Part 3

Extension of time-series: tensor analysis

Yasushi Sakurai (Kumamoto University)
Yasuko Matsubara (Kumamoto University)
Christos Faloutsos (Carnegie Mellon University)
Outline

• Tensor decomposition
• Mining and forecasting of complex time-stamped events
• New challenge: MANT analysis

Multi-Aspect Non-linear Time-series

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/

© 2015 Sakurai, Matsubara & Faloutsos
Outline

• Tensor decomposition
  – Motivation
  – Basic approaches

• Mining and forecasting of complex time-stamped events

• New challenge: MANT analysis
  Multi-Aspect Non-linear Time-series

Motivation 1: 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>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>0</td>
<td>11</td>
<td>22</td>
<td>55</td>
<td>...</td>
</tr>
<tr>
<td>Peter</td>
<td>5</td>
<td>0</td>
<td>6</td>
<td>7</td>
<td>...</td>
</tr>
<tr>
<td>Mary</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Nick</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<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>
<th>...</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>13</td>
<td>11</td>
<td>22</td>
<td>55</td>
<td>...</td>
</tr>
<tr>
<td>Peter</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>...</td>
</tr>
<tr>
<td>Mary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nick</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<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>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>13</td>
<td>11</td>
<td>22</td>
<td>55</td>
<td>...</td>
</tr>
<tr>
<td>Peter</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>...</td>
</tr>
<tr>
<td>Mary</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Nick</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
### Examples of Matrices: Documents and terms

<table>
<thead>
<tr>
<th></th>
<th>data</th>
<th>mining</th>
<th>classif.</th>
<th>tree</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper#1</td>
<td>13</td>
<td>11</td>
<td>22</td>
<td>55</td>
<td>...</td>
</tr>
<tr>
<td>Paper#2</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>...</td>
</tr>
<tr>
<td>Paper#3</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Paper#4</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Examples of Matrices: Authors and terms

<table>
<thead>
<tr>
<th></th>
<th>data</th>
<th>mining</th>
<th>classif.</th>
<th>tree</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>13</td>
<td>11</td>
<td>22</td>
<td>55</td>
<td>...</td>
</tr>
<tr>
<td>Peter</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>...</td>
</tr>
<tr>
<td>Mary</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Nick</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<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>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>t=1</td>
<td>13</td>
<td>11</td>
<td>22</td>
<td>55</td>
<td>...</td>
</tr>
<tr>
<td>t=2</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>...</td>
</tr>
<tr>
<td>t=3</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>t=4</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Motivation 2: Why tensors?

• Q: what is a tensor?
Motivation 2: Why tensors?

- A: N-D generalization of matrix:

<table>
<thead>
<tr>
<th>sigmod’07</th>
<th>data</th>
<th>mining</th>
<th>query</th>
<th>tree</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>13</td>
<td>11</td>
<td>22</td>
<td>55</td>
<td>...</td>
</tr>
<tr>
<td>Peter</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>...</td>
</tr>
<tr>
<td>Mary</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Nick</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Motivation 2: Why tensors?

- A: N-D generalization of matrix:

<table>
<thead>
<tr>
<th></th>
<th>data</th>
<th>mining</th>
<th>query</th>
<th>tree</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>13</td>
<td>11</td>
<td>22</td>
<td>55</td>
<td>...</td>
</tr>
<tr>
<td>Peter</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>...</td>
</tr>
<tr>
<td>Mary</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Nick</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Tensors are useful for 3 or more modes

Terminology: ‘mode’ (or ‘aspect’):

Mode #1

<table>
<thead>
<tr>
<th>data</th>
<th>mining</th>
<th>query</th>
<th>tree</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>11</td>
<td>22</td>
<td>55</td>
<td>...</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mode #2

Mode #3
Motivating Applications

• Why matrices are important?
• Why tensors are useful?
  – P1: social networks
  – P2: web mining
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
Tensors for time-series analysis

