2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA), 28-30 January 2019

Performance evaluation of sampling-based large-scale clustering algorithms

Peter O. Olukanmi; Fulufhelo Nelwamondo; Tshilidzi Marwala

Abstract:
Using benchmark datasets, we study the performances of three efficient clustering algorithms which find cluster centers using a fixed number of random samples. The algorithms are also compared with two other (well-known) algorithms, namely k-means and PAM. One of the efficient algorithms, CLARA, is well-known while the other two, k-means-lite and PAM-lite, were introduced recently. CLARA and PAM-lite are based on the k-medoids approach, while k-means-lite adopts the k-means approach. The study shows that k-means-lite is the most efficient, followed by PAM-lite which is faster than CLARA. PAM-lite exhibits the best balance of efficiency and accuracy; it produces the most competitive results relative to PAM which is the most accurate but most inefficient.