Testing and Improving RL Policies via Rule Learning

Ignacio D. Lopez-Miguel, Martin Tappler, Ezio Bartocci

TU Wien, Vienna, Austria

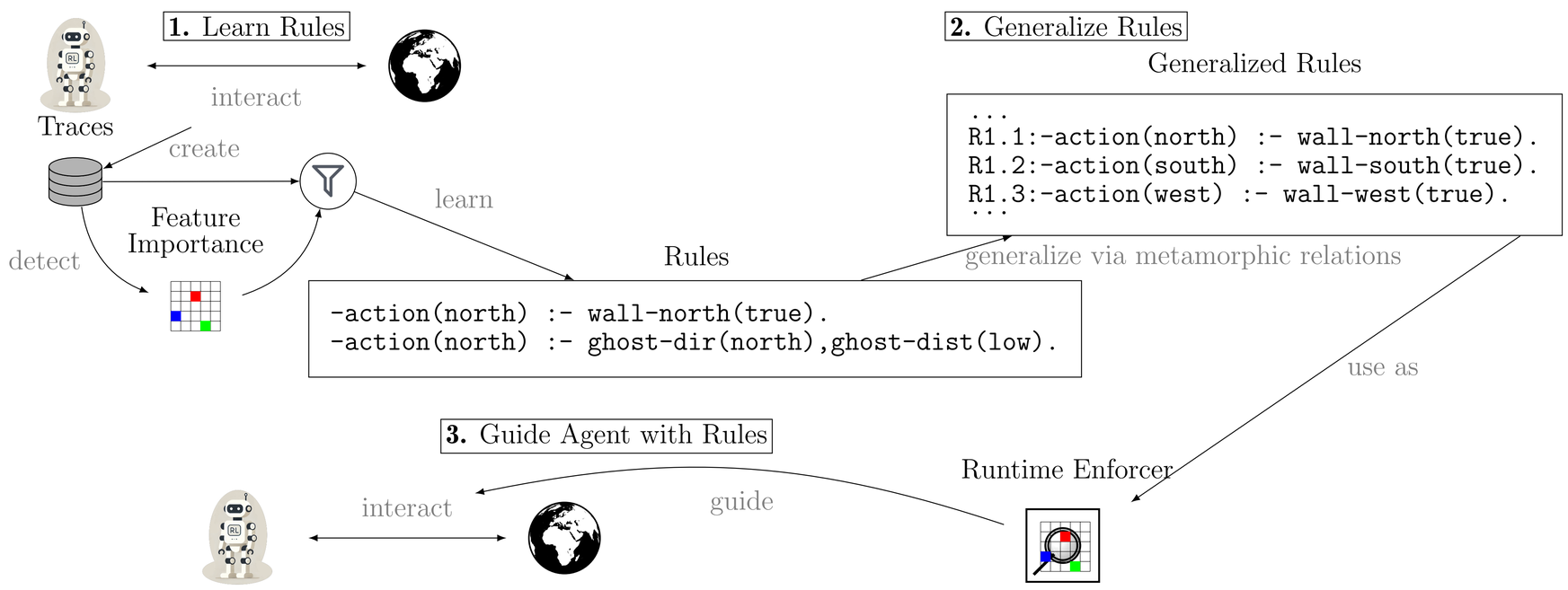
Abstract

Using **domain knowledge** to improve deep RL policies is a current challenge. **LEGIBLE** mines rules from an RL policy, constituting a partially **symbolic representation**. These rules describe which decisions the RL policy makes and which it avoids making. It then **generalizes** the mined rules using domain knowledge. Finally, it evaluates generalized rules to determine which generalizations **improve performance** when enforced. These improvements show weaknesses in the policy, where it has not learned the general rules and thus can be improved by rule guidance. We show the efficacy of our approach by demonstrating that it effectively finds weaknesses, accompanied by **explanations** of these weaknesses in several RL environments.