• Time-stamped events
  – e.g., web clicks

<table>
<thead>
<tr>
<th>Time</th>
<th>URL</th>
<th>User</th>
</tr>
</thead>
<tbody>
<tr>
<td>08-01-12:00</td>
<td>CNN.com</td>
<td>Smith</td>
</tr>
<tr>
<td>08-02-15:00</td>
<td>YouTube.com</td>
<td>Brown</td>
</tr>
<tr>
<td>08-02-19:00</td>
<td>CNET.com</td>
<td>Smith</td>
</tr>
<tr>
<td>08-03-11:00</td>
<td>CNN.com</td>
<td>Johnson</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

Represent as $M^{th}$ order tensor ($M=3$)

$\mathbf{X} \in \mathbb{N}^{u \times v \times n}$

Element $x$: # of events

- e.g., ‘Smith’, ‘CNN.com’, ‘Aug 1, 10pm’; 21 times
Tensors for time-series analysis

- Individual-sequence mining
- Create a set of \((u \times v)\) sequences of length \(n\)
- Apply the mining algorithm for each sequence

URL \(u\)

user \(v\)

time \(n\)

\(u\) \(v\) \(n\)
Tensors for time-series analysis

• Multi-aspect time-series analysis

URL

user

v

n

time

Web clicks

\( \mathbf{X} \)

Topic A (business)

Topic B (news)

Topic C (media)
Outline

• Tensor decomposition
  – Motivation
  – Basic approaches

• Mining and forecasting of complex time-stamped events

• New challenge: MANT analysis

Multi-Aspect Non-linear Time-series
Reminder: SVD

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

– Best rank-k approximation in L2
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 \( \geq 3 \) modes

\[
\mathbf{X} \approx [\lambda ; A, B, C] = \sum_r \lambda_r \ a_r \circ b_r \circ c_r
\]
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

**Tucker Tensor**

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

**Kruskal Tensor**

\[
\mathbf{X} = \sum_r \lambda_r \mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r \equiv \left[ \lambda ; \mathbf{U}, \mathbf{V}, \mathbf{W} \right]
\]
Specially Structured Tensors

• Tucker Tensor
\[ X = \mathcal{G} \times_1 U \times_2 V \times_3 W = \sum_r \sum_s \sum_t g_{rst} \ u_r \odot v_s \odot w_t \equiv [\mathcal{G} ; U, V, W] \]
In matrix form:
\[ X_{(1)} = U G_{(1)} (W \otimes V)^\top \]
\[ X_{(2)} = V G_{(2)} (W \otimes U)^\top \]
\[ X_{(3)} = W G_{(3)} (V \otimes U)^\top \]
\[ \text{vec}(X) = (W \otimes V \otimes U) \text{vec}(\mathcal{G}) \]

• Kruskal Tensor
\[ X = \sum_r \lambda_r \ u_r \odot v_r \odot w_r \equiv [\lambda ; U, V, W] \]
In matrix form:
\[ X_{(1)} = U \Lambda (W \odot V)^\top \]
\[ X_{(2)} = V \Lambda (W \odot U)^\top \]
\[ X_{(3)} = W \Lambda (V \odot U)^\top \]
\[ \text{vec}(X) = (W \odot V \odot U) \lambda \]
Tucker Decomposition - intuition

- author x keyword x conference
- A: author x author-group
- B: keyword x keyword-group
- C: conf. x conf-group
- $G$: how groups relate to each other
Intuition behind core tensor

- 2-d case: co-clustering
- [Dhillon et al. Information-Theoretic Co-clustering, KDD’03]
eg, terms x documents
med. doc  cs doc

[0.05 0.05 0.05 0.0 0.0]
[0.05 0.05 0.05 0.0 0.0]
[0.0 0.0 0.0 .05 0.05 0.05]
[0.0 0.0 0.0 .05 0.05 0.05]
[0.04 0.04 0.04 0.04 0.04]
[0.04 0.04 0.04 0.04 0.04]

| med. terms |
| cs terms |
| common terms |

term group x doc. group

doc x doc group

term x term group
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)

