Offline Reinforcement Learning for Informative Path Planning

Srikar Babu Gadipudi[1], Srujan Deolasee[2], Woojun Kim[2], Katia Sycara[2]

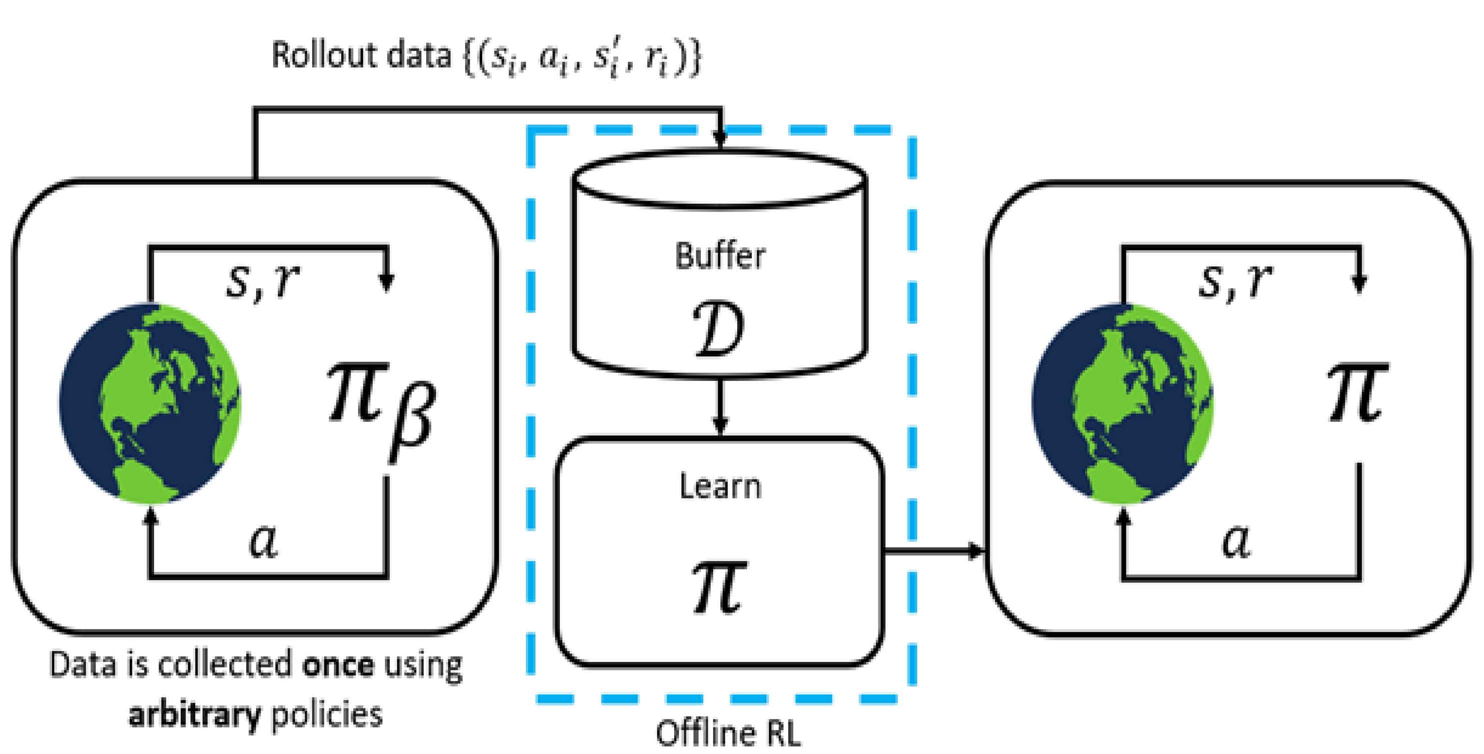
[1]Indian Institute of Technology Madras [2]Carnegie Mellon University

1.Introduction

- **IPP**: Design efficient paths to gather information within resource constraints.

$$\psi^* = \arg\max_{\psi \in \Psi} I(\psi), \text{ s.t. } C(\psi) \leq B$$

- **Traditional approaches**: High planning time and computation cost; low performance.
- **RL-based approaches**: Faster planning time and lesser computation cost; better performance.
- **Limitations**: Requires extensive environment interactions; risky and raises safety concerns.
- **Offline RL:** Leverage pre-collected datasets to train policy, without environmental interactions.



