CMU SCS

15-826: Multimedia Databases and Data Mining

Lecture #21: Tensor decompositions

C. Faloutsos

CMU SCS

Must-read Material

- Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. Technical Report SAND2007-6702, Sandia National Laboratories, Albuquerque, NM and Livermore, CA, November 2007

2

CMU SCS

Outline

Goal: 'Find similar / interesting things'

- Intro to DB
- Indexing - similarity search
- Data Mining

3

CMU SCS

Indexing - Detailed outline

- primary key indexing
- secondary key / multi-key indexing
- spatial access methods
- fractals
- text
- Singular Value Decomposition (SVD)
  - …
  - Tensors
- multimedia
- ...

4

CMU SCS

Most of foils by

- Dr. Tamara Kolda (Sandia N.L.)
- csmr.ca.sandia.gov/~tgkolda

- Dr. Jimeng Sun (CMU -> IBM)
- www.cs.cmu.edu/~jimeng

3h tutorial: www.cs.cmu.edu/~christos/TALKS/SDM-tut-07/

5

CMU SCS

Outline

- Motivation - Definitions
- Tensor tools
- Case studies

6

CMU SCS

Motivation 0: Why "matrix"?

- Why matrices are important?

7

CMU SCS

Examples of Matrices: Graph - social network

	John	Peter	Mary	Nick	...
John	0	11	22	55	...
Peter	5	0	6	7	...
Mary	...	...	...	...	...
Nick	...	...	...	...	...
...	...	...	...	...	...

8

CMU SCS

Examples of Matrices: cloud of n-d points

	chol#	blood#	age	..	...
John	13	11	22	55	...
Peter	5	4	6	7	...
Mary	...	...	...	...	...
Nick	...	...	...	...	...
...	...	...	...	...	...

9

CMU SCS

Examples of Matrices: Market basket

- **market basket** as in Association Rules

	milk	bread	choc.	wine	...
John	13	11	22	55	...
Peter	5	4	6	7	...
Mary	...	...	...	...	...
Nick	...	...	...	...	...
...	...	...	...	...	...

10

CMU SCS

Examples of Matrices: Documents and terms

	data	mining	classif.	tree	...
Paper#1	13	11	22	55	...
Paper#2	5	4	6	7	...
Paper#3	...	...	...	...	...
Paper#4	...	...	...	...	...
...	...	...	...	...	...

11

CMU SCS

Examples of Matrices: Authors and terms

	data	mining	classif.	tree	...
John	13	11	22	55	...
Peter	5	4	6	7	...
Mary	...	...	...	...	...
Nick	...	...	...	...	...
...	...	...	...	...	...

12

CMU SCS

Examples of Matrices: sensor-ids and time-ticks

	temp1	temp2	humid.	pressure	...
t1	13	11	22	55	...
t2	5	4	6	7	...
t3	...	...	...	...	...
t4	...	...	...	...	...
...	...	...	...	...	...

13

CMU SCS

Motivation: Why tensors?

- Q: what is a tensor?

14

CMU SCS

Motivation 2: Why tensor?

- A: N-D generalization of matrix:

KDD'07	data	mining	classif.	tree	...
John	13	11	22	55	...
Peter	5	4	6	7	...
Mary	...	...	...	...	...
Nick	...	...	...	...	...
...	...	...	...	...	...

15

Motivation 2: Why tensor?

- A: N-D generalization of matrix:

KDD'05

KDD'06

KDD'07

	data	mining	classif.	tree	...
John	13	11	22	55	...
Peter	5	4	6	7	...
Mary	...	...	...	...	...
Nick	...	...	...	...	...
...	...	...	...	...	...

Tensors are useful for 3 or more modes

Terminology: 'mode' (or 'aspect'):

Mode#3

data	mining	classif.	tree	...
13	11	22	55	...
5	4	6	7	...
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

Mode#2

Mode (== aspect) #1

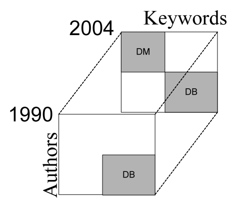
Motivating Applications

- Why matrices are important?
- Why tensors are useful?
  - P1: social networks
  - P2: web mining

CMU SCS

P1: Social network analysis

- Traditionally, people focus on static networks and find community structures
- We plan to monitor the change of the community structure over time



19

CMU SCS

P2: Web graph mining

- How to order the importance of web pages?
  - Kleinberg's algorithm HITS
  - PageRank
  - Tensor extension on HITS (TOPHITS)
    - context-sensitive hypergraph analysis

