18. Fast classifier-based goodness of fit test for online data quality monitoring

Author(s)

Name	Institution	Mail Address	Social Contacts
Gaia Grosso	INFN Padova, University of Padova, CERN	gaia.grosso@cern.ch	N/A
Nicolò Lai	University of Padova	nicolo.lai@studenti.unipd.it	N/A
Marco Letizia	MaLGa-DIBRIS, Università di Genvoa e INFN, Sezione di Genova	marco.letizia@edu.unige.it	N/A
Jacopo Pazzini	INFN Padova, University of Padova	jacopo.pazzini@unipd.it	N/A
Marco Rando	MaLGa-DIBRIS, Università di Genvoa	marco.rando@edu.unige.it	N/A
Lorenzo Rosasco	MaLGa-DIBRIS, Università di Genvoa, IIT Genova, MIT	Lorenzo.Rosasco@unige.it	N/A
Andrea Wulzer	IFAE, ICREA	andrea.wulzer@cern.ch	N/A
Marco Zanetti	INFN Padova, University of Padova	marco.zanetti@cern.ch	N/A

How to Obtain Support

Mail	marco.letizia@edu.unige.it
Social	N/A

General Information

ML/DL Technologies	Classification, Logistic Regression, Kernel Methods
Science Fields	High Energy Physics
Difficulty	Medium
Language	English
Type	Fully annotated, runnable

Software and Tools

Programming Language	Python
ML Toolset	Pytorch
Additional libraries	Falkon
Suggested Environments	Python virtual environment available at main repository

Needed datasets

Data Creator	INFN Legnaro laboratories
Data Type	Real acquisition
Data Size	39 MB
Data Source	https://doi.org/10.5281/zenodo.7128223

Short Description of the Use Case

A machine learning approach for monitoring particle detectors in real-time. The goal is to assess the compatibility of incoming experimental data with a reference dataset, characterising the data behaviour under normal circumstances, via a likelihood-ratio hypothesis test. The model is based on a modern implementation of kernel methods, nonparametric algorithms that can learn any continuous function given enough data. The resulting approach is efficient and agnostic to the type of anomaly that may be present in the data. Our study demonstrates the effectiveness of this strategy on multivariate data from drift tube chamber muon detectors.

How to execute it

Code to reproduce the experiments in [1] with an illustrative Jupyter notebook are available at https://github.com/FalkonHEP/DQM.

References

[1] G. Grosso, N. Lai, M. Letizia, J. Pazzini, M. Rando, L. Rosasco, A. Wulzer, M. Zanetti, *Fast kernel methods for Data Quality Monitoring as a goodness-of-fit test,* [2303.05413].

[2] M. Letizia, G. Losapio, M. Rando, G. Grosso, A. Wulzer, M. Pierini, M. Zanetti, L. Rosasco, *Learning new physics efficiently with nonparametric methods,* Eur. Phys. J. C 82 (2022) 879 [2204.02317].

Presentation made on 15 May 2023 : https://agenda.infn.it/event/35822/contributions/198103/attachments/105222/147811/MarcoLetizia_MLINFN_May2023.pdf