TECHNISCHE UNIVERSITÄT WIEN
Vienna | Austria

Testing and Improving RL Policies via Rule Learning

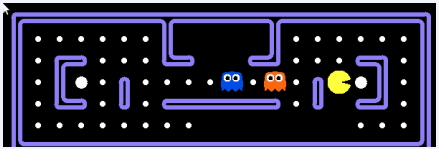
Ignacio D. Lopez-Miguel[1], Martin Tappler[1], and Ezio Bartocci[1]

[1]TU Wien, Vienna, Austria

March 30, 2026

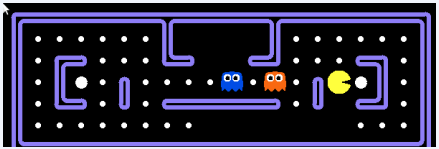
Introduction

Trained agent using RL



Introduction

Trained agent using RL

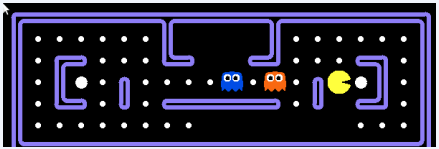


Rule learning

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

Introduction

Trained agent using RL



Rule learning

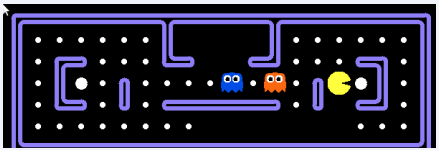
- $action(south) \leftarrow food\text{-}south(yes)$

Rule generalization with domain knowledge

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

Introduction

Trained agent using RL



Rule learning

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

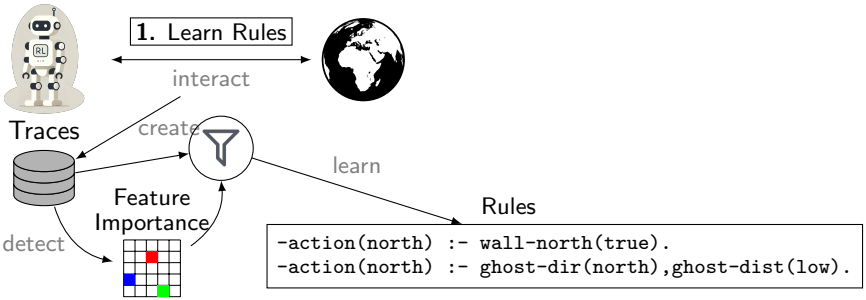
Rule generalization with domain knowledge

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

Goals

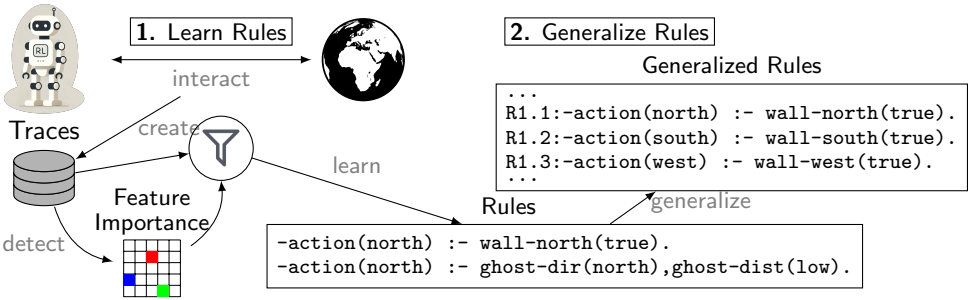
- Explain behavior of RL agents. $\rightarrow$ Rule learning.
- Identify weaknesses in policies (testing). $\rightarrow$ Rule generalization.

