UNIVERSITY
OF AMSTERDAM

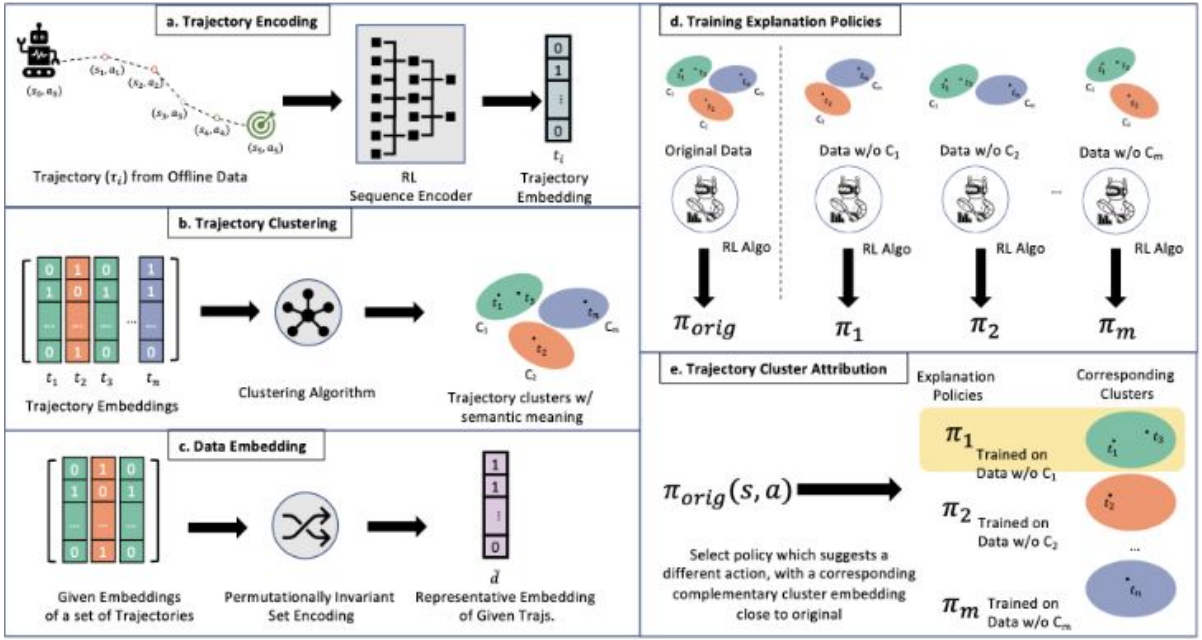
NEURAL INFORMATION
PROCESSING SYSTEMS

‘Explaining RL Decisions with Trajectories’: A Reproducibility Study

Karim Abdel Sadek, Matteo Nulli, Joan Velja, and Jort Vincenti

Motivation & Original Paper

- Previous work focused on salient features of the state of the agent
- **Novelty**:
  Look at trajectories encountered during training by the Offline RL agent
- New framework introduced by the authors [1].



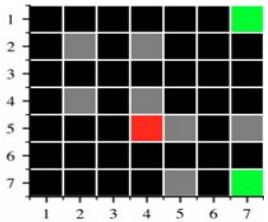
[1] Deshmukh, Shripad Vilasrao, et al. "Explaining RL decisions with trajectories." *arXiv preprint arXiv:2305.04073* (2023).

Claims of the Authors

- Removing Trajectories induces a lower **Initial State Value**
- Clusters present **High Level behaviours**
- **Distant Trajectories** influence Decision of the agents
- Humans correctly identify determinant trajectories

