Breadth-Frist Exploration on Adaptive Grid for Reinforcement Learning

ICML 2024

Youngsik Yoon, Gangbok Lee, Sungsoo Ahn, Jungseul Ok

Pohang University of Science and Technology (POSTECH)

{ysyoon97, gangbok2, sungsoo.ahn, jungseul}@postech.ac.kr

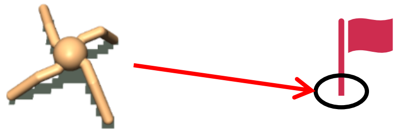
https://youngsikyoon.github.io/BEAG/

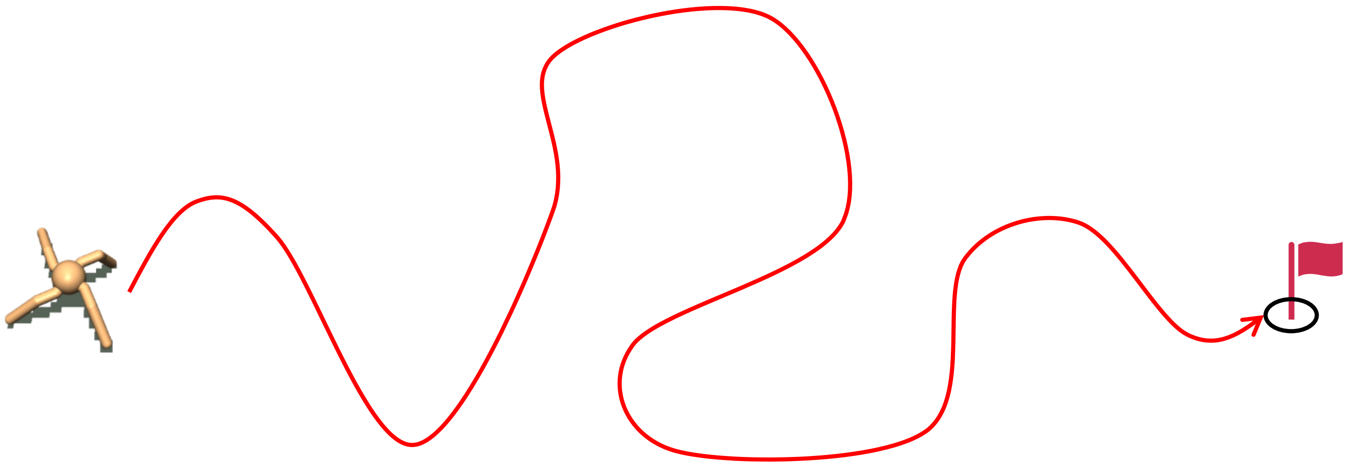
Machine Learning LABoratory

ICML International Conference On Machine Learning

Goal-conditioned Reinforcement Learning

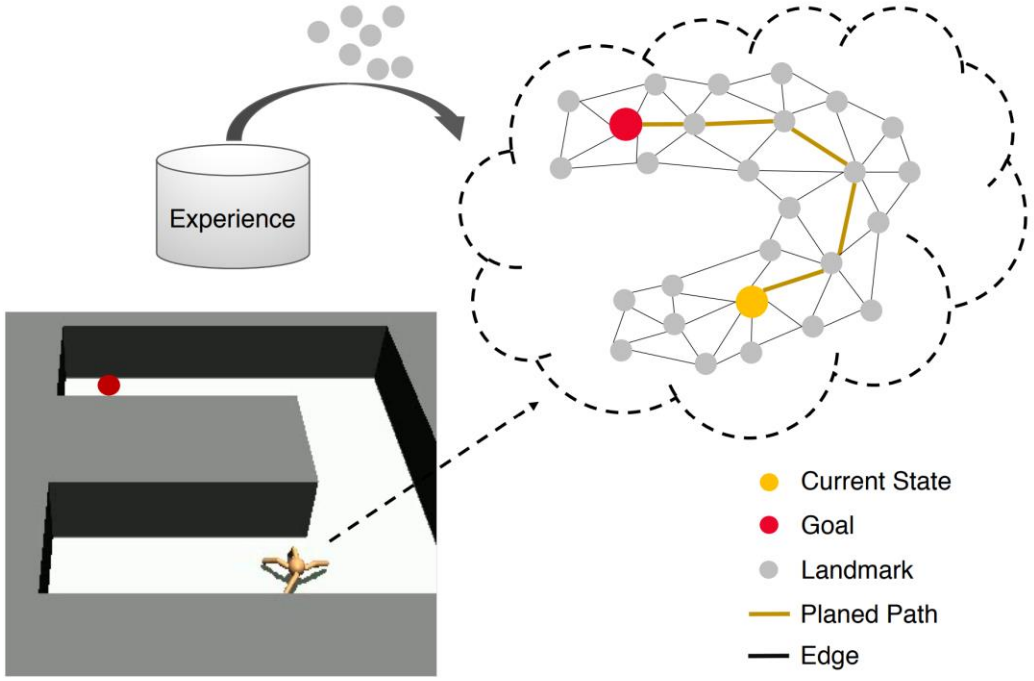
- Goal-conditioned Reinforcement Learning (GCRL) aims to solve the task of targeting a given goal.