G is not diagonal

Not unique

Recall the equations for converting a tensor to a matrix

\[
X(1) = AG_1(C \otimes B)^T \\
X(2) = BG_2(C \otimes A)^T \\
X(3) = CG_3(B \otimes A)^T \\
\text{vec}(X) = (C \otimes B \otimes A)\text{vec}(G)
\]

\[X \approx [G ; A, B, C]\]

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

\[G = [X ; A^\dagger, B^\dagger, C^\dagger]\]
CANDECOMP/PARAFAC Decomposition

\[ \mathcal{X} \approx [\lambda ; A, B, C] = \sum_{r} \lambda_r \mathbf{a}_r \odot \mathbf{b}_r \odot \mathbf{c}_r \]

- 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 \( \text{rank}(\mathcal{X}) > \min\{I,J,K\} \)
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

---


© 2015 Sakurai, Matsubara & Faloutsos
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

- Tensor decomposition
- Mining and forecasting of complex time-stamped events
- New challenge: MANT analysis

Multi-Aspect Non-linear Time-series

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/

© 2015 Sakurai, Matsubara & Faloutsos
Fast Mining and Forecasting of Complex Time-Stamped Events

Yasuko Matsubara (Kyoto University)
Yasushi Sakurai (NTT)
Christos Faloutsos (CMU)
Tomoharu Iwata (NTT)
Masatoshi Yoshikawa (Kyoto University)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/

© 2015 Sakurai, Matsubara & Faloutsos
Motivation

- Complex time-stamped events
  \{timestamp + multiple attributes\}

E.g., web click events:
  \{timestamp, URL, user ID, access devices, http referrer,...\}

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>URL</th>
<th>User</th>
<th>Device</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012-08-01-12:00</td>
<td>CNN.com</td>
<td>Smith</td>
<td>iphone</td>
</tr>
<tr>
<td>2012-08-02-15:00</td>
<td>YouTube.com</td>
<td>Brown</td>
<td>iphone</td>
</tr>
<tr>
<td>2012-08-02-19:00</td>
<td>CNET.com</td>
<td>Smith</td>
<td>mac</td>
</tr>
<tr>
<td>2012-08-03-11:00</td>
<td>CNN.com</td>
<td>Johnson</td>
<td>ipad</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Motivation

Q1. Are there any topics?
- news, tech, media, sports, etc...

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>URL</th>
<th>User</th>
<th>Device</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012-08-01-12:00</td>
<td>CNN.com</td>
<td>Smith</td>
<td>iphone</td>
</tr>
<tr>
<td>2012-08-02-15:00</td>
<td>YouTube.com</td>
<td>Brown</td>
<td>iphone</td>
</tr>
<tr>
<td>2012-08-02-19:00</td>
<td>CNET.com</td>
<td>Smith</td>
<td>mac</td>
</tr>
<tr>
<td>2012-08-03-11:00</td>
<td>CNN.com</td>
<td>Johnson</td>
<td>ipad</td>
</tr>
</tbody>
</table>

e.g.,

CNN.com, CNET.com -> news topic
YouTube.com -> media topic
Motivation

Q2. Can we group URLs/users accordingly?

e.g.,

**CNN.com & CNET.com** (related to news topic)

**Smith & Johnson** (related to news topic)
Motivation

Q3. Can we forecast future events?
- How many clicks from ‘Smith’ tomorrow?
- How many clicks to ‘CNN.com’ over next 7 days?

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>URL</th>
<th>User</th>
<th>Device</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012-08-01-12:00</td>
<td>CNN.com</td>
<td>Smith</td>
<td>iphone</td>
</tr>
<tr>
<td>2012-08-02-15:00</td>
<td>YouTube.com</td>
<td>Brown</td>
<td>iphone</td>
</tr>
<tr>
<td>2012-08-02-19:00</td>
<td>CNET.com</td>
<td>Smith</td>
<td>mac</td>
</tr>
<tr>
<td>2012-08-03-11:00</td>
<td>CNN.com</td>
<td>Johnson</td>
<td>ipad</td>
</tr>
<tr>
<td>2012-08-05-12:00</td>
<td>CNN.com</td>
<td>Smith</td>
<td>iphone</td>
</tr>
<tr>
<td>2012-08-05-19:00</td>
<td>CNET.com</td>
<td>Smith</td>
<td>iphone</td>
</tr>
</tbody>
</table>
Motivation

Web click events – can we see any trends?

