Web Performance Optimization: Analytics

Wim Leers

Promotor: Prof. dr. Jan Van den Bussche
Web Performance Optimization

• Speed matters!

Web Performance Optimization

• Speed matters!

• 0.1 s \rightarrow \text{direct manipulation}

Web Performance Optimization

• Speed matters!

• 0.1 s → direct manipulation

• 1 s → good navigation

Web Performance Optimization

- Speed matters!

- 0.1 s → direct manipulation
- 1 s → good navigation
- 10 s → attention kept

Web Performance Optimization

• Speed matters!

• 0.1 s  →  direct manipulation

• 1 s   →  good navigation

• 10 s  →  attention kept

• >10 s →  bye bye!

How to Measure? Episodes
How to Measure? Episdoes

- Measures “episodes” during page loading
How to Measure? **Episodes**

- Measures “episodes” during page loading

- **Real measurements**: JS in browser, for *each* visitor
How to Measure? **Episodes**

- Measures “episodes” during page loading

- **Real measurements**: JS in browser, for *each* visitor

- Result: Episodes log file
Analytics
Analytics

- Automatically pinpoint causes of slow page loads
Analytics

• Automatically pinpoint causes of slow page loads

• e.g.:
Analytics

• Automatically pinpoint causes of slow page loads

• e.g.:

  • “http://uhasselt.be/ is slow in Belgium, for users of the ISP Telenet”
Analytics

• Automatically pinpoint causes of slow page loads

• e.g.:
  • “http://uhasselt.be/ is slow in Belgium, for users of the ISP Telenet”
  • “http://uhasselt.be/studenten/dossier has slowly loading CSS”
Analytics

- Automatically pinpoint causes of slow page loads

- e.g.:
  - “http://uhasselt.be/ is slow in Belgium, for users of the ISP Telenet”
  - “http://uhasselt.be/studenten/dossier has slowly loading CSS”
  - “http://uhasselt.be/bib has slowly loading JS in Firefox 3”
Analytics

- Automatically pinpoint causes of slow page loads

- e.g.:
  - “http://uhasselt.be/ is slow in Belgium, for users of the ISP Telenet”
  - “http://uhasselt.be/studenten/dossier has slowly loading CSS”
  - “http://uhasselt.be/bib has slowly loading JS in Firefox 3”
  - ...
Literature Study Subjects
Literature Study Subjects

- Data Stream Mining
Literature Study Subjects

- Data Stream Mining
- Anomaly Detection
Literature Study Subjects

- Data Stream Mining
- Anomaly Detection

Data Mining: finding patterns in data
Literature Study Subjects

- Data Stream Mining
- Anomaly Detection
- OLAP: Data Cube

**Data Mining:** finding patterns in data
Literature Study Subjects

- Data Stream Mining
- Anomaly Detection
- OLAP: Data Cube

Data Mining: finding patterns in data
OLAP: querying multidimensional data
Data Stream Mining
Data Stream Mining

- Constraints
Data Stream Mining

- Constraints

- Possibly infinite data stream $\Rightarrow$ approximation
Data Stream Mining

- Constraints
  - Possibly infinite data stream $\Rightarrow$ approximation
- Window model
Data Stream Mining

• Constraints

  • Possibly infinite data stream ⇒ approximation

• Window model

  - Landmark: from beginning until now
Data Stream Mining

- Constraints
  - Possibly infinite data stream $\Rightarrow$ approximation

- Window model
  - Landmark: from beginning until now
  - Tilted-time: per-hour window, 24 “hour windows” $\Rightarrow$ “day window”, etc.
Data Stream Mining

• Constraints

• Possibly infinite data stream ⇒ approximation

• Window model

  - Landmark: from beginning until now

  - Tilted-time: per-hour window, 24 “hour windows” → “day window”, etc.

• Algorithms studied
Data Stream Mining

• Constraints
  • Possibly infinite data stream ⇒ approximation

• Window model
  - Landmark: from beginning until now
  - Tilted-time: per-hour window, 24 “hour windows” → “day window”, etc.

• Algorithms studied
  • Frequent Item Mining: 7
Data Stream Mining

- Constraints
  - Possibly infinite data stream ⇒ approximation

- Window model
  - Landmark: from beginning until now
  - Tilted-time: per-hour window, 24 “hour windows” → “day window”, etc.

