

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

PAM-lite: Fast and accurate k-medoids clustering for massive datasets

Peter O. Olukanmi; Fulufhelo Nelwamondo; Tshilidzi Marwala

Abstract:

The Partitioning Around Medoids (PAM) clustering algorithm is well-known for its robustness and accuracy, but it is computationally expensive. This paper proposes a fast and accurate version, named PAM-lite. Like CLARA which also addresses PAM's inefficiency, PAM-lite applies PAM to random samples. However, unlike CLARA, it does not choose one of the obtained medoid sets (which would involve evaluating each set), but simply applies PAM again to the combination of all the obtained medoids. This simple change yields accuracy and speed improvement. We discuss the rationale behind PAM-lite's approach and evaluate the algorithm on benchmark datasets. In all cases tested, PAM-lite achieves better speed-up and clustering quality than CLARA; the speed-up margin increasing with problem size. PAM-lite competes so closely with the clustering quality produced by the full PAM algorithm, that in one high cluster variance case, it beats PAM's clustering quality slightly.