

Incorporating stability and error-based constraints for a novel partitional clustering algorithm

K. Aparna, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920533310&lokasi=lokal>

Abstrak

Data clustering is one of the major areas in data mining. The bisecting clustering algorithm is one of the most widely used for high dimensional dataset. But its performance degrades as the dimensionality increases. Also, the task of selection of a cluster for further bisection is a challenging one. To overcome these drawbacks, we developed a novel partitional clustering algorithm called a HB-K-Means algorithm (High dimensional Bisecting K-Means). In order to improve the performance of this algorithm, we incorporate two constraints, such as a stability-based measure and a Mean Square Error (MSE) resulting in CHB-K-Means (Constraint-based High dimensional Bisecting K-Means) algorithm. The CHB-K-Means algorithm generates two initial partitions. Subsequently, it calculates the stability and MSE for each partition generated. Inference techniques are applied on the stability and MSE values of the two partitions to select the next partition for the re-clustering process. This process is repeated until K number of clusters is obtained. From the experimental analysis, we infer that an average clustering accuracy of 75% has been achieved. The comparative analysis of the proposed approach with the other traditional algorithms shows an achievement of a higher clustering accuracy rate and an increase in computation time.