Graph-based Reinforcement Learning

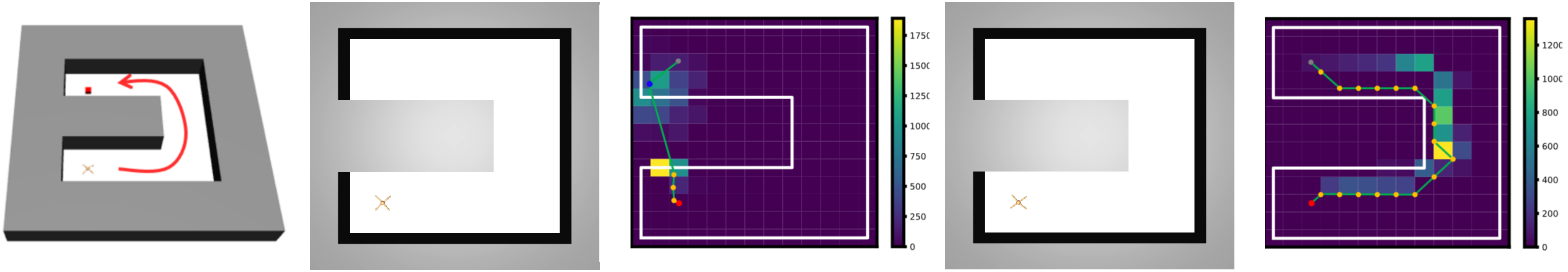
- Graph-based Reinforcement Learning (GBRL) decomposes long-horizon goals into a series of manageable short-horizon subgoals using a graph structure.



Mapping State Space using Landmarks for Universal Goal Reaching, Huang et al., 2019

Motivation

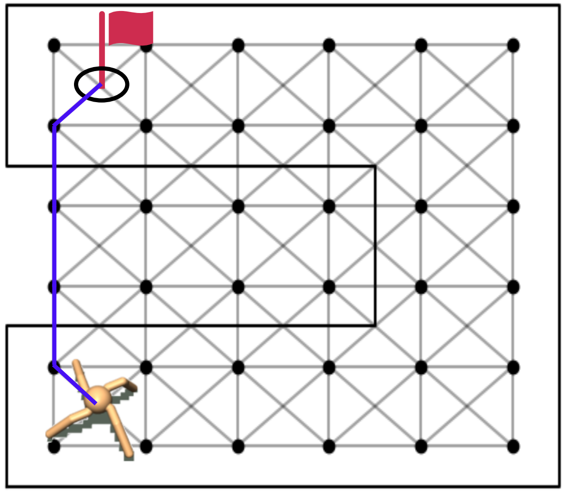
- Existing GBRL methods only manage successful subgoals sampled from the replay buffer without containing failed subgoals and unexplored subgoals, which leads to the problem of repeating the same failure.



