Special Notice

Probabilistic Programming for Advancing Machine Learning (PPAML)

DARPA-SN-13-30

March 19, 2013

Defense Advanced Research Projects Agency
675 North Randolph Street
Arlington, VA 22203-2114

DARPA-SN-13-30
DARPA Information Innovation Office
Probabilistic Programming for Advancing Machine Learning (PPAML)
Proposers' Day – April 10, 2013
Arlington, VA

The Defense Advanced Research Projects Agency (DARPA) Information Innovation Office (I2O) will conduct a briefing in support of the anticipated Broad Agency Announcement (BAA) for the Probabilistic Programming for Advancing Machine Learning (PPAML) program. When released, the BAA will be posted on the Federal Business Opportunities (FBO) website, http://www.fedbizopps.gov. ***This Proposers' Day is unclassified***.

Attendance at the PPAML Proposers' Day is voluntary and is not required to propose to subsequent solicitations (if any) on this topic. The Proposers' Day does not constitute a formal solicitation for proposals or abstracts. This announcement is issued solely for information and program planning purposes and is not a Request for Information (RFI). Since this is not an RFI, no submissions against this notice will be accepted by the Government. DARPA will not provide reimbursement for costs incurred to participate in this Proposers' Day. Interested parties to this notice are cautioned that nothing herein obligates the Government to issue a solicitation.

BACKGROUND

Machine learning is at the heart of modern approaches to artificial intelligence. The field posits that teaching computers how to learn can be significantly more effective than programming them explicitly. This idea has revolutionized what computers can do in a wide range of domains, including Intelligence, Surveillance, and Reconnaissance (ISR); Natural Language Processing (NLP); Predictive Analytics; Cyber; and various scientific disciplines. Example applications include: self-driving cars, image search and activity detection, object tracking, topic models, spam filters, recommender systems, predictive databases, and gene sequencing. Unfortunately, building effective machine learning applications currently requires Herculean efforts on the part of highly trained experts in machine learning.

Probabilistic Programming is a new programming paradigm for managing uncertain information. The goal of PPAML is to advance machine learning by using probabilistic programming to: (1) dramatically increase the number of people who can successfully build machine learning applications, (2) make machine learning experts radically more effective, and (3) enable new applications that are impossible to conceive of using today's technology. In support of this overarching goal, PPAML has a number of sub-goals. Specifically, the sub-goals are: (1) to make machine learning model code shorter, (2) to reduce development time, (3) to facilitate the construction of richer models, (4) to require lower levels of expertise in building machine learning applications, and (5) to support the construction of integrated models.

Achieving this goal requires research breakthroughs in several areas. For the "front end" of the system,(i.e., the portion of the system that application developers will see) these challenges

reduce to answering the following questions: (1) How to balance the expressive power of the language with the corresponding difficulty of producing an efficient solver?; (2) How to make the system maximally usable by users from a range of different backgrounds and skill levels?; and (3) What kinds of profilers, debuggers, and model verification/checking tools should the system have so end users can understand the performance, correctness, and accuracy ramifications of their modeling choices?

For the "back end" of the system, (i.e., the portion of the system responsible for solving the models and answering queries) the key research challenge is improving system performance and predictability. Improvements on the order of two to four orders of magnitude over the state of the art are likely necessary to produce an effective system, with simpler models and those requiring less data likely to fall at the low end of the spectrum while richer models and those needing lots of data at the high end. Advances that are likely necessary to achieve this improvement include: (1) developing analyses that select the most appropriate solver or set of solvers given a particular model, query, and set of prior data; (2) improving the performance of existing solvers by incorporating ideas from the compiler optimization community; (3) compiling specific solvers to diverse hardware platforms in ways that optimize the resources of the hardware including multi-core machines, GPUs, cloud infrastructures, and potentially custom hardware; (4) developing new solvers; and (5) developing an API so that new solvers can be slotted into the solving infrastructure easily.

Advances in basic machine learning technology will also be required. Anticipated research challenges in this area include: (1) advancing the theory of probabilistic programming; (2) discovering new inference algorithms that are more efficient, more accurate, more predictable, or more generalizable; (3) discovering novel representations that support more efficient, more accurate, more predictable, or more generalizable inference; (4) developing inference algorithms that work over streaming data or have better scaling properties; and (5) developing techniques for assessing model fitness for a particular data set.

The systems developed under PPAML will be evaluated using a collection of challenge problems that span the range of machine learning applications. Each system will be evaluated on how well it performs on all challenge problems. In addition, to evaluate the effectiveness of program technologies in enabling the rapid creation of new machine learning applications by domain experts, the program will include annual "Summer Schools" of two to four weeks in duration. These Summer Schools will bring together PPAML performers and potential customers from the Government and commercial sectors with the goal of enabling the customers to develop useful machine learning applications in their domain of expertise within the duration of the Summer School. As PPAML technology matures, Summer School attendees should be able to work with greater levels of independence with the end goal of requiring no assistance from PPAML performers.

PURPOSE

The purpose of the PPAML Proposers' Day is threefold:

1. To familiarize participants with DARPA's interest in innovative approaches to advancing machine learning by using probabilistic programming.
2. To identify potential proposers and promote understanding of the anticipated PPAML BAA proposal requirements.
3. To promote discussion of synergistic capabilities among potential program participants.

Additional information regarding the anticipated PPAML BAA will be available at http://www.darpa.mil/Opportunities/Solicitations/I2O_Solicitations.aspx, following the publication of the BAA on FBO. It is anticipated that the PPAML BAA will be released by the end of March 2013. If a BAA is released, materials presented at the Proposers' Day, as well as a frequently asked questions (FAQ) document compiling questions and answers received to date, may be made available at http://www.darpa.mil/Opportunities/Solicitations/I2O_Solicitations.aspx.

The PPAML Proposers' Day will be recorded, and a copy of the video will be posted on DARPA's website after the Proposers' Day.

Individuals interested in participating in the PPAML program may give presentations during the afternoon session to publicize their capabilities and solicit teaming. Presentations will be limited to two minutes in length and a single content slide. Presenters should note their intent to provide a presentation when registering for the Proposers' Day. Presentation slides should be submitted to PPAML@darpa.mil by April 5, 2013.

TENTATIVE AGENDA

Registration: 0900 – 1000 EST
Government presentations: 1000 – 1130 EST
Lunch: 1130 – 1300 EST
Proposers' presentations: 1300 – 1400 EST
Government response to questions from morning session: 1400 – 1500 EST

LOCATION

Executive Conference Center at Liberty Center
4075 Wilson Boulevard, Suite 350
Arlington, VA 22203
http://ecc.sainc.com/index.aspx

REGISTRATION INFORMATION

Interested parties must register at http://www.solers.com/BAAinfo-reg/ppaml no later than COB on April 5, 2013.

Non-US citizens are required to submit a DARPA Form 60 "Foreign National Visit Request" to DARPA no later than April 3, 2013 to PPAML@darpa.mil. All attendees will be required to present Government-issued photo identification upon entry to the event. Attendance at the Proposers' Day is voluntary and is not required to propose to this solicitation. DARPA will not provide reimbursement for costs incurred in participating in this Proposers' Day. Only two (2) members from each organization may physically attend the Proposers' Day. A maximum of 130 registrants will be allowed to attend.

POINT OF CONTACT

All questions regarding the Proposers' Day should be sent to PPAML@darpa.mil.