華東師範大學
EAST CHINA NORMAL UNIVERSITY
计算机科学与软件工程学院
School of Computer Science and Software Engineering

EMERGENT TOOL USE FROM MULTI-AGENT AUTOCURRICULA

捉迷藏

Jarvis

Multi-Agent Artificial Intelligence Laboratory
East China Normal University

大纲

Environment AutoCurriculum Evaluation Discussion

Environment

Environment

Two groups of agents(seeker and hider)

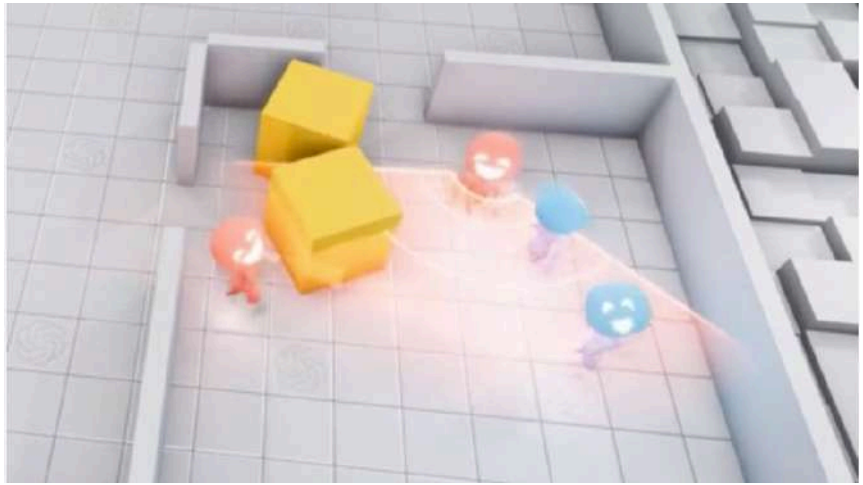
Observation: frontal cone

Actions: Move(x,y,z), grab, lock

Reward: +1 for seekers to find a hider

-1 for found no hiders

While the hiders are opposite



Auxiliary penalty for get too far from play zone

* The locked box can only be unlocked by the same team

AutoCurricula

Background

* single agent RL with specific reward → Overfit, Hard to improve

* unsupervised exploration(i.e. intrinsic reward) → Scale poorly to complex env

But the co-evolution and competition in earth between organisms generate more robust strategies and scale to complex environment.

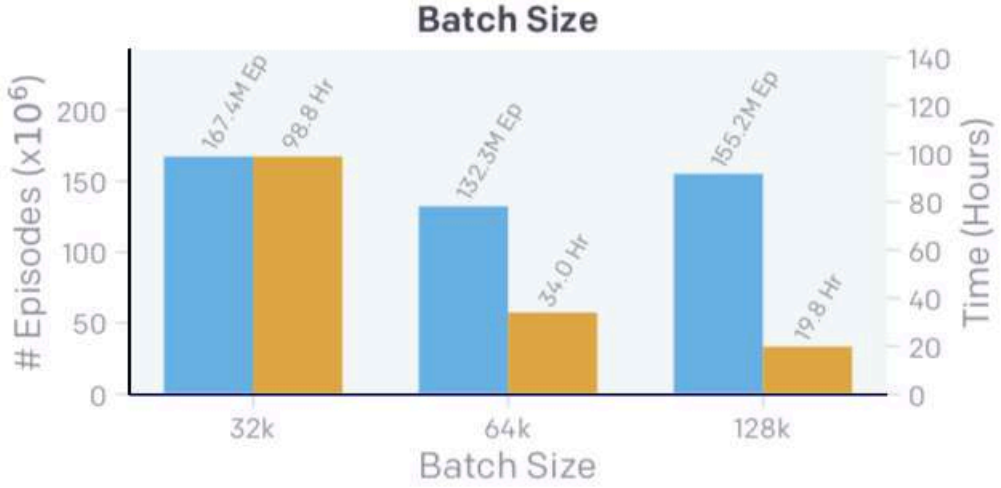
AutoCurricula

Definition: competing agents continually create new tasks for each other.

https://openai.com/blog/emergent-tool-use/

AutoCurricula

Effect of batch size



Large batch size, quicker training.

Reduce the number of optimization steps.

Fair? Confuse.

* 8k and 16k batch never converged

AutoCurricula

* The objects in the env are placed randomly, make the autocurriculum fairly robust

* Test on many other settings (add food), also lead to emergent tool use.

Because the behavior learning is unsupervised guided, how to Evaluate it will be a problem

Evaluation

Evaluation

Metrics:

1. Reward tracking → Ambiguous of improving evenly or stagnated

2. ELO or TrueSkills → No insight of the improvement from new adaption or new skills

3. Environment statistics → Sufficient design for complex env is hard.

1. ELO and TrueSkills are two ranking methods in MS games.

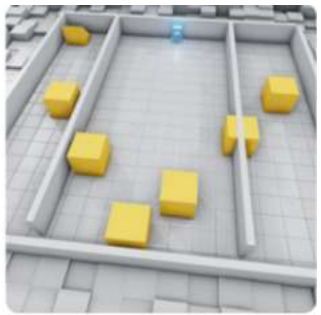
Evaluation

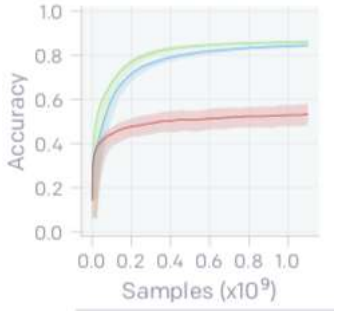
New Metrics: benchmark intelligence tests

(test the transfer ability in test tasks)

Cognition and Memory Task:

Object Counting: watch the boxes movement, predict how many boxes are gone at time t.





