

LEVEL 0 SUMMARY TEMPLATE

Name of student: LECOMTE Pauline

Name of your Level 1: GOLVINDEN Selvina

Source (e.g. scholars.google.com): athena.u-pec.fr

Paper title: AI Enabled Data Quality Monitoring with Hydra

Keywords specific to the paper: data quality monitoring, artificial intelligence, Inception v3, RootSpy, data labeling, deep learning models, image analysis, GlueX.

Summary of the main contributions:

This article is talking about the development and implementation of "Hydra", which is an AI system. The document talks about the challenges of manual monitoring, the advantages of AI automation, the process of data labeling, and the future development goals for Hydra.

In the introduction, the article is explaining the importance of data quality monitoring in high energy and nuclear physics experiments. Indeed, collecting data is very expensive and maintaining data integrity is crucial for the success of scientific projects.

The part 2 of the document explains the traditional data quality monitoring processes employed by the GlueX experiment (scientific project) before the development of Hydra.

First, there was the reliance on shift crews to monitor alarms and ensure data quality. However, a problem occurs since a human couldn't catch every possible issue, and it required a lot of work from, which could result in mistakes or missing some problems.

Second, they use RootSpy for data stream monitoring. This tool is helping to monitor by gathering and displaying statistics from detector systems, mainly in histogram forms. Shift crews were tasked with regularly monitoring these histograms. However, there was no mechanism to ensure comprehensive analysis of these plots by the shift crews, and the monitoring wasn't so efficient because the crew was too busy and not always focused. Histograms generated by RootSpy were archived as PNG images and made accessible for review through webpages and daily emails. This system generated a large volume of data: very challenging to check everything, they had to focus on the main data and the smaller data was forgotten.

Finding certain problems was hard, checking everything by hand took a lot of work, and needing people to ensure data quality showed the need of a faster and more efficient automated system. That's why the Hydra system was developed.

Hydra is an AI system which is in charge of data quality monitoring thanks to the GlueX experiment. The goal is to operate a transition from the traditional human monitored processes to an automated artificial intelligent monitoring. To do so, Hydra will use the Inception V3 network: it's a deep learning tool developed by Google for classifying images with more precision than humans.

To conceptualize Hydra, in order to answer the limits of monitoring depending on human, several steps must be followed. It's hard to label lots of images to train AI. Then, they introduced HydraLabeler, a website tool that makes marking images process more efficient. They built a database for Hydra. The database helps to train AI by storing image links and details, making it easy to find and use them for training and other work. To teach Hydra, they had to prepare data, choose data carefully to avoid bias, well focus on important data to improve Hydra accuracy and efficiency. Finally, Hydra must have been tested by comparing its work to correct answers. Experts check the mistakes to improve how the AI labels things. By testing and fixing errors, the AI gets better at finding problems in data, it's like a virtuous circle

Hydra provides two main forms of feedback: a real-time view of the current status of detectors and a retrospective 24-hour view highlighting any detected issues. These features of Hydra help check data quality better, allow to find and fix problems quickly and make managing data quality more accurate.

Indeed, the goal of using Hydra is to improve the efficiency and accuracy of detecting faults in experimental data. The next goal for Hydra is to be able to make a diagnosis automatically when a problem occurs, instead of only finding it.