(a) *U-maze* task (b) DHRL (Lee et al., 2022) (c) BEAG (ours)

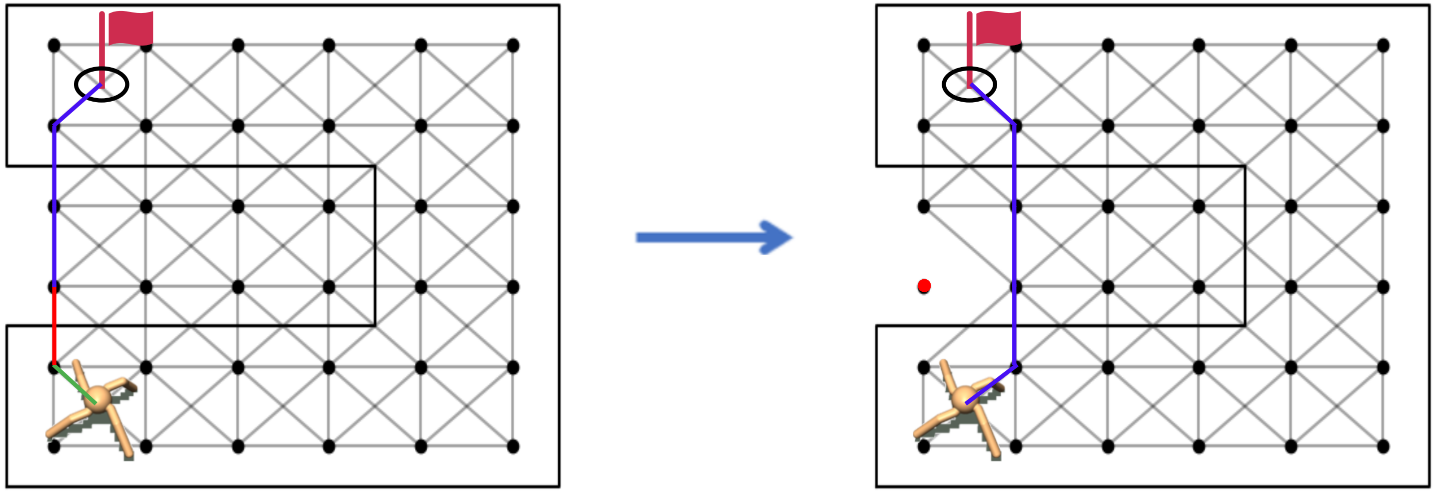
BEAG: Breadth-First Exploration on Adaptive Grid

- We propose BEAG, which employs a grid covering the entire map instead of relying on graphs generated from the replay buffer, which may include unexplored or even impossible subgoals.



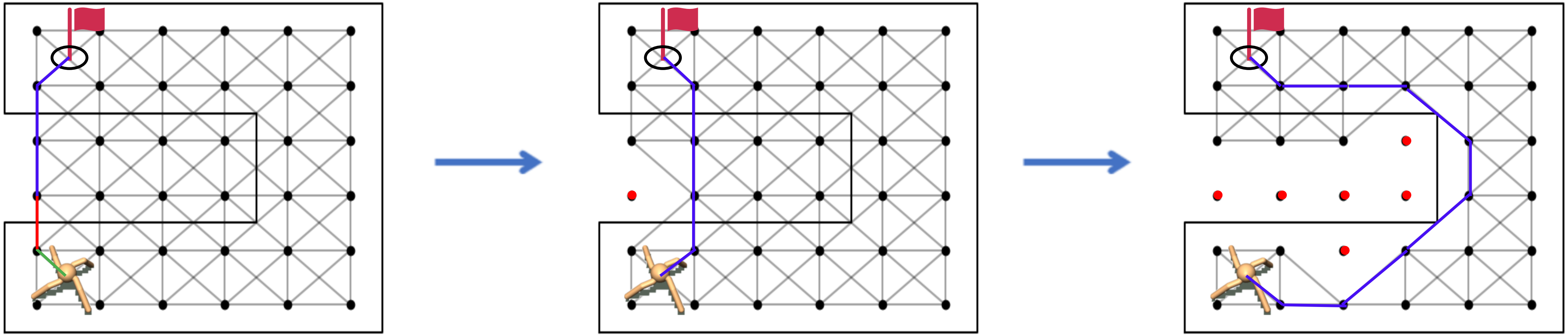
BEAG: Breadth-First Exploration on Adaptive Grid

- BEAG follows paths generated from the graph and explores promising subgoals by removing edges connected to subgoals that have been experienced repeated failures.



