Balancing Familiarity and Curiosity in Data Exploration with Deep Reinforcement Learning

Aurélien Personnaz, Sihem Amer-Yahia (CNRS), Laure Berti-Equille (IRD)
Maximilian Fabricius, Srividya Subramanian (MPE)

aiDM workshop,
June 25, 2021

Motivation

- Exploring very large datasets requires to provide user guidance
- Existing works on Exploratory Data Analysis (EDA) focus on SQL operators allowing roll-up and drill-down (ATENA, User Groups)
- Deep Reinforcement Learning appears as a good solution to guided EDA, but most works focus on **roll-ups** and **drill-downs** with simple reward designs
- In RL, the reward defines the incentive leading an agent to achieve a task
- Existing work in data exploration relies on **extrinsic reward**, an objective reward determined by an evaluation of the success of the agent
- RL community has developed the notion of **intrinsic reward**, representing a subjective motivation like curiosity

Goal

- Study the impact of new operators on the training of DRL agents for guided EDA
- Study the impact of reward methods based on balancing extrinsic and intrinsic rewards on the training of DRL agents for guided EDA

Contributions

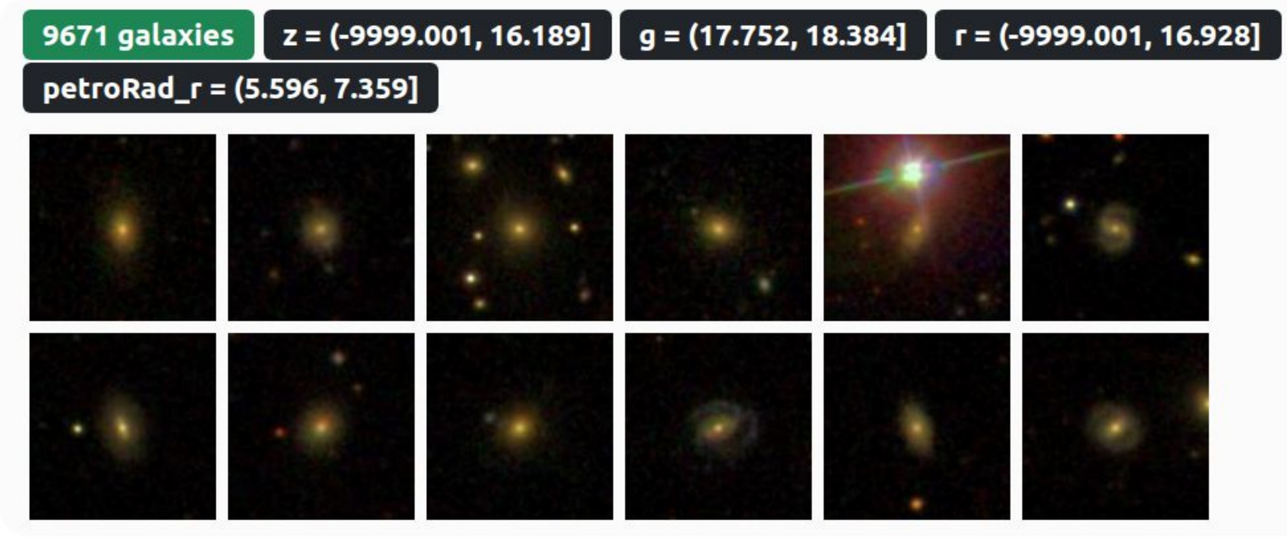
- Propose the **BCF Pipeline Generation Problem** that finds a policy maximizing a combination of extrinsic (familiarity) and intrinsic (curiosity) rewards
- Develop **DORA The Explorer**, a data exploration system that leverages state-of-art A3C curiosity-based learning and expressive data exploration operators
- Run experiments on **real-world SDSS data** (a very large astrophysics dataset), showing that curiosity-based DRL combined with expressive data exploration operators outperforms existing RL and DRL approaches for data exploration

Extrinsic and intrinsic rewards

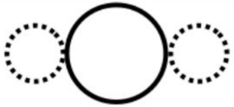
- Extrinsic reward comes from the environment
  - An objective reward determined by an evaluation of the success of the agent
  - Familiarity reward consists in rewarding the agent when it finds some predefined target items
  - Used in many previous works in RL for EDA
  - Limited by the knowledge of the person defining the target items
- Intrinsic reward depends on the agent itself and on its experience
  - Initially defined in N. Chentanez, A. Barto, and S. Singh. 2005. “Intrinsically Motivated Reinforcement Learning”
  - Curiosity reward consists in rewarding the agent when it ventures into new and unexplored states
  - Successful applications to games to compensate a lack or absence of extrinsic reward

An example itemset of galaxies in DORA The Explorer

- An itemset is defined with a conjunction of predicates
- Exploration operators semantics is closed under itemsets

