Generalization in Offline RL: Comparing Implicit Q-Learning with Behavioral Cloning

Author: Juan José Tarazona Rodríguez (j.j.tarazonarodriguez@student.tudelft.nl)
Supervisors: Dr. Matthijs Spaan, Max Weltevrede

CSE3000: Research Project
Final Presentation Poster

(1) Introduction

Offline Reinforcement Learning is a field of RL where the agent learns a policy without interactions with the environment. Existing offline RL algorithms have shown poor generalization performance [1]; they are not able to outperform the **behavioral policy** on similar but new tasks.

Implicit Q-Learning (IQL) [2] is an offline RL method that avoids querying **out-of-dataset** actions and shows outstanding performance on the **D4RL benchmark**.

(2) Research Question

Existing evaluation of generalization [1] focuses on a **limited set of environments** and collection policies. Our research question aims to both **reproduce** existing results and examine **new collection policies:**

*To what extent does Implicit Q-Learning enable **generalization**, and how is this capacity **influenced by the dataset composition**?*

(3) Methodology

We will use Behavioral Cloning (BC) as a benchmark, since it mimics the behavioral policy. The aim of offline RL is to **outperform the behavioral policy**. We will compare IQL and BC in an **experimental evaluation**:

- Generate datasets with different policies
- Train algorithms and hyper-parameters on datasets
- Evaluate on different topologies (reachable vs unreachable)
- Establish a direct comparison in terms of average rewards obtained

We use an **existing implementation for BC** from the *d3rlpy* library [3] and adapt to discrete control an implementation from CORL [4] for IQL.

(4) Environment

We use a simple 4-room environment (Fig. 1) where the agent must **reach the goal** by turning left, right or heading forward. The goal, agent and wall **positions are different** per task. This environment allows us to easily distinguish between reachable (by a sequence of actions from the agent) or unreachable tasks, when starting from the training set. We expect to see **generalization to reachable tasks perform better**.

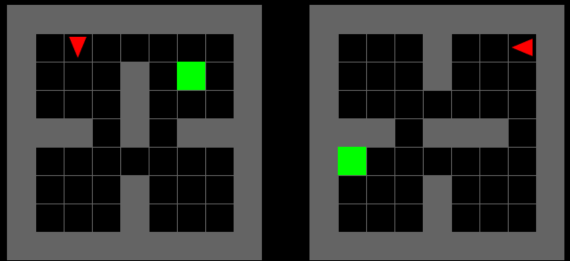

Fig 1: 4-room environment. The tasks are unreachable from each other due to different goal and wall positions

(5) Evaluation and Conclusions

To evaluate the generalization of **IQL**, we compared it to **BC** by training both algorithms on all dataset collection policies and measuring **average rewards** for different numbers of training steps on **reachable, unreachable and training set topologies**.

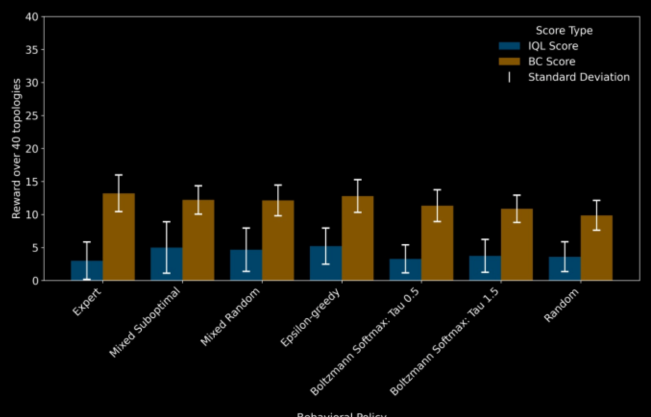

Fig 2: IQL, BC scores on unreachable topologies, 50000 steps

BC outperforms **IQL** even on unreachable test topologies, no matter the behavioral policy. We also observe that it is possible to learn from completely random data. ε-greedy appears to be the best policy (Fig. 2).

IQL reaches its peak reward faster than **BC,** but they are still lower. Unless training is limited, our results indicate that **BC** offers superior performance regardless of reachability, dataset composition and tuning (Fig. 3).

The **random seed** is the **most significant** hyper-parameter, suggesting high instability in the environment. This limits the reliability of data. Commonly used environments such as MuJoCo could improve this but incur higher computational complexity.

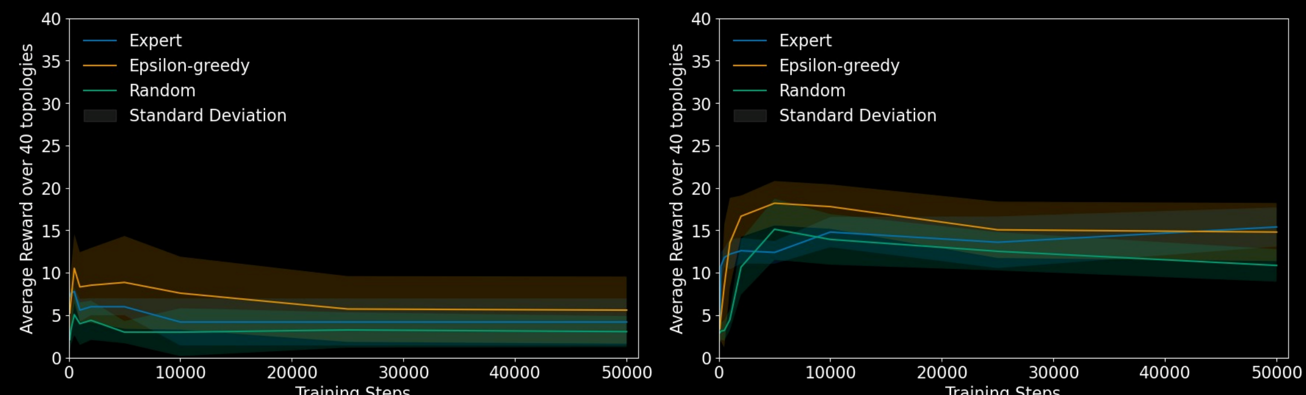

Fig 3: IQL (left), BC (right) scores on reachable topologies, different numbers of training steps

IQL lacks the ability to yield satisfactory results in generalization. We hope **future work** bridges this generalization gap and attempts to expand on environments as well as exploring the **impact of network architectures** and other parameters on our results.

References

[1] I. Mediratta, Q. You, M. Jiang, and R. Raileanu, "The generalization gap in offline reinforcement learning," *CoRR*, vol. abs/2312.05742, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2312.05742

[2] I. Kostrikov, A. Nair, and S. Levine, "Offline reinforcement learning with implicit q-learning," in *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022.* OpenReview.net, 2022. [Online]. Available: https://openreview.net/forum?id=68n2s9ZJWF8

[3] T. Seno and M. Imai, "d3rlpy: An offline deep reinforcement learning library," *J. Mach. Learn. Res.*, vol. 23, pp. 315:1–315:20, 2022. [Online]. Available: http://jmlr.org/papers/v23/22-0017.html

[4] D. Tarasov, A. Nikulin, D. Akimov, V. Kurenkov, and S. Kolesnikov, "CORL: research-oriented deep offline reinforcement learning library," in Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., 2023. [Online]. Available: http://papers.nips.cc/paper_files/paper/2023/hash/62d2cec62b7fd46dd35fa8f2d4aeb52d-Abstract-Datasets_and_Benchmarks.html

TU Delft