Original access counts of each URL

- 100 random users
- 1 week (window size = 1 hour)

URL: money site

URL: blog site
Motivation

Web click events – can we see any trends?
Original access counts of each URL

We cannot see any trends!!

© 2015 Sakurai, Matsubara & Faloutsos
Our goals

Q1: Hidden topics
Q2: Groups
Q3: Forecasting

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
Problem definition

**Given:** a set of complex time-stamped events

1. **Find:** major topics/trends
2. **Forecast:** future events

**Original web-click events**

**URL in topic space**

**User in topic space**

"Hidden topics" wrt each aspect

(URL, user, time)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
Main idea (1): M-way analysis

Complex time-stamped events

*Example*: web clicks

<table>
<thead>
<tr>
<th>Time</th>
<th>URL</th>
<th>User</th>
<th>Time</th>
<th>URL</th>
<th>User</th>
</tr>
</thead>
<tbody>
<tr>
<td>08-01-12:00</td>
<td>CNN.com</td>
<td>Smith</td>
<td>08-02-15:00</td>
<td>YouTube.com</td>
<td>Brown</td>
</tr>
<tr>
<td>08-02-19:00</td>
<td>CNET.com</td>
<td>Smith</td>
<td>08-03-11:00</td>
<td>CNN.com</td>
<td>Johnson</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

Represent as $M^{th}$ order tensor ($M=3$)

$$ \mathbf{X} \in \mathbb{N}^{u \times v \times n} $$

Element $x$: # of events

*Example*: ‘Smith’, ‘CNN.com’, ‘Aug 1, 10pm’; 21 times
Main idea (1) : M-way analysis

A. decompose to a set of 3 topic vectors:

- **Object vector**  **Actor vector**  **Time vector**

### Diagram:
- **Object** vector: Web clicks
- **Actor** vector: Time
- **Time** vector: Topic A (business), Topic B (news), Topic C (media)

1. **Object vector**
2. **Actor vector**
3. **Time vector**

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/

© 2015 Sakurai, Matsubara & Faloutsos
Main idea (1) : M-way analysis

A. decompose to a set of 3 topic vectors:

- **Object vector**  
- **Actor vector**  
- **Time vector**

Higher value:
Highly related topic

Web

e.g., Business topic vectors

Object/URL

Actor/user

Money.com

CNN.com

Smith
Johnson

Mon-Fri  Sat-Sun

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/

© 2015 Sakurai, Matsubara & Faloutsos 48
Main idea (1) : M-way analysis (details)

- M-way decomposition (M=3)
  - [Gibbs sampling] infer $k$ hidden topics for each non-zero element of $X$, according to probability $p$

$$p(z_{i,j,t} = r|X, O', A', C', \alpha, \beta, \gamma)$$

\[\propto \frac{\phi_{i,r} + \alpha}{\sum_r \phi_{i,r} + \alpha k} \cdot \frac{a_{r,j} + \beta}{\sum_j a_{r,j} + \beta v} \cdot \frac{c_{r,t} + \gamma}{\sum_t c_{r,t} + \gamma n}\]
Main idea (2): Multi-scale analysis (details)

- Tensors with multiple window sizes

\[
\chi = \chi^{(0)}
\]

1. Infer O, A, C at highest level

Hourly pattern

Daily pattern

Weekly pattern

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/

© 2015 Sakurai, Matsubara & Faloutsos
Main idea (2) : Multi-scale analysis (details)

- Tensors with multiple window sizes

2. Share O & A for all levels

3. Compute C for each level
TriMine-Forecasts

Our final goal: “forecast future events”!

Q. How can we generate a realistic events?

e.g., estimate the number of clicks for user “smith”, to URL “CNN.com”, for next 10 days

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
Why not naïve?