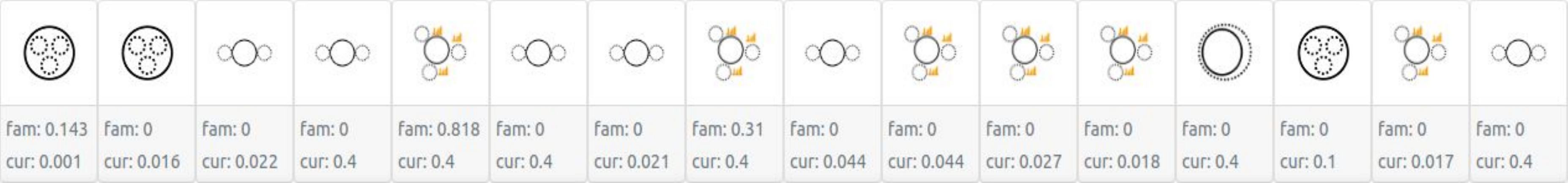


Exploration operators

Operator	RCC8 Formalism [25]		Output description
by-facet(D, A)	`NTPPi`		returns as many subsets of D as there are combinations of values of attributes in A
by-superset(D, k)	`NTPP`		returns the k smallest supersets of input set D (k is application-dependent)
by-distribution(D)	`DC`		returns all sets that are distinct from the input set D and whose attribute value distribution is similar to D
by-neighbors(D, a)	`EC`		returns 2 sets that are distinct from the input set D and that have the previous (smaller) and next (larger) values for attribute a

Exploration operators and pipelines

- EDA session = a pipeline of operators



- We model exploration pipelines as policies trainable by some RL agents
- We formalize intrinsic curiosity reward to complement the usual extrinsic familiarity reward

Familiarity reward

- Familiarity targets are obtained by sampling from classified data in the Galaxy Zoo project
- “Finding” an item in a very large dataset is not trivial, since it can be “drowned” in a big itemset
- We define familiarity as a function of the **concentration ratio** of target objects in an itemset
- The familiarity reward of a state is the sum of the familiarity score of each itemset displayed in this state

$$Familiarity(s_i, T) = \Sigma_{O \in sets(s_i)} \frac{|O \cap T|^2}{|O| \times |T|}$$

Curiosity reward

- Previous works on curiosity applied to games, like Curiosity-driven Exploration by Self-supervised Prediction, use a complex multi-model curiosity module to filter and recognize features of interest
- In DORA The Explorer, every feature can be of potential interest and we can keep track of the states the agent went through
- We keep an occurrence counter for each state the agent goes through and the curiosity reward is inversely proportional to its value

