



Smart Analytics for Big Time-series Data

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

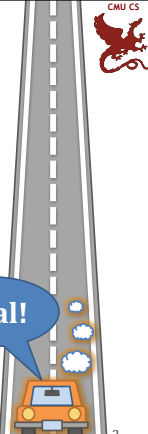
http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 1



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**

Goal!

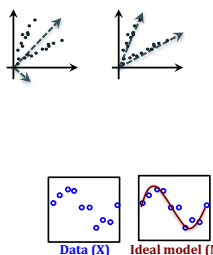


http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 2






Conclusions – Part 1

- Similarity search:
 - Euclidean/time-warping; feature extraction and SAMs
- Feature extraction
 - DFT, DWT, SVD and ICA
- Linear forecasting
 - auto-regression (AR)
 - RLS for streams
- Stream mining
 - RLS, multi-scale windows
- Automatic mining
 - MDL

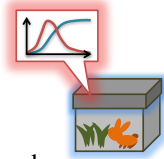


http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 3






Conclusions – Part 2

- Non-linear forecasting
 - Black box: lag-plots + k-nearest neighbors
 - Gray box: with equations, domain knowledge
 - differential equations
 - Logistic function
 - Lotka-Volterra equations, etc.
 - Epidemics, SI & SIR models
 - Hawkes Poisson process, Power law

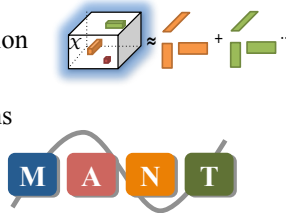


http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 4






Conclusions – Part 3

- Fundamentals for MANT (Multi-Aspect Non-linear Time-series)
 - Tucker/PARAFAC/ tensor decomposition
 - Gibbs sampling
 - Non-linear equations



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

Future direction

- MANT forecasting

M
A
N
T


“MANT (Multi-Aspect Non-linear Time-series)”

 - **Web mining:** e.g., web clicks
{time, user, url, access device, http referrer}
 - **Sensor data monitoring:** e.g., automobile
{time, location, velocity, longitudinal/lateral acceleration}
 - **Medical data analysis:** e.g., EHR (Electronic Health Record)
{time, patient, medical institution, medicine}
- Ideal method for big time-series data
 - Scalable and automatic

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

Kumamoto U CMU CS

Acknowledgements







Thanks to: JSPS KAKENHI Grant-in-Aid for Scientific Research Number JP15H02705, JP17H04681, JP16K12430, PRESTO JST, the MIC/SCOPE #162110003. NSF IIS-0705359, IIS-0534205, CTA-INARC; Yahoo, LLNL, IBM, SPRINT, Google, INTEL, HP, iLab

Disclaimer: All opinions are mine; not necessarily reflecting the opinions of the funding agencies

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


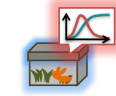
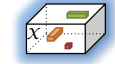
Kumamoto U CMU CS

Questions?

(yasuko, yasushi)[at]es.kumamoto-u.ac.jp christos[at]cs.cmu.edu

URL <http://www.es.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/>

R1	R2	R3
<p>Automatic mining (no magic numbers!)</p> 	<p>Non-linear (gray-box) modeling</p> 	<p>Large-scale tensor analysis</p> 

M
A
N
T

Multi-Aspect Non-linear Time-series

http://www.es.kumamoto-u.ac.jp/~yasuko/TALKS/17-KDD-tut/ © 2017 Sakurai, Matsubara & Faloutsos 8