20

CMU SCS

Outline

- Motivation – Definitions
- Tensor tools
- Case studies

{
- Tensor Basics
- Tucker
- PARAFAC

21

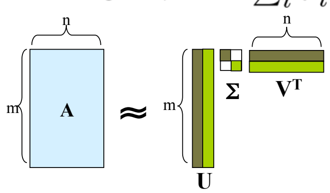
CMU SCS

Tensor Basics

CMU SCS

Reminder: SVD

$$\mathbf{A} \approx \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_i \sigma_i \mathbf{u}_i \circ \mathbf{v}_i$$



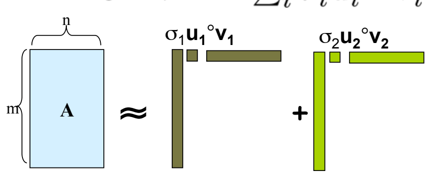
– Best rank-k approximation in L2

23

CMU SCS

Reminder: SVD

$$\mathbf{A} \approx \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_i \sigma_i \mathbf{u}_i \circ \mathbf{v}_i$$

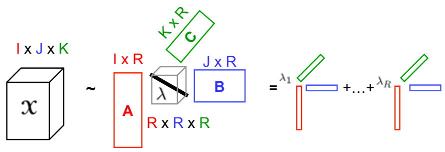


– Best rank-k approximation in L2

24

CMU SCS

Goal: extension to >=3 modes



$$\mathfrak{X} \approx [\![\lambda \ ; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] = \sum_r \lambda_r \ \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

25

CMU SCS

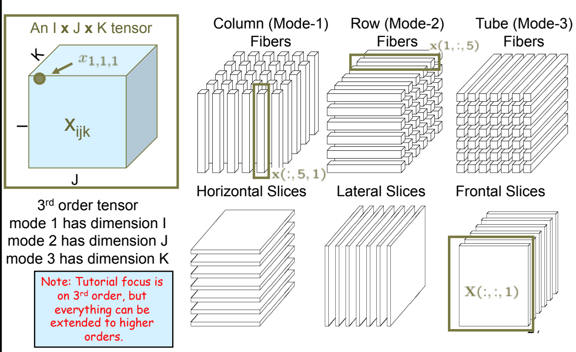
Main points:

- 2 major types of tensor decompositions: PARAFAC and Tucker
- both can be solved with ``alternating least squares'' (ALS)
- Details follow – we start with terminology:

26

CMU SCS

A tensor is a multidimensional array

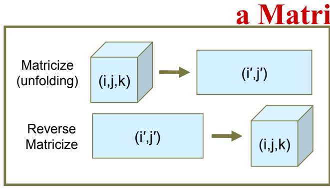


3rd order tensor
mode 1 has dimension I
mode 2 has dimension J
mode 3 has dimension K

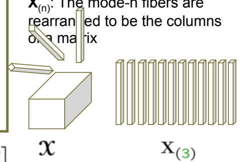
Note: Tutorial focus is on 3rd order, but everything can be extended to higher orders.

CMU SCS

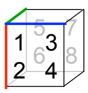
Matricization: Converting a Tensor to a Matrix



$\mathbf{X}_{(n)}$: The mode-n fibers are rearranged to be the columns of a matrix



$\mathcal{X} =$



$$\mathbf{X}_{(1)} = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{bmatrix}$$

$$\mathbf{X}_{(2)} = \begin{bmatrix} 1 & 2 & 5 & 6 \\ 3 & 4 & 7 & 8 \end{bmatrix}$$

$$\mathbf{X}_{(3)} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}$$

Vectorization

$$\text{vec}(\mathcal{X}) = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{bmatrix}$$

28

CMU SCS

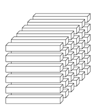
Tensor Mode-n Multiplication

$$\mathcal{X} \in \mathbb{R}^{I \times J \times K}, \ \mathbf{B} \in \mathbb{R}^{M \times J}, \ \mathbf{a} \in \mathbb{R}^{I}$$

- Tensor Times Matrix

$$\mathcal{Y} = \mathcal{X} \times_2 \mathbf{B} \in \mathbb{R}^{I \times M \times K}$$
$$y_{imk} = \sum_j x_{ijk} \, b_{mj}$$
$$\mathbf{Y}_{(2)} = \mathbf{B}\mathbf{X}_{(2)}$$

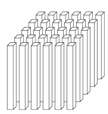
Multiply each row (mode-2) fiber by **B**



- Tensor Times Vector

$$\mathbf{Y} = \mathcal{X} \,\bar{\times}_1\, \mathbf{a} \in \mathbb{R}^{J \times K}$$
$$y_{jk} = \sum_i x_{ijk} \, a_i$$

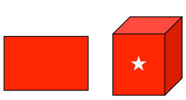
Compute the dot product of **a** and each column (mode-1) fiber



29

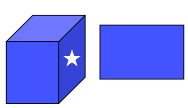
CMU SCS

Pictorial View of Mode-n Matrix Multiplication



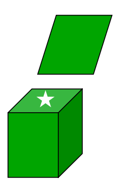
Mode-1 multiplication (frontal slices)

