



Contribution ID : 148

Type : Poster

Deep Learning Method for Inferring Cause of Data Anomalies

Tuesday, 22 August 2017 16:30 (15)

Daily operation of a large-scale experiment is a resource consuming task, particularly from perspectives of routine data quality monitoring. Typically, data comes from different channels (subdetectors or other subsystems) and the global quality of data depends on the performance of each channel. In this work, we consider the problem of prediction which channel has been affected by anomalies in the detector behaviour.

We introduce a generic deep learning model and prove, that, under reasonable assumptions, the model learns to identify 'channels' affected by an anomaly. Such model could be used for data quality manager cross-check and assistance and identifying good channels in anomalous data samples.

The main novelty of the method is that the model does not require ground truth labels for each channel, only global flag is used. This effectively distinguishes the model from classical classification methods.

Evaluation of the method on data collected by the CERN CMS experiment is presented.

Primary author(s) : BORISYAK, Maxim (Yandex School of Data Analysis (RU))

Co-author(s) : DERKACH, Denis (Yandex School of Data Analysis (RU)); USTYUZHANIN, Andrey (Yandex School of Data Analysis (RU)); KOVAL, Olga (Yandex); RATNIKOV, Fedor (Yandex School of Data Analysis (RU))

Presenter(s) : RATNIKOV, Fedor (Yandex School of Data Analysis (RU))

Session Classification : Poster Session

Track Classification : Track 2: Data Analysis - Algorithms and Tools