IML news

Michele Floris, Sergei Gleyzer, Lorenzo Moneta,
Steven Schramm, Paul Seyfert

IML Meeting
February 24, 2017
First IML workshop: March 20 to 22, 2017

Please register here if you plan to attend

- Already 100 registrants!

We strongly encourage people to attend in person if possible

- Vidyo will be provided for the discussion sessions
- Vidyo probably not available for tutorials and hands-on sessions

Please send abstracts to iml.coordinators@cern.ch by March 3

- Abstracts are for giving talks in the tagging session, not participation
- When submitting an abstract, specify the desired length (20 or 30 min)
IML workshop schedule

- Monday morning: introduction and industry session (including panel)
- Monday afternoon: community white paper - the future of ML in HEP
- Tuesday morning: object identification session #1
- Tuesday early afternoon: tutorials #1
- Tuesday late afternoon: object identification challenge / hands-on
- Wednesday morning: object identification session #2
- Wednesday early afternoon: tutorials #2
- Wednesday late afternoon: challenge results, close-out
Other news

- HSF Google Summer of Code
  - CERN-SFT Google Summer of Code → HSF Google Summer of Code 2017
  - New webpage: http://hepsoftwarefoundation.org/activities/gsoc.html
  - Have a ML software development project idea? Want to be a mentor? Student who wants to participate?
    - Contact Sergei (sergei@cern.ch) and Enric (etejedor@cern.ch)

- HSF support position in SFT
  - Software developer position open in SFT to support HSF activities
  - More details: https://jobs.web.cern.ch/job/12259

- CWP Preparation
  - 2nd part of CWP-ML Software and Tools discussions will be on Monday afternoon during IML workshop
  - Interested in contributing? Join the google group
  - Pre-workshop preparatory meetings to be announced next week
**IML Machine Learning Working Group - Parallelized/Distributed Machine Learning**

**Friday 24 Feb 2017, 15:00 → 17:35**  
Europe/Zurich  
40-S2-C01 - Salle Curie (CERN)

**Videoconference Rooms**  
IML-MachineLearning-WG

---

15:00 → 15:10 **News and group updates**  
**Speakers:** Lorenzo Moneta (CERN), Michele Floris (CERN), Paul Seyfert (Universita di INFN, Milano-Bicocca (IT)), Dr. Sergei Gleyzer (University of Florida (US)), Steven Randolph Schramm (Universite de Geneve (CH))

15:10 → 15:30 **Internally-Parallelized Boosted Decision Trees**  
**Speaker:** Andrew Mathew Carne (University of Florida (US))

15:30 → 15:50 **Rapid development platforms for machine learning**  
**Speaker:** Dr. Andrew Lowe (Hungarian Academy of Sciences (HU))

15:50 → 15:55 **Distributed Deep Learning using Apache Spark and Keras (see materials)**

Data parallelism is an inherently different methodology of optimizing parameters. The general idea is to reduce the training time by having n workers optimizing a central model by processing n different shards (partitions) of the dataset in parallel. In this setting we distribute n model replicas over n processing nodes, i.e., every node (or process) holds one model replica. Then, the workers train their local replica using the assigned data shard. However, it is possible to coordinate the workers in such a way that, together, they will optimize a single objective during training and as a result, reduce the wall clock training time. There are several approaches to achieve this, and these will be discussed in greater detail in the materials below.

**Speaker:** Joeri Hermans (Maastricht University (NL))

---

15:55 → 16:25 **Parallelization in Machine Learning with Multiple Processes**  
**Speaker:** Omar Andrea Zapata Mesa (University of Antioquia & Metropolitan Institute of Technology)