Policy evaluation guided by rules



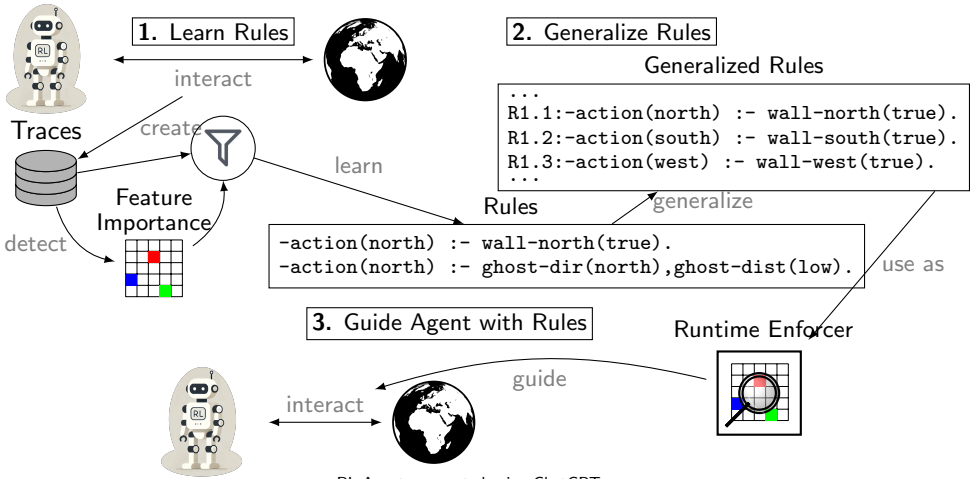
sources: RL Agent: generated using ChatGPT,
World Map: https://pixabay.com/de/vectors/afrika-asien-erde-europa-globus-1299545/

Policy evaluation guided by rules



sources: RL Agent: generated using ChatGPT,
World Map: https://pixabay.com/de/vectors/afrika-asien-erde-europa-globus-1299545/

Policy evaluation guided by rules



sources: RL Agent: generated using ChatGPT,
World Map: https://pixabay.com/de/vectors/afrika-asien-erde-europa-globus-1299545/

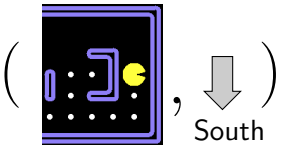
Rule Types

- Positive rules:

- Negative rules:

Rule Types

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

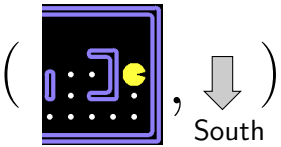


$action(south) \leftarrow food\text{-}south(yes)$

- Negative rules:

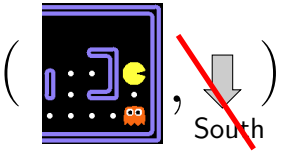
Rule Types

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



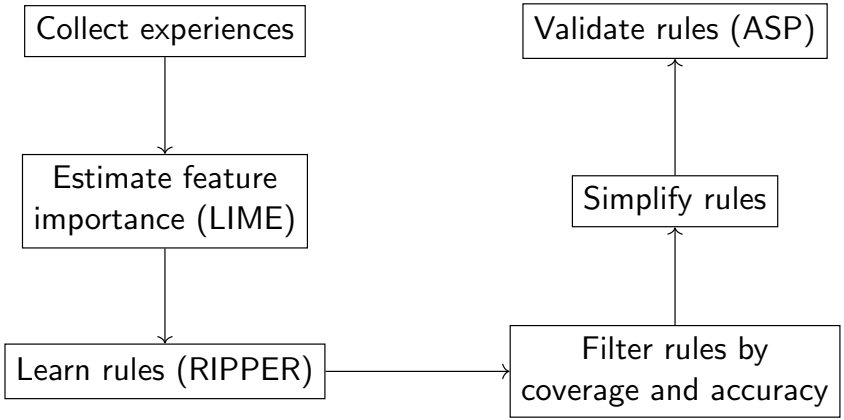
$action(south) \leftarrow food\text{-}south(yes)$