$$Curiosity(s_i) = \frac{1}{Counter_{s_i}}$$

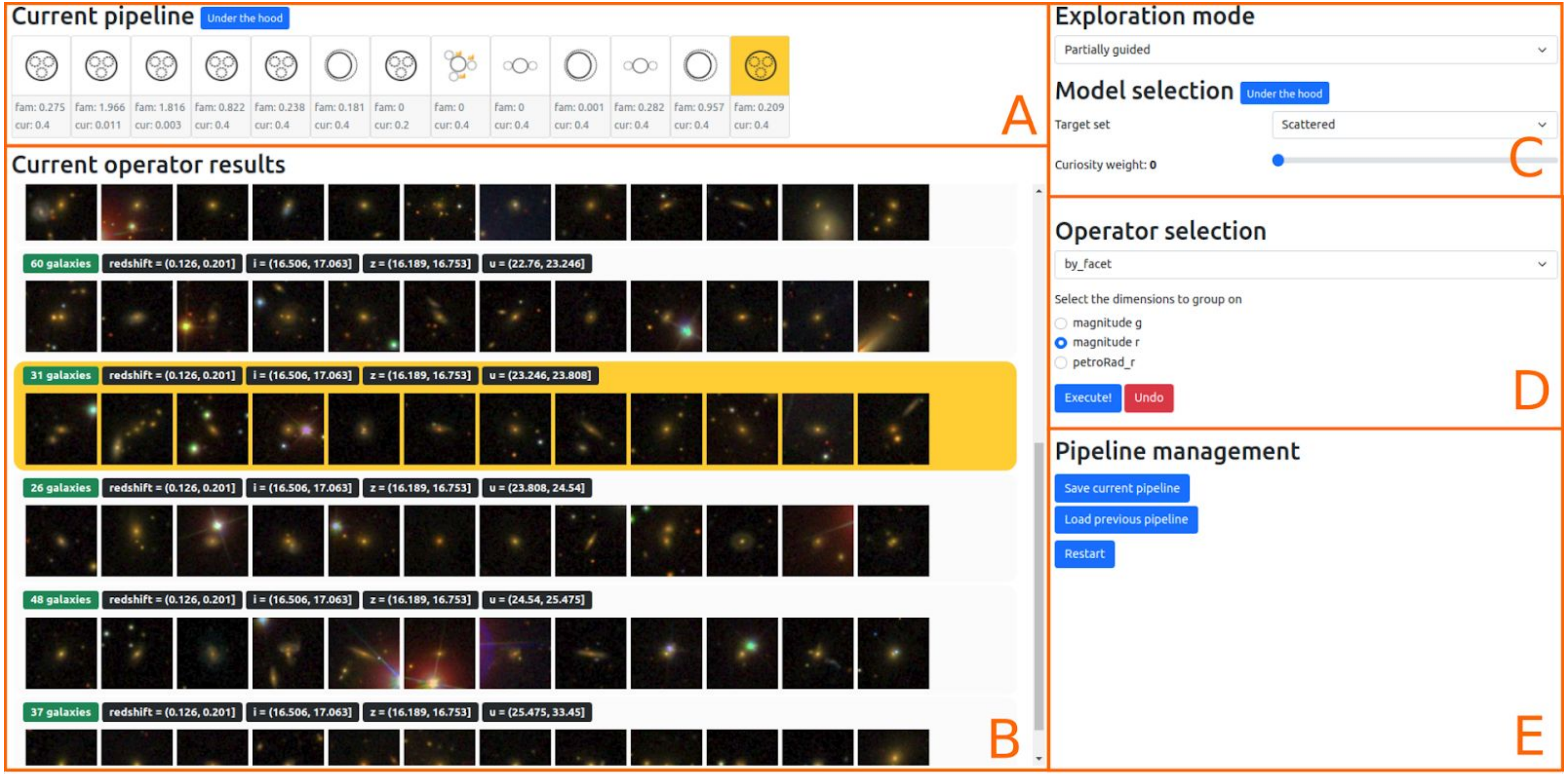
Reward definition

- The reward of applying action e_i on state s_i causing a transition to state s_{i+1} is:

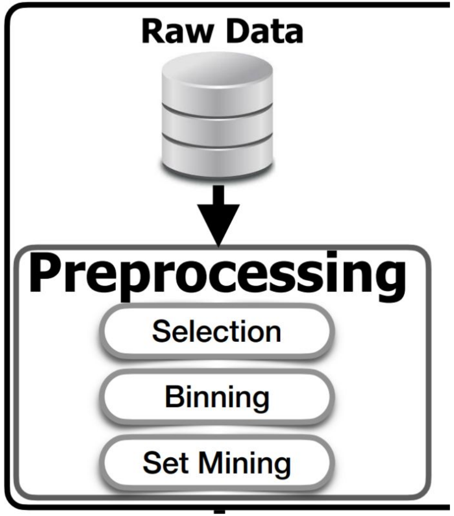
$$R(s_i, e_i, s_{i+1}) = \delta.Familiarity(s_{i+1}, T) + \beta.Curiosity(s_{i+1})$$