- **Individual-sequence forecasting**
  - Create a set of \((u \times v)\) sequences of length \(n\)
  - apply the forecasting algorithm for each sequence
Why not naïve?

- **Individual-sequence forecasting**
  - Create a set of $(u \times v)$ sequences of length $(n)$
  - apply the forecasting algorithm for each sequence

- **Scalability**: time complexity is at least $O(uvn)$
- **Accuracy**: each sequence “looks” like noise, (e.g., {0, 0, 0, 1, 0, 0, 2, 0, 0, ...}) -> hard to forecast

TriMine-F

Our approach:

–Step 1: Forecast time-topic matrix:

–Step 2: Generate events using 3 matrices

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/

© 2015 Sakurai, Matsubara & Faloutsos
Forecast ‘time-topic matrix’ (details)

Q. How to capture multi-scale dynamics?
   e.g., bursty pattern, noise, multi-scale period

Multi-scale forecasting

Forecast $\hat{C}_{r,t}^{(0)}$ using multiple levels of matrices

\[
\begin{align*}
\text{Forecasted value} & \quad \hat{C}_{r,t}^{(0)} \\
\text{w=1} & \quad c_r^{(0)} \\
\text{w=2} & \quad c_r^{(1)} \\
\text{w=4} & \quad c_r^{(2)}
\end{align*}
\]

\[
c_{r,t}^{(0)} = \sum_{h=0}^{[\log n]} \sum_{i=1}^{w} \lambda_{i,r}^{(h)} c_{r,t-i}^{(h)} + \epsilon_t.
\]

(Details in paper)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
Our goals

☑ Q1: Hidden topics
☑ Q2: Groups
☑ Q3: Forecasting
Q1&2. WebClick data

URL-topic matrix (O)

Three hidden topics: “drive”, “business”, “media”

* Red point : each web site

Car & bike site is related to travel site

Money site & Finance site have similar trends

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/

© 2015 Sakurai, Matsubara & Faloutsos
Q1&2. WebClick data

User-topic matrix (A)

Three hidden topics: “drive”, “business”, “media”

* Red point: each user

Very clear user groups along the spokes
Q1&2. WebClick data

Time-topic matrix (C)

Three hidden topics: “drive”, “business”, “media”

* Each sequence: each topic over time

“Business” topic:
Less access during weekend

“Drive” topic:
Spikes during weekend

Q3. Forecasting accuracy

- Benefit of multiple time-scale forecasting

Original sequence of matrix (C)
Forecast C' using single level -> failed
Multi-scale forecast -> captured cyclic patterns

Q3. Forecasting accuracy

Temporal perplexity (entropy for each time-tick)

Lower perplexity: higher predictive accuracy

(a) WebClick

(b) Ondemand TV T2: [Hong et al. KDD’11]
Q3. Forecasting accuracy

Accuracy of event forecasting

RMSE between original and forecasted events
(lower is better)

PLiF [Li et al. VLDB’10], T2: [Hong et al. KDD’11]
Q3. Scalability

• Computation cost (vs. AR)

• **TriMine** provides a reduction in computation time (up to 74x)
Outline

- Tensor decomposition
- Mining and forecasting of complex time-stamped events
- New challenge: MANT analysis

Multi-Aspect Non-linear Time-series

Non-linear tensor analysis

New research directions
1. Automatic mining
2. Non-linear (gray-box) modeling
3. Large-scale tensor analysis

Put all together

New challenge: MANT analysis

Multi-Aspect Non-linear Time-series

FUNNEL: Automatic Mining of Spatially Coevolving Epidemics

Yasuko Matsubara, Yasushi Sakurai (Kumamoto University)
Willem G. van Panhuis (University of Pittsburgh)
Christos Faloutsos (CMU)

Motivation

Given: Large set of epidemiological data
e.g., Measles cases in the U.S.

