Mining and Forecasting of Big Time-series Data

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

Roadmap

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

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

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

Conclusions – Part 3

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

Future direction

- MANT forecasting
  “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
Acknowledgements

Thanks to: JSPS KAKENHI Grant-in-Aid for Scientific Research Number 26730060, 24500138, 26280112, 257946. NSF IIS-0705539, IIS-0934205, CTA-INARC, Yahoo, LLNL, IBM, SPRINT, Google, INTEL, HP, iLab

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

Questions?

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/

Automatic mining (no magic numbers!)

Non-linear (gray-box) modeling

Large-scale tensor analysis

URL

http://www.cs.kumamoto-u.ac.jp/~yasuko/TALKS/15-SIGMOD-tut/