Tools for data visualization and quality validation

Tutorials

Introduction presentation - N-Dimensional analysis pipeline O(20 min)

Jupyter tutorials:

Part 1. (1 h)

- Data visualization (AliDrawStyle, AliPainter and CSS styles)
- Materialized views (AliExternalInfo)

Part 2. (1 h)

- Selection of good quality data (N-dimension analysis, TMVA)
- Data set comparison based on parametrizations

Tutorials using Jupyter notebook:

- first attempt for Alice tutorial
- tutorial code in git-hub (currently in AliRoot github)
    - https://github.com/alisw/AliRoot/tree/master/STAT/Notebooks

- **In the future (I) assume Jupyter notebook (local or on SWAN) to be more used for the troubleshouting, resp for the period performance comparison with user selected data inputs**

Ndimensional analysis pipeline (0)

Overview of tools developed for the QA, performance and calibration monitoring and for the N-dimensional tune on data MC

- Developped within AliRoot
- Planned to be independent package for RUN12 and Run3 analysis
- see latest version of presentations:
  https://indico.cern.ch/event/686140/contributions/3011435/attachments/1653608/2646159/MultiDimensionalVisualization_O2meeting22052018.pdf

Tools for experts, but also for “standard analysis”

- expert providing aggregated information for “analyzers”

Strategy - experts indicates possible problems (MC/Anchor mismatch) and provides recipe - analysis should show “sensitivity” in physics observable

Ndimensional analysis pipeline ()

Detector conditions and corresponding performance change is space and time

- Not all effects described in the MC
  - distortion fluctuation, pile-up, ion tail
  - → correlated efficiency loss, correlated dEdx bias

Recipes under preparation:

- **Quality dependent run list**
- **Quality dependent time series**
- See examples in presentation to follow

Strategy:

Experts indicates possible problems (MC/Anchor mismatch) and provides recipe - analysis should show “sensitivity” in physics observable

Tutorials

Tutorials in form of the Jupyter C++ Notebooks

New AliRoot release with ROOT6

- v5-09-34-01_ROOT6-1

CVMFS ROOT6+Jupyter setup distribution not (yet) fully operational

- laptop and SWAN setup not operational

Tutorial to be running using AliRoot/AliPhysics from afs according instructions in JIRA:

- https://alice.its.cern.ch/jira/browse/ATO-448
- see description part
    - starting servers on lxplus7 using predefined port

Running tutorial from afs (temporary solution)

Recipe to load environment - should work for everybody

It is similar as for the cvmfs installation for alice users

- Choose port to work on, e.g:

```
export JupyPORT=8899
```

- example login as a tpcdrop user

```
ssh -Y -L 127.0.0.1:$JupyPORT:127.0.0.1:$JupyPORT tpcdrop@lxplus7.cern.ch
```

- source environment form user mivanov

```
export JupyPORT=8899
export ALIBUILD_WORK_DIR=/afs/cern.ch/work/m/mivanov/alicesw/sw
export WORKON_HOME="/afs/cern.ch/user/m/mivanov/.virtualenvs/"
export AliExternalInfoCache="/afs/cern.ch/work/m/mivanov/AliExternalInfoCache"
source virtualenvwrapper.sh
workon our_new_env
/afs/cern.ch/user/m/mivanov/.virtualenvs/our_new_env/bin/alienv enter AliPhysics/latest-master-root6
export AliRoot_SRC="$ALIBUILD_WORK_DIR/../AliRoot"
export AliPhysics_SRC="$ALIBUILD_WORK_DIR/../AliPhysics"
```

- downolad tutorials

```
cd your/turial/directory
rsync -avzt  $AliRoot_SRC/STAT/Notebooks .
cd Notebooks
```

- run jupyter

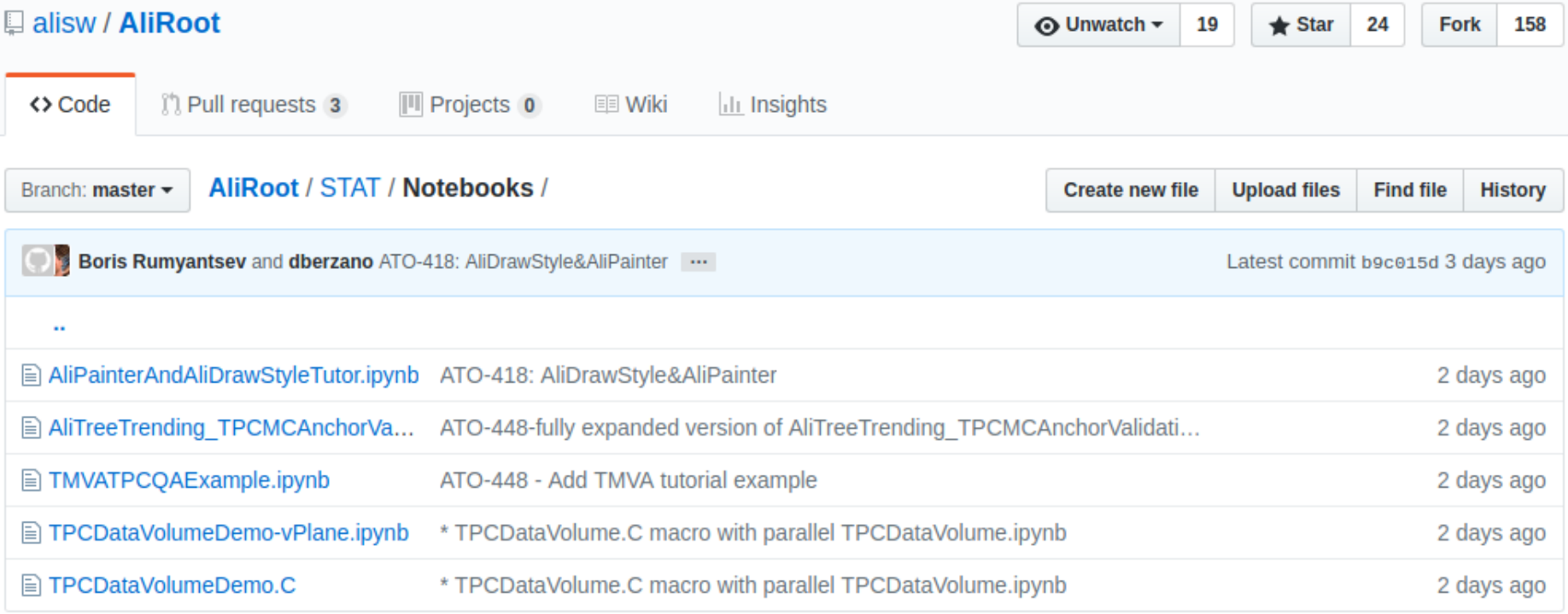
```
jupyter notebook --no-browser --ip=127.0.0.1 --port=$JupyPORT
```