- Each agent is trained with predefined weights $\delta + \beta = 1$

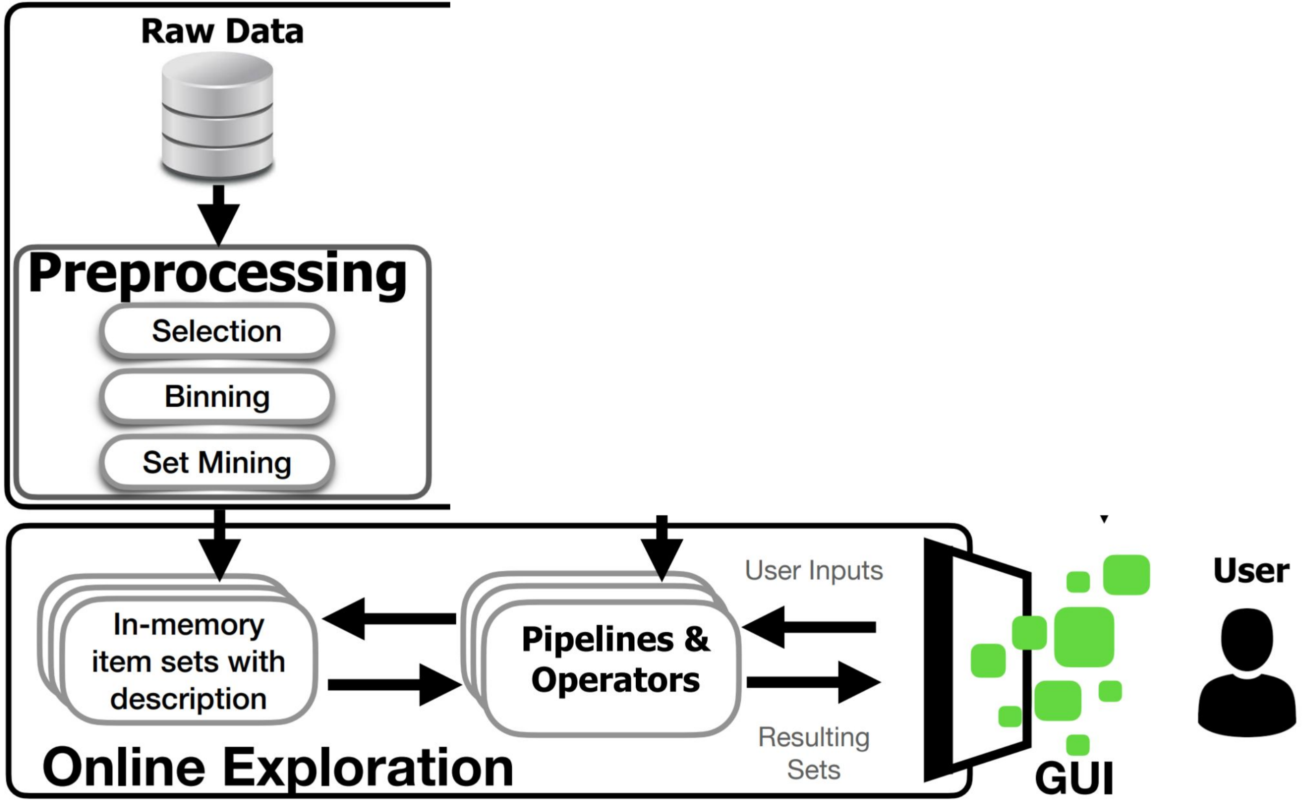
Dora The Explorer



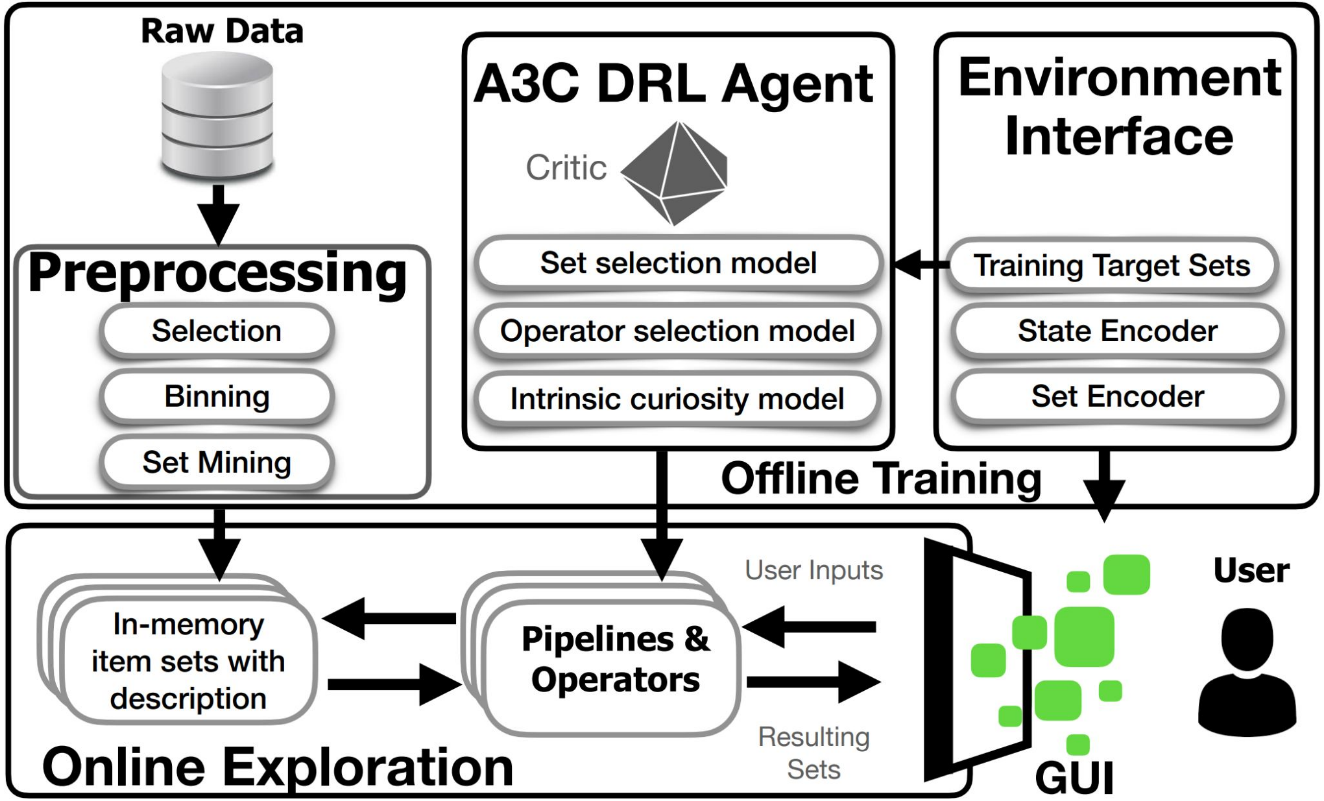
Architecture of Dora The Explorer



Architecture of Dora The Explorer



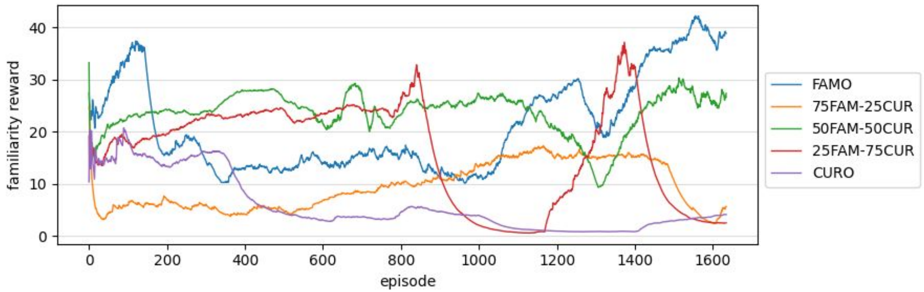
Architecture of Dora The Explorer



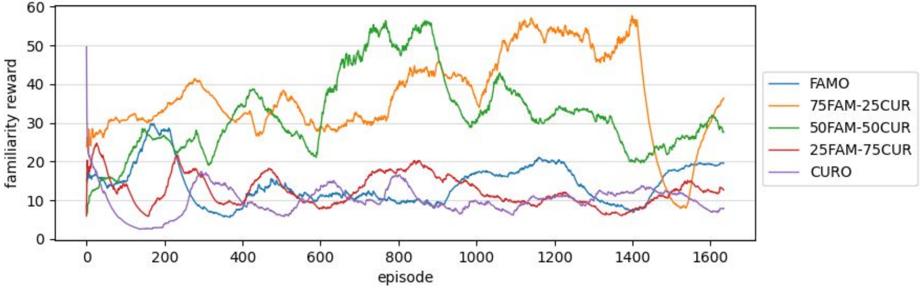
Experiments

- 2.6 million galaxies dataset with 7 attributes
- Comparison of two sets of operators:
  - TRADITIONAL = by_subset + by_superset (drill-down and roll-up)
  - ALL-OPERATOR = TRADITIONAL + by_neighbors + by_distribution
- Comparison of different combinations of extrinsic and intrinsic rewards
  - FAMO for familiarity-only (this mimics exiting data exploration work)
  - CURO for curiosity-only
  - 50FAM-50CUR for 50% familiarity and 50% curiosity
  - 75FAM-25CUR for 75% familiarity and 25% curiosity
  - 25FAM-75CUR for 25% familiarity and 75% curiosity
- We trained the agents for 100 hours, then studied their reward evolution in training and their behavior online

