Web Performance Optimization: Analytics

Wim Leers

Promotor: Prof. dr. Jan Van den Bussche

Web Performance Optimization

- Speed matters!

Source: http://www.useit.com/alertbox/response-times.html, Jakob Nielsen, June 21, 2010

Web Performance Optimization

- Speed matters!
  - 0.1 s → direct manipulation

Source: http://www.useit.com/alertbox/response-times.html, Jakob Nielsen, June 21, 2010

Web Performance Optimization

- Speed matters!
  - 0.1 s → direct manipulation
  - 1 s → good navigation

Source: http://www.useit.com/alertbox/response-times.html, Jakob Nielsen, June 21, 2010

Web Performance Optimization

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

Source: http://www.useit.com/alertbox/response-times.html, Jakob Nielsen, June 21, 2010

Web Performance Optimization

- Speed matters!
  - 0.1 s → direct manipulation
  - 1 s → good navigation
  - 10 s → attention kept
  - >10 s → *bye bye!*

Source: http://www.useit.com/alertbox/response-times.html, Jakob Nielsen, June 21, 2010

How to Measure? **Episodes**

How to Measure? **Episodes**

- 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

} **Data Mining:** finding patterns in data

- OLAP: Data Cube

Literature Study Subjects

- Data Stream Mining
- Anomaly Detection

} **Data Mining:** finding patterns in data

- OLAP: Data Cube

} **OLAP:** querying multidimensional data

Data Stream Mining

Data Stream Mining

- Constraints

Data Stream Mining

- Constraints
  - Possibly infinite data stream ⇒ approximation

Data Stream Mining

- Constraints
  - Possibly infinite data stream ⇒ 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 ⇒ approximation
  - Window model
    - Landmark: from beginning until now
    - Tilted-time: per-hour window, 24 “hour windows” → “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

