Contribution ID: **202**   Type: **Oral**

Boosted Decision Tree Reweighter

Tuesday 19 January 2016 16:10 (25 minutes)

Machine learning tools are commonly used in high energy physics (HEP) nowadays.
In most cases, those are classification models based on ANN or BDT which are used to select the "signal" events from data. These classification models are usually trained using Monte Carlo (MC) simulated events.

A frequently used method in HEP analyses is reweighting of MC to reduce the discrepancy between real processes and simulation. Typically this is done via so-called "histogram division" approach. While being very simple, this method has strong limitations in the applications. Recently classification ML tools were successfully applied to this problem [1]. Also in sociology ML-based survey reweighting is used to reduce non-respose bias [2].

In my talk I will present the novel method of reweighting, a modification of BDT algorithm, which alters the procedures of boosting and decision tree building.
This method outperforms known reweighting approaches and makes it possible to reweight dozen of variables. When compared on the same problems, it requires less data.

The other part of my talk is devoted to proper usage of reweighting in physical analysis, in particular, to correctly measuring the quality of reweighting.

[1] Martschei, D., et al. "Advanced event reweighting using multivariate analysis." Journal of Physics: Conference Series. Vol. 368. No. 1. IOP Publishing, 2012.

[2] Kizilcec, R. "Reducing non-response bias with survey reweighting: Applications for online learning researchers." Proceedings of the first ACM conference on Learning @ scale conference. ACM, 2014.

Author: ROGOZHNIKOV, Aleksei (Yandex School of Data Analysis (RU))

Presenter: ROGOZHNIKOV, Aleksei (Yandex School of Data Analysis (RU))

Session Classification: Track 2

Track Classification: Data Analysis - Algorithms and Tools