Multifrontal Non-negative Matrix Factorization

Piyush Sao, Ramakrishnan Kannan
Oak Ridge National Laboratory, USA
{saopk,kannanr}@ornl.gov

Non-negative matrix factorization(Nmf) is an important tool in high-performance large scale data analytics with applications ranging from community detection, recommender system, feature detection and linear and non-linear unmixing. While traditional NMF works well when the data set is relatively dense, however, it may not extract sufficient structure when the data is extremely sparse. Specifically, traditional NMF fails to exploit the structured sparsity of the large and sparse data sets resulting in dense factors. We propose a new algorithm for performing NMF on sparse data that we call multi-frontal NMF (Mf-Nmf) since it borrows several ideas from the multi-frontal method for unconstrained factorization (e.g. LU and QR). We also present an efficient shared memory parallel implementation of Mf-Nmf and discuss its performance and scalability. We conduct several experiments on synthetic and real-world datasets and demonstrate the usefulness of the algorithm by comparing it against standard baselines. We obtain a speedup of 1.2x to 19.5x on 24 cores with an average speedup of 10.3x across all the real-world datasets.

Keywords: NMF, recommender-system, sparse linear algebra, matrix factorization, graph algorithm.