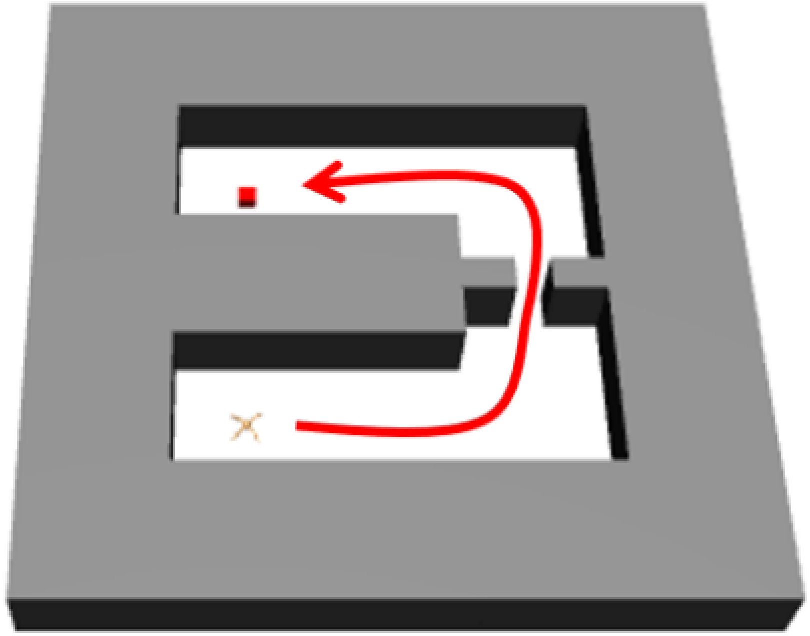
BEAG: Breadth-First Exploration on Adaptive Grid

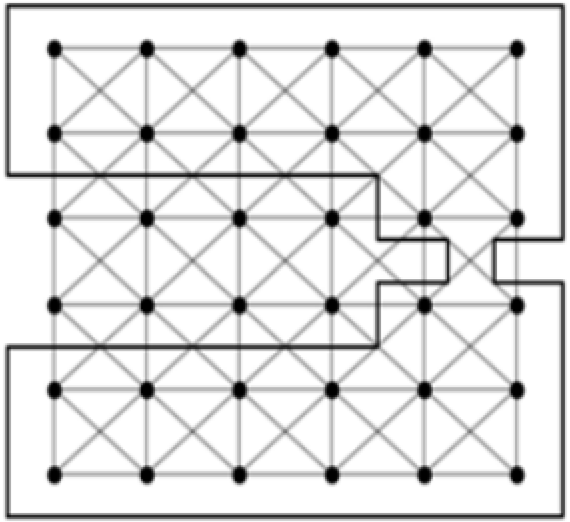
- After repeating this process a few times, BEAG will be able to find a successful path.



Bottleneck Environment

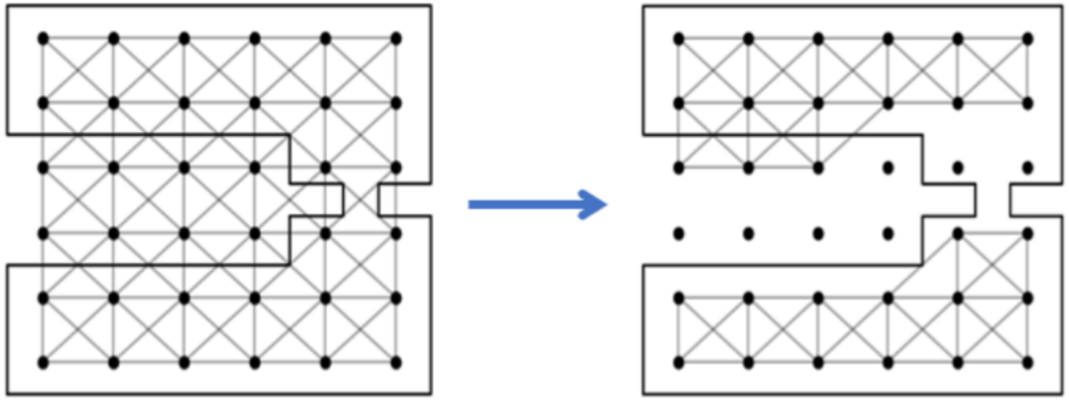
- While generating the grid, there may be issues in creating paths depending on the initial interval.





