

Modeling multi-way data can be accomplished using tensors, which are data structures indexed along three or more dimensions. Tensor factorization is increasingly used to analyze extremely large and sparse multi-way datasets in life sciences, engineering, and business.

Tensor factorization is a computationally challenging task. The time and memory required to factor sparse tensors limits the size and dimensionality of the tensors that can be solved on a typical workstation, often making distributed solution approaches the only viable option. Existing tools are either totally unable to factor large tensors or can require days or even weeks to complete.

To that end, we present SPLATT, a software toolkit for large-scale sparse tensor factorization. SPLATT is a hybrid MPI+OpenMP code designed with performance from the start. SPLATT uses a compressed data structure that reduces memory requirements while reducing the operation count and improving cache locality. In result, SPLATT can factor tensors with billions of non-zeros in only a few minutes time on a supercomputer.