Methodology and Code Setup

Grid-World

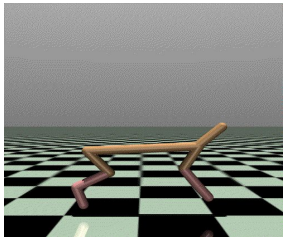


Code ✅

SeaQuest

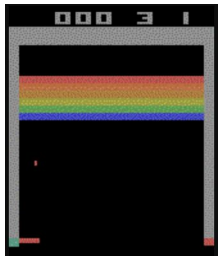
Code ❌

Half-Cheetah



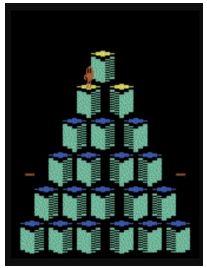
Code ❌

Breakout



Code ❌

Q*Bert

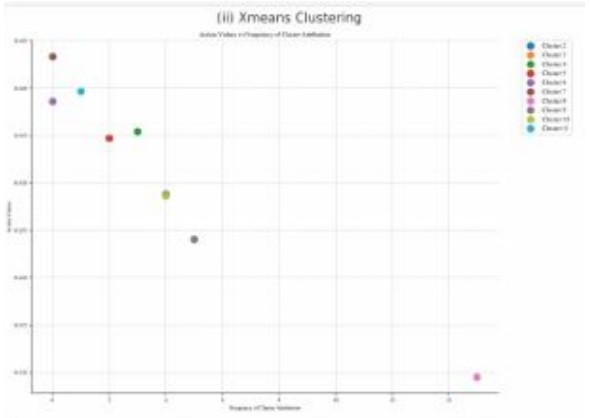


Code ❌

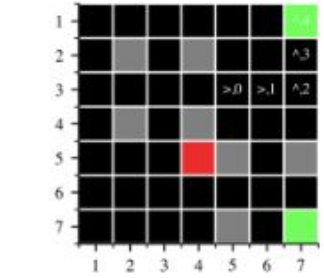
Claim 1:
Removing trajectories reduces ISV

- Removing trajectories reduces Initial State Value.
- Reproducibility: Varied results in different environments, supporting the original claim.
- Extra experiments to see a correlation between ISV and trajectory attribution.

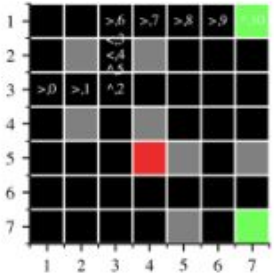
π	$\mathbb{E}[V(s_0)]$	$\mathbb{E}[\lvert\Delta Q_{\pi_{\text{orig}}}(s)\rvert]$	$\mathbb{E}[1(\pi_{\text{orig}}(s) \neq \pi_j(s))]$	$W_{\text{dist}}(d, d_j)$
Mean Clusters (Original Paper)	0.3027	0.0231	0.0821	0.0301
Mean Clusters(Reproduced)	0.3029	0.0230	0.0714	0.1098
$\lvert\Delta\rvert$	0.0002	0.0001	0.0107	0.0797



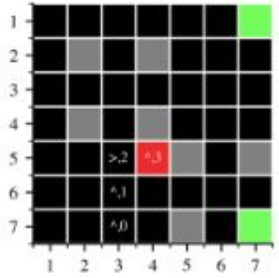
Claim 2:
Clusters have High-Level Behaviours