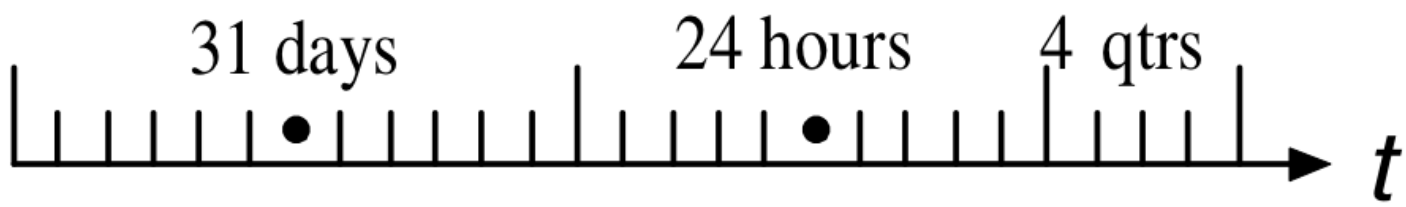
Data Stream Mining

- 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

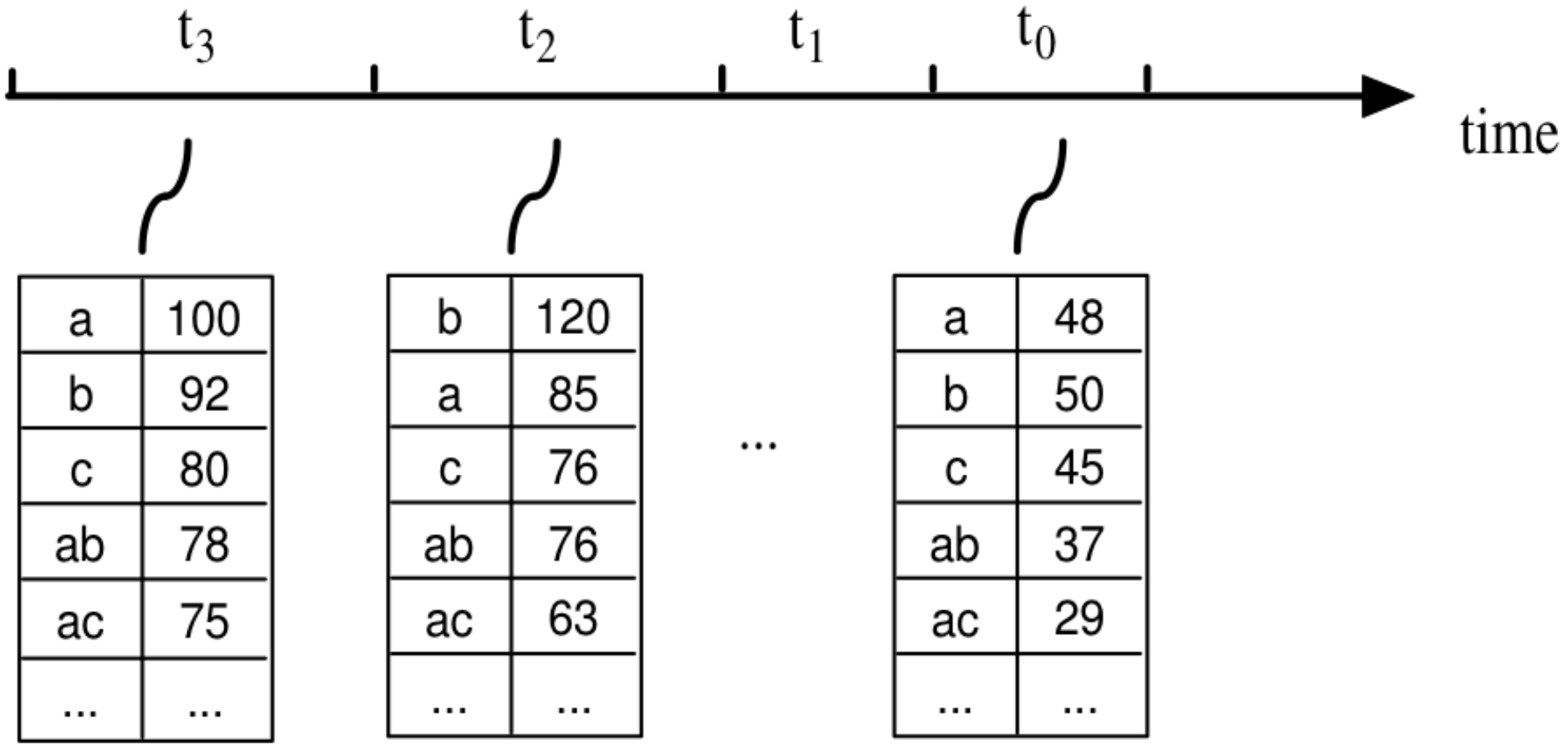
- Constraints
  - Possibly infinite data stream ⇒ approximation
  - Window model
    - Landmark: from beginning until now
    - Tilted-time: per-hour window, 24 “hour windows” → “day window”, etc.
- Algorithms studied
  - Frequent Item Mining: 7
  - Frequent Pattern Mining: 2

