Sensor Mining at work: Principles and a Water Quality Case-Study

Christos Faloutsos
School of Computer Science, Carnegie Mellon University
christos@cs.cmu.edu

Jeanne Vanbriesen
Department of Civil and Environmental Engineering
Carnegie Mellon University
jeanne@cmu.edu

(KDD-06 Tutorial proposal)

1 INTENDED DURATION

3 hours

2 MOTIVATION - BASIC INFORMATION

How can we find patterns in a collection of measurements, say, on water quality sensors? Is the water safe to drink? Are we under biological attack? How many sensors do we need to place, and where?

The instructors have been collaborating on exactly these problems for the past 3 years. The tutorial will report our experiences. Specifically, the tutorial surveys the related areas and has two goals: (a) to review the main principles and main data base tools for sensor data analysis (b) to showcase them on a real, important application, namely drinking water quality.

The first part will examine the state of the art in time series indexing and mining. We will cover feature extraction, powerful tools from signal processing (Fourier, Wavelets), and traditional methods for mining and forecasting: the Box-Jenkins (AutoRegressive) methodology. We will also cover powerful methods for discovering correlations across co-evolving time sequences, like Singular Value Decomposition (SVD) and Blind Source Separation (BSS), also known as Independent Component Analysis (ICA).

The second part will review the state of the art of water sensors, their capabilities and limitations, the corresponding research challenges they induce to the sensor mining algorithms. It will also describe the problem of monitoring

multiple substances that chemically react with each other, and how to infer concentrations of one substance, given measurements of the other. Finally, it will describe the 'control' problem of where to optimally place a fixed number of sensors, to maximize the monitoring benefits.

The tutorial ends with a list of future directions for database research, motivated by the water quality monitoring application.

3 INTENDED AUDIENCE

Researchers and practitioners that want a concise, intuitive overview of the major tools in sensor mining, motivated by the vital problem of water quality monitoring.

4 COVERAGE

- Problem definition [Faloutsos]
- Main tools [Faloutsos]
  - Time series and Forecasting
    - Time series indexing and feature extraction
    - Fourier, Wavelets, Time Warping
    - Linear forecasting, ARIMA, recursive least squares
  - Lag correlations
  - Cross correlations
    - Hidden variable detection
    - Singular value decomposition
    - Sketches and random projections
    - Independent Component Analysis
- Intro to water quality [Vanbriesen]
  - Types of sensors now and in the future
  - Direct and surrogate data collection
  - Integration of sensors into control systems
  - Modeling Distribution systems with and without potential sensors
  - Recent achievements in multi-species modeling modeling
  - Future directions in sensor data for water systems.
- Conclusions [Faloutsos]
  - Database research directions for water quality monitoring.

5 SPECIAL EQUIPMENT

None

6 OTHER INFORMATION

- The tutorial is a spin-off of a large NSF joint project on water quality monitoring, which we reported in [1].
- The first part of the tutorial, has overlap with older tutorials by Faloutsos. The differences are (a) coverage of additional methods, like Blind Source Separation, lag correlation discovery, and on-line SVD, and (b) emphasis on water-quality applications.

7 BIOGRAPHICAL NOTES

Christos Faloutsos is a Professor at Carnegie Mellon University. He has received the Presidential Young Investigator Award by the National Science Foundation (1989), six "best paper" awards, and several teaching awards. He served as a member of the executive committee of SIGKDD; he has published over 130 refereed articles, one monograph, and holds five patents. His research interests include data mining for streams, sensors and networks, fractals, indexing for multimedia and bio-informatics data bases, and database performance.

Jeanne Vanbriesen is an Associate Professor at Carnegie Mellon University in the Department of Civil and Environmental Engineering. She has received the George Tallmann Ladd Award for outstanding research and professional accomplishments from the College of Engineering and is the first recipient of the Paul and Norene Christiano Faculty Fellowship. Her research interests include detection of pathogens in water, applications of thermodynamics to bacterial systems, and biodegradation of recalcitrant organic compounds.

References

[1] A. Ailamaki, C. Faloutsos, P. S. Fischbeck, M. J. Small, and J. VanBriesen. An environmental sensor network to determine drinking water quality and security. *SIGMOD Record*, 32(4):47–52, 2003.