Reward evolution during training

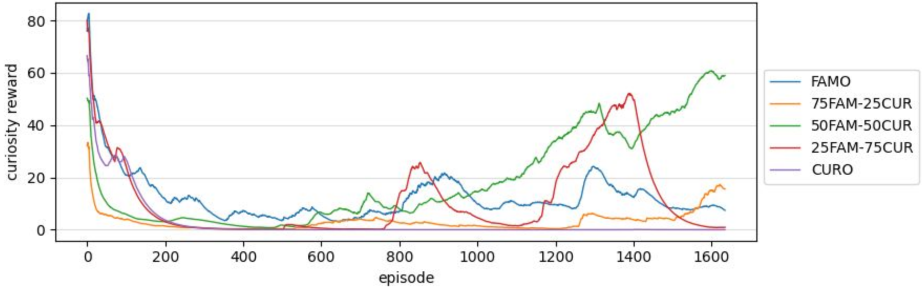


TRADITIONAL

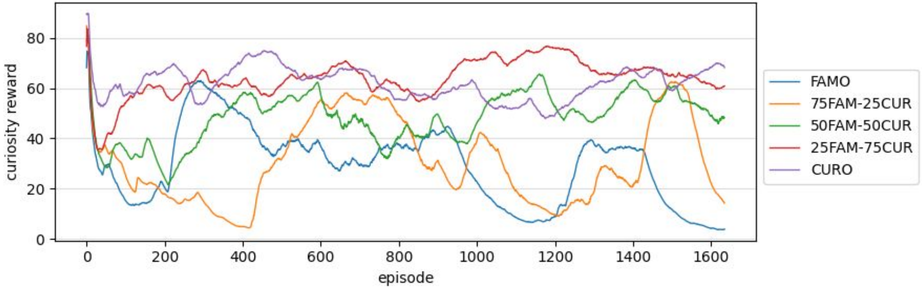


ALL-OPERATOR

Evolution of extrinsic reward (familiarity-based) during training



TRADITIONAL



ALL-OPERATOR

Evolution of intrinsic reward (curiosity-based) during training

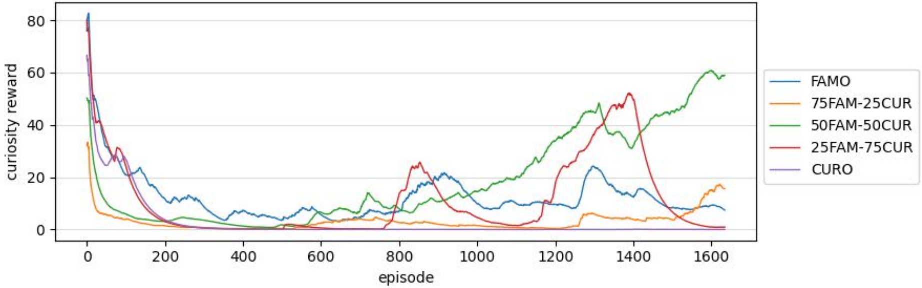
Reward evolution during training

- **Curiosity** reward is difficult to obtain in **TRADITIONAL**

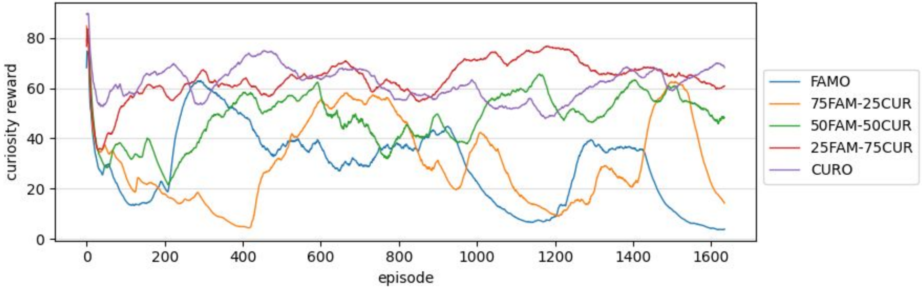
TRADITIONAL

ALL-OPERATOR

Evolution of extrinsic reward (familiarity-based) during training



TRADITIONAL

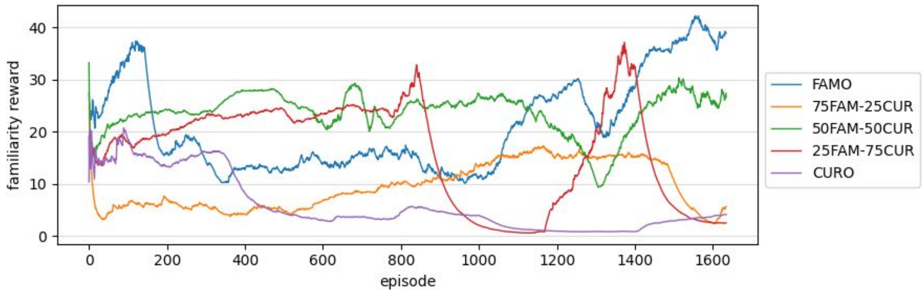


ALL-OPERATOR

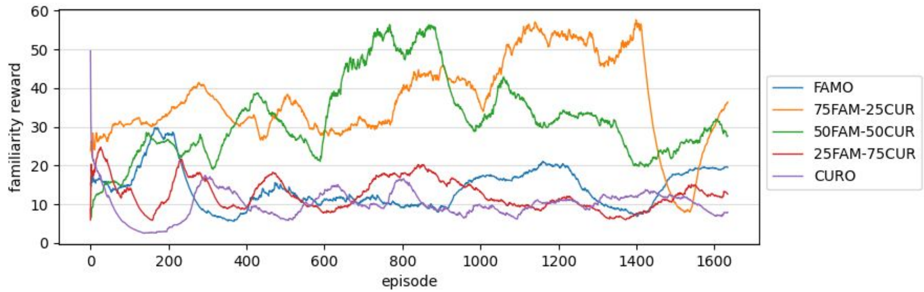
Evolution of intrinsic reward (curiosity-based) during training

Reward evolution during training

- **Curiosity only (CURO)** is not adapted for EDA



TRADITIONAL



ALL-OPERATOR

Evolution of extrinsic reward (familiarity-based) during training

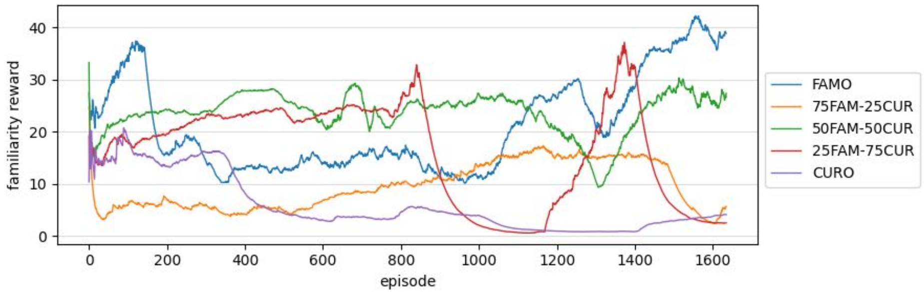
TRADITIONAL

ALL-OPERATOR

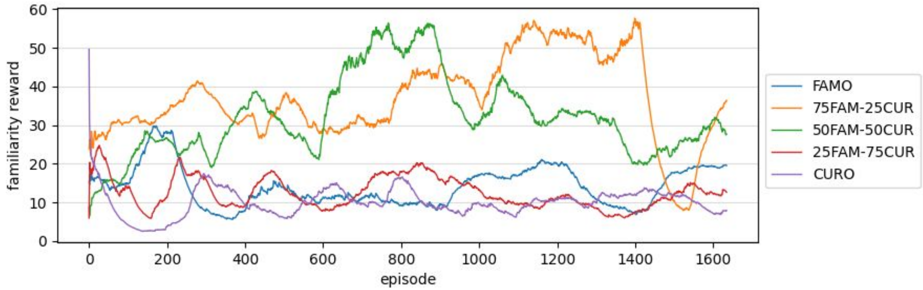
Evolution of intrinsic reward (curiosity-based) during training

Reward evolution during training

- Except for FAMO-TRADITIONAL, the best results are obtained by mixed reward agents