(Weekly)
Motivation

**Given**: Large set of epidemiological data
e.g., Measles cases in the U.S.

- Yearly periodicity
- (Weekly)

[Graph showing yearly and weekly periodicity of Measles cases from 1930 to 1980]
Motivation

**Given**: Large set of epidemiological data
e.g., Measles cases in the U.S.

![Graph showing yearly periodicity and vaccine effect](http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/)

© 2015 Sakurai, Matsubara & Faloutsos
Motivation

**Given**: Large set of epidemiological data
e.g., Measles cases in the U.S.

- Yearly periodicity
- Shocks, e.g., 1941
- Vaccine effect

© 2015 Sakurai, Matsubara & Faloutsos
Motivation

**Given:** Large set of epidemiological data
e.g., Measles cases in the U.S.

**Goal:** summarize all the **epidemic time-series, “fully-automatically”**
Data description

Project Tycho: infectious diseases in the U.S.

50 states

1888

Time (weekly)

 (> 125 years)

56 diseases

© 2015 Sakurai, Matsubara & Faloutsos
Data description

Project Tycho: infectious diseases in the U.S.

50 states

56 diseases

1888

Time (weekly)

n

Time (weekly)

 (> 125 years)

Element \( x \) : # of cases

e.g., ‘measles’, ‘NY’, ‘April 1-7, 1931’, ‘4000’

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/

© 2015 Sakurai, Matsubara & Faloutsos
Problem definition

Given:

Tensor $\mathbf{X}$ (disease x state x time)
Problem definition

Given:
Tensor $\mathcal{X}$ (disease x state x time)

Find:
Compact description of $\mathcal{X}$, “automatically”

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/

© 2015 Sakurai, Matsubara & Faloutsos
Problem definition

Given:
Tensor $X$ ($\text{disease} \times \text{state} \times \text{time}$)

Find:
Compact description of $X$, "automatically"

Compact description of $X = P_1 \ P_2 \ P_3 \ P_4 \ P_5$

Seasonality

Discontinuities
Problem definition

Given:
Tensor $X$ (disease x state x time)

Find:
Compact description of $X$, "automatically"

$X = P_1 P_2 P_3 P_4 P_5$

NO magic numbers!

Parameter-free!
Modeling power of FUNNEL

Questions about epidemics

Q1  Q2  Q3  Q4  Q5

$\chi$
Questions

Q1

Are there any periodicities? If yes, when is the peak season?
Answers

P1  Seasonality

Influenza in Feb. Detected by FUNNEL (strong seasonality)
Seasonality

Measles (children’s) in spring

Detected!
Detected!

Lyme-disease (tick-borne) in summer
Answers

P1 Seasonality

Gonorrhea (STD) no periodicity

Detected!

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/

© 2015 Sakurai, Matsubara & Faloutsos
Can we see any discontinuities?
Disease reduction effect

Measles

1963: Vaccine licensure

1965: Detected by FUNNEL

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/

© 2015 Sakurai, Matsubara & Faloutsos
Q3

What’s the difference between measles in NY and in FL?
P3  Area sensitivity

FUNNEL’s guess of susceptibles (measles)

- CA: Detected!
- TX: NY, PA (more children)
- FL: (fewer children)
Q4

Are there any external shock events, like wars?
Answers

External shock events

Funnel can detect external shocks “fully-automatically”!

Scarlet fever Detected by FUNNEL

World war II Detected!

Questions

Q5

How can we remove mistakes and incorrect values?

© 2015 Sakurai, Matsubara & Faloutsos
It can also detect typos, "automatically"!!

Typhoid fever cases

Detected!
Modeling power of FUNNEL

Our model can capture 5 properties

- **P1** Seasonality
- **P2** Disease reductions
- **P3** Area sensitivity
- **P4** External events
- **P5** Mistakes
Problem definition

Given:
Tensor $\mathbf{X} (\text{disease x state x time})$

Find:
Compact description of $\mathbf{X}$, “automatically”
Main ideas

1. Automatic mining (no magic numbers!)
2. Non-linear (gray-box) modeling
3. Tensor analysis

New challenge: MANT analysis

Multi-Aspect Non-linear Time-series

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/

© 2015 Sakurai, Matsubara & Faloutsos
Problem definition

Given:
Tensor $\mathbf{X}$ (disease x state x time)

Find:
Compact description of $\mathbf{X}$, “automatically”
Problem definition

**Automatic mining**

**Gray-box (non-linear)**

**Tensor analysis**

**Compact description of**

\[ \chi = P_1 P_2 P_3 P_4 P_5 \]

© 2015 Sakurai, Matsubara & Faloutsos
Two main ideas

Idea #1: Grey-box model

Idea #2: MDL for fitting

NO magic numbers!

