2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), Durban, South Africa, 6-7 August 2020 (Virtual)

HP: A light-weight hybrid algorithm for accurate data partitioning

Peter Olukanmi
Institute for Intelligent Systems University of Johannesburg, Johannesburg, South Africa
polukanmi@uj.ac.za

Fulufhelo Nelwamondo
Institute for Intelligent Systems, University of Johannesburg, Johannesburg, South Africa.
Council for Scientific and Industrial Research, Next Generation Enterprises and Institutions, Pretoria, South Africa
fnelwamondo@csir.co.za

Tshilidzi Marwala
Institute for Intelligent Systems, University of Johannesburg, Johannesburg, South Africa.
tmarwala@uj.ac.za

Abstract

This paper introduces a hybridization of the k-means and k-medoids paradigms. The new algorithms is named HP (hybrid partitioning) algorithm. Specifically, we improve on a recently developed scalable version of k-means (k-means-lite), by introducing the PAM algorithm into it in such a way that the high accuracy of the latter is absorbed without inheriting its high inefficiency. K-means-lite runs standard k-means on the combination of intermediate centroids obtained by initially feeding n samples into k-means. In HP, instead of k-means, PAM is used to cluster the combination of centroids obtained from the samples. This PAM component is fast because it is run on very small data, precisely of size nk. Experiments show that this modification improves not only the accuracy of k-means-lite but also outperforms the accuracy of k-means, without losing much k-means-lite's efficiency.