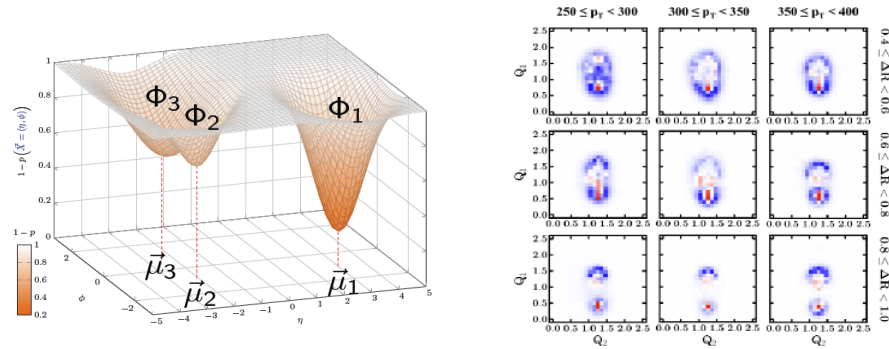


Machine Learning Physics Jamboree



Ariel Schwartzman
SLAC

Machine learning in jet physics

- **Machine learning is widely used at hadron colliders**
 - **Use physical intuition to reduce information down to a small set of input variables into a multivariate discriminant**
 - Do not make use of all information available
 - Require the invention of variables that capture the main features of signals and backgrounds
 - No guidance when the optimal performance achieved
- **A new approach:**
 - Connect the fields of jets and jet substructure with machine learning, computer vision, statistics, and information theory
 - Apply advanced machine learning and pattern recognition methods to to **analyze LHC data in new ways**
 - Use machine learning to **understand the physics within jets**, to help design more powerful analysis methods

Stanford Data Science Initiative (SDSI)

<https://sdsi.stanford.edu>

Stanford | Stanford Data Science Initiative



Funded Research Projects
in Data Science

OCTOBER 2015

Stanford | Stanford Data Science Initiative

Physics Event Reconstruction at the Large Hadron Collider



Data scientists and physicists both work with numbers. But they seldom work together. Stanford's Ariel Schwartzman, a particle physicist, and Lester Mackey, a professor of statistics, think that collaborating will give better understanding of the information that is being produced by the Large Hadron Collider (the LHC), the huge particle accelerator in Geneva, Switzerland. It might even help find new subatomic particles.

By applying big data analysis techniques to the petabytes of data that the LHC generates every year, the scientists expect to be able to more accurately identify and differentiate particles like W, Z, and Higgs bosons, as well as top quarks. Bosons and top quarks are important to study because they are predicted by many models of new physics, such as supersymmetry, or models with extra dimensions of space.

ML in HEP

Higgs challenge  **the HiggsML challenge**
 May to September 2014
 When High Energy Physics meets Machine Learning



info to participate and compete : <https://www.kaggle.com/c/higgs-boson>

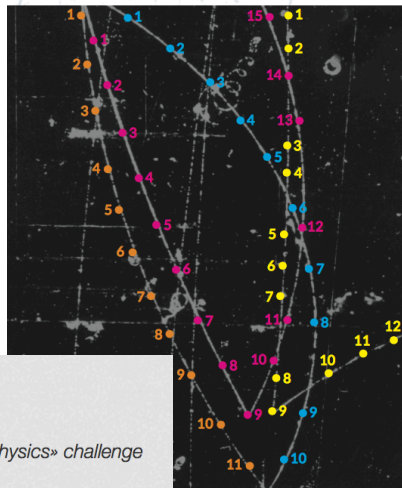


Connecting The Dots 2015

A Workshop on Pattern Recognition in Sparsely Sampled Data
 The Berkeley Experimental Particle Physics Center Workshop Series

WHEN: February 9–11, 2015
 WHERE: University of California, Berkeley and Berkeley Lab

