaling patterns can detect similarity of pattern.