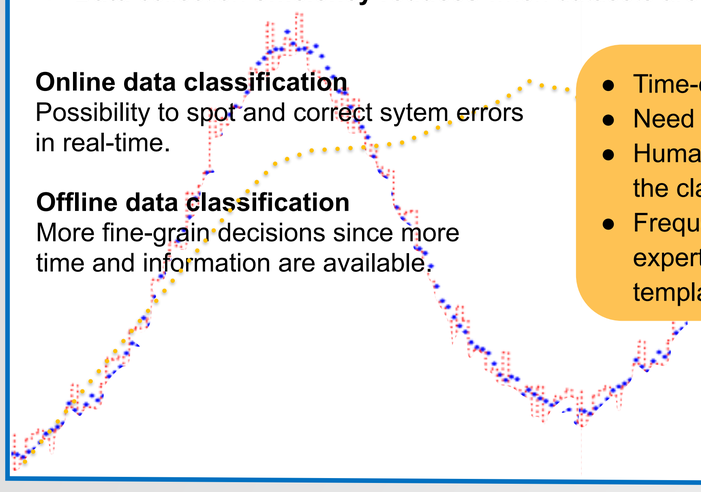
Reinforcement learning for automatic data quality monitoring in HEP experiments

Olivia Jullian Parra, Lorenzo Del Pianta Pérez, Julián García Pardiñas, Suzanne Klaver, Thomas Lehéricy, Maximilian Janisch, Nicola Serra

Motivation

Data quality (DQ) monitoring is a **crucial phase**. However, The **detectors are imperfect** and may **bias** the collected **data.**

We **Need to label properly** each collected dataset **as 'good' or 'bad':**
- **Measurements** are **biased** when datasets are **incorrectly classified** as good.
- **Data** collection **efficiency reduces** when datasets are **incorrectly classified** as bad.



Online data classification
Possibility to spot and correct sytem errors in real-time.

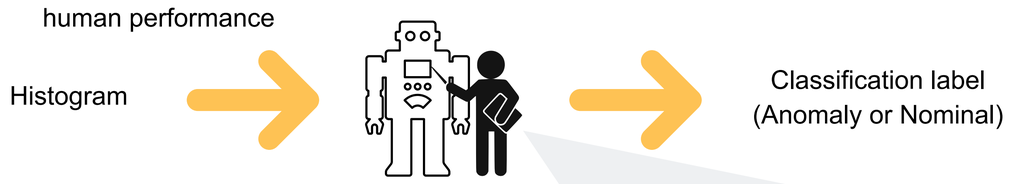
Offline data classification
More fine-grain decisions since more time and information are available.

- Time-consuming task
- Need for appropriate training
- Human errors lead to inaccuracies in the classification
- Frequent changes make it hard for experts to keep updated reference templates

Reinforcement Learning solution

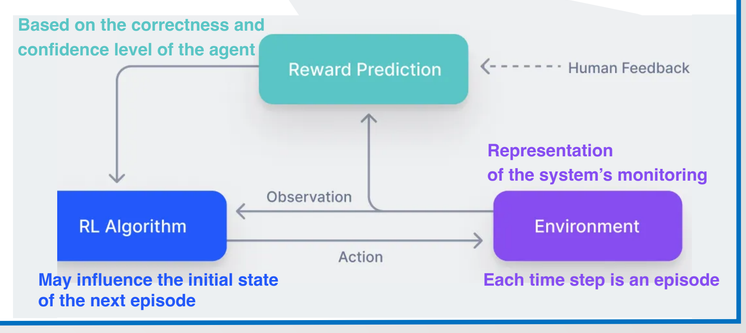
Implementation of an algorithm based on **Reinforcement Learning (RLHF) with human feedback**, which learns at the same time as humans do.

- The learning agent models an 'expert shifter'.
- Mutual feedback between human and machine can hopefully lead to super-human performance



Main objectives

1. **Adapt to changing conditions**
2. **Improve** shifter **accuracy**
3. **Reduce shifters intervention**
4. **Enhance** the **reliability** of the system



Independent studies

Offline

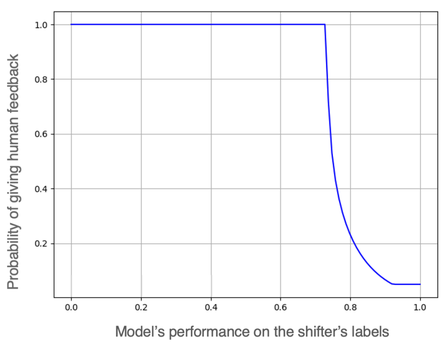
Histograms are presented sequentially in a dataset and the initial state of the episode is independent of the agent's actions.