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{A}$$
$$\mathbf{Y}_{::k} = \mathbf{X}_{::k}\mathbf{A}^{\mathsf{T}}$$



Mode-2 multiplication (lateral slices)

$$\mathcal{Y} = \mathcal{X} \times_2 \mathbf{B}$$
$$\mathbf{Y}_{:j:} = \mathbf{X}_{:j:}\mathbf{B}^{\mathsf{T}}$$



Mode-3 multiplication (horizontal slices)

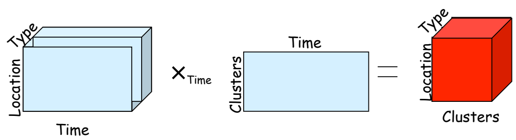
$$\mathcal{Y} = \mathcal{X} \times_3 \mathbf{C}$$
$$\mathbf{Y}_{i::} = \mathbf{X}_{i::}\mathbf{C}^{\mathsf{T}}$$

30

CMU SCS

Mode-n product Example

- Tensor times a matrix



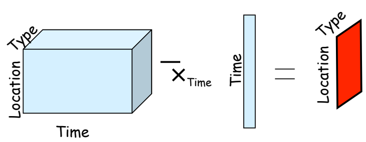
31

CMU SCS

details

Mode-n product Example

- Tensor times a vector



32

CMU SCS

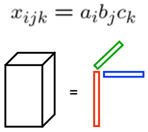
details

Outer, Kronecker, & Khatri-Rao Products

3-Way Outer Product

$$\mathcal{X} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$$

$$x_{ijk} = a_i b_j c_k$$



Review: Matrix Kronecker Product

$$\underset{M \times N}{\mathbf{A}} \otimes \underset{P \times Q}{\mathbf{B}} = \underset{MP \times NQ}{\begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1N}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2N}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1}\mathbf{B} & a_{M2}\mathbf{B} & \cdots & a_{MN}\mathbf{B} \end{bmatrix}}$$

$$= \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_1 \otimes \mathbf{b}_2 & \cdots & \mathbf{a}_N \otimes \mathbf{b}_Q \end{bmatrix}$$

Matrix Khatri-Rao Product

$$\underset{M \times R}{\mathbf{A}} \odot \underset{N \times R}{\mathbf{B}} = \underset{MN \times R}{\begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_2 \otimes \mathbf{b}_2 & \cdots & \mathbf{a}_R \otimes \mathbf{b}_R \end{bmatrix}}$$

Observe: For two vectors **a** and **b**, **a** ± **b** and **a** - **b** have the same elements, but one is shaped into a matrix and the other into a vector.

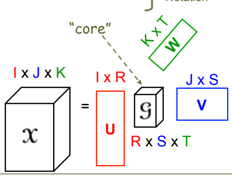
33

Specially Structured Tensors

Specially Structured Tensors

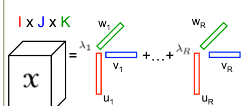
- Tucker Tensor

$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}$$
$$= \sum_r \sum_s \sum_t g_{rst}\, \mathbf{u}_r \circ \mathbf{v}_s \circ \mathbf{w}_t$$
$$\equiv [\![\mathcal{G}\,;\mathbf{U},\mathbf{V},\mathbf{W}]\!] \Big\} \text{ Our Notation}$$



- Kruskal Tensor

$$\mathcal{X} = \sum_r \lambda_r\, \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r$$
$$\equiv [\![\lambda\,;\mathbf{U},\mathbf{V},\mathbf{W}]\!] \Big\} \text{ Our Notation}$$



35

details

Specially Structured Tensors

- Tucker Tensor

$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}$$
$$= \sum_r \sum_s \sum_t g_{rst}\, \mathbf{u}_r \circ \mathbf{v}_s \circ \mathbf{w}_t$$
$$\equiv [\![\mathcal{G}\,;\mathbf{U},\mathbf{V},\mathbf{W}]\!]$$

In matrix form:

$$\mathbf{X}_{(1)} = \mathbf{U}\mathbf{G}_{(1)}(\mathbf{W} \otimes \mathbf{V})^{\mathsf{T}}$$
$$\mathbf{X}_{(2)} = \mathbf{V}\mathbf{G}_{(2)}(\mathbf{W} \otimes \mathbf{U})^{\mathsf{T}}$$
$$\mathbf{X}_{(3)} = \mathbf{W}\mathbf{G}_{(3)}(\mathbf{V} \otimes \mathbf{U})^{\mathsf{T}}$$

$$\mathrm{vec}(\mathcal{X}) = (\mathbf{W} \otimes \mathbf{V} \otimes \mathbf{U})\,\mathrm{vec}(\mathcal{G})$$