(parameter-free)
Two main ideas

Idea #1: Grey-box model - domain knowledge

(SIRS+) : 6 parameters

\[
\begin{align*}
S(t+1) &= S(t) - \beta(t)\epsilon(t)S(t)I(t) \\
I(t+1) &= I(t) + \beta(t)\epsilon(t)S(t)I(t) \\
V(t+1) &= V(t) + \delta I(t) - \gamma V(t) 
\end{align*}
\]
Two main ideas

Idea #2: Fitting with MDL -> parameter free!

Cost function

\[ Cost_T(\mathcal{X}; \mathcal{F}) = \log^*(d) + \log^*(l) + \log^*(n) + Cost_M(\mathcal{B}) + Cost_M(\mathcal{R}) + Cost_M(\mathcal{N}) + Cost_M(\mathcal{E}) + Cost_M(\mathcal{M}) + Cost_C(\mathcal{X}|\mathcal{F}) \]
Proposed model: **FUNNEL**

(a) **FUNNEL-single**

(b) **FUNNEL-full (tensor)**
Proposed model: FUNNEL

(a) FUNNEL-single

(b) FUNNEL-full (tensor)
FUNNEL – with a single epidemic

Given:
“single” epidemic sequence

e.g., measles in NY

Find:
nonlinear equation, model parameters

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/

© 2015 Sakurai, Matsubara & Faloutsos
FUNNEL – with a single epidemic

With a single epidemic: Funnel-RE

People of 3 classes
- **S**: Susceptible
- **I**: Infected
- **V**: Vigilant/vaccinated

© 2015 Sakurai, Matsubara & Faloutsos
FUNNEL – with a single epidemic

With a single epidemic: Funnel-RE

\[
\begin{align*}
S(t+1) &= S(t) - \beta(t)\epsilon(t)S(t)I(t) + \gamma V(t) - \theta(t)S(t) \\
I(t+1) &= I(t) + \beta(t)\epsilon(t)S(t)I(t) - \delta I(t) \\
V(t+1) &= V(t) + \delta I(t) - \gamma V(t) + \theta(t)S(t)
\end{align*}
\]

\( S(t) \): susceptible
\( I(t) \): Infected
\( V(t) \): Vigilant / Vaccinated
FUNNEL – with a single epidemic

With a single epidemic: Funnel-RE

\[
\begin{align*}
S(t+1) &= S(t) - \beta(t) \epsilon(t) S(t) I(t) + \gamma V(t) - \theta(t) S(t) \\
I(t+1) &= I(t) + \beta(t) \epsilon(t) S(t) I(t) - \delta I(t) \\
V(t+1) &= V(t) + \delta I(t) - \gamma V(t) + \theta(t) S(t)
\end{align*}
\]

(3)

\(\beta(t)\): strength of infection (yearly periodic func)

\[
\beta(t) = \beta_0 \cdot \left(1 + P_a \cdot \cos\left(\frac{2\pi}{P_p}(t + P_s)\right)\right)
\]

\(P_p = 52\)

\(\epsilon(t)\)

\(\theta(t)\)

\(\delta\)
With a single epidemic: Funnel-RE

\[ S(t+1) = S(t) - \beta(t)\epsilon(t)S(t)I(t) + \gamma V(t) - \theta(t)S(t) \]
\[ I(t+1) = I(t) + \beta(t)\epsilon(t)S(t)I(t) - \delta I(t) \]
\[ V(t+1) = V(t) + \delta I(t) - \gamma V(t) + \theta(t)S(t) \]  

