

SIGNIFICANT GENES PATTERNS IDENTIFICATION IN HYPOXIA-INDUCED MODULATION GENE EXPRESSION DATA WITH K-MEANS ALGORITHM BASED ON NORMALIZED AND WEIGHTED MAHALANOBIS DISTANCE

Nwayyin Najat Mohammed^{1, 2} and Michael Anthony Cawthorn³

¹University of Sulaymaniyah, Iraq

²University of Zakho, Iraq

³University of Buckingham, UK

Partitioning clustering method breaks the original data set into k partitions. It performs certain optimal criteria iteratively. The partition clustering uses a criterion function, in which the squared error criterion is widely used with K-Means algorithm. The K-Means algorithm is a process of partitioning n-Dimensional objects into k sets, and it produces subsets which are reasonably coherent in sense of within-class variance. K-Means algorithm is computationally economical and it can be used on large datasets and microarray data that are considered as a big and complex data. However, the K-Means performance on hypoxia-induced modulation dataset which is pre-processed prior to the analyses, was not sufficient when used with default Euclidian distance. Therefore, we have used normalized and weighted Mahalanobis distance with K-Means algorithm. The proposed algorithm produced an ultimate clusters solution in the dataset, and the quality of generated clusters have been measured using Dunn validity index. The generated ultimate genes patterns reveal significant biological results in which significant pathways in hypoxia induced modulation dataset have been identified.

Biography

Nwayyin Najat Mohammed has completed her PhD at Zakho University, Kurdistan, Iraq as split site student with Buckingham University, UK. Her PhD was on Microarray data analysis using machine learning approaches. She has completed her MSc degree on Agent system in Babylon University, Iraq. She is now a Lecturer at the University of Sulaymaniyah, She has published 4 papers in reputed journals and conferences.

Nawing1@gmail.com