- Kruskal Tensor

$$\mathcal{X} = \sum_r \lambda_r\, \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r$$
$$\equiv [\![\lambda\,;\mathbf{U},\mathbf{V},\mathbf{W}]\!]$$

In matrix form:

Let $\Lambda = \mathrm{diag}(\lambda)$

$$\mathbf{X}_{(1)} = \mathbf{U}\,\Lambda\,(\mathbf{W} \odot \mathbf{V})^{\mathsf{T}}$$
$$\mathbf{X}_{(2)} = \mathbf{V}\,\Lambda\,(\mathbf{W} \odot \mathbf{U})^{\mathsf{T}}$$
$$\mathbf{X}_{(3)} = \mathbf{W}\,\Lambda\,(\mathbf{V} \odot \mathbf{U})^{\mathsf{T}}$$

$$\mathrm{vec}(\mathcal{X}) = (\mathbf{W} \odot \mathbf{V} \odot \mathbf{U})\,\lambda$$

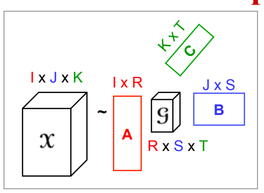
36

CMU SCS

Tensor Decompositions

CMU SCS

Tucker Decomposition - intuition



- author x keyword x conference
- A: author x author-group
- B: keyword x keyword-group
- C: conf. x conf-group
- $\mathcal{G}$: how groups relate to each other

38

CMU SCS

Intuition behind core tensor

- 2-d case: co-clustering
- [Dhillon et al. Information-Theoretic Co-clustering, KDD'03]

39

CMU SCS

$$
\begin{array}{cc}
 & n \\
m & \begin{bmatrix} .05 & .05 & .05 & 0 & 0 & 0 \\ .05 & .05 & .05 & 0 & 0 & 0 \\ 0 & 0 & 0 & .05 & .05 & .05 \\ 0 & 0 & 0 & .05 & .05 & .05 \\ .04 & .04 & 0 & .04 & .04 & .04 \\ .04 & .04 & .04 & 0 & .04 & .04 \end{bmatrix}
\end{array}
$$

eg, terms x documents

$$
\begin{array}{cc}
 & k \\
m & \begin{bmatrix} .5 & 0 & 0 \\ .5 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & .5 \\ 0 & 0 & .5 \end{bmatrix}
\end{array}
\begin{array}{cc}
 & l \\
k & \begin{bmatrix} .3 & 0 \\ 0 & .3 \\ .2 & .2 \end{bmatrix}
\end{array}
\begin{array}{cc}
 & n \\
l & \begin{bmatrix} 36 & .36 & .28 & 0 & 0 & 0 \\ 0 & 0 & 0 & .28 & .36 & .36 \end{bmatrix}
\end{array}
=
\begin{bmatrix} .054 & .054 & .042 & 0 & 0 & 0 \\ .054 & .054 & .042 & 0 & 0 & 0 \\ 0 & 0 & 0 & .042 & .054 & .054 \\ 0 & 0 & 0 & .042 & .054 & .054 \\ .036 & .036 & 028 & .028 & .036 & .036 \\ .036 & .036 & .028 & .028 & .036 & .036 \end{bmatrix}
$$

40

CMU SCS

med. doc

cs doc

$$
\begin{bmatrix} .05 & .05 & .05 & 0 & 0 & 0 \\ .05 & .05 & .05 & 0 & 0 & 0 \\ 0 & 0 & 0 & .05 & .05 & .05 \\ 0 & 0 & 0 & .05 & .05 & .05 \\ .04 & .04 & 0 & .04 & .04 & .04 \\ .04 & .04 & .04 & 0 & .04 & .04 \end{bmatrix}
$$

med. terms

cs terms

common terms

term group x
doc. group

$$
\begin{bmatrix} .5 & 0 & 0 \\ .5 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & .5 \\ 0 & 0 & .5 \end{bmatrix}
\begin{bmatrix} .3 & 0 \\ 0 & .3 \\ .2 & .2 \end{bmatrix}
\begin{bmatrix} 36 & .36 & .28 & 0 & 0 & 0 \\ 0 & 0 & 0 & .28 & .36 & .36 \end{bmatrix}
=
\begin{bmatrix} .054 & .054 & .042 & 0 & 0 & 0 \\ .054 & .054 & .042 & 0 & 0 & 0 \\ 0 & 0 & 0 & .042 & .054 & .054 \\ 0 & 0 & 0 & .042 & .054 & .054 \\ .036 & .036 & 028 & .028 & .036 & .036 \\ .036 & .036 & .028 & .028 & .036 & .036 \end{bmatrix}
$$

doc x
doc group

term x
term-group

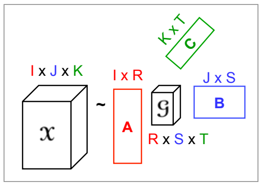
41

CMU SCS

42

CMU SCS

Tucker Decomposition



$$\mathcal{X} \approx [\![\mathcal{G} \,;\, \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

