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 Univ.)
Motivation

Complex time-stamped events

consists of \{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>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

e.g., CNN.com, CNET.com -> news topic

YouTube.com -> media topic
Motivation

Q2. Can we group URLs/users accordingly?

<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 (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

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

URL: money site
URL: blog site

Noisy 😞  Sparse 😞  Bursty 😞

We cannot see any trends !!
Outline

- Motivation
- Problem definition
- Proposed method: TriMine
- TriMine-F forecasting
- Experiments
- Conclusions
Problem definition

Given: a set of complex time-stamped events

Original web-click events
Problem definition

Given: a set of complex time-stamped events

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

Original web-click events

![Graphs showing click patterns over time]
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)

Time evolution
Outline

- Motivation
- Background
- Proposed method: TriMine
- TriMine-F forecasting
- Experiments
- Conclusions
Main idea (1): M-way analysis

Complex 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>
Main idea (1) : M-way analysis

Complex 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 ($M=3$)

$$\mathcal{X} \in \mathbb{N}^{u \times v \times n}$$
Main idea (1): M-way analysis

Complex 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
Main idea (1) : M-way analysis

Undesirable properties

- High dimension 😞
- Categorical data 😞
- Sparse tensor 😞
- Look like noise 😞

e.g., \( x = \{0, 1, 0, 2, 0, 0, 0, \ldots\} \)
Main idea (1): \( M \)-way analysis

Undesirable properties

- High dimension 
- Categorical data 
- Sparse tensor 
- Look like noise

\[ \text{e.g., } x = \{0, 1, 0, 2, 0, 0, 0, \ldots \} \]

Questions:

How to find meaningful patterns?
**Main idea (1): M-way analysis**

A. decompose to a set of 3 topic vectors:

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

- **Object/URL**
- **Actor/user**
- **Time**

Web clicks $\chi$

Topic A (business)
Topic B (news)
Topic C (media)
Main idea (1) : M-way analysis

A. decompose to a set of 3 topic vectors:
   - **Object vector**
   - **Actor vector**
   - **Time vector**

 e.g., business topic vectors

Higher value: Highly related topic
Main idea (1) : M-way analysis

A set of 3 topic vectors = 3 topic matrices

- \([O]\) Object-topic matrix \((u \times k)\)
- \([A]\) Actor-topic matrix \((k \times v)\)
- \([C]\) Time-topic matrix \((k \times n)\)
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 $p$

\[
p(z_{i,j,t} = r | \mathcal{X}, O', A', C', \alpha, \beta, \gamma) \propto \frac{o_{i,r} + \alpha}{\sum_r o_{i,r} + \alpha k} \cdot \frac{a_{r,j} + \beta}{\sum_j a_{r,j} + \beta v} \cdot \frac{c_{r,t} + \gamma}{\sum_t c_{r,t} + \gamma n}
\]
Main idea (2) : Multi-scale analysis

Q: What is the right window size to capture meaningful patterns?
  ... minute? hourly?
  ... daily?
Main idea (2): Multi-scale analysis

Q: What is the right window size to capture meaningful patterns?

A. Our solution: **Multiple window sizes**
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

KDD 2012

Y. Matsubara et al.

24
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
Main idea (2): Multi-scale analysis (details)

Tensors with multiple window sizes

2. Share O & A

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

TriMine is linear on the input size \( N \), i.e.,

\[ O(N \log n) \rightarrow O(N) \]

\( N \): counts of events in \( X \), \( n \): duration of \( X \)

KDD 2012

Y. Matsubara et al.
Outline

- Motivation
- Background
- Proposed method: TriMine
  - TriMine-F forecasting
- Experiments
- Conclusions
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
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., \(\{0, 0, 0, 1, 0, 0, 2, 0, 0, \ldots\}\)) \(
\rightarrow \) hard to forecast
Our approach:
- [Step 1] Forecast time-topic matrix: \( \hat{C} \)
- [Step 2] Generate events using 3 matrices
Q. How to capture multi-scale dynamics?
  e.g., bursty pattern, noise, multi-scale period

A. Multi-scale forecasting

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

\[ c_{r,t}^{(0)} = \sum_{h=0}^{[\log n]} \sum_{i=1}^{w} \lambda_{i,r}^{(h)} c_{r,t-i}^{(h)} + \epsilon_t. \] (Details in paper)
[Step 2] Generate events using O A Ĉ (details)

We propose 2 solutions:

A1. Count estimation
Use O A Ĉ matrices

\[ \hat{x}_{i,j,t} = n \bar{x}_i \sum_{r=1}^{k} o_{i,r} \cdot a_{r,j} \cdot \hat{c}_{r,t}, \]

A2. Complex event generation
Use sampling-based approach (Details in paper)
Outline

- Motivation
- Background
- Proposed method: TriMine
- TriMine-F forecasting
- Experiments
- Conclusions
Experimental evaluation

The experiments were designed to answer:

- **Effectiveness**
  Q1. How successful is TriMine in spotting patterns?
- **Forecasting accuracy**
  Q2. How well does TriMine forecast events?
- **Scalability**
  Q3. How does TriMine scale with the dataset size?
Experimental evaluation

Datasets

- **WebClick data**
  
  *Click: \{URL, user ID, time\}*
  
  - 1,797 URLs, 10,000 heavy users, one month

- **Ondemand TV data**
  
  *View: \{channel ID, viewer ID, time\}*
  
  - 13,231 TV program, 100,000 users, 6 month
Q1. Effectiveness

Result of three matrices O, A, C

Visualization: “TriMine-plots”

- **URL-topic matrix O**
- **User-topic matrix A**
- **Time-topic matrix C**

Tensor X → O → A → C
Q1-1. 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
**Q1-1. WebClick data**

**User-topic matrix (A)**

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

* Red point : each user

Very clear user groups along the spokes
Q1-1. 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
Q1-1. WebClick data

Other topics

Three topics: “Communication”, “food”, “blog”

Three related sites: route-map, diet, restaurant
i.e., users check out
1. Restaurants
2. route map in their area
3. Calories of their meals
Other topics

Three topics: “Communication”, “food”, “blog”

- **4pm**: Food related sites: visited in the early evening before users go out
- **11pm**: Communication sites: Used in the late evening for private purposes

Time-topic matrix \( C \)
Q1-2. Ondemand TV data

TV program-topic matrix (O)

Three topics: “sports”, “action”, “romance”

* Red point: each TV program

Several clusters (LOST, tennis etc.)
Q1-2. Ondemand TV data

Time-topic matrix (C)

Three hidden topics: “sports”, “action”, “romance”

* Each sequence: each topic over time

Daily & weekly periodicities

“Action”: High peaks on weekends
Q2-1. 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]
Q2-2. 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]
Q2-3. 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. Scalability

Computation cost (vs. AR)

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

- Motivation
- Problem definition
- Proposed method: TriMine
- TriMine-F forecasting
- Experiments
- Conclusions
Conclusions

- TriMine has following properties:
  • **Effective**
    – It finds meaningful patterns in real datasets
  • **Accurate**
    – It enables forecasting
  • **Scalable**
    – It is linear on the database size
Thank you

**Code:** http://www.kecl.ntt.co.jp/csl/sirg/people/yasuko/software.html

**Email:** matsubara.yasuko @ lab.ntt.co.jp