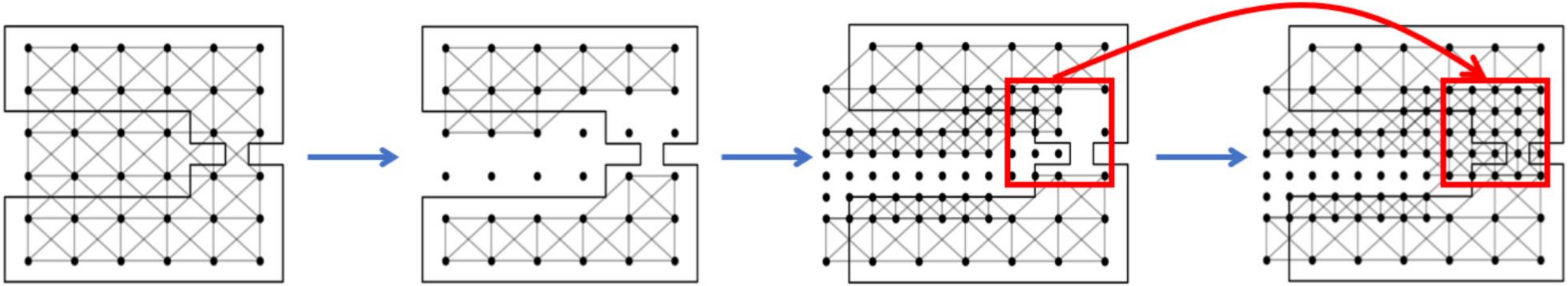
Adaptive Grid Refinement

- BEAG identifies unattainable subgoals until failing to generate paths.



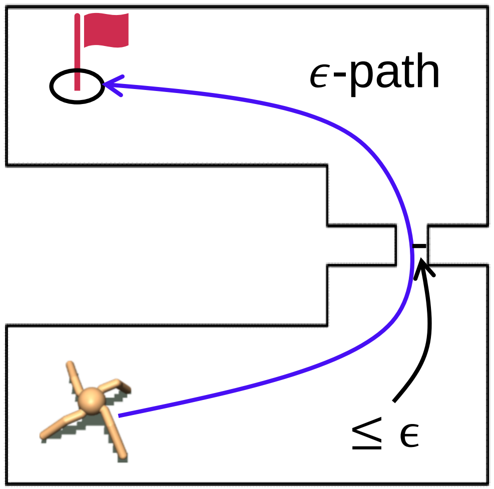
Adaptive Grid Refinement

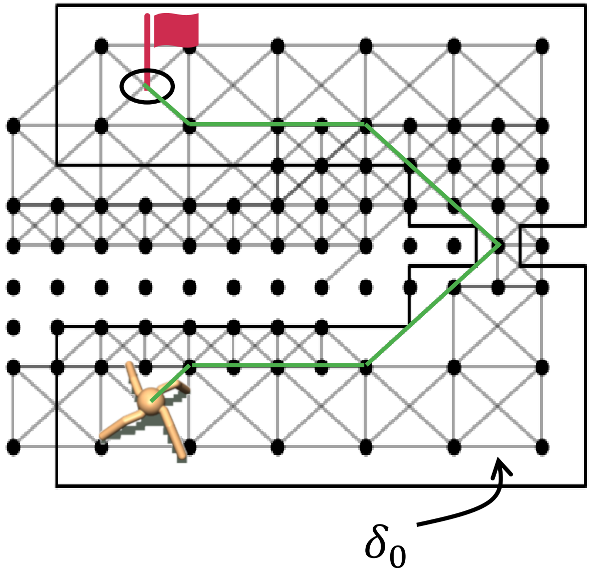
- After that, when a path to the goal cannot be generated, we select one of the failed nodes and perform more dense refinement around it.



Analysis

- According to our analysis,
  1. BEAG can always generate a possible path if there exists an ϵ-path.
  2. The length of the grid path is at most $\left(\left\lceil \frac{l}{\delta_0} \right\rceil + 1\right) \delta_0 \sqrt{K}$, where l is the length of the ϵ-path.