- Algorithms studied
  - Frequent Item Mining: 7
  - Frequent Pattern Mining: 2
Data Stream Mining: **FP-Stream**

Source: Mining Frequent Patterns in Data Streams at Multiple Time Granularities, Giannella; Han et al., 2003
Data Stream Mining: **FP-Stream**

Source: Mining Frequent Patterns in Data Streams at Multiple Time Granularities, Giannella; Han et al., 2003
Data Stream Mining: **FP-Stream**

<table>
<thead>
<tr>
<th>frequent pattern</th>
<th>support</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>100</td>
</tr>
<tr>
<td>b</td>
<td>92</td>
</tr>
<tr>
<td>c</td>
<td>80</td>
</tr>
<tr>
<td>ab</td>
<td>78</td>
</tr>
<tr>
<td>ac</td>
<td>75</td>
</tr>
<tr>
<td>bc</td>
<td>70</td>
</tr>
<tr>
<td>abc</td>
<td>63</td>
</tr>
</tbody>
</table>

Frequent Patterns

Pattern Tree

Source: Mining Frequent Patterns in Data Streams at Multiple Time Granularities, Giannella; Han et al., 2003
Data Stream Mining: **FP-Stream**

Source: Mining Frequent Patterns in Data Streams at Multiple Time Granularities, Giannella; Han et al., 2003
Anomaly Detection
Anomaly Detection

- Types
Anomaly Detection

• Types

• **Point**: e.g. rainfall in mm
Anomaly Detection

• **Types**

  • **Point**: e.g. rainfall in mm

  • **Contextual**: point + contextual attributes, e.g. rainfall in mm + lat/lon
Anomaly Detection

• Types

  • **Point**: e.g. rainfall in mm

  • **Contextual**: point + contextual attributes, e.g. rainfall in mm + lat/lon

• **Contextual anomaly detection** algorithms categories
Anomaly Detection

• Types
  • **Point**: e.g. rainfall in mm
  • **Contextual**: point + contextual attributes, e.g. rainfall in mm + lat/lon

• **Contextual anomaly detection** algorithms categories
  • Reduction: 1) certain context, 2) point anomaly algorithm
Anomaly Detection

• Types

• **Point**: e.g. rainfall in mm

• **Contextual**: point + contextual attributes, e.g. rainfall in mm + lat/lon

• **Contextual anomaly detection** algorithms categories

  • Reduction: 1) certain context, 2) point anomaly algorithm

  • Model: 1) learn through training, 2) compare: *observed* vs. *expected*
Anomaly Detection

• Types

  • **Point**: e.g. rainfall in mm

  • **Contextual**: point + contextual attributes, e.g. rainfall in mm + lat/lon

• **Contextual anomaly detection** algorithms categories

  • Reduction: 1) certain context, 2) point anomaly algorithm

  • Model: 1) learn through training, 2) compare: *observed* vs. *expected*

• Algorithms studied: 2
Anomaly Detection: Vilalta/Ma
Anomaly Detection: Vilalta/Ma

- Based on frequent pattern mining
Anomaly Detection: Vilalta/Ma

- Based on frequent pattern mining
- Find all frequent itemsets that precede anomalies
OLAP: Data Cube

Source: Introduction to Data Mining, Tan; Steinbach; Kumar, 2006
OLAP: Data Cube

Source: Introduction to Data Mining, Tan; Steinbach; Kumar, 2006
OLAP: Data Cube: **Range-Sum Performance**
OLAP: Data Cube: **Range-Sum Performance**

- Very common type of query
OLAP: Data Cube: **Range-Sum Performance**

- Very common type of query
- Algorithms studied: 3
OLAP: Data Cube: **Dynamic Data Cube**

Source: Data Cubes in Dynamic Environments, Geffner; Riedewald; Agrawal, 1999
OLAP: Data Cube: **Dynamic Data Cube**

---

### Level 2 (root), k=4

<table>
<thead>
<tr>
<th>11</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>33</td>
</tr>
<tr>
<td>40</td>
<td>48</td>
</tr>
<tr>
<td>15</td>
<td>29</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>24</td>
<td>31</td>
</tr>
<tr>
<td>42</td>
<td>47</td>
</tr>
<tr>
<td>12</td>
<td>26</td>
</tr>
</tbody>
</table>

### Level 1, k=2

<table>
<thead>
<tr>
<th>8</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
</tr>
</tbody>
</table>

### Level 0 (leaves), k=1

<table>
<thead>
<tr>
<th>3</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>8</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

---

Source: Data Cubes in Dynamic Environments, Geffner; Riedewald; Agrawal, 1999
Outlook
Outlook

- Further literature study, especially: data cubes over data streams
Outlook

• Further literature study, especially: data cubes over data streams

• Implementation
Outlook

- Further literature study, especially: data cubes over data streams
- Implementation

<table>
<thead>
<tr>
<th>Month</th>
<th>Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>September 2010</td>
<td>further literature study + episodes log mining</td>
</tr>
<tr>
<td>October 2010</td>
<td>data stream mining</td>
</tr>
<tr>
<td>November 2010</td>
<td>OLAP + initial UI</td>
</tr>
<tr>
<td>December 2010</td>
<td>finish UI + anomaly detection</td>
</tr>
</tbody>
</table>
Questions?

Thanks for your time!