The NIST IAD Data Science Evaluation Series Workshop

Evaluation Breakouts
Craig Greenberg
March 17-18, 2016

Breakout Logistics

- There are three breakout sessions, please attend one
- Breakouts consist of discussions led by moderator
- Slides from each breakout will be presented

Evaluation Breakout

In your domain of interest, given a problem:

- Challenges:
  - What workflow challenges (data dependencies) arise in your domain?
  - What is ground truth and how is it obtained?
  - Other challenges?
- Evaluation:
  - Are there specific types of measures of effectiveness in your domain?
  - What workflows need to be in place to ensure a constructive evaluation?
  - How to evaluate without ground truth?
  - Other evaluation issues?
- The Future:
  - What would you like a new evaluation to address/to look like?
  - What other questions do we need to answer?

Benchmarking Breakout

- What traits are desired in a software monitor?
- What performance metrics are important to obtain?
- What are good benchmark suites for data science? What makes a benchmark suite good?
- What aspects of hardware, software, or system architecture might influence benchmarking performance? What practices might be used to control or characterize their influence on performance?

Infrastructure Breakout

- What would be beneficial to you in terms of:
  - virtualization (VM images, Containers, Bare Metal)
  - specialized hardware (e.g., GPUs or Xeon Phis)
  - benchmark models and metrics
  - data access methodology (S3 buckets, Local Storage with Map-Reduce, NFS drive, etc.)?
  - compute paradigm support
  - Amount of temporary storage
- Evaluation workflow and software submission suggestions

Breakout Sessions

- Evaluation
  - Lecture Room D
- Infrastructure
  - Room B113
- Benchmarking
  - Executive Conference Room