Goal

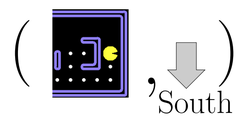
Learn rules to **explain** and **improve** a trained deep RL policy.

Overview



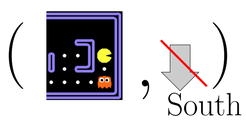
Rules

- Positive rules:
  - $action(a) \leftarrow \bigwedge_i b_i$
  - Perform a if $\bigwedge_i b_i$ holds in current state



$action(south) \leftarrow food\text{-}south(yes)$

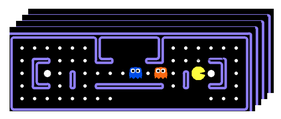
- Negative rules:
  - $-action(a) \leftarrow \bigwedge_i b_i$
  - Avoid a if $\bigwedge_i b_i$ holds in current state



$-action(south) \leftarrow ghostdir\text{-}south(yes) \wedge ghostdistance\text{-}low(yes)$

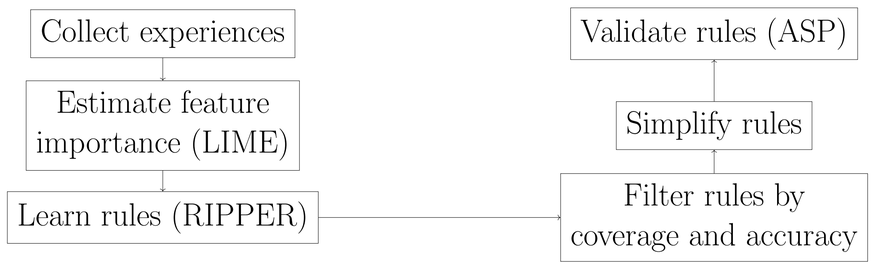
Collect experiences

The agent is executed in the environment and traces are collected. Atable with all these experiences is created and we will create univariate classification problems, where each target is one (positive or negative) action, and the states are the features.



Feature 1	Feature 2	...	Action
yes	no	...	north
no	yes	...	east
...	...	...	...

Workflow to learn rules

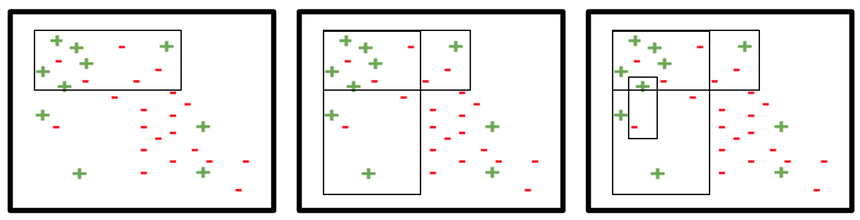


1. christophm.github.io/interpretable-ml-book/lime.html.

Rule learning

Iterative process. Growing rules greedily until no information gain (FOIL) is possible. Then, pruning to reduce overfitting.

$$\text{FOIL}(L) = p_{positive} \cdot \left(\log_2 \frac{p_{positive}}{p_{positive} + n_{positive}} - \log_2 \frac{p_{negative}}{p_{negative} + n_{negative}} \right)$$



2. medium.com/data-science/how-to-perform-explainable-machine-learning-classification-without-any-trees-873db4192c68.

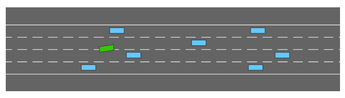
Results

Policy evaluation:

- Enforce generalized rules per learned rule.
- $\sim$ 15% of the learned rules lead to detection of weaknesses in Pac-Man.

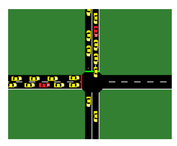
Rule-based improvement:

- Look for the best set of generalized rules.
- Returns increase $\sim \times 4$ for Highway environment.



Extended work

- Inclusion of rules into deep RL policies.



- Mixed continuous-discrete environment with continuous actions.