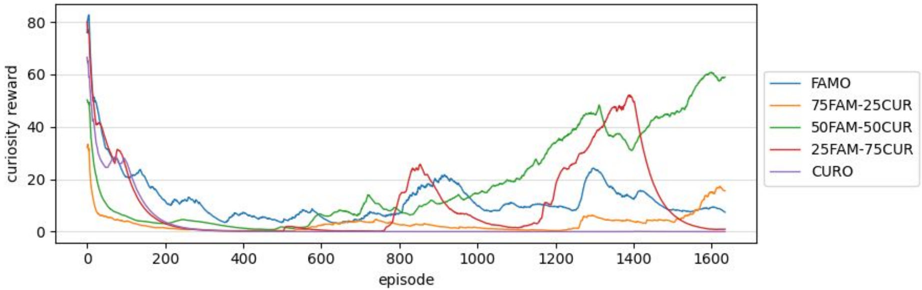


TRADITIONAL

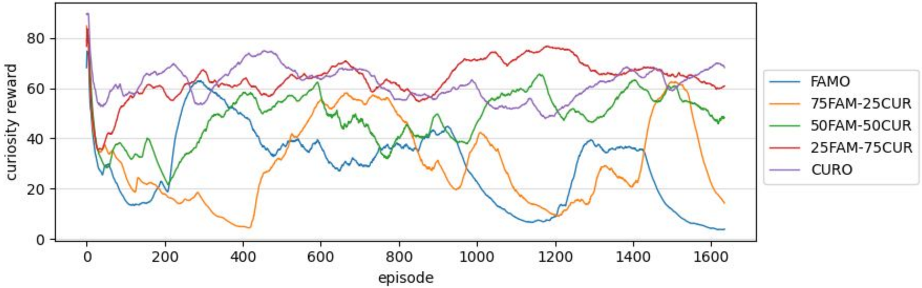


ALL-OPERATOR

Evolution of extrinsic reward (familiarity-based) during training



TRADITIONAL



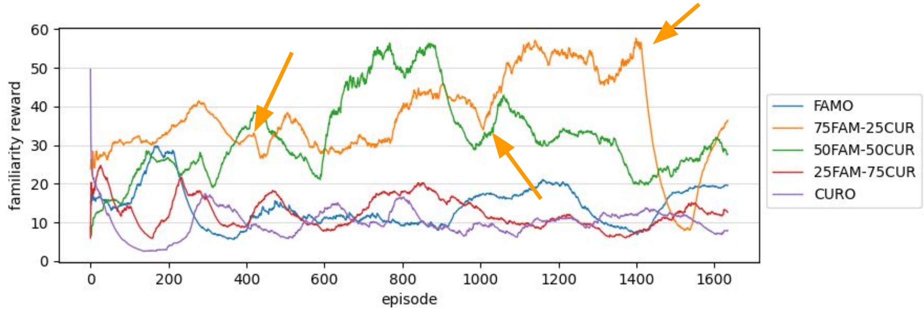
ALL-OPERATOR

Evolution of intrinsic reward (curiosity-based) during training

Reward evolution during training

- We can observe **policies switch** when an agent changes its priority

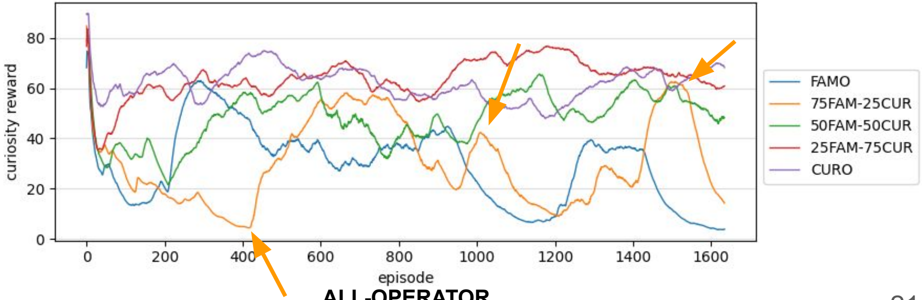
TRADITIONAL



ALL-OPERATOR

Evolution of extrinsic reward (familiarity-based) during training

TRADITIONAL

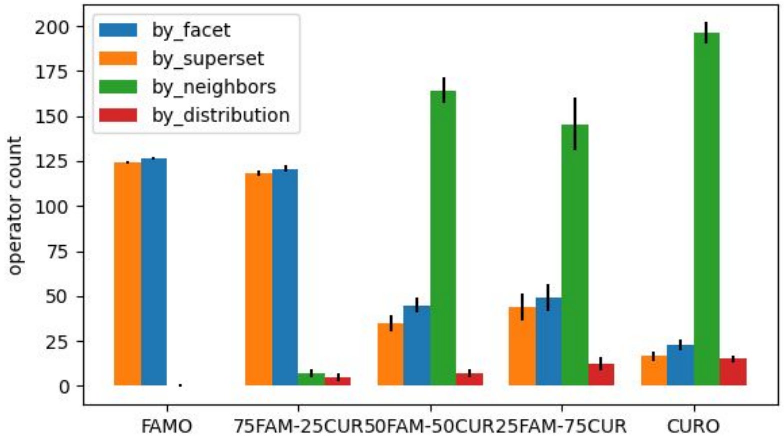


ALL-OPERATOR

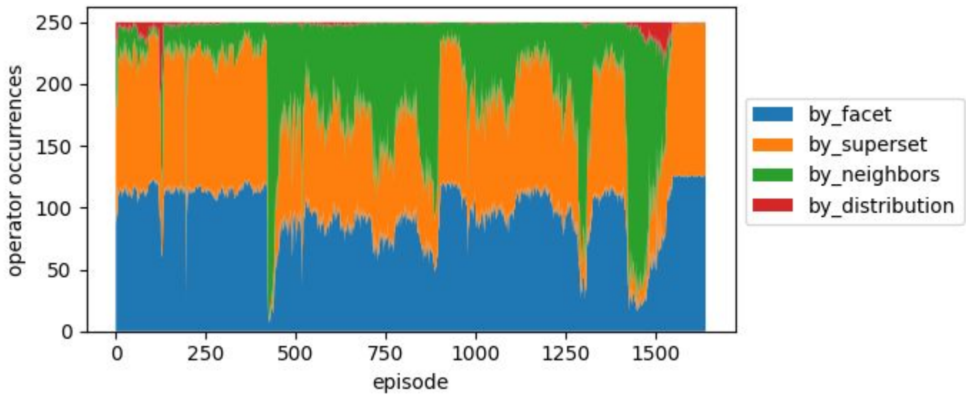
Evolution of intrinsic reward (curiosity-based) during training

Operator usage

- Agent strategies are different and depend on the type of reward they seek
- Mixed reward agents shift their strategies multiple times during training



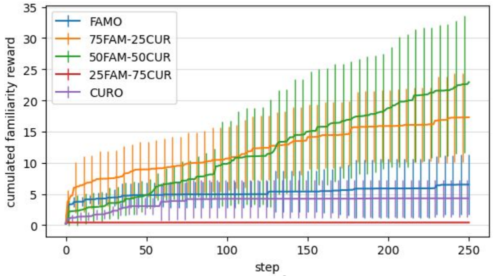
Operator distribution in online pipelines with **ALL-OPERATOR**



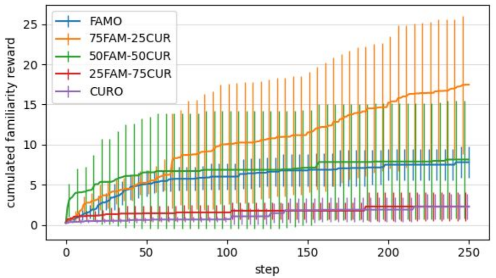
Operator distribution during training for **75FAM-25CUR** with **ALL-OPERATOR**

Reward evolution during the online phase

- Best results are obtained by agents with **mixed-reward**
- **Curiosity** reward is difficult to obtain in **TRADITIONAL**, and easier to obtain in **ALL-OPERATOR**

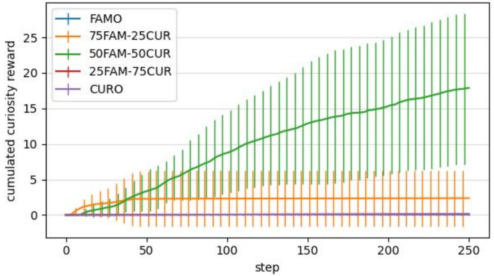


TRADITIONAL

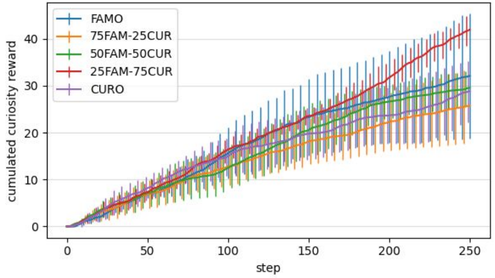


ALL-OPERATOR

Cumulative extrinsic reward (familiarity-based) in online pipelines