**Details:**
- \( \delta \): healing rate
- \( \theta(t) \): disease reduction effect

\[ \theta(t) = \begin{cases} 
0 & (t < t_\theta) \\
\theta_0 & (t \geq t_\theta) 
\end{cases} \]
FUNNEL – with a single epidemic

With a single epidemic: Funnel-RE

\[
\begin{align*}
S(t+1) &= S(t) - \beta(t)\varepsilon(t)S(t)I(t) + \gamma V(t) - \theta(t)S(t) \\
I(t+1) &= I(t) + \beta(t)\varepsilon(t)S(t)I(t) - \delta I(t) \\
V(t+1) &= V(t) + \delta I(t) - \gamma V(t) + \theta(t)S(t)
\end{align*}
\]

\(\varepsilon(t)\) : temporal susceptible rate

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/

© 2015 Sakurai, Matsubara & Faloutsos
Proposed model: FUNNEL

(a) FUNNEL-single

(b) FUNNEL-full (tensor)
FUNNEL-full

\[ \chi = \begin{array}{c}
\text{diseases} \\
\text{states} \\
\text{time}
\end{array}
\]

\[ \begin{array}{c}
N, \beta_0, \delta, \gamma, P_a, P_s \\
\theta_0, t_0
\end{array}
\]

\[ \begin{array}{c}
E: \text{shocks} \\
M: \text{mistakes}
\end{array}
\]
FUNNEL-full

Details

Global

Base matrix $B$ (d x 6)

Disease reduction matrix $R$ (d x 2)

States $l$

Diseases $d$

Time $n$

$X = (P1, P2)$

$B, R$

$N, \beta_0, \delta, \gamma, P_a, P_s$

$\theta_0, t_\theta$

$\text{(global/country)}$

Local

Geo-disease matrix $\mathbf{N}$ (d x l)

$\mathbf{N} = \{N_{ij}\}_{i,j=1}^{d,l}$ : potential population of disease $i$ in state $j$
FUNNEL-full

Details

Extra

External shock tensor $\mathcal{E}$

Mistake tensor $\mathcal{M}$

$\mathcal{E}$

$\mathcal{M}$

extra - $\mathcal{E}$: shocks & $\mathcal{M}$: mistakes

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/

© 2015 Sakurai, Matsubara & Faloutsos
Details

$\mathcal{E} = \{ E^{(D)}, E^{(T)}, E^{(S)} \}$

Disease matrix

Time matrix

State matrix
Challenges

Q1. How to automatically
   - find “external shocks”?  
   - ignore “mistakes” (i.e., typos)?

Q2. How to efficiently estimate model parameters?
Challenges

Q1. How to automatically
   - find “external shocks”?
   - ignore “mistakes” (i.e., typos)?

   Idea (1): Model description cost

Q2. How to efficiently estimate model parameters?

   Idea (2): Multi-layer optimization - $O(dln)$
FUNNEL at work - forecasting
Forecasting future epidemics

Train: 2/3 sequences
Forecast: 1/3 following years

(a) Influenza

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/

© 2015 Sakurai, Matsubara & Faloutsos
FUNNEL at work - forecasting
Forecasting future epidemics

Train: 2/3 sequences
Forecast: 1/3 following years

Funnel can capture future epidemics (AR: fail)

(a) Influenza

Count (log)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/
© 2015 Sakurai, Matsubara & Faloutsos
FUNNEL at work - forecasting
Forecasting future epidemics

Train: \( \frac{2}{3} \) sequences
Forecast: \( \frac{1}{3} \) following years

Funnel can capture future epidemics (AR: fail)
Generality of FUNNEL

Epidemics on computer networks

Funnel is general: it fits computer virus very well!
Part 3 Conclusions

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

Multi-Aspect Non-linear Time-series
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
Part 3

Extension of time-series: tensor analysis

Yasushi Sakurai (Kumamoto University)
Yasuko Matsubara (Kumamoto University)
Christos Faloutsos (Carnegie Mellon University)