Data Quality Monitoring (DQM) and Data Certification (DC) At CMS

- Goal: Ensure proper detector functionality and data integrity [1].
- DQM data is divided into runs (continuous data taking period).
- A run is divided into units of ~23 seconds called lumisections.
- Shifters look at many per-run monitoring element (ME) plots that describe what happened during the run.
- Challenges:
  - Vulnerable to human error
  - Labor intensive
  - Limited granularity with current tools

Overcoming DQM Challenges With ML

- Tools are being developed to integrate ML into the DQM workflow to aid shifters in DC [2].
- Models will flag anomalous lumisections for shifters to review.
- Types of models:
  - 1D autoencoders
  - 2D residual neural networks
  - Non-negative matrix factorization
- New challenges:
  - What data will we use to train these models?
  - How will shifters inspect the problematic lumisections in a convenient way?

Reference Run Ranking (RRR)

- Rank candidate reference runs for a target run by minimizing differences in data-taking conditions.
- Use of PCA for feature reduction and prioritization.
- Offer curated DQM ML training datasets and partially automate reference run selection.

Data Inspection Tools For DIALS

- Development of software tools for data inspection, including interactive visualization tools that utilize the DIALS Python API to fetch DQM data.
- These tools are being incorporated into DIALS to facilitate per-lumisection data inspection for shifters.

Acknowledgements

- NSF award 2111134
- This work has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

Bibliography

[1] https://doi.org/10.1016/j.cernregul.2021-034

1 University of Puerto Rico, Mayagüez