Offline RL Pipeline

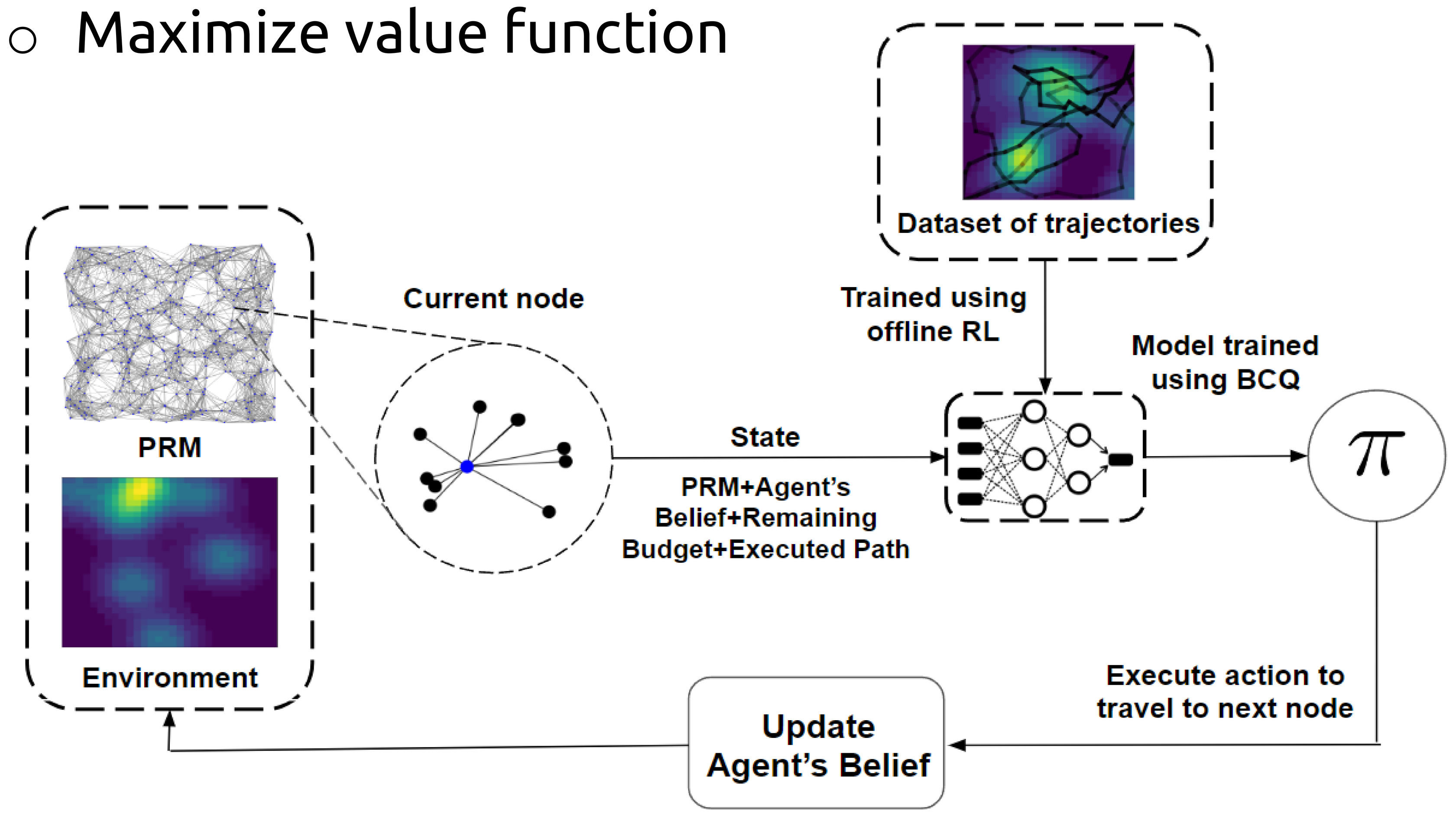
- **Challenge**: Extrapolation error due to distribution shift.
- Use **Gaussian Processes** (GP) to model robot's *belief*. GP allows interpolation between observations to represent interest.

