Distributed Large-Scale Tensor Decompositions in Collaborative Networks

Prof. André de Almeida

Wireless Telecom Research Group (GTEL)
Federal University of Ceará (UFC), Fortaleza, Brazil

Abstract

In many disciplines, data inherently has more than two axes of variation and can be arranged as tensors (i.e. multi-way arrays). Computing tensor decompositions of multi-way datasets is particularly useful to extract hidden patterns and structure in data analytics problems. The traditional approach to compute tensor decomposition assumes a single machine with enough memory and processing power to process the whole data tensor. However, in large-scale (big data) applications where tensors can have billions of entries, tensor factorizations imply huge storage and processing capabilities that call for distributed algorithms affording some degree of parallel processing. The PARAFAC decomposition, also known as canonical polyadic (CP) decomposition, is a very popular tool for multi-way analysis, which has found innumerous applications in several domains including data mining, signal processing, and machine learning. In this talk, we discuss a fully distributed framework to compute the PARAFAC decomposition of a large-scale three-way data tensor across a network of collaborative machines with limited storage and computation resources. The big tensor is partitioned into several smaller sub-tensors that are allocated to different machines. Each machine operates on its sub-tensor to partially compute the factor matrices. The distributed computation of the PARAFAC decomposition is based on a “multi-way collaboration architecture": collaborative sub-networks of machines are established for each slicing direction, or mode, of the data tensor to estimate the factor matrices in a distributed and parallel fashion using alternating least squares. Such a multi-way network allows a meaningful reconstruction of the global factor matrices as for traditional (centralized) PARAFAC-ALS algorithms at the expense of some communication cost between machines. Insights on extensions of this approach to other tensor models will be briefly shown.