

Penerapan algoritma partisi som self organizing maps dalam metode hopach clustering = Implementation of som self organizing maps partitioning algorithm in hopach clustering method

Septian Wulandari, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20423273&lokasi=lokal>

Abstrak

Sejak adanya penemuan tentang struktur DNA yang berupa double helix, terdapat perkembangan tentang interaksi kompleks yang dibutuhkan untuk clustering (mengelompokkan) DNA menjadi clusters (kelompok-kelompok) yang memiliki kesamaan sifat ataupun fungsinya. Clustering DNA dapat dilakukan dengan metode partisi maupun metode hirarki. Dua metode tersebut dapat dipadukan dengan melakukan tahap partisi dan hirarki secara bergantian yang dikenal dengan nama HOPACH clustering. Tahap partisi dapat dilakukan dengan algoritma SOM, PAM, dan K-Means. Algoritma SOM dipilih karena menggunakan metode unsupervised learning dan efisien untuk digunakan pada data yang besar. Proses partisi dilanjutkan dengan proses ordering kemudian dilakukan collapsing dengan proses agglomerative, sehingga hasil clustering yang diperoleh menjadi lebih akurat.

Penentuan cluster utama dilakukan dengan menghitung nilai kehomogenan hasil clustering menggunakan MSS (Mean Split Silhouette). Kriteria penentuan cluster utama adalah pilih nilai MSS yang terkecil. Barisan 136 DNA EVD (Ebola Virus Disease) diperoleh dari Genbank NCBI dengan proses melakukan ekstraksi ciri DNA, selanjutnya melakukan normalisasi, dan dilanjutkan dengan menghitung jarak genetik menggunakan Jarak Euclidean. Matriks jarak genetik dapat dijadikan dasar untuk melakukan partisi serta clustering dengan menggunakan algoritma partisi SOM dalam metode HOPACH clustering. Proses ekstraksi ciri, normalisasi, dan penerapan algoritma partisi SOM dalam metode HOPACH clustering menggunakan bantuan program open source. Pada hasil clustering penerapan algoritma partisi SOM dalam metode HOPACH clustering diperoleh 9 cluster dengan nilai MSS sebesar 0,50280. Cluster yang dihasilkan dapat diidentifikasi berdasarkan spesies dan tahun pertama kali mewabah.

.....

Since the discovery of DNA structure in form of double helix, there is a development about the complex interaction required, DNA clustering into clusters which have the same features or functions. DNA clustering can be done by applying partitional or hierarchical method. Those two methods can be combined by doing partitional and hierarchical stage alternately known as HOPACH clustering. The partitional stage can be done by using SOM Algorithm, PAM, and K-Means. SOM algorithm is chosen because it uses unsupervised learning method and efficient to be used for large data. The partitional process is continued by ordering process and then performed collapsing with agglomerative process, so that the clustering result which is obtained will be more accurate. The determination of the main cluster done by calculating homogeneous value of the clustering result uses MSS (Mean Split Silhouette).

The determination criteria of the main cluster is choosing the smallest MSS value. 136 sequences of DNA EVD (Ebola Virus Disease) are obtained from NCBI Genbank by applying extraction of DNA sequence, after that doing normalization, and then calculating the genetic distance use Euclidean Distance. Genetic distance matrix can be used as a basis to do partitional and clustering by implementation SOM partitioning algorithm in HOPACH clustering method. The extraction of DNA sequence, normalization, and the

implementation of SOM partitioning algorithm in HOPACH clustering method use open source program . On the result of implementation SOM partitioning algorithm in HOPACH clustering method retrieved 9 clusters with MSS value of 0,50280. The cluster which is obtained can be identified according to species and the first year of becoming an epidemic.