Active Learning in Networks

Mark Coates

Frederic Thouin, Michael Rabbat

Sponsored by National Science and Engineering Research Council of Canada (NSERC)
Network Estimation

Performance Monitoring
- Traffic Matrices
- Latencies, Loss Rates, Topology
- Available Bandwidth
- Traffic Classification

Sensor and Actuator Networks
- Target tracking
- Anomaly detection
- Causal effect analysis
Why active learning?

Network monitoring is costly

- Traffic matrices:
  - Installing & maintaining measurement devices
  - Very large volumes of data
- Loss, latency, topology, available bandwidth
  - Bandwidth overhead, time-sensitivity
  - Potential to disturb the system they are measuring

Sensor and Actuator Networks

- Measurement energy (sensor activation and operation)
- Communication overhead
- Node wake-up overhead
- Processing and memory cost
Three Examples

- **Out-of-sequence measurements**
  - Estimate mutual information (Extended Kalman Smoother)
  - Decide whether to process and what type of processing

- **Traffic Flow Classification**
  - Measure flow characteristics: #packets, size, rate, spacing
  - Request application label for limited number of packets

- **Topology Identification**
  - Noisy pairwise similarity metrics between nodes
  - How many do you need to reconstruct topology?
Case Study: Available Bandwidth

- Traditional definition: unused capacity on a network path
Available Bandwidth

Available bandwidth of a link

\[ A_i(t, t + \tau) = \frac{1}{\tau} \int_{t}^{t+\tau} C_i(x) - \lambda_i(x) \, dx \]

Available bandwidth of a path

\[ A(t, t + \tau) = \min_{i=1,\ldots,H} A_i(t, t + \tau) \]

Tight link determines available bandwidth
Probabilistic Available Bandwidth

- What do we really want to know?
- What can we really measure?

- Largest ingress rate so that egress is nearly the same

\[ A_p(\delta, \varepsilon) = \max_{R_{in}} \Pr(R_{out} > R_{in} - \varepsilon) > 1 - \delta \]
Available Bandwidth Techniques

Pathload [Jain and Dovrolis, 2002]
- Sender transmits periodic stream of rate $P$
- Receiver measures one-way delay $D(k)$
- Calculate one-way delay variations $\Delta(k) = D(k) - D(k-1)$

Transmit: 
Receive: (below AB)
Available Bandwidth Techniques

Pathload [Jain and Dovrolis, 2002]
- Sender transmits periodic stream of rate $P$
- Receiver measures one-way delay $D(k)$
- Calculate one-way delay variations $\Delta(k) = D(k) - D(k-1)$

Transmit: \[ \square \square \square \square \square \square \]
Receive: \[ \square \square \square \square \square \square \square \] (above AB)
Available Bandwidth Techniques

Pathload [Jain and Dovrolis, 2002]

- Sender transmits periodic stream of rate P
- Receiver measures one-way delay $D(k)$
- Calculate one-way delay variations $\Delta(k) = D(k) - D(k-1)$

- Ideally, (stationary, fluid-model cross-traffic), if $P > B$ then $\Delta(k) > 0$ for all $k$

- Binary bisection search to determine upper and lower bounds
Delay Measurements

- Green line: $R=45.9$, $PCT=0.43$
- Red line: $R=114.75$, $PCT=0.96$

Graphs showing OWD versus Packet ID (Arrival Time) and PCT versus Probing Rate $P$ (Mbps).
Alternative Metric: Rate Differential

![Graphs showing receiving rate and rate differential vs probing rate](image)
**Bandwidth Estimation Algorithm**

- **Goal:**
  - Calculate marginal posterior of PAB for each path

- **Initialization**
  - Create factor graph from known topology

- **Bandwidth Estimation**
  - Determine which path to probe
  - Determine probing rate P
  - Update distributions using belief propagation
  - Repeat until stopping criterion is met

