Limestone: Tensor Overview

Tensor Preliminaries

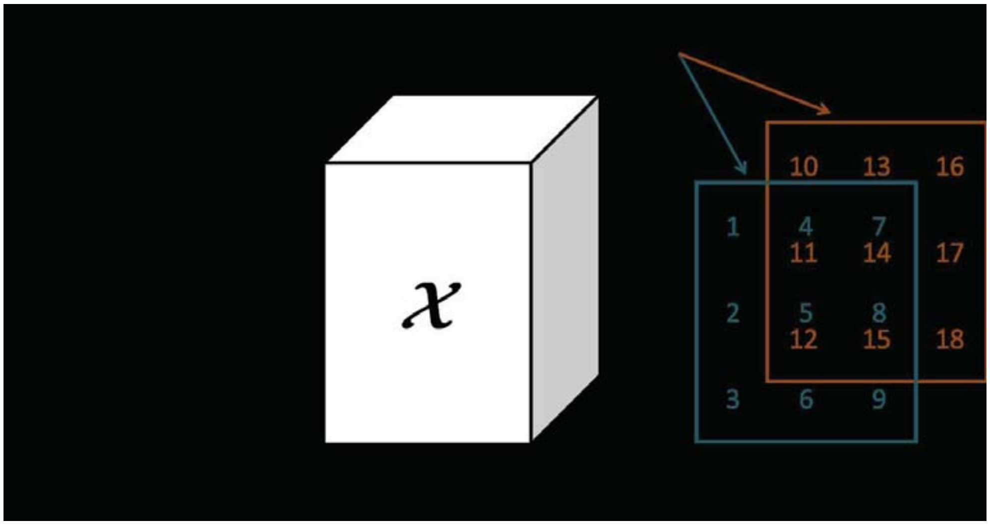

Figure 1: A third-order (or 3-mode) tensor of size $3 \times 3 \times 2$ with the tensor elements shown on the right side.

A tensor is a generalization of the matrix to multiway arrays. The number of dimensions of a tensor is known as the *order* or *mode* of the tensor. Figure 1 illustrates an example of a third-order (or 3-mode) tensor $\boldsymbol{\mathcal{X}}$ with the modes $1, 2$, and 3. Element (a, b, c) of $\boldsymbol{\mathcal{X}}$ is denoted by x_{abc}; e.g. $x_{2,3,2} = 17$, $x_{3,1,1} = 3$, and $x_{2,2,1} = 5$. For the rest of the paper, vectors (first-order tensors) are denoted by lowercase boldface letters ($\mathbf{x}$), matrices (second-order tensors) by uppercase boldface letters ($\mathbf{X}$), and higher-order (order three or higher) tensors by Euler boldface letters ($\boldsymbol{\mathcal{X}}$).

Slices are two-dimensional sections of a tensor, defined by fixing all but two modes. Figure 1 shows the two slices obtained from fixing modes 1 and 2.

The tensor $\boldsymbol{\mathcal{X}}$ can be unfolded (or flattened) into a matrix, which reorders the tensor elements into a matrix. $\mathbf{X}_{(n)}$ symbolizes the *mode-n matricization*. "Although matricization is conceptually simple, the formal notation is clunky" [1]. Thus, to orient the reader we provide the three mode-n matrices for our example tensor (Figure 1).

$$\mathbf{X}_{(1)} = \begin{bmatrix} 1 & 4 & 7 & 10 & 13 & 16 \\ 2 & 5 & 8 & 11 & 14 & 17 \\ 3 & 6 & 9 & 12 & 15 & 18 \end{bmatrix}$$
$$\mathbf{X}_{(2)} = \begin{bmatrix} 1 & 2 & 3 & 10 & 11 & 12 \\ 4 & 5 & 6 & 13 & 14 & 15 \\ 7 & 8 & 9 & 16 & 17 & 18 \end{bmatrix}$$
$$\mathbf{X}_{(3)} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 \end{bmatrix}$$

Matrix Algebra

Here we detail the matrix algebra used in the paper.

Definition 1. *The outer product of N vectors, $\mathbf{a}^{(1)} \circ \mathbf{a}^{(2)} \circ \cdots \circ \mathbf{a}^{(N)}$, produces a Nth order tensor $\boldsymbol{\mathcal{X}}$ where each element $x_{\vec{i}} = x_{i_1,i_2,\cdots,i_N} = a^{(1)}_{i_1} a^{(2)}_{i_2} \cdots a^{(N)}_{i_N}$.*

Definition 2. *The element-wise multiplication (and division) of two same-sized matrices $\mathbf{A} * \mathbf{B}$ ($\mathbf{A} \oslash \mathbf{B}$) produces a matrix $\mathbf{Z}$ of the same size such that the element $c_{\vec{i}} = a_{\vec{i}} b_{\vec{i}}$ ($c_{\vec{i}} = a_{\vec{i}} / b_{\vec{i}}$) for all $\vec{i}$.*

Definition 3. *The Khatri-Rao product of two matrices $\mathbf{A} \odot \mathbf{B}$ of sizes $I_A \times R$ and $I_B \times R$ respectively, produces a matrix $\mathbf{Z}$ of size $I_A I_B \times R$ such that $Z = \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \cdots & \mathbf{a}_R \otimes \mathbf{b}_R \end{bmatrix}$, where $\otimes$ represents the Kronecker product. The Kronecker product of two vectors* $\mathbf{a} \otimes \mathbf{b} = \begin{bmatrix} a_1\mathbf{b} \\ a_2\mathbf{b} \\ \vdots \\ a_{I_A}\mathbf{b} \end{bmatrix}$.

