Title
BARD: Bayesian Assisted Resource Discovery

Permalink
https://escholarship.org/uc/item/2735x02h

Authors
Fred Stann
John Heidemann

Publication Date
2004
BARD: Bayesian Assisted Resource Discovery
Fred Stann, John Heidemann
fstann@usc.edu, johnh@isi.edu, http://www.isi.edu/ilense

Problem Description: Flooding Limitation with Attribute-Based Data Dissemination

Efficient Data Resource Discovery

• **Goal:** Flooding Limitation
  - Without a priori knowledge, location of data resources/sinks is expensive
  - Flooding in Directed Diffusion
  - Even non-queried data is constantly moved with Data-Centric Storage
• **Proposed Solution**
  - Efficient data resource discovery should be available to applications independent of type or number of data attributes
  -Flooding-limitation schemes rely on specific attributes (e.g., geography)
  - Exploit Spatial Correlation
  - Data can be spatially correlated by a large variety of problem aspects: topography, target tendencies, geography, sink locations, link reliability, inactive regions, etc...

Approach: Per-attribute history of data paths as a statistical predictor of future success

Probabilistic Approach

• Use prior routing history to avoid flooding for similar queries
• Collect per-attribute statistics about which neighbors lead to data resources in a sliding window of time
• Use the collected statistics (prior distribution) to limit flooding for 4/5ths of subsequent queries
• Use occasional flooding (1/5th) to maintain prior distribution

Bayesian Estimation

• **Naïve Bayes** can build probabilities incrementally from the set of attributes that defines a query or data
• Even partial matches contribute to probability calculation
• Limit flooding to high probability candidates (neighbors)
• Fully distributed scheme without inter-node transfer of meta-data

Proposed Solution: Use Bayesian Estimation to Limit Flooding in Diffusion

Conclusions

• Improvement from BARD is in proportion to spatial correlation of events in environment and size (fan-out) of routing tree
• BARD is a general technique applicable to a wide variety of applications with arbitrary attributes for their named data
• Bayesian estimation allows for partial matches where there is still a high probability of data and query matching
• Because it is implemented as a diffusion filter it is a selectable service to users of diffusion