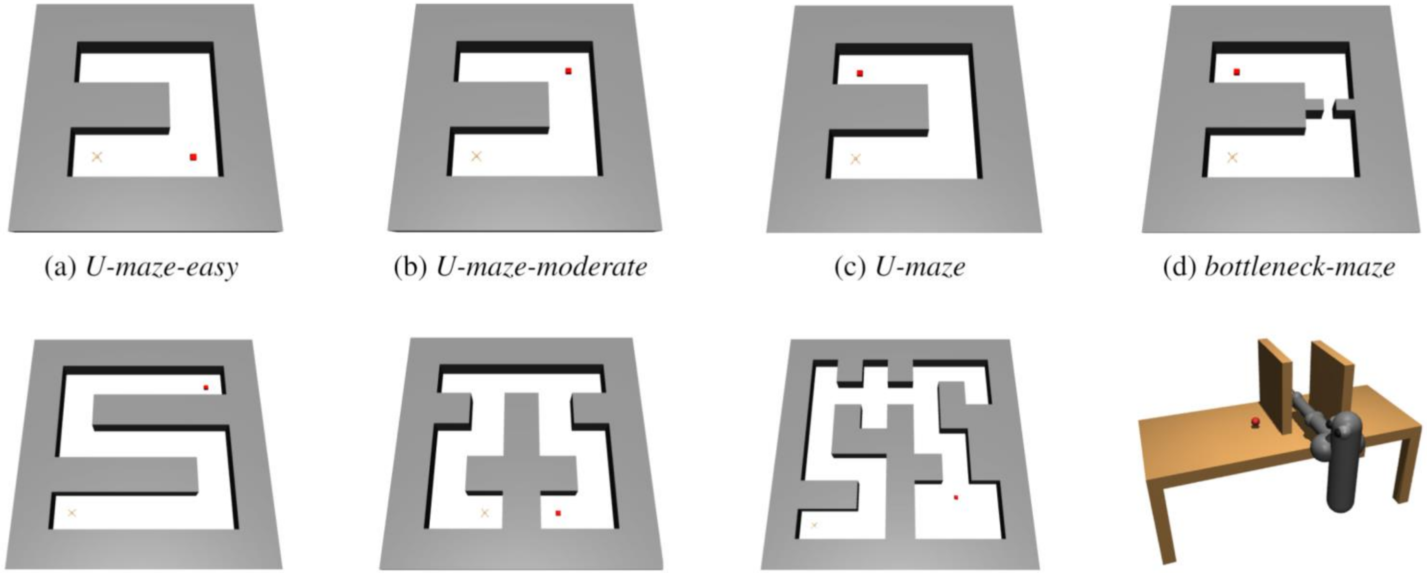




K: dimension of the goal space

Experimental Environments

- We conduct experiments in the various maps of the AntMaze environment (a-g) and the Reacher3D environment (h).



