15-826: Multimedia Databases and Data Mining

Lecture #20: Tensor decompositions
C. Faloutsos

Must-read Material


Outline

Goal: ‘Find similar / interesting things’
• Intro to DB
• Indexing - similarity search
• Data Mining
Indexing - Detailed outline

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

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/

Outline

- Motivation - Definitions
- Tensor tools
- Case studies
Motivation 0: Why “matrix”?

• Why matrices are important?

Examples of Matrices:
Graph - social network

<table>
<thead>
<tr>
<th></th>
<th>John</th>
<th>Peter</th>
<th>Mary</th>
<th>Nick</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>0</td>
<td>11</td>
<td>22</td>
<td>55</td>
</tr>
<tr>
<td>Peter</td>
<td>5</td>
<td>0</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Mary</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nick</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Examples of Matrices:
cloud of n-d points

<table>
<thead>
<tr>
<th></th>
<th>chol#</th>
<th>blood#</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>13</td>
<td>11</td>
<td>22</td>
</tr>
<tr>
<td>Peter</td>
<td>5</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Mary</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nick</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Examples of Matrices:**

**Market basket**

- **market basket** as in Association Rules

<table>
<thead>
<tr>
<th></th>
<th>milk</th>
<th>bread</th>
<th>choc.</th>
<th>wine</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>13</td>
<td>11</td>
<td>22</td>
<td>55</td>
</tr>
<tr>
<td>Peter</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Mary</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Nick</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

**Documents and terms**

<table>
<thead>
<tr>
<th></th>
<th>data</th>
<th>mining</th>
<th>classif.</th>
<th>tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper#1</td>
<td>13</td>
<td>11</td>
<td>22</td>
<td>55</td>
</tr>
<tr>
<td>Paper#2</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Paper#3</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Paper#4</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

**Authors and terms**

<table>
<thead>
<tr>
<th></th>
<th>data</th>
<th>mining</th>
<th>classif.</th>
<th>tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>13</td>
<td>11</td>
<td>22</td>
<td>55</td>
</tr>
<tr>
<td>Peter</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Mary</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Nick</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Examples of Matrices: sensor-ids and time-ticks

<table>
<thead>
<tr>
<th></th>
<th>temp1</th>
<th>temp2</th>
<th>humid</th>
<th>pressure</th>
</tr>
</thead>
<tbody>
<tr>
<td>t1</td>
<td>13</td>
<td>11</td>
<td>22</td>
<td>55</td>
</tr>
<tr>
<td>t2</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>t3</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Motivation: Why tensors?

• Q: what is a tensor?

Motivation 2: Why tensors?

• A: N-D generalization of matrix:

<table>
<thead>
<tr>
<th></th>
<th>data</th>
<th>mining</th>
<th>classif.</th>
<th>tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>13</td>
<td>11</td>
<td>22</td>
<td>55</td>
</tr>
<tr>
<td>Peter</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Mary</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Nick</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Motivation 2: Why tensor?

- A: N-D generalization of matrix:

<table>
<thead>
<tr>
<th></th>
<th>Mode#3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>data</td>
<td>mining</td>
</tr>
<tr>
<td>John</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>Peter</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Mary</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>Nick</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

Terminology: ‘mode’ (or ‘aspect’):

- Mode (== aspect) #1

Tensors are useful for 3 or more modes

Motivating Applications

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

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

Outline

- Motivation – Definitions
- Tensor tools
  - Tensor Basics
  - Tucker
  - PARAFAC
- Case studies
Tensor Basics

Reminder: SVD

\[ A \approx U \Sigma V^T = \sum_i \sigma_i u_i \circ v_i \]

- Best rank-k approximation in L2
Goal: extension to >=3 modes

\[ \mathcal{X} \approx [\lambda; A, B, C] = \sum_{r} \lambda_r a_r \circ b_r \circ c_r \]

Main points:
- 2 major types of tensor decompositions: PARAFAC and Tucker
- both can be solved with "alternating least squares" (ALS)
- Details follow

Specially Structured Tensors
Specially Structured Tensors

- **Tucker Tensor**
  \[
  X = G \circ U \times_2 V \times_3 W \\
  = \sum_{i} \sum_{j} U_{i,j} \otimes V_{i,j} \otimes W_{i,j} \\
  = [G; U, V, W]
  \]

- **Kruskal Tensor**
  \[
  X = \sum_{i} \lambda_i U_{i} \otimes V_{i} \otimes W_{i} \\
  = [\lambda; U, V, W]
  \]

\[
I \times J \times K = U \times R \times J \times S \times T
\]

Our Notation

\[
X = (W \otimes V \otimes U) \lambda
\]

Tensor Decompositions
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

Intuition behind core tensor

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

- 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)
- $S$ is not diagonal
- Not unique

$\mathbf{X} \approx \mathbf{G} ; \mathbf{A}, \mathbf{B}, \mathbf{C}$

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

$\mathbf{G} = \langle \mathbf{X} ; \mathbf{A}^T, \mathbf{B}^T, \mathbf{C}^T \rangle$

Recall the equations for converting a tensor to a matrix:

$X_{(1)} = AG_{(2)}(C \otimes B)^T$

$X_{(2)} = BG_{(2)}(C \otimes A)^T$

$X_{(3)} = CG_{(2)}(B \otimes A)^T$

$\text{vec}(\mathbf{X}) = (C \otimes B \otimes A)\text{vec}(\mathbf{S})$
Outline

- Motivation – Definitions
- Tensor tools
- Case studies

Tensor tools

- CANDECOMP = Canonical Decomposition (Carroll & Chang, 1970)
- PARAFAC = Parallel Factors (Harshman, 1970)
- Core is diagonal (specified by the vector $\lambda$)
- 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($\bar{\mathcal{X}}$) > min(I,J,K)

CANDECOMP/PARAFAC Decomposition

\[ \mathcal{X} \approx \sum_{r} \lambda_r \ a_r \ b_r \ c_r \]

- 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

Tucker vs. PARAFAC Decompositions

IMPORTANT
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/

Outline

- Motivation - Definitions
- Tensor tools
  - Case studies

P1: Web graph mining

- How to order the importance of web pages?
  - Kleinberg’s algorithm HITS
  - PageRank
  - Tensor extension on HITS (TOPHITS)
P1: Web graph mining


Kleinberg’s Hubs and Authorities (the HITS method)

Sparse adjacency matrix and its SVD:
\[ x_{ij} = \begin{cases} 1 & \text{if page } i \text{ links to page } j \\ 0 & \text{otherwise} \end{cases} \]

\[ X \approx \sum_{r} \lambda_r h_r \otimes a_r \]

Hub scores for 1st topic
Authority scores for 1st topic

Authority scores for 2nd topic
Hub scores for 2nd topic

Kleinberg, JACM, 1999

HITS Authorities on Sample Data

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

Three-Dimensional View of the Web

Topical HITS (TOphits)

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

\[
\mathbf{X} \approx \sum_{r=1}^{R} \lambda_r \mathbf{h}_r \odot \mathbf{b}_r
\]

Topical HITS (TOphits)

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

\[
\mathbf{X} \approx \sum_{r=1}^{R} \lambda_r \mathbf{h}_r \odot \mathbf{b}_r \odot \mathbf{t}_r
\]
TOPHITS Terms & Authorities on Sample Data

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

authority scores

hub scores

hub scores

term scores

term scores

Tensor PARAFAC

wk = # unique links using term k

NELL (Never Ending Language Learner)

Nonzeros = 144M

GigaTensor: Scaling Tensor Analysis Up By 100 Times – Algorithms and Discoveries

U Kang
Evangelos Papalexakis
Abhay Harpale
Christos Faloutsos

KDD 2012

P2: N.E.L.L. analysis

• NELL: Never Ending Language Learner
  – Q1: dominant concepts / topics?
  – Q2: synonyms for a given new phrase?

“Eric Clapton plays guitar”

“Barrack Obama is the president of U.S.”

verbs (48M)

subjects (26M)

objects (26M)

NELL (Never Ending Language Learner)

Nonzeros = 144M
A1: Concept Discovery

• Concept Discovery in Knowledge Base

<table>
<thead>
<tr>
<th>Concept 1: &quot;Web Protocol&quot;</th>
<th>Concept 2: &quot;Credit Cards&quot;</th>
<th>Concept 3: &quot;Health System&quot;</th>
<th>Concept 4: &quot;Family Life&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>internet</td>
<td>credit</td>
<td>health</td>
<td>life</td>
</tr>
<tr>
<td>protocol</td>
<td>information</td>
<td>provider</td>
<td>rest</td>
</tr>
<tr>
<td>file</td>
<td>debt</td>
<td>number</td>
<td>part</td>
</tr>
<tr>
<td>software</td>
<td>report</td>
<td>library</td>
<td>years</td>
</tr>
<tr>
<td>suite</td>
<td>'np1'</td>
<td>'np2'</td>
<td>'np3'</td>
</tr>
<tr>
<td>'np1' marketing</td>
<td>'np1'</td>
<td>'np3' service</td>
<td>'np2' of 'np3'</td>
</tr>
<tr>
<td>'np2' dating</td>
<td>'np2'</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A2: Synonym Discovery

• Synonym Discovery in Knowledge Base
A2: Synonym Discovery

<table>
<thead>
<tr>
<th>Given Noun Phrase</th>
<th>Discovered Potential Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>pollutants</td>
<td>dioxin, sulfur dioxide,</td>
</tr>
<tr>
<td></td>
<td>greenhouse gases, particulates,</td>
</tr>
<tr>
<td></td>
<td>nitrogen oxide, all pollutants, cholesterol</td>
</tr>
<tr>
<td>disabilities</td>
<td>infections, dizziness, injuries, diseases, drowsiness, stiffness, injuries</td>
</tr>
<tr>
<td>vodafone</td>
<td>verizon, comcast</td>
</tr>
<tr>
<td>Christian history</td>
<td>European history, American history, Islamic history, history</td>
</tr>
<tr>
<td>disbelief</td>
<td>dismay, disgust, astonishment</td>
</tr>
<tr>
<td>cyberpunk</td>
<td>online-gaming</td>
</tr>
<tr>
<td>soul</td>
<td>body</td>
</tr>
</tbody>
</table>

Conclusions

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

\[ \mathbf{X} = a_1 \mathbf{A}_1 + a_2 \mathbf{A}_2 + \ldots + a_R \mathbf{A}_R \]

References

- Inderjit S. Dhillon, Subramanyam Mallela, Dharmendra S. Modha: Information-theoretic co-clustering. KDD 2003: 89-98
- 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