Given A, B, C, the optimal core is:

$$\mathcal{G} = [\![\mathcal{X} \,;\, \mathbf{A}^{\dagger}, \mathbf{B}^{\dagger}, \mathbf{C}^{\dagger}]\!]$$

- Proposed by Tucker (1966)
- AKA: Three-mode factor analysis, three-mode PCA, orthogonal array decomposition
- **A**, **B**, and **C** generally assumed to be orthonormal (generally assume they have full column rank)
- $\mathcal{G}$ is not diagonal
- Not unique

Recall the equations for converting a tensor to a matrix

$$\mathbf{X}_{(1)} = \mathbf{A}\mathbf{G}_{(1)}(\mathbf{C} \otimes \mathbf{B})^{\mathsf{T}}$$
$$\mathbf{X}_{(2)} = \mathbf{B}\mathbf{G}_{(2)}(\mathbf{C} \otimes \mathbf{A})^{\mathsf{T}}$$
$$\mathbf{X}_{(3)} = \mathbf{C}\mathbf{G}_{(3)}(\mathbf{B} \otimes \mathbf{A})^{\mathsf{T}}$$
$$\text{vec}(\mathcal{X}) = (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})\text{vec}(\mathcal{G})$$

43

CMU SCS

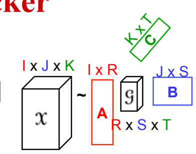
details

Solving for Tucker

$$\mathcal{X} \approx [\![\mathcal{G} \,;\, \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

Given A, B, C orthonormal, the optimal core is:

$$\mathcal{G} = [\![\mathcal{X} \,;\, \mathbf{A}^{\mathsf{T}}, \mathbf{B}^{\mathsf{T}}, \mathbf{C}^{\mathsf{T}}]\!]$$



Tensor norm is the square root of the sum of all the elements squared

Eliminate the core to get:

$$\underbrace{\| \mathcal{X} - [\![\mathcal{G} \,;\, \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2}_{\substack{\text{Minimize} \\ \text{s.t. } \mathbf{A},\mathbf{B},\mathbf{C} \text{ orthonormal}}} = \| \mathcal{X} \|^2 - 2\langle \mathcal{X}, [\![\mathcal{G} \,;\, \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \rangle + \| \mathcal{G} \|^2$$

$$= \underbrace{\| \mathcal{X} \|^2}_{\text{fixed}} - \underbrace{\left\| [\![\mathcal{X} \,;\, \mathbf{A}^{\mathsf{T}}, \mathbf{B}^{\mathsf{T}}, \mathbf{C}^{\mathsf{T}}]\!] \right\|^2}_{\text{maximize this}}$$

If B & C are fixed, then we can solve for A as follows:

$$\left\| [\![\mathcal{X} \,;\, \mathbf{A}^{\mathsf{T}}, \mathbf{B}^{\mathsf{T}}, \mathbf{C}^{\mathsf{T}}]\!] \right\| = \underbrace{\left\| \mathbf{A}^{\mathsf{T}}\mathbf{X}_{(1)}(\mathbf{C} \otimes \mathbf{B}) \right\|}$$

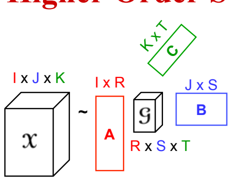
Optimal **A** is R left leading singular vectors for $\boxed{\mathbf{X}_{(1)}(\mathbf{C} \otimes \mathbf{B})}$

44

CMU SCS

details

Higher Order SVD (HO-SVD)



Not optimal, but often used to initialize Tucker-ALS algorithm.

(Observe connection to Tucker1)

$\mathbf{A}$ = leading R left singular vectors of $\mathbf{X}_{(1)}$

$\mathbf{B}$ = leading S left singular vectors of $\mathbf{X}_{(2)}$

$\mathbf{C}$ = leading T left singular vectors of $\mathbf{X}_{(3)}$

$$\mathcal{G} = [\![\mathcal{X} \,;\, \mathbf{A}^{\mathsf{T}}, \mathbf{B}^{\mathsf{T}}, \mathbf{C}^{\mathsf{T}}]\!]$$

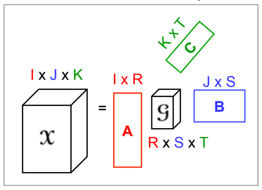
De Lathauwer, De Moor, & Vandewalle, SIMAX, 1980

45

CMU SCS

Tucker-Alternating Least Squares (ALS)

Successively solve for each component (**A**,**B**,**C**).