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

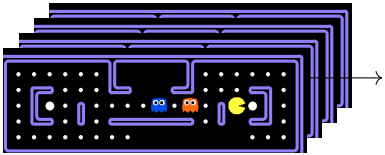


$-action(south) \leftarrow ghostdir\text{-}south(yes)$
$\wedge\ ghostdistance\text{-}low(yes)$

Rule learning



Collect experiences

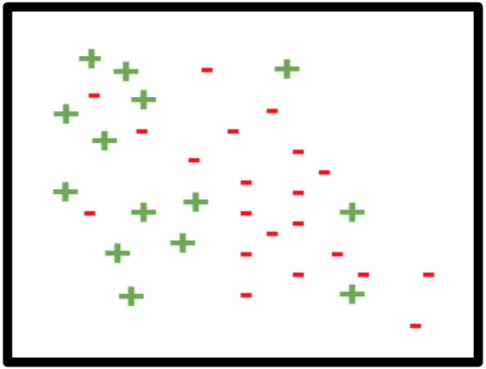


Feature 1	Feature 2	...	Action
yes	no	...	north
no	yes	...	east
...	...	...	...

Learn rules

Action

Ruleset



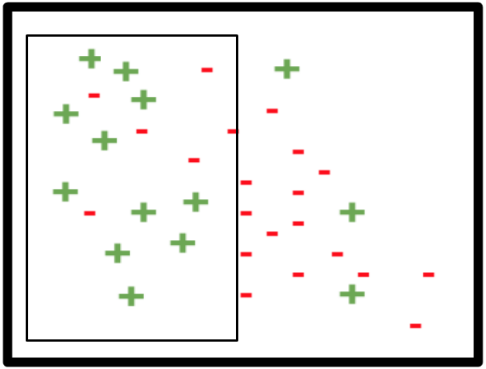
Learn rules

Action

Add conditional

Ruleset

Cond 1



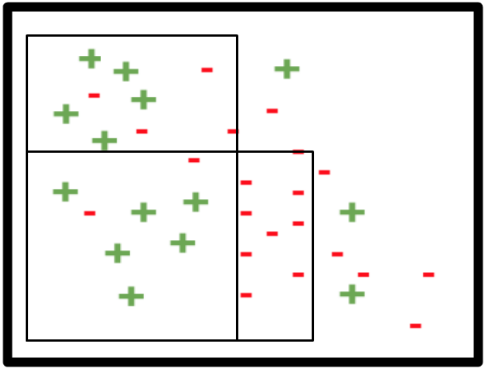
Learn rules

Action

Add conditional

Ruleset

Cond 1 & Cond 2



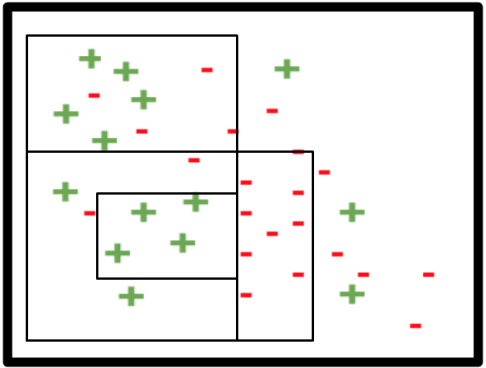
Learn rules

Action

Add conditional

Ruleset

Cond 1 & Cond 2 & Cond 3



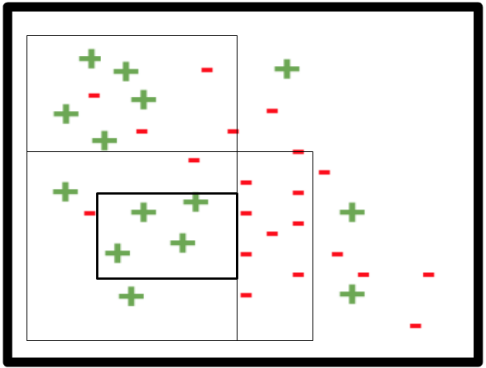
Learn rules

Action

Done growing rule

Ruleset

Cond 1 & Cond 2 & Cond 3



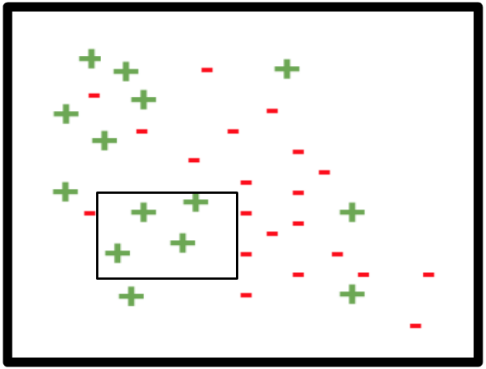
Learn rules

Action

Grown rule

