

ABSTRACT

Distributed-Memory Parallel Symmetric Nonnegative Matrix Factorization

Authors:

Srinivas Eswar, Dept. of Computational Science and Engineering, Georgia Tech

Koby Hayashi, Dept. of Computational Science and Engineering, Georgia Tech

Grey Ballard, Dept. of Computer Science, Wake Forest University

Ramakrishnan Kannan, Computer Science and Mathematics Division, ORNL

Richard Vuduc, Dept. of Computational Science and Engineering, Georgia Tech

Haesun Park, Dept. of Computational Science and Engineering, Georgia Tech

We develop the first distributed-memory parallel implementation of Symmetric Nonnegative Matrix Factorization (SymNMF), a key data analytics kernel for clustering and dimensionality reduction. Our implementation includes two different algorithms for SymNMF, which give comparable results in terms of time and accuracy. The first algorithm is a parallelization of an existing sequential approach that uses solvers for nonsymmetric NMF. The second algorithm is a novel approach based on the Gauss-Newton method. It exploits second-order information without incurring large computational and memory costs. We evaluate the scalability of our algorithms on the Summit system at Oak Ridge National Laboratory, scaling up to 128 nodes (4,096 cores) with 70% efficiency. Additionally, we demonstrate our software on an image segmentation task.