

The future of the machine learning working group

Sergei Gleyzer, Steven Schramm

October 27, 2015



- **I**nter-experimental **M**achine **L**earning *working group* (**IML**):
 - A new (informal) effort to facilitate communication between the LHC experiments and the rapidly evolving world of Machine Learning (ML)
 - ML algorithms from a couple years ago are already outdated
 - Need to determine which new ML techniques are applicable to HEP
 - A centralized effort will make it much easier for everyone to benefit
- We want this to be a *working group*, not just a series of seminars
- In order to make this group effective, we must:
 1. Identify a set of tasks to define the scope and priorities of the group
 2. Determine what information can be shared between experiments
 3. Understand the current group composition and needs
 4. Define the meeting frequency and structure
- Input is welcome throughout, as are suggestions of missed topics
 - We want to ensure this group is useful to everyone

Scope: communication-related topics

1. Provide a forum for the transfer of knowledge between experiments
 - Some techniques will undoubtedly remain private until published
 - Many instances where sharing of general information could be useful
 - Trying new ML techniques: were they beneficial in the HEP context?
 - Comparing different algorithms in benchmark topologies
 - Sharing problems encountered such that others can suggest solutions
2. Facilitate communication between the HEP and ML communities
 - The HEP and ML communities can mutually benefit from interactions
 - ML experts work with advanced algorithms on huge datasets
 - HEP problems don't always fit into standard ML techniques
 - For example, problems with a variable number of inputs and outputs
 - We can provide a unique environment for testing new algorithms
 - Aim to invite experts from both communities to present the latest ideas

Scope: software-related topics

3. Maintenance and development of core ML software for HEP
 - Large amount of development of R/Py/TMVA in recent months
 - Significantly simplifies the use of modern ML tools in HEP
 - Perfect example of the type of work benefiting members of this forum
 - All of the different experiments can make use of new functionality
 - ML experts more likely to collaborate with HEP with R/python support
 - While a good example, TMVA is not the only relevant package
 - Many good options for us to use: [scikit-learn](#), [REP \(Yandex\)](#), etc
 - Maintain a list of open HEP software tasks related to ML
 - Each experiment will benefit from the implementation of these items
4. Tutorials for use of ML tools in HEP (TMVA, scikit-learn, REP, etc)
 - Tutorials are a great way to increase the use of ML in HEP
 - Hope to hold multiple tutorials per year, not just at DS@LHC2015
 - We hope to incorporate info on new features into the tutorials

Scope: ML challenges

5. Provide support and guidance for future HEP-ML challenges
 - Provide introductions to HEP concepts on the IML website
 - Would expand as more challenges covering different areas are created
 - Build a repository on what works for the benefit of future challenges
 - Figures of merit, experiences with kaggle vs other platforms, etc
 - Assist in coordination to avoid repetition and build on past studies
 - Too many binary classification problems risks losing ML expert interest
 - Help plan a series of consecutive challenges to address advanced topics
 - Help with importing winning entries back into HEP contexts
 - Point 3: “Maintenance and development of core ML software for HEP”
 - Note that this is all optional - we are not taking control of challenges
 - The idea is to instead provide assistance where possible *if desired*

Inter-experimental communication on ML

- LHC Dark Matter forum activities in early 2015 are a good example
 - Agreement to compare plots containing only signal MC
 - Established a common direction between ATLAS and CMS for Run-II
- The LHC top working group is another long-lasting example
 - Long-term effort improve our understanding of the top quark
 - One of five **LPCC groups** (Lhc Physics Center at Cern)
 - If possible, LPCC status is probably a good goal for this forum
- Regardless of the path, some things will have to be done
 - Obtain the support of the management of each experiment
 - Identify benchmark(s) where ML techniques can be compared without giving away experimental details
 - Maybe consider some common signal MC sample(s)?
 - Maybe work with public data? (from challenges, etc)
- Open dialogue between HEP experiments is difficult, but possible!

Current list of participants

- There are currently 106 people on the mailing list
- The following groups are represented:
 - ALICE (PH-UAI), ATLAS (PH-UAT and PH-ADP), CMS (PH-UCM and PH-CMG), LHCb (PH-ULB and PH-LBC)
 - CERN software, computing, and ML (PH-SFT and GS-SIS-OA)
 - A few people from other areas
- This is not to suggest official support from each collaboration
 - Rather it expresses interest from the members of each experiment
- Currently there is a large ATLAS/CMS bias
- Ideally, this will be a truly inter-experimental forum
 - LHCb *flavours of physics* challenge indicates strong interest in ML
 - We hope to involve ALICE and LHCb much more in the future
 - We also want to encourage the participation of the ML community
- For those not from ATLAS and CMS: what would you like to see?

Meeting structure

- The current idea is to hold informal meetings \sim once per month
- Meetings would ideally be \sim 2 hours long, capped at 3 hours
- Self-submitted contributions will be accepted at each meeting
 - These are particularly encouraged as a part of the *working group*
- We hope to regularly have invited talks from the ML community
- Occasionally the meeting will have a primary topic
 - For example, this meeting focused on tutorials and DS@LHC2015
 - This should be the exception, not the rule
- The next meeting will tentatively be during Nov 30 - Dec 4

Summary

- An informal working group scope was proposed:
 1. Provide a forum for inter-experimental discussions about ML
 2. Facilitate communication between the HEP and ML communities
 3. Maintenance and development of core ML software for HEP
 4. Tutorials for use of ML tools in HEP (TMVA, scikit-learn, etc)
 5. Provide support and guidance for future HEP-ML challenges
- Discussed possible paths for ML communication between experiments
 - Will require discussion with and approval of experiment coordination
- Defined the meeting frequency and structure
 - The next meeting will tentatively be during Nov 30 - Dec 4
- This is intended to be a *working group*, please help where you can
 - Together, we can ensure all of HEP benefits from ML developments
- If you are interested in following this effort, please join us!

egroup: [lhc-machinelearning-wg](#)

website: [iml.cern.ch](#)