

The Enhancement of Rebuilding Centroid in K-Means Algorithm through Minimum Graph Forest

A. Maududie¹ and W. C. Wibowo²

Abstract: K-Means is one of famous algorithms based on partitioning clustering method, particularly on sum-of-squared error criterion. It generates a single partition data for a single group of data that has high degree in similarity. This method have some advantages such as linear complexity, ease of interpretation, simplicity of implementation, speed of convergence and adaptability to sparse data, and versatile in almost every aspect. However, this method also has some weaknesses, such as very sensitive to initial centroids (center) that drives the quality of clustering output. Although there is a recommendation to make some runs with different initial centroids and select the initial centroid that produces cluster with minimum error, frequently, this procedure does not achieve a satisfy result.

This paper introduces a new technique to overcome this problem through enhancing the refinement mechanism in K-Means algorithm. This technique focuses on rebuilding new centroids using minimum graph forest to reproduces better model in the refinement mechanism. In addition, the quality of clustering result was measured using two, i.e. Dunn index and Information gain.

¹ Department of Information System, Jember University
Kalimantan 37 Jember, East Java, Indonesia
maududie@gmail.com

² Department of Computer Science, The University of Indonesia
Kampus UI Depok 16424, West Java Indonesia
wibowo@cs.ui.ac.id