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**

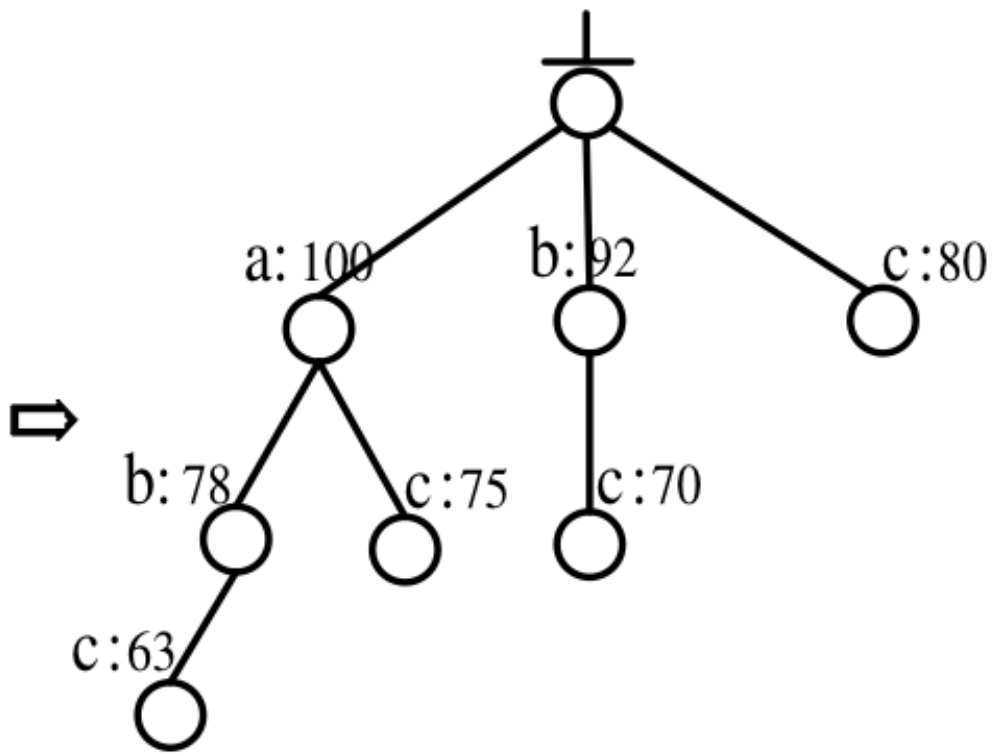


Source: Mining Frequent Patterns in Data Streams at Multiple Time Granularities, Giannella; Han et al., 2003