Ruleset

Rule 1



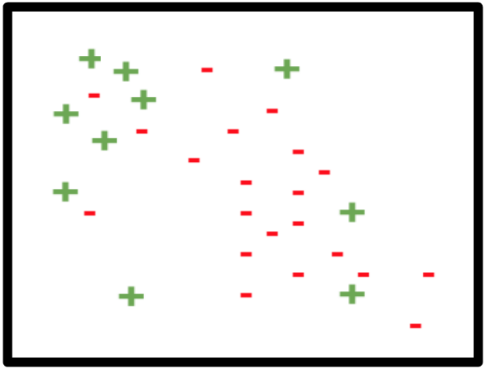
Learn rules

Action

Remove covered examples

Ruleset

Rule 1



Learn rules

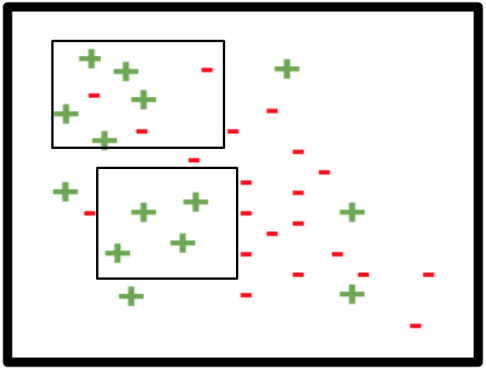
Action

Final model

Ruleset

Rule 1 or

Rule 2



Learn rules

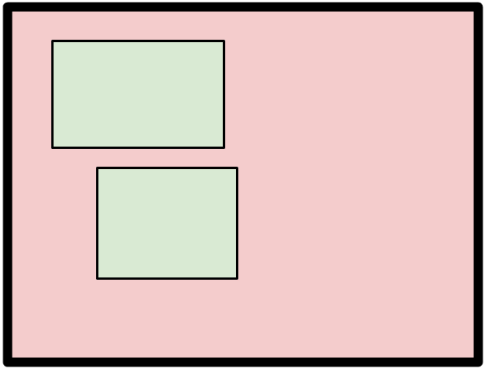
Action

Classified examples

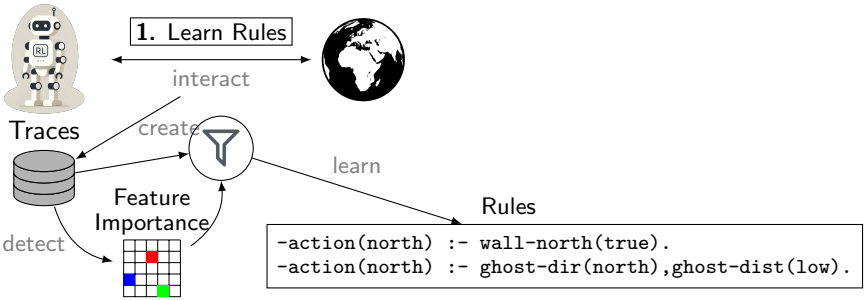
Ruleset

Rule 1 or

Rule 2

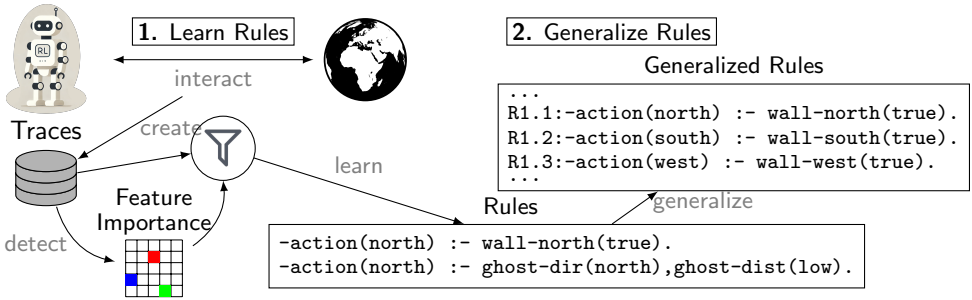


Policy evaluation guided by rules



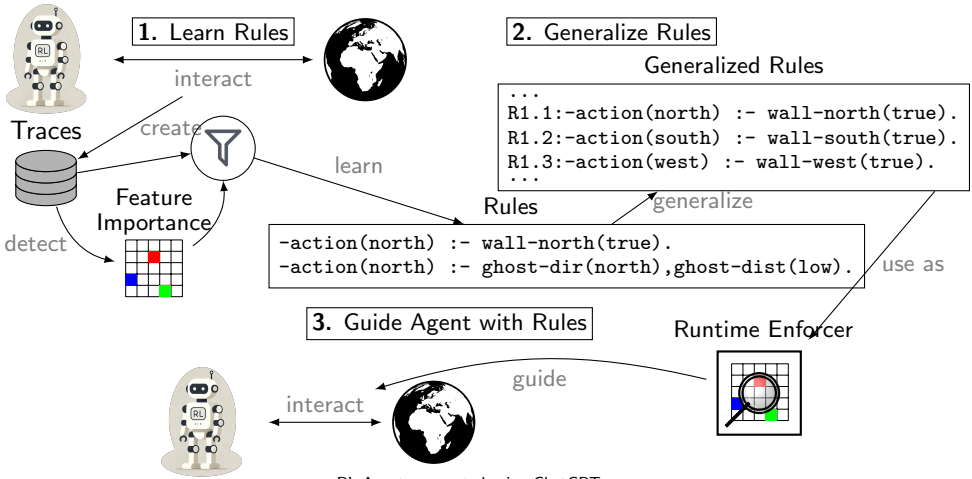
sources: RL Agent: generated using ChatGPT,
World Map: https://pixabay.com/de/vectors/afrika-asien-erde-europa-globus-1299545/

Policy evaluation guided by rules



sources: RL Agent: generated using ChatGPT,
World Map: https://pixabay.com/de/vectors/afrika-asien-erde-europa-globus-1299545/

Policy evaluation guided by rules



sources: RL Agent: generated using ChatGPT,
World Map: https://pixabay.com/de/vectors/afrika-asien-erde-europa-globus-1299545/

Results

Policy evaluation

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

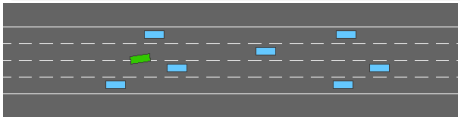
Results

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

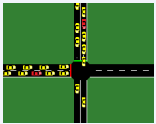
- Look for the best set of generalized rules.
- Returns increase $\sim \times 4$ for Highway environment.



Extended work

On-the-fly Norm Teaching

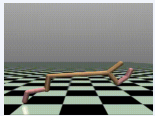
- Inclusion of rules into deep RL policies.



Continuous actions

- Mixed continuous-discrete environment with continuous actions.

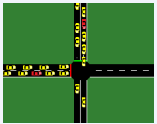




Extended work

On-the-fly Norm Teaching

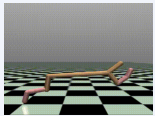
- Inclusion of rules into deep RL policies.



Continuous actions

- Mixed continuous-discrete environment with continuous actions.





Thank you! Questions?