- Initialize
  - Choose R, S, T
  - Calculate **A**, **B**, **C** via HO-SVD
- Until converged do…
  - **A** = R leading left singular vectors of $\mathbf{X}_{(1)}(\mathbf{C} \otimes \mathbf{B})$
  - **B** = S leading left singular vectors of $\mathbf{X}_{(2)}(\mathbf{C} \otimes \mathbf{A})$
  - **C** = T leading left singular vectors of $\mathbf{X}_{(3)}(\mathbf{B} \otimes \mathbf{A})$
- Solve for core:

$$\mathcal{G} = [\![\mathcal{X} \,; \mathbf{A}^{\mathsf{T}}, \mathbf{B}^{\mathsf{T}}, \mathbf{C}^{\mathsf{T}}]\!]$$

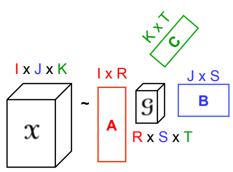
Kroonenberg & De Leeuw, Psychometrika, 1980

46

CMU SCS

details

Tucker in Not Unique



Tucker decomposition is *not* unique. Let Y be an RxR orthogonal matrix. Then…

$$\mathcal{X} \approx \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = \left(\mathcal{G} \times_1 \mathbf{Y}^{\mathsf{T}}\right) \times_1 (\mathbf{A}\mathbf{Y}) \times_2 \mathbf{B} \times_3 \mathbf{C}$$

$$\mathbf{X}_{(1)} \approx \mathbf{A}\mathbf{G}_{(1)}(\mathbf{C} \otimes \mathbf{B})^{\mathsf{T}} = \mathbf{A}\mathbf{Y}\mathbf{Y}^{\mathsf{T}}\mathbf{G}_{(1)}(\mathbf{C} \otimes \mathbf{B})^{\mathsf{T}}$$

47

CMU SCS

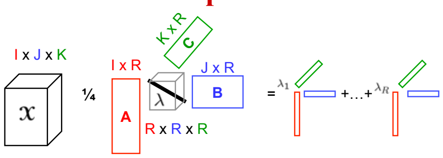
Outline

- Motivation – Definitions
- Tensor tools
- Case studies

- Tensor Basics
- Tucker
- PARAFAC

48

CANDECOMP/PARAFAC Decomposition



$$\mathfrak{X} \approx [\![\lambda \,;\, \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] = \sum_r \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

- CANDECOMP = Canonical Decomposition (Carroll & Chang, 1970)
- PARAFAC = Parallel Factors (Harshman, 1970)
- Core is diagonal (specified by the vector λ)
- Columns of **A**, **B**, and **C** are not orthonormal
- If R is *minimal*, then R is called the **rank** of the tensor (Kruskal 1977)
- Can have rank($\mathfrak{X}$) > min{I,J,K}

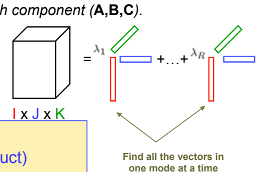
details

PARAFAC-Alternating Least Squares (ALS)

*Successively solve for each component (**A**,**B**,**C**).*

$$\mathfrak{X} \approx [\![\lambda \,;\, \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

$$\mathbf{X}_{(1)} \approx \mathbf{A}\Lambda(\mathbf{C} \odot \mathbf{B})^{\mathsf{T}}$$



Khatri-Rao Product
(column-wise Kronecker product)

$$\mathbf{C} \odot \mathbf{B} \equiv \left[\mathbf{c}_1 \otimes \mathbf{b}_1 \quad \mathbf{c}_2 \otimes \mathbf{b}_2 \quad \cdots \mathbf{c}_R \otimes \mathbf{b}_R\right]$$

$$(\mathbf{C} \odot \mathbf{B})^{\dagger} \equiv (\mathbf{C}^{\mathsf{T}}\mathbf{C} \underset{\text{Hadamard Product}}{*} \mathbf{B}^{\mathsf{T}}\mathbf{B})^{\dagger}(\mathbf{C} \odot \mathbf{B})^{\mathsf{T}}$$

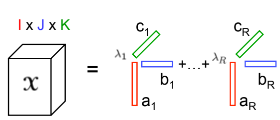
If **C**, **B**, and **Λ** are fixed, the optimal A is given by:

$$\mathbf{A} = \mathbf{X}_{(1)}(\mathbf{C} \odot \mathbf{B})(\mathbf{C}^{\mathsf{T}}\mathbf{C} * \mathbf{B}^{\mathsf{T}}\mathbf{B})^{\dagger}\Lambda^{-1}$$

*Repeat for **B**,**C**, etc.*

details

PARAFAC is often unique



Assume PARAFAC decomposition is exact.

$$\mathfrak{X} = [\![\lambda \,;\, \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] = \sum_r \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