- if running sucessfull open http as indicated in the log bellow: e.g

```
[I 19:30:04.060 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 19:30:04.062 NotebookApp]

    Copy/paste this URL into your browser when you connect for the first time,
    to login with a token:
        http://127.0.0.1:8898/?
token=cef64e43b25560e89ad4f64f5111dd452d5baf46f77e5576&token=cef64e43b25560e89ad4f64f5111dd452d5baf46f77e5576
```

Tutorial list.



4 demo tutorials prepared planned to be shown today

- More to come to demonstrate full functionality of NDimension pipeline

1.1) Data visualization (AliDrawStyle, AliPainter and CSS styles)

See:

https://github.com/alisw/AliRoot/blob/master/STAT/Notebooks/AliPainterAndAliDrawStyleTutor.ipynb

AliDrawStyle - CSS style instead of ROOT TStatyle as an analogue to CSS in HTML

- separation of concerns
- separate styling code (configuration file) and drawing
- possibility to apply re-apply different styles for the same data using CSS functionality
  - presentation style, publication style, QA style
  - http queries
  - working with array of primitives

AliPainter:

- Canvas support
- CSS support
- THn drawing/slicing/fitting
  - interface insipired by ROOT TH and Python

1.2) Materialized view - AliExternalInfo

https://github.com/alisw/AliRoot/blob/master/STAT/Notebooks/TPCDataVolumeDemo-vPlane.ipynb

AliExternalInfo class to interface set of materialized views (root trees/tables)

- See presentation slides 17-25

https://indico.cern.ch/event/686140/contributions/3011435/attachments/1653608/2646159/MultiDimensionalVisualization_O2meeting22052018.pdf

Root tree based interface for information query form different data sources:

- Extension on top of the standard root (in classes described in presentation)
- Drawing, support for metadata (e.g variable description, axis,title description per data ….)
- Set of predefined data inputs
  - QA.<subsystem>, QA.<rawDetector>, Lobook.<xxx>, <Monalisa> …
  - User defined data sources options
- modifying configuration file
  - adding custom input as friend tree (e.g DCS sensors exprted as an tree)
  - used often for TPC calibration fitting, trending, troubleshouting

Example demo:

- query TPC data volume and fit it as function of rate
- Test: do the same studies using TRD infromation

2.1) Selection of good quality data (N-dimension analysis, TMVA)

Notebook in github:

- https://github.com/alisw/AliRoot/blob/master/STAT/Notebooks/TMVATPCQAExample.ipynb

Demo usage of the information from the TMVA interface fitting TPC QA variables

- TPC QA variable (resoluation, efficiency, separationPower) as function of explanatory variables (interaction rate, bz, gain)
- Load TMVA interface function
- Load tree and defining derived information (TTree aliases) and metadata
- CacheTree input variables to tree format usable by TMVA
- Register example methods used for regression
- Emulation of the bootstrap - training repeated several time
- Load array of regression -used later in the array regression evaluation (mean, median, rms)
- Example draw/outlier queries

2.2 Data set comparison based on parametrizations

Jupyter notebook:

- https://github.com/alisw/AliRoot/blob/master/STAT/Notebooks/AliTreeTrending_TPCMCAnchorValidation_Demo.ipynb

Goal

- **Compare MC with the Anchored data period.** Only some part of the functionality as defined in the macro:AliPhysics_SRC/PWGPP/TPC/macros/tpcMCValidation.C+

Alarms definition

- compare data with expectation - "invariants" defiend differnces, ratio
- hierarchy of alarms used

Algorithm to show

- Load input data
- Define alarm aliases in InitTPCMCValidation
- Alarms TPC/ITS/TRD specific
- Example- Redefine some alarms
- Make Status plot draw example using AliTreeTrending
- Make example plot using AliTreeTrending
- In particular case of the TPC MC/Anchor trending ("LHC15k1a1","passMC","LHC15o", "pass3_lowIR_pidfix") outlier detected in DCA
- outlier bacause of differnt DCA for B+ and B-

Future usage:

- Jupyter notebook to be used for troubleshouting