- Active Learning
Noisy Active Learning

- Use previous data to guide choice of measurements
Noisy Active Learning

- Use previous data to guide choice of measurements
Noisy Active Learning

- Use previous data to guide choice of measurements
Noisy Active Learning

- Use previous data to guide choice of measurements
Noisy Active Learning

- Use previous data to guide choice of measurements
Noisy Active Learning

- Probe at the median of the posterior distribution
  - Intuitively most informative measurement
Likelihood Model

\[ \Pr(Z_{\text{RDT}} = 1) \]

\[ r_p - \hat{A}_p \text{ (Mbps)} \]

- \text{data}
- \text{best fit: logsig(}-\alpha \, x\text{)}
Likelihood Model
Updating the distributions

$$z = 0$$

$$z = 1$$

$p(A | z)$

Available Capacity $A_y$ (Mbps)
Noisy Active Learning

- Noiseless case (binary search) [Dasgupta ‘04]
  - Active: $O(\log n)$ samples
  - Passive: $O(n)$ samples

- Bounded noise [Balcan, Beygelzimer, Langford ‘06]
  - Excess risk decays exponentially
  - Rate depends on the noise margin

- Unbounded noise [Castro and Nowak ‘08]
  - Less improvement, but still important gain
  - Passive $O(n^{2/3})$ vs. active $O(n^{-1})$
Network-wide Measurement

- **Multiple paths**
  - High load when measuring multiple paths.
  - Simultaneous measurement can bias results.
  - Sequential rate-scanning is a slow process.

- **Exploit correlations**
  - Paths share tight links
  - Use information from measurements on other paths
Factor Graph

Link variables

Path variables \( y_1 = \min(x_1, x_3, x_4) \)

\[
\begin{align*}
y_1 &= x_1 - x_3 - x_4 \\
y_2 &= x_1 - x_3 - x_5 \\
y_3 &= x_2 - x_3 - x_4 \\
y_4 &= x_2 - x_3 - x_5
\end{align*}
\]
**Measurements**

- **Binary measurements**
  \[ I(R_{in} - R_{out} < \varepsilon) \]

- **Measurement model (likelihood function)**
  \[ f(z) = L(z|y_1) \]

---

RSS Workshop
Updating the distributions

\[ f(y_1) \]

\[ y \]

\[ x_1 \quad x_2 \quad x_3 \]

Probing Rate

\[ t \]

RSS Workshop
Updating the distributions

$x_1$, $x_2$, $x_3$

$f(y_1)$

$y_1$

$y$

$f(z)$

Probing Rate

$t$

$t+1$

RSS Workshop
Algorithm

Initialization
- Create factor graph from known topology

Bandwidth Estimation
- Determine which path to probe
- Determine probing rate $P$
- Update distributions using belief propagation
- Repeat until stopping criterion is met
Choosing the path

Goal: choose most informative path

Possible methods:

- Choose path with largest expected information gain
  - Simulate all outcomes
- Probabilistic choice weighted by
  - Path entropy (WE)
  - Path confidence interval (WCI)
Planetlab Experiments

- 20 paths, 32 links, 25 nodes
- **End nodes:** echo.cs.princeton.edu, planetlabone.ccs.neu.edu, planet2.scs.cs.nyu.edu, pl2.csl.utoronto.ca, pl1.bit.uoit.ca.

- Measurement: 3 trains of 150 packets of 1000 bytes (Median of 3 rates)
- Stopping criterion: 95% Confidence interval < 10 Mbps

- Testing:
  - Train of 2400 packets of 1000 bytes (60 secs video at 320kbps)
  - Test at lower-bound, mean, upper-bound, upper-bound+5
Topology
How Many Packets Per Train?

- Repeat experiment with different length trains
  - Execute until satisfying the same stopping criterion
    (95% confidence in 10Mbps)
Train Length and Accuracy

Results are consistent across wide range of train lengths
Some loss in accuracy when probing above the estimated available bandwidth
Experimental Results
Summary

- Significant number of networking problems where active learning is very attractive
  - Multiple situations where acquiring data has a cost

- Currently we strive to approximate the expected information gain by fast, low-cost calculations
  - Weighted confidence interval
  - Posterior median
  - Mutual information under Gaussian approximation

- More effective techniques?