Tensor Factorization

Tensor factorization or decomposition is a natural extension of matrix factorization and utilizes information from the multiway structure that is lost when modes are collapsed to use matrix factorization algorithms [2, 3, 4]. The two common tensor decompositions, CANDECOMP / PARAFAC [5, 6] and Tucker [7], are considered higher-order generalizations of singular value decomposition and PCA [1].

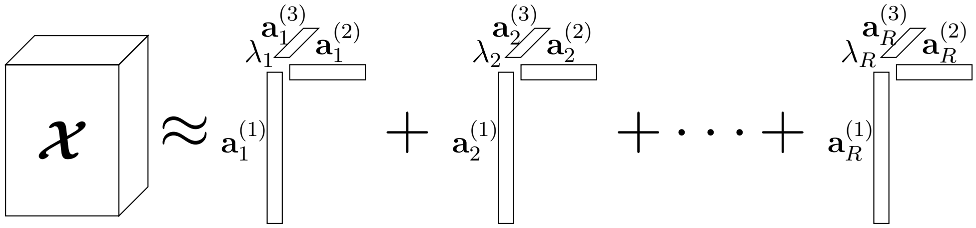

Figure 2: CANDECOMP/PARAFAC tensor decomposition

The CANDECOMP / PARAFAC (CP) model approximates the original tensor $\boldsymbol{\mathcal{X}}$ as a sum of R rank-one tensors

$$\begin{aligned} \boldsymbol{\mathcal{X}} &\approx \sum_{r=1}^{R} \lambda_r \mathbf{a}_r^{(1)} \circ \ldots \circ \mathbf{a}_r^{(N)} \\ &= \ \boldsymbol{\lambda}; \mathbf{A}^{(1)}; \ldots; \mathbf{A}^{(n)} \ \ . \end{aligned}$$

Note that $\ \boldsymbol{\lambda}; \mathbf{A}^{(1)}; \ldots; \mathbf{A}^{(n)}\ $ is shorthand notation to describe the CP decomposition, where $\boldsymbol{\lambda}$ is a vector of the weights λ_r and $\mathbf{a}_r^{(n)}$ is the rth column of $\mathbf{A}^{(n)}$.

The Tucker model decomposes a tensor into a core tensor $\mathcal{G}$ multiplied by a factor matrix $\mathbf{U}^{(n)}$ along each mode

$$\begin{aligned} \boldsymbol{\mathcal{X}} &\approx \sum_{j_1=1}^{J_1} \ldots \sum_{j_N=1}^{J_N} g_{j_1 \ldots j_N} (u_{j_1}^{(1)} \circ \ldots \circ u_{j_N}^{(N)}) \\ &= \boldsymbol{\mathcal{G}} \times_1 \mathbf{U}^{(1)} \times_2 \ldots \times_N \mathbf{U}^{(N)} \\ &= \ \boldsymbol{\mathcal{G}}; \mathbf{U}^{(1)}; \ldots; \mathbf{U}^{(N)} \ \ . \end{aligned}$$

The Tucker decomposition model is shown in Figure 3. Thus, CP is a special case of Tucker, where the size of each factor matrix has the same column dimension ($J_1 = J_2 = \ldots = J_N$) and the core tensor is superdiagonal [1].

References

[1] T. G. Kolda, B. W. Bader, Tensor decompositions and applications, SIAM Review 51 (2009) 455–500.

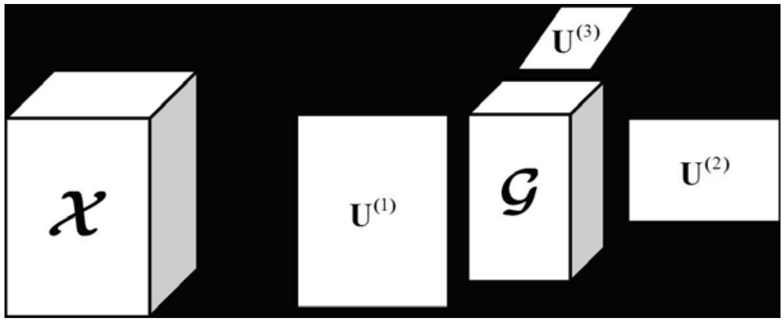

Figure 3: Tucker tensor decomposition

[2] H. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, A survey of multilinear subspace learning for tensor data, Pattern Recognition 44 (2011) 1540–1551.

[3] M. Mørup, Applications of tensor (multiway array) factorizations and decompositions in data mining, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1 (2011) 24–40.

[4] D. Wang, S. Kong, Feature selection from high-order tensorial data via sparse decomposition, Pattern Recognition Letters 33 (2012) 1695–1702.

[5] J. D. Carroll, J.-J. Chang, Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition, Psychometrika 35 (1970) 283–319.

[6] R. A. Harshman, Foundations of the PARAFAC procedure: models and conditions for an" explanatory" multimodal factor analysis, UCLA Working Papers in Phonetics 16 (1970) 1–84.

[7] L. R. Tucker, Some mathematical notes on three-mode factor analysis, Psychometrika 31 (1966) 279–311.