Data Stream Mining: **FP-Stream**

frequent pattern	support
a	100
b	92
c	80
ab	78
ac	75
bc	70
abc	63

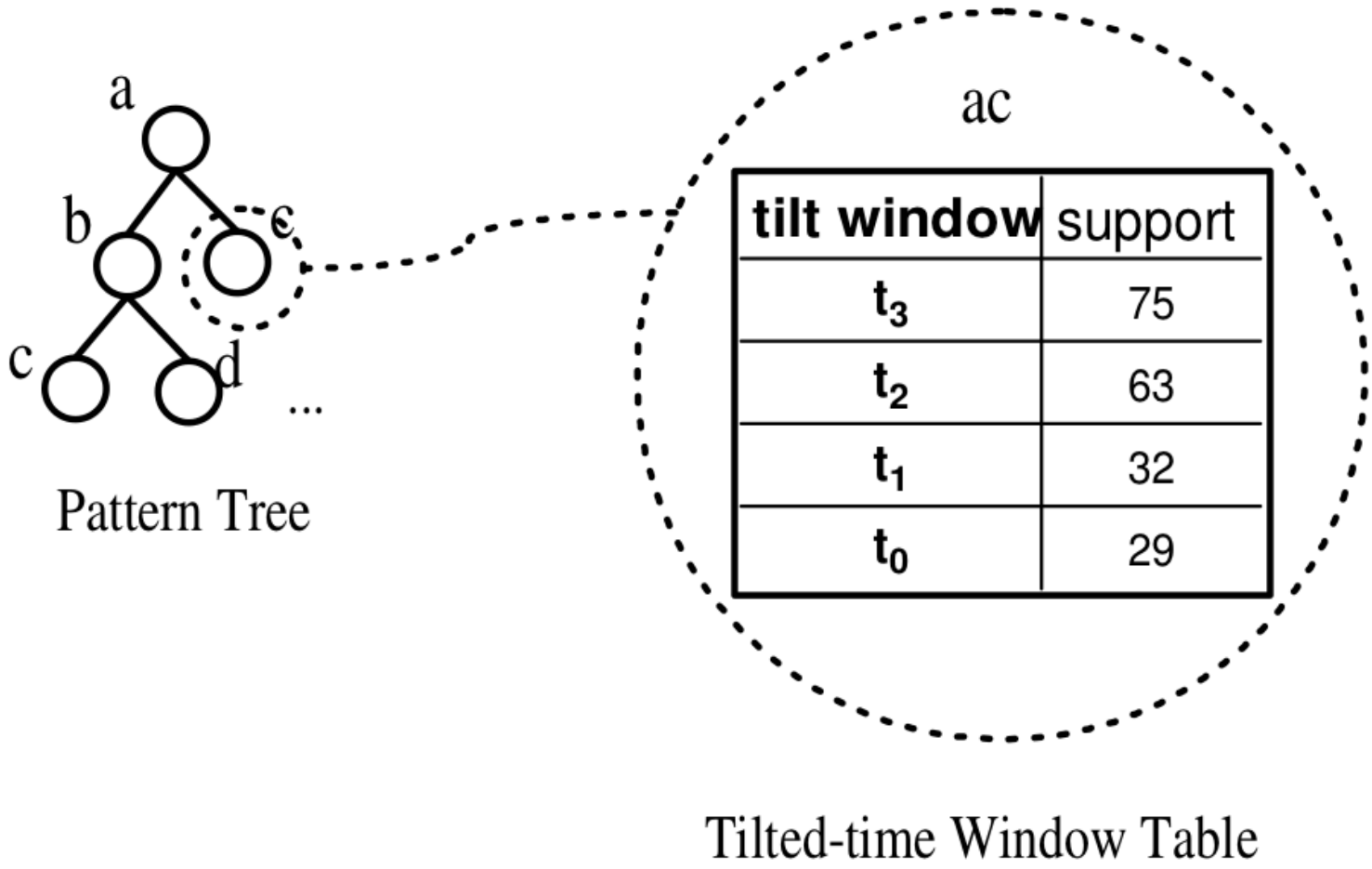
Frequent Patterns



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

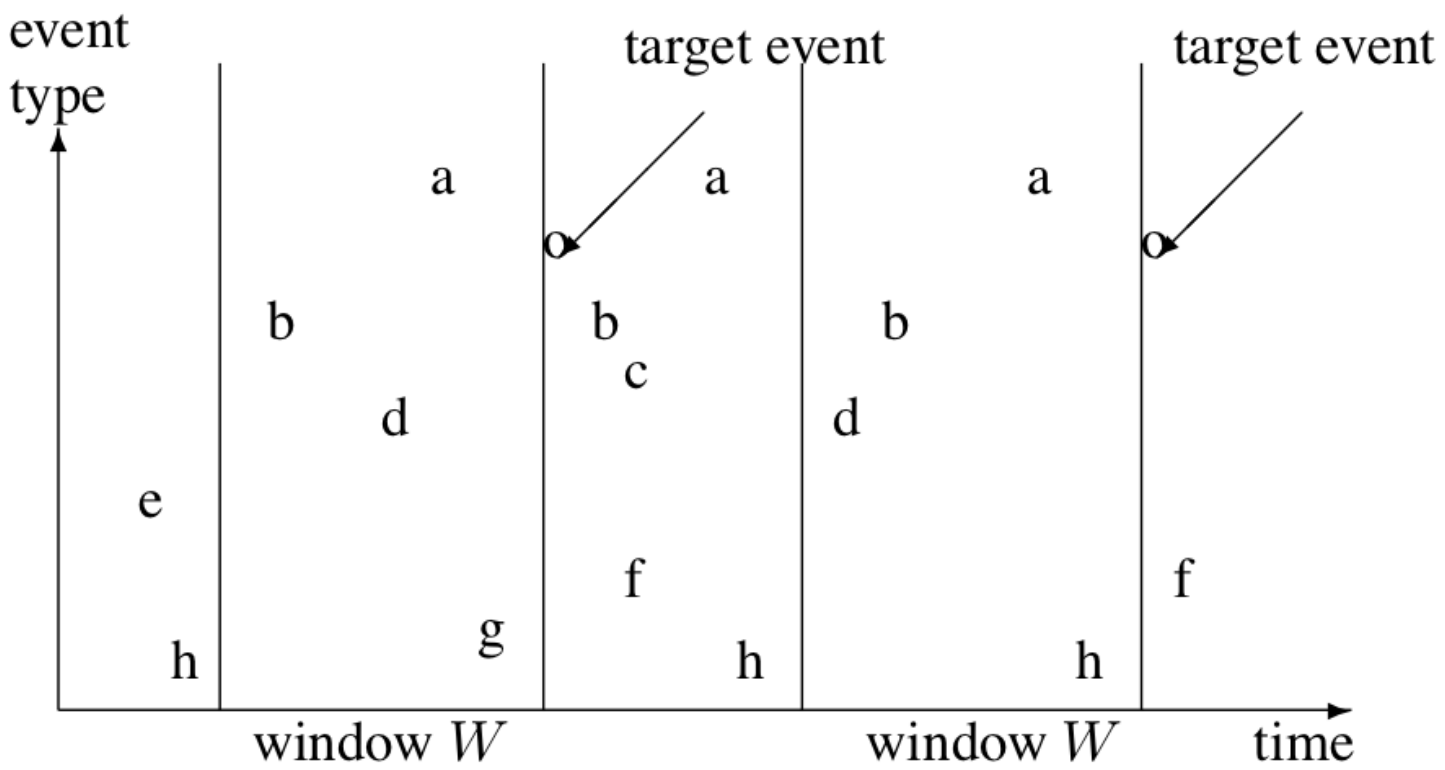
