

MaxD K-Means: A clustering algorithm for auto-generation of centroids and distance of data points in clusters

Wan Maseri Wan Mohd; Abul Hashem Beg; Tutut Herawan; Khandakar Fazley Rabbi
Faculty of Computer Systems & Software Engineering University Malaysia
Pahang Kuantan Malaysia

ABSTRACT

K-Means is one of the unsupervised learning and partitioning clustering algorithms. It is very popular and widely used for its simplicity and fastness. The main drawback of this algorithm is that user should specify the number of cluster in advance. As an iterative clustering strategy, K-Means algorithm is very sensitive to the initial starting conditions. In this paper, we propose a clustering technique called MaxD K-Means clustering algorithm. MaxD K-Means algorithm auto generates initial k (the desired number of cluster) without asking for input from the user. MaxD K-means also used a novel strategy of setting the initial centroids. The experiment of the Max-D means has been conducted using synthetic data, which is taken from the Lloyd's K-Means experiments. The results from the new algorithm show that the number of iteration improves tremendously, and the number of iterations is reduced by confirming an improvement rate is up to 78%.

KEYWORDS:

K-means Algorithm; Partitioning Algorithm; Clustering; MaxD K-means; Data Mining

REFERENCES

1. Zhou, H., Liu, Y.: Accurate integration of multi-viewrange images using k-means clustering. *Pattern Recognition* 41, 152–175 (2008)
2. Bandyopadhyay, S., Maulik, U.: An evolutionary technique based on K-Means algorithm for optimal clustering. *Information Sciences* 146, 221–237 (2002)
3. Herawan, T., Yanto, I.T.R., Deris, M.M.: Rough Set Approach for Categorical Data Clustering. In: Ślęzak, D., Kim, T.-h., Zhang, Y., Ma, J., Chung, K.-i. (eds.) DTA 2009. CCIS, vol. 64, pp. 179–186. Springer, Heidelberg (2009)
4. Yanto, I.T.R., Herawan, T., Deris, M.M.: Data clustering using Variable Precision Rough Set. *Intelligent Data Analysis* 15(4), 465–482 (2011)
5. Yanto, I.T.R., Vitasari, P., Herawan, T., Deris, M.M.: Applying Variable Precision Rough Set Model for Clustering Student Suffering Study's Anxiety. *Expert System with Applications* 39(1), 452–459 (2012)