Acknowledgement

We would like to acknowledge the support of the entire RISS community, Dr. John M. Dolan, Mrs. Rachel Burcin and Ms. Morgan Grimm.

2. Proposed Method

- Perform IPP employing model trained using offline RL.
- Encoder-Decoder architecture with **attention mechanism** and LSTM, inspired by CAtNIPP[1].
- **Probabilistic roadmap** (PRM) to discretize state space.
- State include **location, agent's belief** (extracted from GP), **budget** and **executed trajectory**.
- Using offline RL algorithm **Batch Constrained Q-learning**[2](BCQ).
  - Minimize distance between chosen action, during TD update, and behaviour policy
  - Maximize value function



Model	Budget 6	Budget 8	Budget 10	Budget 12
Greedy Planning	73.21 ± 99.80	65.00 ± 102.84	60.46 ± 104.41	57.11 ± 105.74
RAOr	49.47 ± 20.29	19.87 ± 7.71	12.54 ± 5.13	12.27 ± 4.99
BC - Expert	30.84 ± 14.02	9.93 ± 5.03	7.02 ± 3.06	5.26 ± 2.30
Our model - Expert	**23.28** ± 5.80	**7.83** ± 2.87	**3.96** ± 1.41	**2.62** ± 1.12
BC - Medium	45.44 ± 26.62	20.62 ± 19.09	12.09 ± 14.20	11.39 ± 15.77
Our model - Medium	**28.70** ± 12.96	**10.15** ± 3.80	**5.12** ± 1.83	**3.68** ± 1.89
BC - Greedy	42.25 ± 27.91	**16.39** ± 10.27	8.67 ± 5.56	4.74 ± 2.76
Our model - Greedy	**39.16** ± 22.42	16.56 ± 12.82	**7.61** ± 5.75	**4.62** ± 2.94

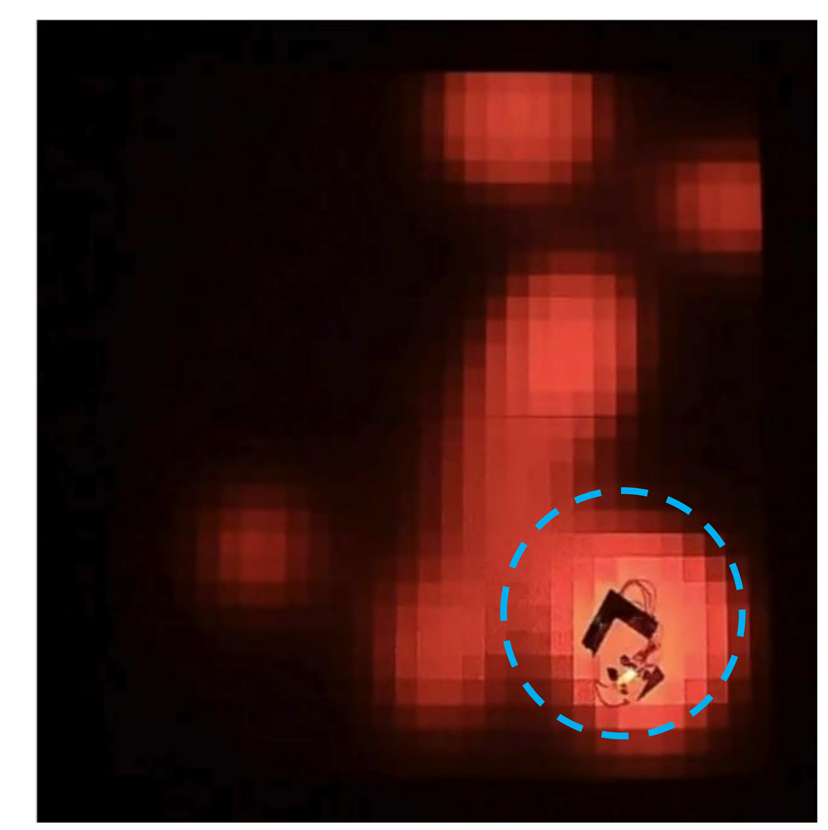
Table of Results (values shown are covariance trace of Gaussian Process)