Pretrained in Hide-and-Seek

Trained From Scratch

Pretrained with Count-Based Intrinsic Motivation

Evaluation

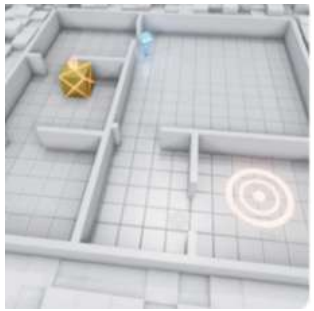
New Metrics: benchmark intelligence tests

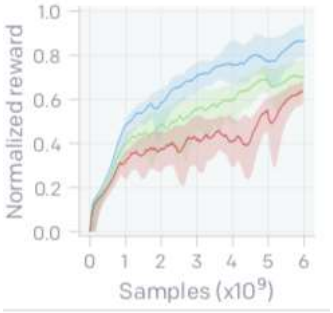
(test the transfer ability in test tasks)

Cognition and Memory Task:

Lock and Return: Lock the box and return.

Can agents remember its original location after performing A task?





Pretrained in Hide-and-Seek

Trained From Scratch

Pretrained with Count-Based Intrinsic Motivation

Evaluation

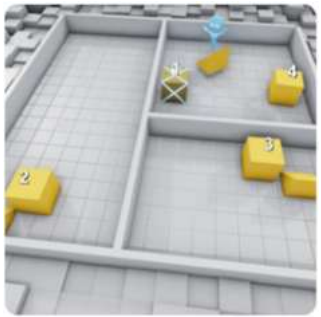
New Metrics: benchmark intelligence tests

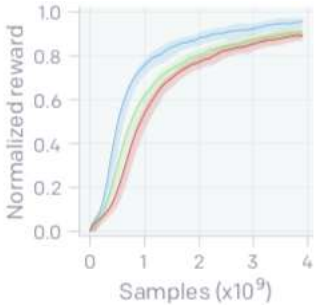
(test the transfer ability in test tasks)

Cognition and Memory Task:

Sequential Lock: Lock the box sequentially.

Can agents discover the order, remember the box status?





Pretrained in Hide-and-Seek

Trained From Scratch

Pretrained with Count-Based Intrinsic Motivation

Multi-Agent Artificial Intelligence Laboratory
East China Normal University

Evaluation

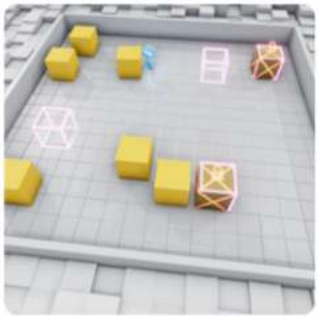
New Metrics: benchmark intelligence tests

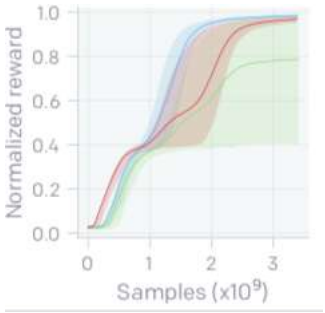
(test the transfer ability in test tasks)

Manipulation Task: Whether the the agents have latent skills.

Construction from blueprints: Move boxes to target place

Can agents discover the order, remember the box status?





Pretrained in Hide-and-Seek

Trained From Scratch

Pretrained with Count-Based Intrinsic Motivation

Multi-Agent Artificial Intelligence Laboratory
East China Normal University

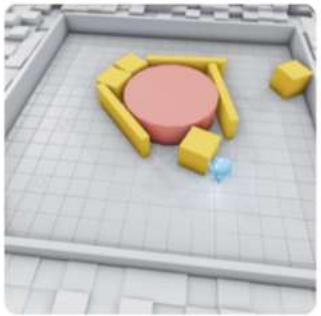
Evaluation

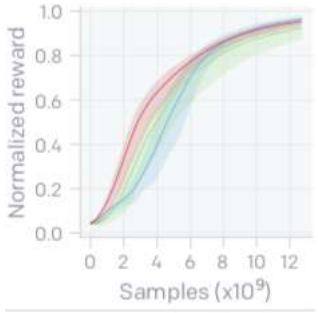
New Metrics: benchmark intelligence tests

(test the transfer ability in test tasks)

Manipulation Task: Whether the the agents have latent skills.

Shelter construction: build a shelter around cylinder.





Pretrained in Hide-and-Seek

Trained From Scratch

Pretrained with Count-Based Intrinsic Motivation

Evaluation

* agents learning skills representations are entangled and hard to fine tune.

The better than baseline are due to reuse of learned feature representations.

While remaining tasks require reuse of learned skills.

Discussion

Discussion

* AutoCurriculum leads to human relevant skills.

* How to design algorithm learn skills (meta?)

* How to prevent unwanted behaviors(surf on the box)?

HAIL
HYDRA