TRADITIONAL



ALL-OPERATOR

Cumulative intrinsic reward (curiosity-based) in online pipelines

Summary of experiments

- Unlike in games, a full curiosity-based intrinsic reward is not adapted for EDA
- Importance of optimizing familiarity and curiosity in tandem
  - The highest levels of familiarity reward were reached by agents with some level of curiosity reward
  - When both reward sources are available, the agents tend to shift priority between curiosity- and familiarity-based policies
- Curiosity-based intrinsic reward is easier to produce with ALL-OPERATOR
  - TRADITIONAL limits the agents to set generalization/specialization, while ALL-OPERATOR allows them to reach sets with similar granularity
- Adding new operators benefits data familiarity-driven EDA for agents with a mixed reward
  - Agents learn to choose the most efficient operators to produce the type of reward they seek

Conclusion

- Our framework exploits the interplay between DRL with familiarity and curiosity rewards and expressive data exploration operators
- Future investigations
  - Examine the relation between curiosity/familiarity and the scattering of target objects
  - Investigate the possible roles of user feedback
  - Determine the ideal weights of familiarity and curiosity based on user feedback

Thank you for listening

DORA The Explorer available at:

https://bit.ly/dora-application

Code freely available at:

https://github.com/apersonnaz/rl-guided-galaxy-exploration