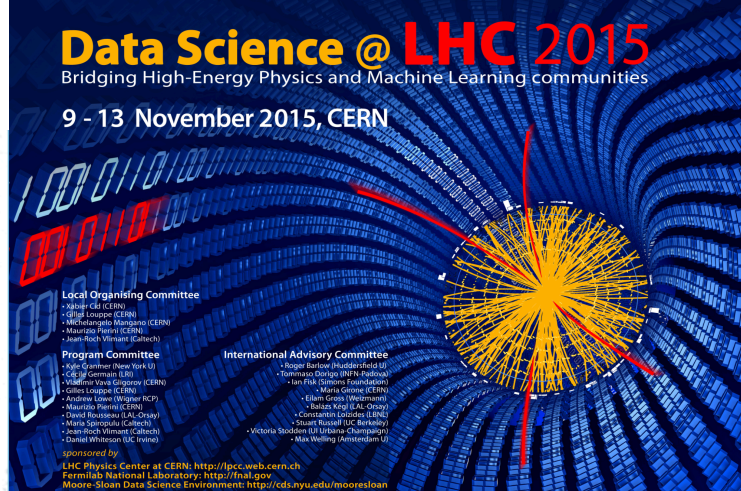
Motivated by the problem of charged particle reconstruction in particle physics experiments, the workshop will focus more generally on pattern recognition in sparsely sampled data. The goal of the workshop is to bring together researchers inclusively, across a variety of disciplines, in hopes that common solutions or new directions may be identified for the greater benefit.



ALEPH Workshop @ NIPS 2015

Applying (machine) Learning to Experimental Physics (ALEPH) and «Flavours of Physics» challenge

Data Science @ LHC 2015
 Bridging High-Energy Physics and Machine Learning communities
 9 - 13 November 2015, CERN



Local Organising Committee

- Xéar O'Leary (CERN)
- Cristof Louppe (CERN)
- Michelangelo Mangano (CERN)
- Maurizio Pippi (CERN)
- Jennifer Wang (CERN)

Program Committee

- Kyle Cranmer (New York U)
- Cole Gersmuth (CERN)
- Valentina Vasy (CERN)
- Andrew Lowe (Wigner RCP)
- Marcin Ptaszyński (CERN)
- David Rousseau (LAL Orsay)
- Maria Spiropulu (CERN)
- Jean-Philippe Brunet (CERN)
- Daniel Whittaker (UC Irvine)

International Advisory Committee

- Roger Barlow (Huddersfield U)
- Tommaso Dorigo (INFN Padova)
- Jan Filip (Silesian Foundation)
- Marijn Giesse (CERN)
- Eitan Gross (Westminster)
- Brian Kilg (U of Oxford)
- Constantin Lidzba (LMU)
- Stuart Russell (UC Berkeley)
- Victoria Stodden (U of Illinois-Champaign)
- Max Welling (Amsterdam U)

sponsored by
 LHC Physics Center at CERN: <http://lpc.web.cern.ch>
 Fermilab National Laboratory: <http://fnal.gov>
 Moore-Sloan Data Science Environment: <http://cds.nyu.edu/moore Sloan>

<http://cern.ch/DataScienceLHC2015>

and neuromorphic approaches
 • Applications and performance evaluation

- Local Organizing Committee:**
- Dave Brown (BNL)
 - Maurice Garcia-Sciveres (BNL)
 - Carl Haber (BNL)
 - Boris Heinemann (UC Berkeley/LBNL)
 - Bob Jacobsen (UC Berkeley/LBNL)
 - Simone Pagan-Griso (LBNL)
 - Marjorie Shapiro (UC Berkeley/LBNL)
 - Lauren Tompkins (Stanford/SLAC)
- Scientific Advisory Committee:**
- Paolo Calafura (LBNL)
 - Aaron Dominguez (U of Nebraska)
 - Mariusus Eising (CERN)
 - Rudi Fruhwirth (HEP Vienna)
 - Luciano Ristori (U of Pisa)
 - David Rousseau (LAL Orsay)
 - Andre Schoening (U of Heidelberg)
 - Arvid Schwartzman (SLAC)
 - Mei Shochet (U of Chicago)
 - Laura Waller (UC Berkeley)

