Tensors Factorizations for Sparse Data

Mathematical Methods for Tensor Decompositions

Tamara G. Kolda
Sandia National Laboratories
P.O. Box 969, MS 9159, Livermore, CA 94550

Abstract

Tensors have found application in a variety of fields, ranging from chemometrics to signal processing and beyond. In this presentation, we consider the problem of multilinear modeling of sparse data, specifically sparse count data. This type of data arises in a variety of contexts. For example, we may have data on computer communications (such as tracking communications in a large-scale simulation over discrete time intervals, perhaps even taking into account the message labels) or even person-to-person communications (such as emails tracked over time). Our goal is to develop a descriptive tensor model of such data.

Let $\mathcal{X}$ represent a 3-way data tensor of size $I \times J \times K$. (Our methods apply to arbitrary N-way models, but we discuss 3-way here for simplicity of notation.) As the data is strictly nonnegative, we will restrict the model to be nonnegative as well. Thus, we are interested in R-component model $\mathcal{M}$ of the form

$$\mathcal{M} = \sum_{r=1}^{R} \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r,$$

where λ_r is a positive scalar, $\mathbf{a}_r$, $\mathbf{b}_r$, and $\mathbf{c}_r$ denote the rth columns of nonnegative matrices of size $I \times R$, $J \times R$, and $K \times R$, respectively. The notation $\circ$ denotes the outer product so that $m_{ijk} = \sum_r \lambda_r \, a_{ir} \, b_{jr} \, c_{kr}$.

In many applications, we fit the model to the data using a least squares fit, which implicitly assumes that the random variation in the tensor data follows a Gaussian distribution. For count data, however, the random variation is better described via a Poission distribution. In other words, we assume

$$x_{ijk} \sim \text{Poisson}(m_{ijk})$$

rather than $x_{ijk} \sim N(m_{ijk}, \sigma_{ijk}^2)$. Correspondingly, we fit the data using the negative log-likelihood cost function rather than least squares. So, the optimal model will minimize

$$\sum_{ijk} m_{ijk} - x_{ijk} \log m_{ijk}.$$

The difficulty of this approach is fitting this more complex objective function. We present a novel alternating method that solves the subproblems via a majorization-minimization (MM) algorithm. Our approach generalizes the Lee-Seung multiplicative updates (even in the matrix case). The advantage of our approach is that we can prove convergence under conditions that are generically (i.e., with probability one) satisfied. We also present numerical results demonstrating the effectiveness of such factorizations as well as the interpretability of the results.

This is joint work with **Eric C. Chi**, who spent two summers at Sandia National Laboratories as part of the DOE Computational Science Graduate Fellowship program. Dr. Chi recently received his Ph.D. in Statistics from Rice University and currently holds a postdoctoral appointment at UCLA.