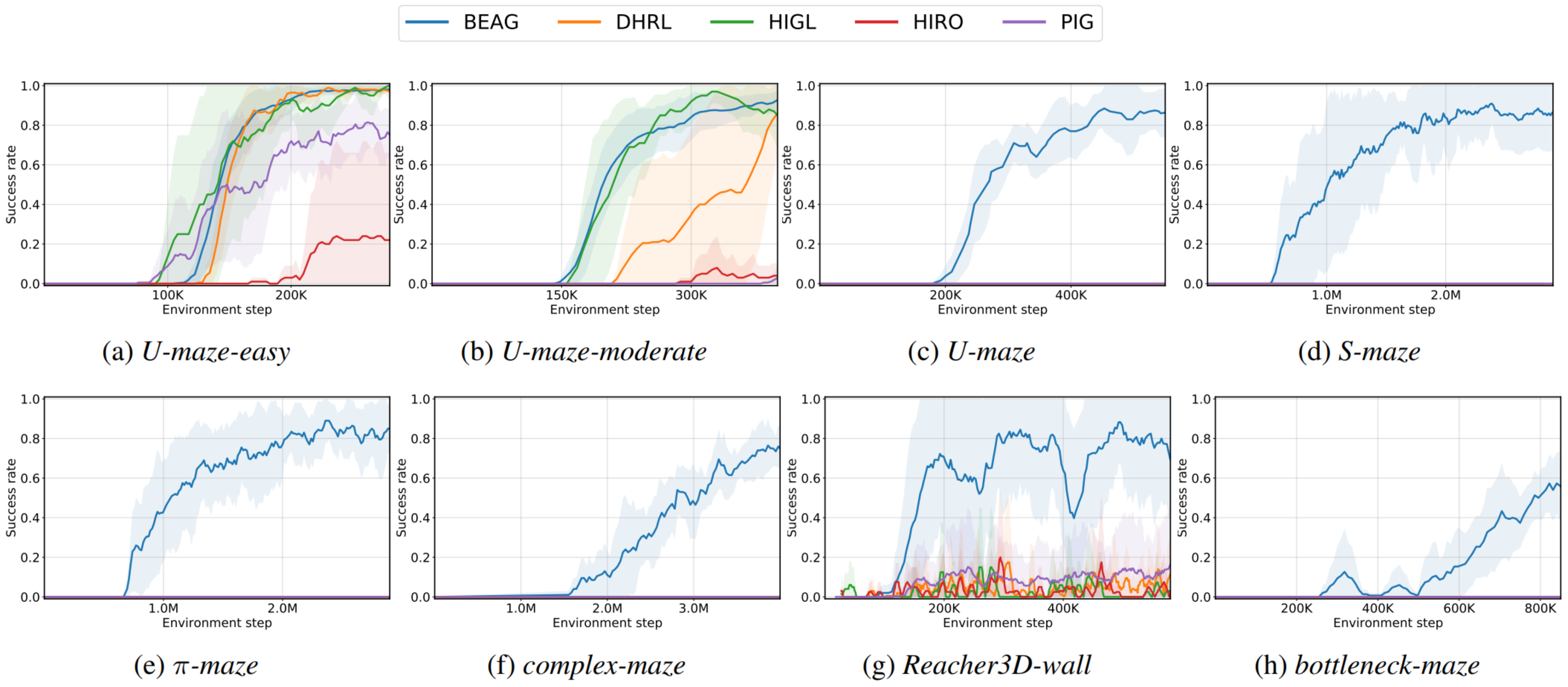
(a) *U-maze-easy* (b) *U-maze-moderate* (c) *U-maze* (d) *bottleneck-maze*

(e) *S-maze* (f) π*-maze* (g) *complex-maze* (h) *Reacher3D-wall*

- We compare our method, BEAG, with the state-of-the-art hierarchical RL and graph-based RL algorithms: HIRO (Nachum et al., 2018), HIGL (Kim et al., 2021), DHRL (Lee et al., 2022), and PIG (Kim et al., 2023)