Sufficient condition for uniqueness (Kruskal, 1977):

$$2R + 2 \le k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}}$$

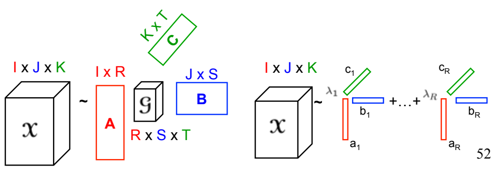
$k_{\mathbf{A}}$ = k-rank of **A** = max number k such that every set of k columns of **A** is linearly independent

CMU SCS

IMPORTANT

Tucker vs. PARAFAC Decompositions

- Tucker
  - Variable transformation in each mode
  - Core G may be dense
  - A, B, C generally orthonormal
  - Not unique
- PARAFAC
  - Sum of rank-1 components
  - No core, i.e., superdiagonal core
  - A, B, C may have linearly dependent columns
  - Generally unique



52

CMU SCS

Tensor tools - summary

- Two main tools
  - PARAFAC
  - Tucker
- Both find row-, column-, tube-groups
  - but in PARAFAC the three groups are identical
- To solve: Alternating Least Squares

- Toolbox: from Tamara Kolda:

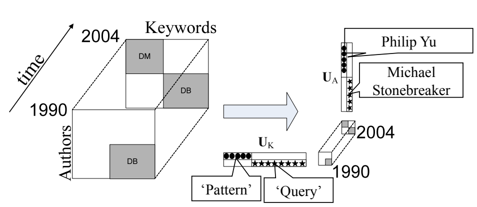
http://csmr.ca.sandia.gov/~tgkolda/TensorToolbox/

53

CMU SCS

P1: Social network analysis

- Multiway latent semantic indexing (LSI)
  - Monitor the change of the community structure over time



54

CMU SCS

P1: Social network analysis (cont.)

Authors	Keywords	Year
michael carey, michael stonebreaker, h. jagadish, hector garcia-molina	queri,parallel,optimization,concurr, ...ent	1995
surajit chaudhuri,mitch cherniack,michael stonebreaker,ugur etintemel	distribut,systems,view,storage,servic,process, cache	2004
jiawei han,jian pei,philip s. yu, jianyong wang,charu c. aggarwal	...tern,support,cluster, ...ner,queri	2004

DB

DM

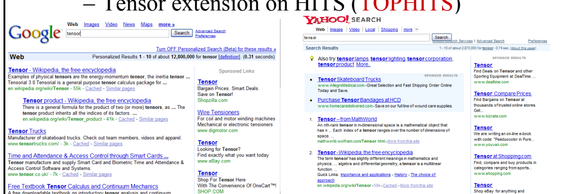
- Two groups are correctly identified: Databases and Data mining
- People and concepts are drifting over time

55

CMU SCS

P2: Web graph mining

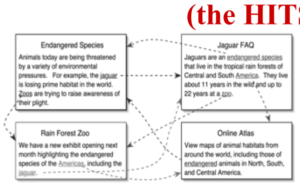
- How to order the importance of web pages?
  - Kleinberg's algorithm HITS
  - PageRank
  - Tensor extension on HITS (TOPHITS)

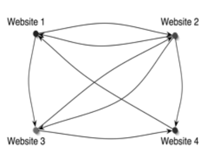


56

CMU SCS

Kleinberg's Hubs and Authorities (the HITS method)

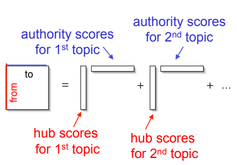




Sparse adjacency matrix and its SVD:

$$x_{ij} = \begin{cases} 1 & \text{if page } i \text{ links to page } j \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{X} \approx \sum_r \sigma_r \, \mathbf{h}_r \circ \mathbf{a}_r$$



57

Kleinberg, JACM, 1999

CMU SCS

HITS Authorities on Sample Data

1st Principal Factor	
.97	www.ibm.com
.24	www.alphaw
.08	www-128.ibm
.05	www.develop
.02	www.researc
.01	www.redbook
.01	news.com.co

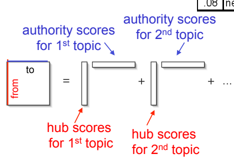
2nd Principal Factor	
.99	www.lehigh.edu
.11	www2.lehigh
.06	www.lehighal
.06	www.lehighs
.02	www.bethleh
.02	www.adobe.c
.02	lewisweb.cc.
.02	www.leo.lehi
.02	www.distanc
.02	fp1.cc.lehigh

3rd Principal Factor	
.75	java.sun.com
.38	www.sun.com
.36	developers.sun.
.24	see.sun.com
.16	www.samag.co
.13	docs.sun.com
.12	blogs.sun.com
.08	sunsolve.sun.c
.08	www.sun-catalo
.08	news.com.com

4th Principal Factor	
.60	www.pueblo.gsa.gov
.45	www.whitehouse.gov
.35	www.irs.gov
.31	travel.state
.22	www.gsa.g
.20	www.ssa.g
.16	www.censu
.14	www.govbe
.13	www.kids.g
.13	www.usdoj

