



# Smart Analytics for 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/17-KDD-tut/ © 2017 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

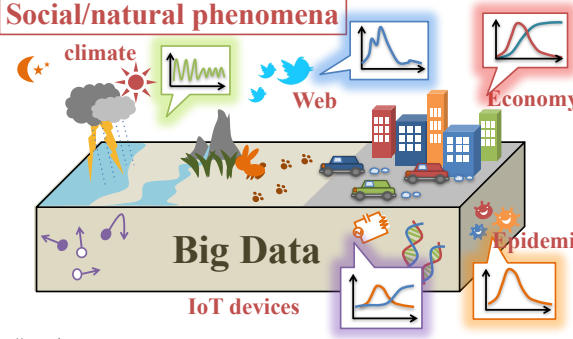


http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 2






## Big time-series data

Social/natural phenomena

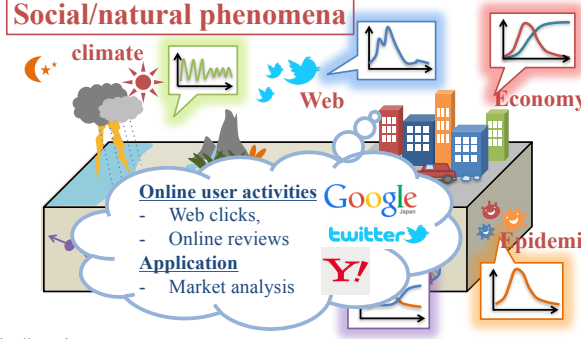


http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 3






## Big time-series data

Social/natural phenomena

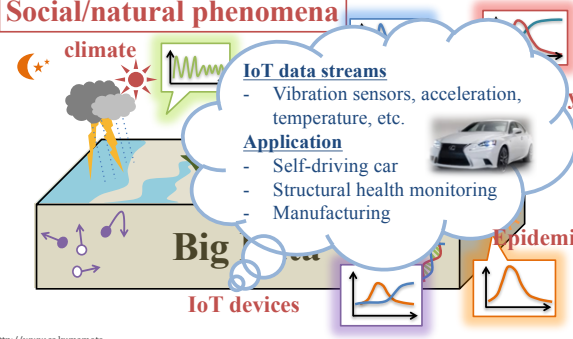


http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 4






## Big time-series data






Social/natural phenomena



http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 5

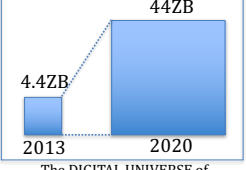
## Motivation

- **Given:** Big time-series data 
- **Goal:**
  - Find important patterns** 
  - Forecast future social activities**
- **At-work:**
  - Online marketing 
  - Sensor monitoring, anomaly detection 
  - Forecasting future events 

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 6

## Motivation

- Time-series analysis for big data
  - Web and social networks
  - IoT data streams
  - Medical and healthcare records
- Digital universe growth
  - 4.4 zettabytes (4.4 trillion gigabytes)
  - 44 zettabytes in 2020

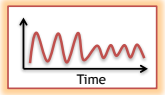


The DIGITAL UNIVERSE of OPPORTUNITIES (IDC 2014)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 7

## Big Time-series analysis

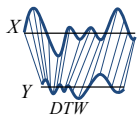
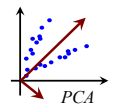
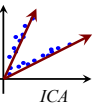
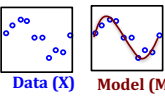
- Volume and Velocity**
  - High-speed processing for large-scale data
  - Low memory consumption
  - Online processing for real-time data management
- Variety of data types**
  - Multi-dimensional time-series data (e.g., IoT device data)
  - Complex time-stamped events (e.g., web-click logs)
  - Time-evolving 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/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 8

## Big Time-series analysis

- Time-series data mining

Indexing, similarity search	Feature extraction	Linear modeling	Stream mining
ED, DTW Correlation	DFT, DWT, SVD, ICA	AR, ARIMA, LDS	StatStream etc...
 DTW	 PCA	 ICA	 Data (X) Model (M)

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 9

## New research directions

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

**NO magic numbers!**



http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 10

## (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/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 11

## (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 analysis for big time-series data

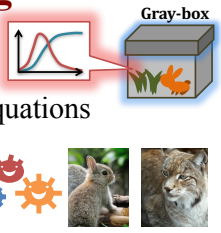


Image courtesy of Tina Phillips and amemic181 at FreeDigitalPhotos.net.  
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 12

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

Diagram: A 3D tensor  $\mathcal{X}$  with axes labeled URL  $u$ , user  $v$ , and Time  $n$ . A red box labeled  $x$  is shown within the tensor.

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

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 13

**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

Diagram: A line graph showing a fluctuating signal over Time. A red 'NO magic #' sign is placed over a graph with a red 'X' and a magnifying glass. Below, a 3D tensor is shown being decomposed into three smaller tensors.

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 14

**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

Diagram: A line graph showing a fluctuating signal over Time. A red 'NO magic #' sign is placed over a graph with a red 'X' and a magnifying glass. Below, a 3D tensor is shown being decomposed into three smaller tensors, labeled Part 1, Part 2, and Part 3.

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 15