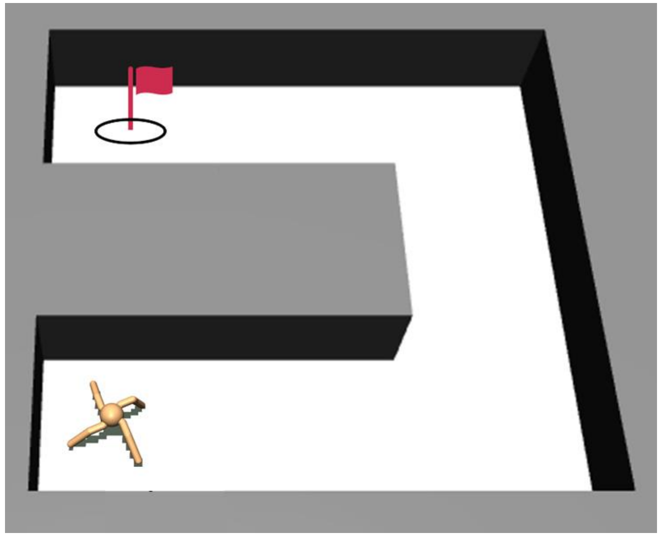
Experimental Results

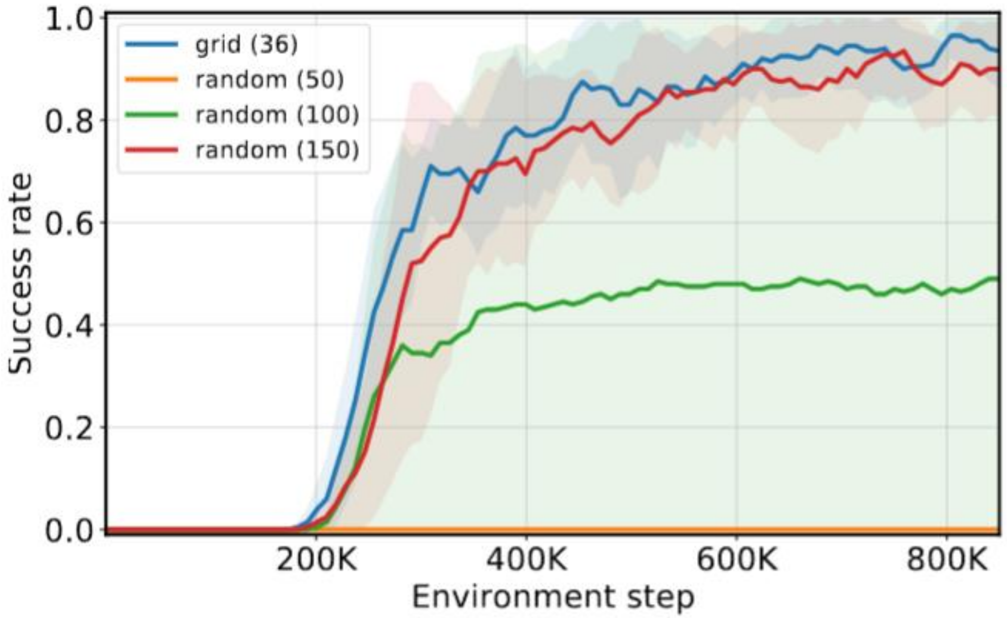
- BEAG remarkably outperforms other methods, especially in complex environments that require exploration, only BEAG achieves success.



Justification for Choosing Grid

- We conduct experiments performing a breadth-first search using a uniformly randomly generated graph over the entire map.

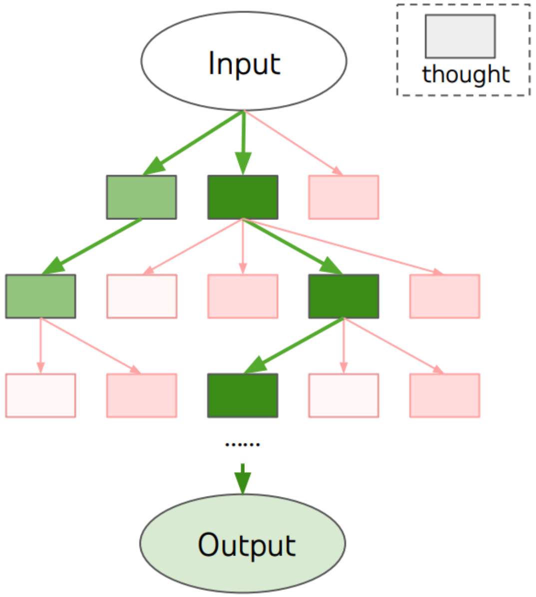




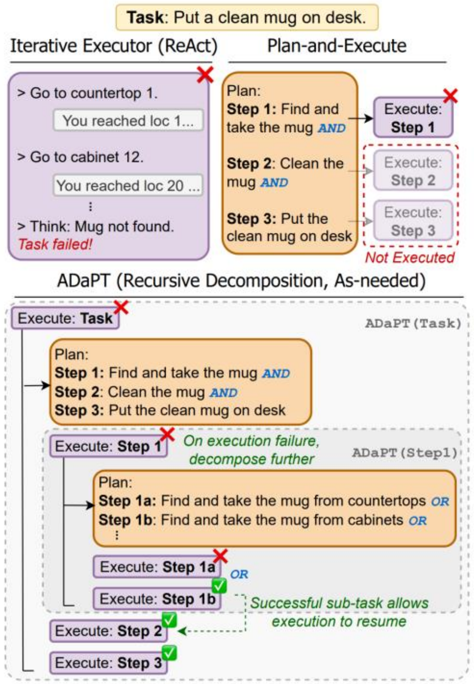
Take Away

- BEAG has demonstrated the effectiveness of breadth-first search in exploration.

- This approach can be applied not only in the robotics but also in various applications, such as prompt engineering in large language models (LLM).



Tree of Thoughts: Deliberate Problem Solving with Large Language Models, Yao et al., 2023



ADAPT: As-Needed Decomposition and Planning with Language Models, Prasad et al., 2024