Aims of the study

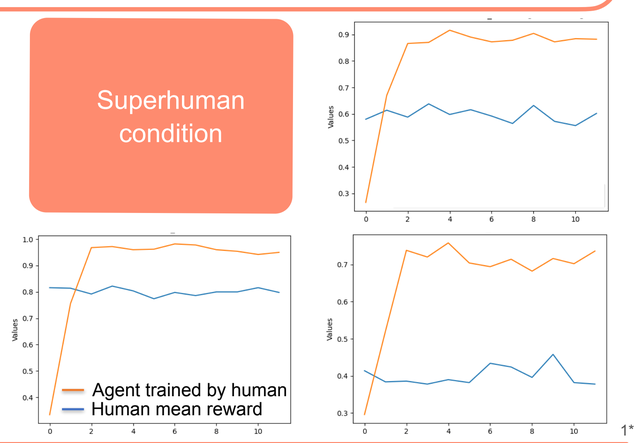
- Classify as nominal
- Classify as anomaly

Improvement of the current shifter's accuracy when detecting system changes. For that, the **agent receives constant feedback**.

Shifter's feedback function

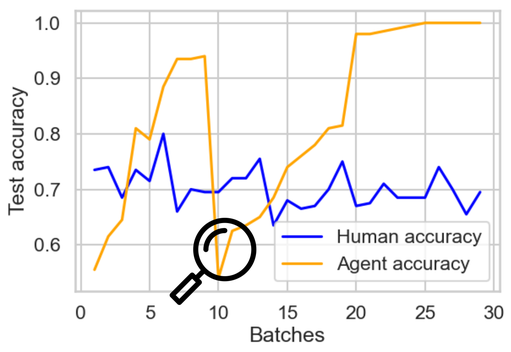


Superhuman condition

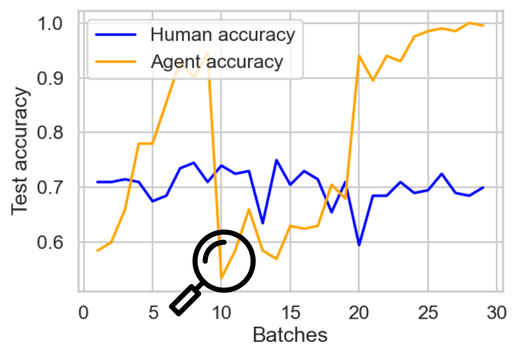


1*

Dynamic evolution



Anomaly type changing over time



Nominal type changing over time

Changing event

*Probability of human failure: 0.3

Online

Histograms are presented sequentially in real-time and the initial state of the episode depends on the agent's actions.

Aims of the study

Maximize the current shifter accuracy while **reducing** its **interactions** with the agent as least as possible. The **agent knows** when to **incorporate human feedback** into its policy.

- Call the shifter
- Classify as nominal
- Classify as anomaly

Environment setup

- Single episode: vector composed by a set of bins (histogram)
- The **initial state is influenced** by the agent
- If there is a new **changing condition undetected** by the agent, this **histogram persists** until its detection

Balancing performance with **number of calls to the shifter and classification accuracy**

Reward scheme shifter check

prediction == groundtruth (either *anomaly* or *ground_truth* label):
- reward the agent for being right
- punish the agent for having called the shifter without a reason

prediction != groundtruth (either *anomaly* or *ground_truth* label):
- reward the agent for being wrong and decide to check -

Reward scheme prediction

we assume that the shifter will solve the problem or check the anomalous histogram every time the agent predicts is an anomaly.

prediction== *ground_truth* label:
- no reward

prediction == *anomaly* label:
- reward the agent if it is right
- punish the agent if it is wrong

Conclusions & future steps

Error	Human		Agent	
	Reward	Acc.	Reward	Acc.
0%	1.00	100%	0.99	99%
10%	0.80	90%	0.95	97.5%
20%	0.60	80%	0.88	94%
30%	0.40	70%	0.76	88%

Ability to **increase** the good-data-collection **efficiency**, learning beyond the average human error.
Superhuman condition achieved!

Interaction between human and machine possible and with clear benefits.

Ability to **adapt** to **slowly-changing experimental conditions**, either anomalies or nominal distribution changes

Future steps:
Capacity to train small datasets with **data augmentation**

Future steps:

Notes & References

1 Study only valid for a static domain where the data taking conditions are not changing (possibility to define a real training scheme in which shifters can see the algorithm's decision in advance and (partially) change their own decision based on that; quantitative study of the convergence of such an approach)

Real-time anomaly detection with RL thesis: https://riunet.upv.es/bitstream/handle/10251/198499/Pianta%20-%20Real-time%20anomaly%20detection%20using%20Reinforcement%20Learning%20at%20LHCb%20CERN%20experiment.pdf?sequence=2&isAllowed=y