6th Principal Factor	
.97	mathpost.asu.edu
.18	math.la.asu.edu
.17	www.asu.edu
.04	www.act.org
.03	www.eas.asu.edu
.02	archives.math.utk.edu
.02	www.geom.uiuc.edu
.02	www.fulton.asu.edu
.02	www.amstat.org
.02	www.maa.org

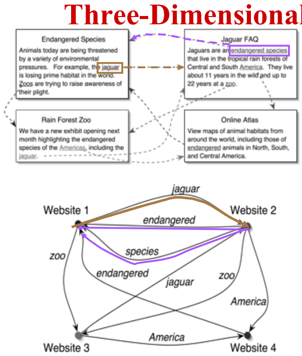
We started our crawl from http://www-neos.mcs.anl.gov/neos, and crawled 4700 pages, resulting in 560 cross-linked hosts.



58

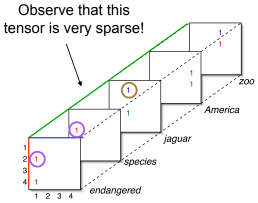
CMU SCS

Three-Dimensional View of the Web



$$x_{ijk} = \begin{cases} 1 & \text{if page } i \rightarrow \text{page } j \text{ with term } k \\ 0 & \text{otherwise} \end{cases}$$

Observe that this tensor is very sparse!



Kolda, Bader, Kenny, ICDM05

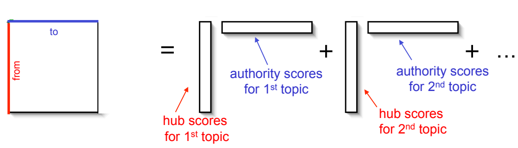
59

CMU SCS

Topical HITS (TOPHITS)

Main Idea: Extend the idea behind the HITS model to incorporate term (i.e., topical) information.

$$\mathfrak{X} \approx \sum_{r=1}^{R} \lambda_r \, \mathbf{h}_r \circ \mathbf{a}_r$$



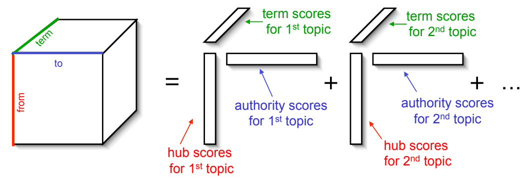
60

CMU SCS

Topical HITS (TOPHITS)

Main Idea: Extend the idea behind the HITS model to incorporate term (i.e., topical) information.

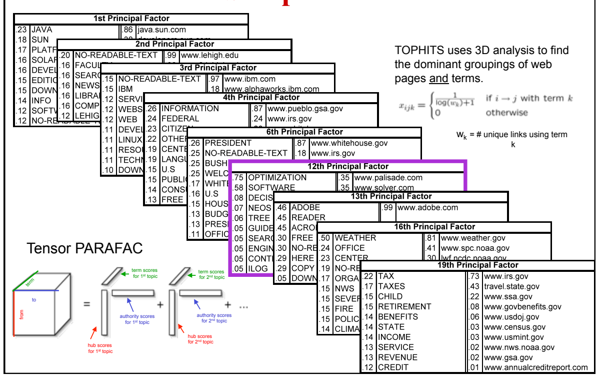
$$\mathcal{X} \approx \sum_{r=1}^{R} \lambda_r \, \mathbf{h}_r \circ \mathbf{a}_r \circ \mathbf{t}_r$$



61

CMU SCS

TOPHITS Terms & Authorities on Sample Data



TOPHITS uses 3D analysis to find the dominant groupings of web pages and terms.

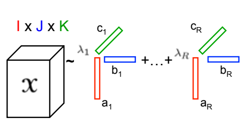
$$x_{ijk} = \begin{cases} \frac{1}{\log(w_k)+1} & \text{if } i \to j \text{ with term } k \\ 0 & \text{otherwise} \end{cases}$$

w_k = # unique links using term k

CMU SCS

Conclusions

- Real data are often in high dimensions with multiple aspects (modes)
- Matrices and tensors provide elegant theory and algorithms



63

CMU SCS

References

- Inderjit S. Dhillon, Subramanyam Mallela, Dharmendra S. Modha: Information-theoretic co-clustering. KDD 2003: 89-98
- T. G. Kolda, B. W. Bader and J. P. Kenny. *Higher-Order Web Link Analysis Using Multilinear Algebra*. In: ICDM 2005, Pages 242-249, November 2005.
- Jimeng Sun, Spiros Papadimitriou, Philip Yu. *Window-based Tensor Analysis on High-dimensional and Multi-aspect Streams*, Proc. of the Int. Conf. on Data Mining (ICDM), Hong Kong, China, Dec 2006

64