


Mining and Forecasting of Big Time-series Data

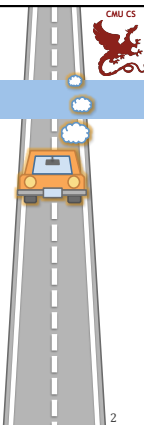
Yasushi Sakurai (Kumamoto University)
Yasuko Matsubara (Kumamoto University)
Christos Faloutsos (Carnegie Mellon University)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 1




Roadmap

- Motivation
- Similarity search, pattern discovery and summarization Part 1
- Non-linear modeling and forecasting Part 2
- Extension of time-series data: tensor analysis Part 3

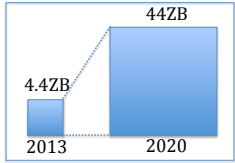


http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 2




Motivation

- Time-series analysis for big data
 - Large-scale sensor networks
 - Web and social networks
 - Medical and healthcare records
- Digital universe growth
 - 4.4 zettabytes (4.4 trillion gigabytes)
 - 44 zettabytes in 2020
 - Data management and data mining



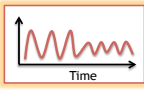
The DIGITAL UNIVERSE of OPPORTUNITIES (IDC 2014)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 3




Time-series analysis for big data

- Volume and Velocity
 - High-speed processing for large-scale data
 - Low memory consumption
 - Online processing for real-time data management
- Variety of data types
 - Multi-dimensional time-series data (e.g., sensor data)
 - Complex time-stamped events (e.g., web-click logs)
 - Time-evolving graph (e.g., social networks)
- Advanced techniques for big data
 - Model estimation, summarization
 - Anomaly detection, forecasting

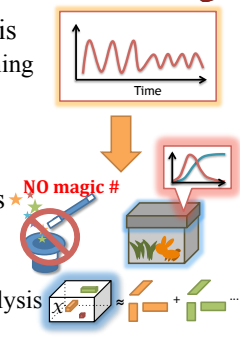


http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 4



New research directions

- Time-series data analysis
 - Indexing and fast searching
 - Sequence matching
 - Clustering
 - etc.
- New research directions
 1. Automatic mining
 2. Non-linear modeling
 3. Large-scale tensor analysis



http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 5



New research directions

- R1. Automatic mining (no magic numbers!)
- R2. Non-linear (gray-box) modeling
- R3. Tensor analysis

NO magic numbers!



http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 6

(R1) Automatic mining

No magic numbers! ... because,

Manual

- sensitive to the parameter tuning
- long tuning steps (hours, days, ...)

Automatic (no magic numbers)

- no expert tuning required

Big data mining:
-> **we cannot afford human intervention!!**

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 7

(R2) Non-linear (gray-box) modeling

- Gray-box mining
 - If we know the equations
- Non-linear (differential) equations
 - Epidemic
 - Biology
 - Physics, Economics, etc.,
- Modeling non-linear phenomena
 - Non-linear time-series analysis for social networks

Image courtesy of Tina Phillips and amenic181 at FreeDigitalPhotos.net. http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 8

(R3) Large-scale tensor analysis

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

Time	URL	User
08-01-12:00	CNN.com	Smith
08-02-15:00	YouTube.com	Brown
08-02-19:00	CNET.com	Smith
08-03-11:00	CNN.com	Johnson
...

Represent as M^{th} order tensor ($M=3$)
 $\mathcal{X} \in \mathbb{N}^{u \times v \times n}$

Element x: # of events
e.g., 'Smith', 'CNN.com', 'Aug 1, 10pm'; 21 times

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 9

New research directions

- Time-series data analysis
 - Indexing and fast searching
 - Sequence matching
 - Clustering
 - etc.
- New research directions
 - R1. Automatic mining
 - R2. Non-linear modeling
 - R3. Large-scale tensor analysis

NO magic #

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 10

New research directions

- Time-series data analysis
 - Indexing and fast searching
 - Sequence matching
 - Clustering
 - etc.
- New research directions
 - R1. Automatic mining
 - R2. Non-linear modeling
 - R3. Large-scale tensor analysis

Part 1
Part 2
Part 3

NO magic #

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/ © 2015 Sakurai, Matsubara & Faloutsos 11