

# Notes

## What is needed?

### 1. evaluation of *"interesting"* states

go explore solution:

- give each cell a subscore, based on:

- attribute a
- weight and power hyperparameter for attribute a
- $\epsilon_1$  as to not divide by zero
- $\epsilon_2$  as a minum value

their attributes:

- number of times a cell has been chosen as starting point
- number of times cell has been visited during exploration
- number of times a cell has been chosen since exploration from it last produced the discovery of a new or better cell

### 2. evaluation of *"reachable"* states (confidence)

- one large mcts tree?
- number of times cell has been target and actually been reached??? (success rate)
- simple visitation count? either based on cell or random network distillation
- if possible look at how direct the fastest recorded path to the taken cell is
- 

### 3. evaluation which state is both

- if cells are used in both cases should be easy

### 4. evaluation which trajectory to that state is faster

- again if cells in both cases should be easy
- isnt that the point were triangle inequality should be used??

### 5. unfortunately also missing: ***whats the point?***

## What do we have?

- continuous state space
- discrete non-stochastic action space
- triangle inequality?????
- atari games I guess?
- go explore approach only works on cells
- cells in their case are various states, that are all represented by same downsampled images,
- so can just be yonked for selection for HER + GO
- new paper idea let HER GO (explore)

## Run through of one iteration

- We have a goal G, Agent O, History H, Statespace S, Actionspace A, Dictionary of Cells(States) D,
- Agent tries to to get to G, we take furthest state in trajectory that is both interesting according to dictionary, as well as reachable according to D,
- add all info to Dictionary,
- make that new goal and try to reach it, go from there?

## Questions for literature research:

- How do we figure out if state is reachable?
- Do we try to find state to reach? or is the goal setting just for in experience replay?
- Is HER combinable with GO explore? As go explore "reduces statespace" but does not guarantee that all states in one Cell are reachable without leaving the cell

## Questions for discussion:

- What situations are we aiming for to improve either HER or GO explore?
- Are we trying to get the 2-phase based GO explore to use one phase iteratively?
- Setting of goals is just for replay buffer or are we actually setting those?

## Papers

- [HER](#)
- [Go Explore](#)
- [Learning Montezuma's Revenge from a Single Demonstration](#)
-