
An OpenMP parallelization of the K-means algorithm accelerated using KD-trees

Wojciech Kwedlo, Michał Łubowicz
Faculty of Computer Science
Białystok University of Technology
Wiejska 45a, 15-351 Białystok, Poland
w.kwedlo@pb.edu.pl

In the paper a KD-tree based filtering algorithm for K-means clustering is considered. A parallel version of the algorithm for shared memory systems, which uses OpenMP tasks both for KD-tree construction and filtering in the assignment step of K-means, is proposed. In our approach, an OpenMP task is created for a recursive call performed by tree construction and filtering procedures. A data partitioning step during the tree construction is also parallelized by OpenMP tasks. In computational experiments we measured runtimes of the parallel and serial version of the filtering algorithm and a parallel version of classical Lloyd's algorithm for six datasets sampled from two distributions. The results of experiments, performed on a 24-core system indicate that our version filtering algorithm has very good parallel efficiency. Its runtime is up to four orders of magnitude shorter than the runtime of parallel Lloyd's algorithm.

Keywords: OpenMP tasks, K-means clustering, KD-trees.