Achieving Goal in Top right corner - Cluster 1

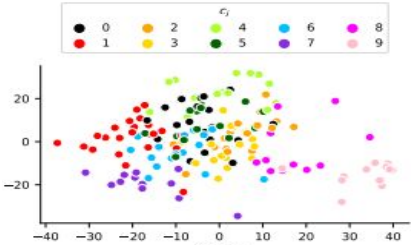


Mid-Grid journey to goal - Cluster 6

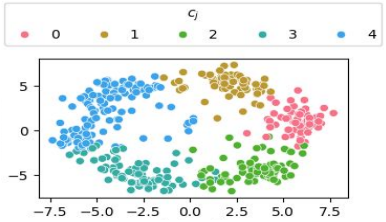


Falling into lava - Cluster 2

Half-Cheetah Clusters



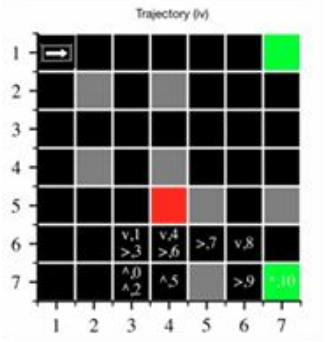
Authors

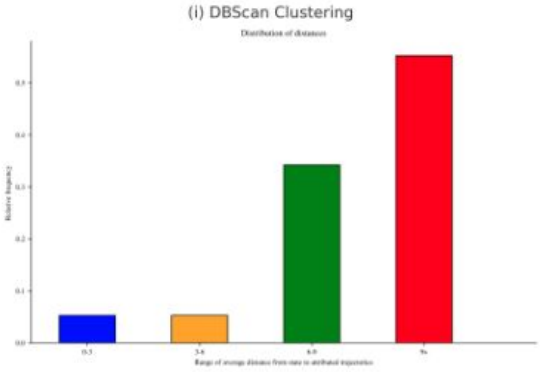


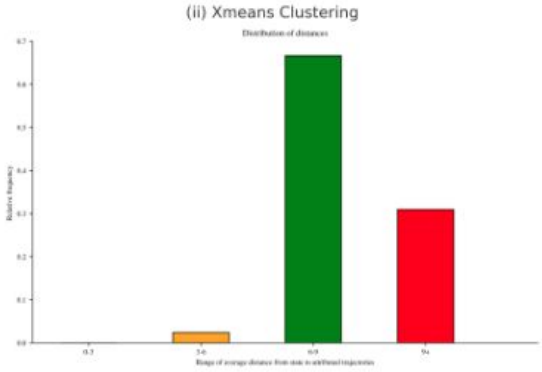
Ours

- Different Clusters represent High-Level Behaviours
- Reproducibility: We can identify similar High-Level Behaviours, but reproduction was not assured.
- We confirm the claim on Grid-World.
- We cannot validate it for Seaquest and Half-cheetah

Claim 3:
Distant trajectories are relevant



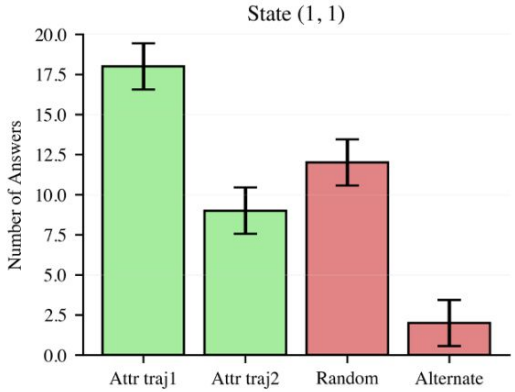


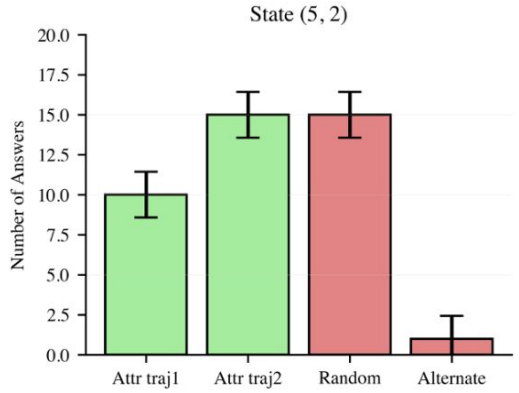


- Distant trajectories influence RL agent's decisions.
- Claim weakly supported by their evidence, but reproducible.
- Extra experiments to further confirm the claim using formal metrics

Claim 4: Human Study

- Method: Interview-based approach, focusing on trajectory identification.
- Humans do have a good understanding. Accuracy is around 63%.
- Claim is not fully supported by the experiments.





Claim Verification Results

	Grid-World	Seaquest	Half-Cheetah	Breakout	Q*Bert
Removing trajectories	✓	✓	✗	✓	✗
Cluster behaviours	✓	✗	✗	?	?
Distant trajectories	✓	?	?	?	?
Human study	?	?	?	?	?

Conclusion

- **Novel approach** towards the understanding and the interpretability of RL decisions, even if at a early stage.

- **Future Work**: Possible extensions are **Online RL** agents, combining **trajectory based** method with **classical** ones, and many others.