

Metode Imputasi Biclustering Terurut Berbasis Bayesian Principal Component Analysis, Mean Squared Residual, dan Jarak Euclidean (SBI-BPCA-MSREimpute) pada Data Ekspresi Gen Dua Dimensi = Biclustering Imputation Method Based on Bayesian Principal Component Analysis, Mean Squared Residual, and Euclidean Distance (SBI-BPCA-MSREimpute) on TwoDimensional Gene Expression Data

Kevin Prawira, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20525194&lokasi=lokal>

Abstrak

Studi mengenai kesehatan terus dikembangkan untuk meneliti dan mengetahui struktur penyusun makhluk hidup dalam dunia ini. Salah satunya adalah studi mengenai gen yang telah dikembangkan melalui teknologi microarray yang dapat menghasilkan data ekspresi gen. Namun sayangnya, studi ini sering terhambat akibat banyaknya informasi yang hilang pada data ekspresi gen tersebut, sehingga analisis lebih lanjut mengenai data ekspresi gen sulit dilanjutkan. Oleh karena itu, dibutuhkan suatu metode imputasi untuk mengisi hilangnya informasi tersebut sehingga analisis dapat dilanjutkan dan dikembangkan dengan lebih luas lagi. Pada penelitian ini, dikembangkan sebuah metode imputasi baru bernama SBI-BPCA-MSREimpute yang menggunakan perpaduan Bayesian Principal Component Analysis (BPCA), Biclustering serta jarak Euclidean dalam melakukan prediksi nilai imputasi. Metode ini menggunakan pendekatan hybrid dalam imputasinya, yaitu dengan metode BPCA (global) sebagai metode pra-imputasi, serta penggunaan metode nearest neighbour (lokal) dalam penentuan bicluster untuk memetakan gen yang memiliki pola sama dengan gen target. Penggunaan BPCA didasarkan pada struktur korelasi data yang besar, sehingga BPCA cocok digunakan untuk mereduksi dimensi data. Adapun penentuan nearest neighbour sebagai bentuk bicluster didasarkan pada nilai mean squared residual serta jarak Euclidean terhadap gen target. Akibatnya, bicluster yang terpilih merupakan gen kandidat yang memiliki sifat sama dengan gen target. Perhitungan nilai imputasi akhir dihitung menggunakan rata-rata terboboti pada anggota bicluster, serta normalized root mean squared error digunakan sebagai pengukuran evaluasi. Percobaan imputasi menggunakan metode SBI-BPCA-MSREimpute dilakukan pada data ekspresi gen sel kanker usus besar dengan percobaan tingkat missing rate 5%, 10%, 15%, 20%, 25%, 30%, 40%, serta 50%. Terdapat beberapa k neighbour gen yang diuji yaitu pada tingkat k = 5%,10%,15%,20%,25% dari banyaknya baris pada data. Pengujian menunjukkan bahwa SBI-BPCA-MSREimpute merupakan metode imputasi yang lebih baik daripada SBI-MSREimpute. Pengujian juga mendapatkan hasil bahwa nilai k = 5%,10%,15%,25% paling optimal digunakan pada data dengan missing rates 15% serta k = 20% paling optimal digunakan pada data dengan missing rate 10%. Berdasarkan missing rates, data dengan missing rates 5%, 10%, 15%, 20%, 25%, 30%, 40% paling optimal diimputasi menggunakan tingkat k = 5%, sedangkan data dengan missing rates 50% paling optimal diimputasi menggunakan tingkat k = 10%.

.....