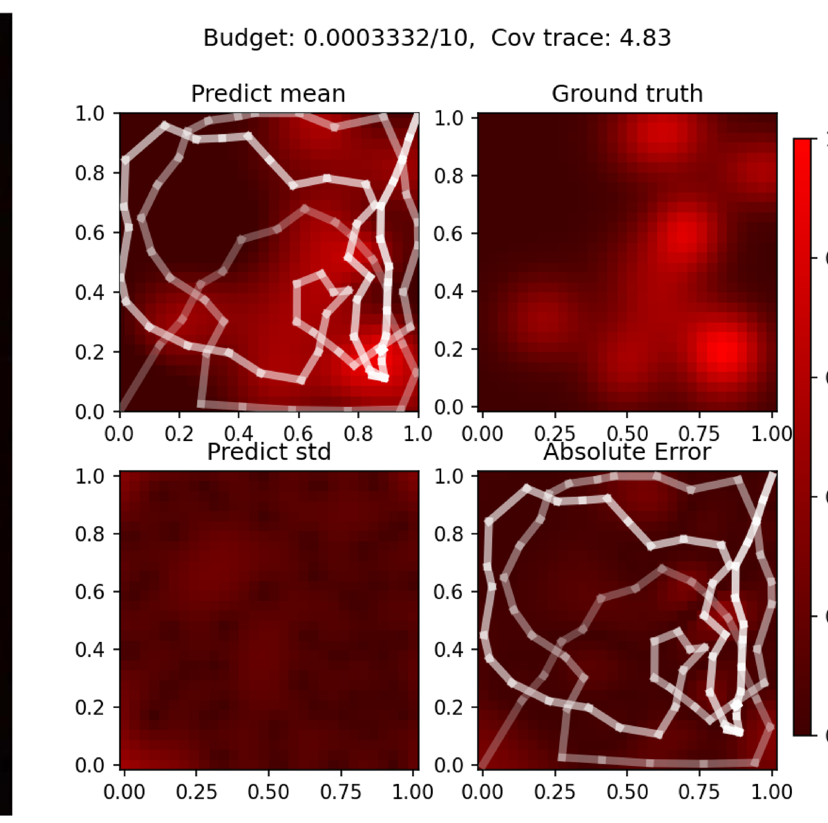
3. Datasets

- **Expert**: Best performing policy – CAtNIPP fully trained online.
- **Medium**: Suboptimal policy – CAtNIPP partially trained online.
- **Greedy**: Entropy based planning.

4. Results

- Environment simulated using randomly generated mixture of Gaussian distributions within a unit square.





Policy executed on robot **Simulation results**

5. Conclusion and Future Work

- **Conclusion**: Learn a planning policy, without environment interactions even with suboptimal datasets.
- **Future work**: Extend to a multi-agent setting and spatio-temporal environments.

References

1. Y. Cao, Y. Wang, A. Vashisth, H. Fan, and G. A. Sartoretti, "Catnipp: Context-aware attention-based network for informative path planning," in 2023.
2. S. Fujimoto, D. Meger, and D. Precup, "Off-policy deep reinforcement learning without exploration," in 2019.