



Contribution ID: 211

Type: Poster

RelMon: A General Approach to QA, Validation and Physics Analysis through Comparison of large Sets of Histograms

Thursday, May 24, 2012 1:30 PM (4h 45m)

The estimation of the compatibility of large amounts of histogram pairs is a recurrent problem in High Energy Physics. The issue is common to several different areas, from software quality monitoring to data certification, preservation and analysis. Given two sets of histograms, it is very important to be able to scrutinize the outcome of several goodness of fit tests, obtain a clear answer about the overall compatibility, easily spot the single anomalies and directly access the concerned histogram pairs. This procedure must be automated in order to reduce the human workload, therefore improving the process of identification of differences which is usually carried out by a trained human mind. Some solutions to this problem have been proposed, but they are experiment specific. RelMon depends only on ROOT and offers several goodness of fit tests (e.g. Chi-squared or Kolmogorov-Smirnov). It produces highly readable web reports, in which aggregations of the comparisons rankings are available as well as all the plots of the single histogram overlays. The comparison procedure is fully automatic and scales smoothly towards ensembles of millions of histograms. Examples of RelMon utilisation within the regular workflows of the CMS collaboration and the advantages therewith obtained are described. Its interplay with the data quality monitoring infrastructure is illustrated as well as its role in the QA of the event reconstruction code, its integration in the CMS software release cycle process, CMS user data analysis and dataset validation.

Summary

The estimation of the compatibility of large amounts of histogram pairs is a recurrent problem in High Energy Physics. The issue is common to several different areas, from software quality monitoring to data certification, preservation and analysis. Given two sets of histograms, it is very important to be able to scrutinize the outcome of several goodness of fit tests, obtain a clear answer about the overall compatibility, easily spot the single anomalies and directly access the concerned histogram pairs. This procedure must be automated in order to reduce the human workload, therefore improving the process of identification of differences which is usually carried out by a trained human mind. Some solutions to this problem have been proposed, but they are experiment specific. RelMon depends only on ROOT and offers several goodness of fit tests (e.g. Chi-squared or Kolmogorov-Smirnov). It produces highly readable web reports, in which aggregations of the comparisons rankings are available as well as all the plots of the single histogram overlays. The comparison procedure is fully automatic and scales smoothly towards ensembles of millions of histograms. Examples of RelMon usage within the regular workflows of the CMS collaboration and the advantages therewith obtained are described. Its interplay with the data quality monitoring infrastructure is illustrated as well as its role in the QA of the event reconstruction code, its integration in the CMS software release cycle process, CMS user data analysis and dataset validation.

Primary author: PIPARO, Danilo (CERN)

Presenter: PIPARO, Danilo (CERN)

Session Classification: Poster Session

Track Classification: Software Engineering, Data Stores and Databases (track 5)