Anomaly Detection: **Vilalta/Ma**

Anomaly Detection: **Vilalta/Ma**

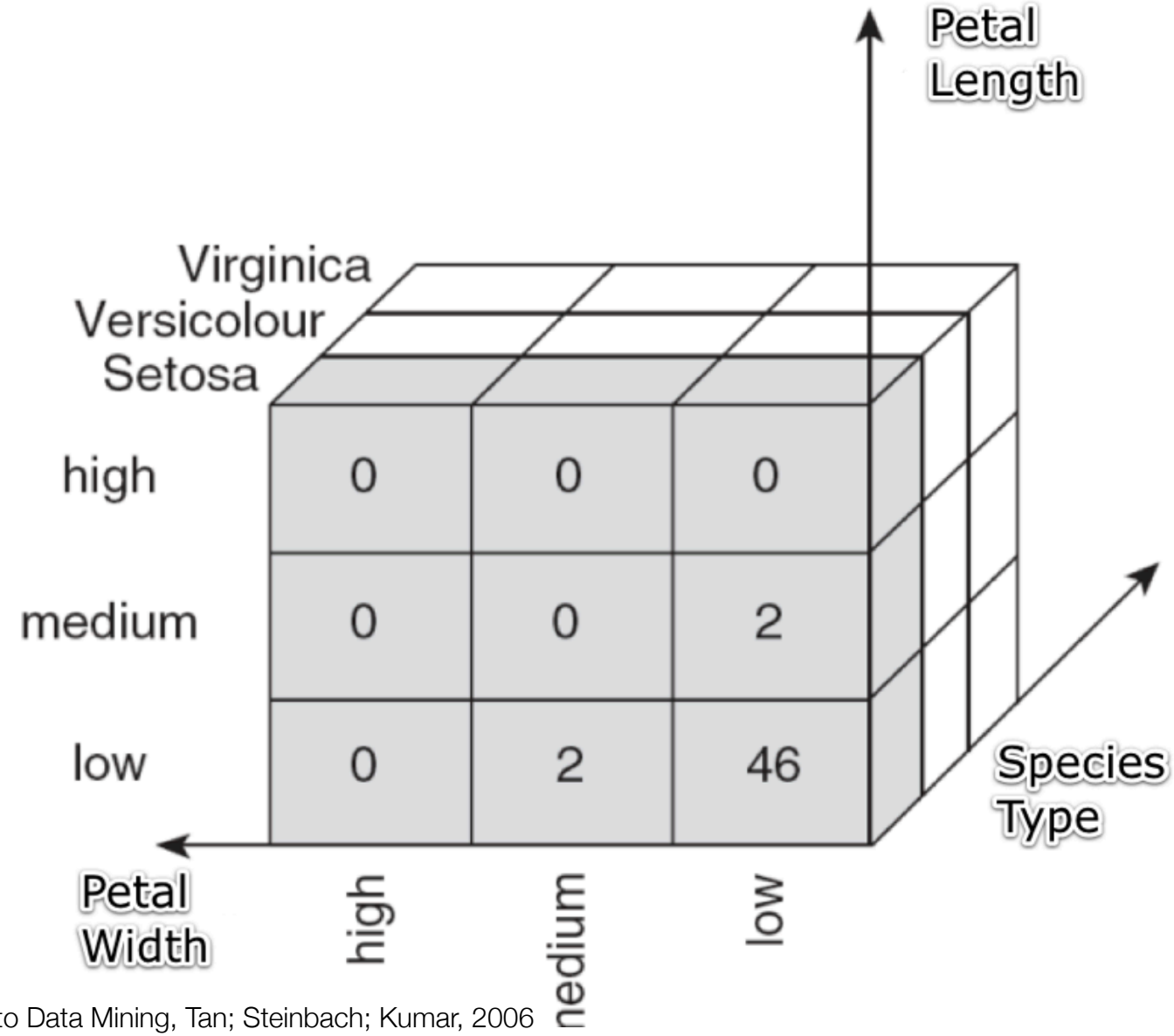
- Based on frequent pattern mining

Anomaly Detection: **Vilalta/Ma**

- Based on frequent pattern mining
- Find all frequent itemsets that precede anomalies

