Adaptive Sampling of Streaming Signals

Tristan Wietsma
Illinois Institute of Informatics
Dept. of Civil & Environmental Engineering
University of Illinois at Urbana-Champaign

Prof. Barbara Minsker
Dept. of Civil & Environmental Engineering
National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign

NCSA
Illinois Informatics Institute
Invent. Imagine. Innovate.
Basic Idea

Sample the events you care about.
More algo at the front, less at the back.
One size does not fit all.
Introduction

Motivation
Definitions
Literature
Motivation

Sensor control can mitigate the data deluge.

Sensor → Database → End User

Opportunity!

Distributed Processing

SQL, NoSQL, NewSQL
Motivation

Case in point: $600 logger or $25 Linux
Definitions

- Adaptive sampling: responsive to dynamics
- Compressive sensing: reconstruction error
- Intelligent sampling: feature extraction
- Hot Moment: period of elevated signal activity
- Heuristic: rule/logic/common sense (art + science)
Literature

- Alippi et al.: *Nyquist-Shannon sampling theorem*
- Feizi et al.: *Locally adaptive state space models*
- Marbini & Sacks: *Adaptive feedback framework*
- Jain & Chang: *Kalman filtering with feedback loop*
- Law et al.: *Box-Jenkins statistical approach*

Gist: Compressive sensing via local adaptation.
Methodology

Nyquist-Shannon Sampling Theorem
Adaptive Sampling Model
Measuring Performance
Total Signal Power
Nyquist-Shannon Sampling Theorem

- Fundamental contribution from signal processing
- Defines a bound on sampling rate
- Algo: Calculate power spectrum for signal
  → Find $f_{\text{max}}$, highest frequency with power $> 0$
  → Double it! This is the Nyquist rate.
Nyquist-Shannon Sampling Theorem
Adaptive Sampling Model

1. Observe initial window
2. Calculate power spectrum
3. Calculate Nyquist rate
4. Schedule next sample
5. Update window
6. Observe next sample
Adaptive Sampling Model

Observe initial window of the signal using uniform sampling

Calculate the Nyquist rate and schedule next sample

Draw scheduled sample

Update recent window using linear interpolation to estimate unobserved times.
Measuring Performance

Models are evaluated w.r.t two metrics:

- Sampling Fraction: measure of efficiency
- Hot Moment Sampling Fraction: qualitative*

* Requires a heuristic to find hot moments.
Measuring Performance

Point C is inferior to A and B. It is less efficient and captures fewer hot samples.

Points A and B are equivalent to each other in the Pareto sense. B is more efficient than A, but captures fewer hot samples. Similarly, A captures more hot samples than B, but at lower efficiency.
Total Signal Power

- Area under the power spectrum.
- Cheap indicator of relative signal complexity.
- (We'll see how in the case study)
Case Study: Soil Moisture

Description of the Data

Hot Moment Heuristic

Model Calibration

Training & Testing Set Performance

Sensitivity Analysis

Reconstruction Error
Description of the Data

- Volumetric soil moisture at 5 cm depth
- Energy Biosciences Institute
- 5020 15-minute samples from Spring of '09
- First 1500 samples used for training set
Hot Moment Heuristic

4-hour rolling window of signal power. Threshold = $\mu + se$; 24 hour duration.
Model Calibration

- Window size from 30 minutes to 48 hours
- Noise threshold (epsilon) from $3.7605 \times 10^{-11}$ to 0.0036
- Just shy of 50,000 parameterizations
- Code: Python (snips @ github.com/tristanwietsma)
- Gear: i7 laptop (batchjobs.py)
Training Results

![Graph showing hot moment sampling fraction against sampling fraction for uniform and adaptive sampling.](image-url)
Testing Set Performance

![Graph showing Testing Set Performance](image)

- **Uniform Sampling**
- **Adaptive Sampling**
Testing Set Performance
"The hot moment heuristic is based on a practical, yet extremely variable, domain-specific estimate of the time required for soil moisture to decay to field capacity. This fuzzy definition is in line with the nature of heuristics."

That said, different heuristic parameters didn't substantially alter the testing set Pareto curve.
- Rolling window was tested out to 24 hours
- Event threshold changes have predictable impact
- Drainage time was tested from 12 to 36 hours
Reconstruction Error

![Graph showing mean square residual error against sampling fraction with lines for uniform and adaptive sampling.](image)
Conclusion

Take Aways

Future Research

Acknowledgements
Take Aways

- Intelligent sampling vs compressive sensing
- Acquisition can use domain specific knowledge too
Future Research

- Code: Open source data logger & sampling logic.
- Theory: How does it scale? Spatial information?
- Multi-scale: local-regional-global and CMD
Acknowledgements

Big thanks to Prof. Carl Bernacchi (USDA ARS) at the Energy Biosciences Institute!
Thank You!