<https://indico.physics.fnl.gov/indico/conferenceDisplay.py?confid=149>



SDSI LHC projects


- Three main frontiers:
 - **Learning discriminative, interpretable representations for jet tagging:**
 - Can we learn more discriminative representations of jets?
 - What do these representations teach us about physics?
 - Goal: improve classification and physical understanding with learned jet representations
 - Jet images (computer vision), deep learning, probabilistic models (fuzzy jets)
 - **Quantifying sub-optimality**
 - For a given representation of a jet and a given family of classifiers, what is the best achievable classification performance?
 - Information-theoretic lower bounds
 - **Suppressing pileup:**
 - Can we learn to identify pileup particles and remove their influence?
 - Probabilistic models to capture correlations structure among particles, regression algorithms for pileup corrections, and image processing noise reduction methods

CS229 and CS221 Projects

- **Color flow tagging**
 - Bonner Lenny, Guy Amdur, Anton Apostolatos
- **Jet charge**
 - Dylan Rueter, Krishna Soni
- **b-quark jet tagging**
 - Qi Zeng, Zihao Jiang
- **Pileup suppression**
 - Aviv Cukierman,
- **Pileup subtraction**
 - Vein Kong, Jiakun Li, Yujia Zhang

Jamboree goals

- **Informal discussions about current projects, insights gained, lessons learned, future ideas and directions**
- **Identify connections with other areas of physics**
 - Particle physics theory, neutrinos, accelerator physics, ...
- **Identify possible new “grand challenges”**
- **Informal workshop: please participate!**

- 13:00 - 13:20 **Introduction 20'**
Speaker: Ariel Gustavo Schwartzman (SLAC National Accelerator Laboratory (US))
  MLPJ_ariel.pdf
- 13:30 - 13:50 **Machine learning with probabilistic models for jet physics 20'**
Speaker: Ben Nachman (SLAC National Accelerator Laboratory (US))
- 14:00 - 14:20 **Machine learning with computer vision for jet physics 20'**
Speakers: Ben Nachman (SLAC National Accelerator Laboratory (US)), Michael Aaron Kagan (SLAC National Accelerator Laboratory (US)), Luke Percival De Oliveira (SLAC National Accelerator Laboratory (US))
- 14:30 - 14:50 **Machine learning with recurrent neural networks for b-quark identification 20'**
Speakers: Qi Zeng (SLAC National Accelerator Laboratory (US)), Michael Aaron Kagan (SLAC National Accelerator Laboratory (US)), Zihao Jiang (Stanford University (US))
- 15:00 - 15:20 **Measuring color flow with machine learning 20'**
Speakers: Bronner Lenny (Computer Science Department, Stanford University), Guy Amdur (Computer Science Department, Stanford University), Anton Apostolatos (Computer Science Department, Stanford University)
- 15:30 - 16:00 **Coffee break**
- 16:00 - 16:20 **Exploring the Limits of Classification Accuracy 20'**
Speakers: Carolyn Kim (Computer Science Department, Stanford University), Lester Mackey (Statistics Department, Stanford University)
- 16:30 - 16:50 **Using artificial intelligence to suppress pileup 20'**
Speaker: Aviv Ruben Cukierman (SLAC National Accelerator Laboratory (US))
- 17:00 - 17:20 **Pileup subtraction using regression methods 20'**
Speakers: Vein Kong (Electrical Engineering Department, Stanford University), Jiakun Li (Electrical Engineering Department, Stanford University), Yujia Zhang (Electrical Engineering Department, Stanford University)
  pileup_presentatio...
- 17:30 - 17:50 **Concluding remarks 20'**
Speakers: Ben Nachman (SLAC National Accelerator Laboratory (US)), Michael Aaron Kagan (SLAC National Accelerator Laboratory (US))