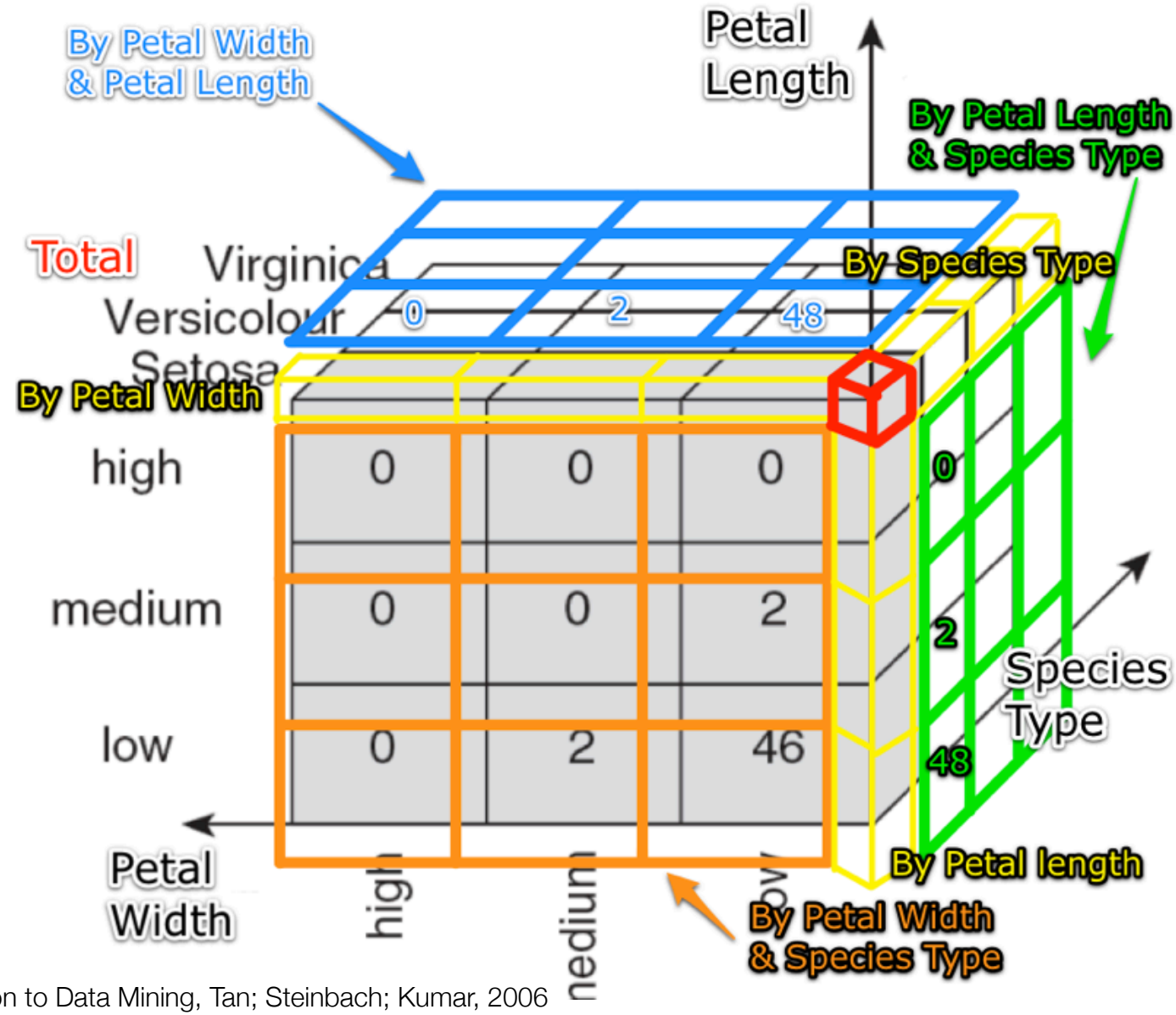


OLAP: Data Cube



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**

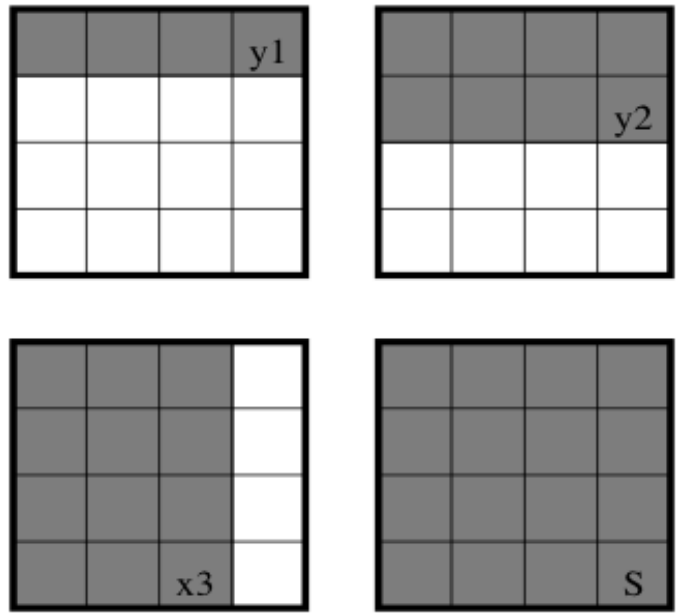
- Very common type of query

OLAP: Data Cube: **Range-Sum Performance**

- Very common type of query
- Algorithms studied: 3

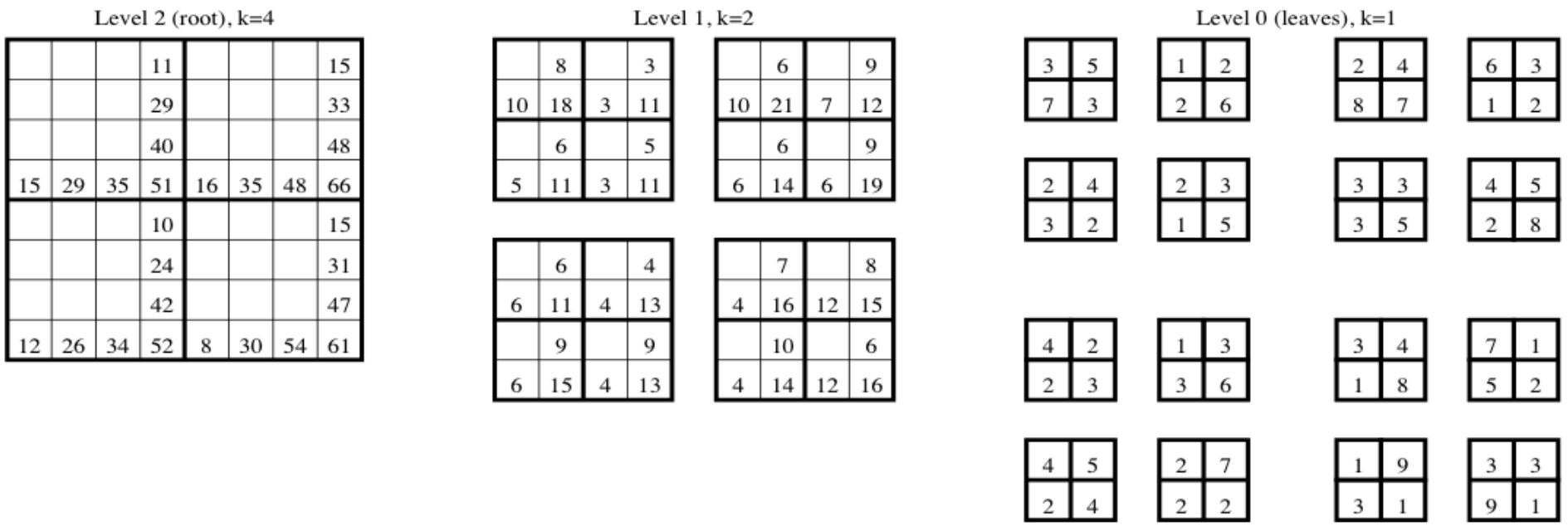
OLAP: Data Cube: **Dynamic Data Cube**

	0	1	2	3	4	5	6	7
0				y1				y1
1				y2				y2
2				y3				y3
3	x1	x2	x3	S	x1	x2	x3	S
4				y1				y1
5				y2				y2
6				y3				y3
7	x1	x2	x3	S	x1	x2	x3	S



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

OLAP: Data Cube: **Dynamic Data Cube**



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

Month	Goals
September 2010	further literature study + episodes log mining
October 2010	data stream mining
November 2010	OLAP + initial UI
December 2010	finish UI + anomaly detection

Questions?

Thanks for your time!