Concluding Remarks

March 18, 2016

What We Learned from You

- **Development data sets** during the evaluation are desirable.
- **Scalability challenges** can be examined with a 150 GB data set.
- **Regression tests** on past evaluation data allow for additional algorithm comparisons.
- **Confidence interval evaluation** would be useful.

What We Learned from You

- **Template Virtual Machines (VMs)** should be provided with sample data to facilitate testing.

- **Measures of effectiveness** need to be defined at a task and domain level.

- **New benchmarks** should consider new technologies, new architectural platforms, new algorithms, and new computing scenarios.

What We Learned from You

- **Small data sets** aid in building systems and reading data before testing

- **Ground truth** should be carefully selected and as close to reality as possible

- **Data provenance** is important and is involved in all of the tasks, and metadata should be provided with the data

- **Data granularity** is important match with the tasks and problems of interest

What Else? Feedback?

- Suggestions accepted **now**.
- Other suggestions later? Contact us at **datascience@nist.gov**

Upcoming Events: **DSE Pilot Evaluation**, Summer 2016