Environments for Lifelong Reinforcement Learning

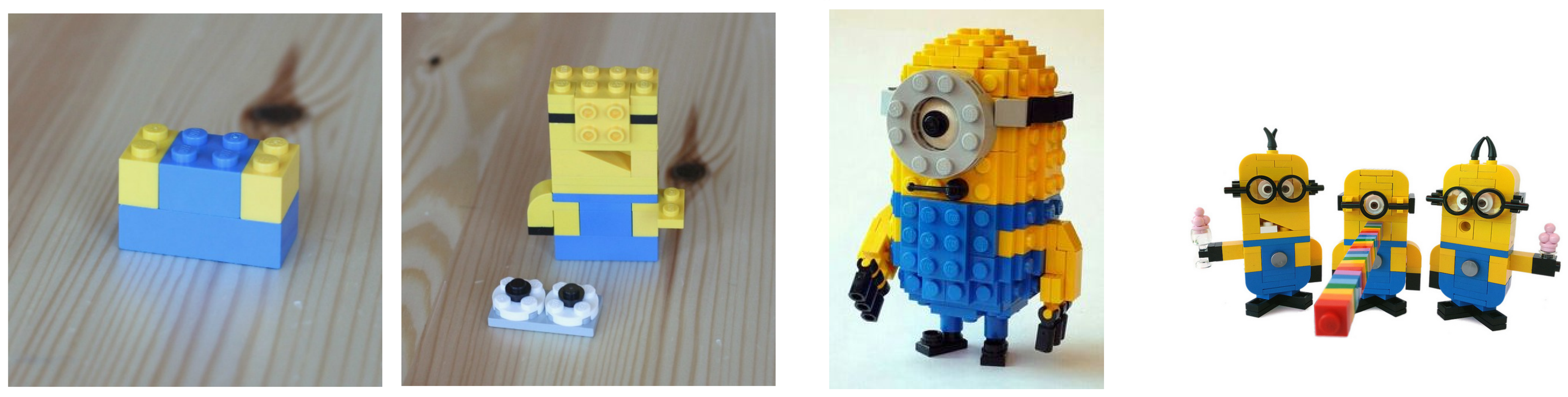
Khimya Khetarpal* [1,2], Shagun Sodhani* [1,3], Sarath Chandar [1,3], Doina Precup[1,2]
Mila[1], McGill University[2], Universite de Montreal[3]

TL;DR

- To achieve general artificial intelligence, reinforcement learning (RL) agents should learn not only to optimize returns for one specific task but also to **constantly build more complex skills** and **scaffold their knowledge** about the world, **without forgetting** what has already been learned.
- We discuss the **desired characteristics of environments** that can support the training and evaluation of lifelong RL agents.
- We argue that **virtual-embodiment is perhaps the most natural setup** for training and evaluating lifelong reinforcement learning agents.
- We **propose recommendations** for devising suitable lifelong RL environments in the future.

What constitutes a Lifelong Learning agent?

- Learn **behaviors**, **skills**, and **predictions** about the environment while solving the given tasks.
- Learn **incrementally** throughout its lifetime.
- **Combine and compose previously learned skills** to solve increasingly complex tasks.
- **Plan** for both short-term **and** long-term goals



Current RL Benchmarks

Benchmark	Vision	Environment Artifacts	Stochasticity	Difficulty Levels
ALE-V1	2D	Synthetic	✗	✗
ALE-V2	2D	Synthetic	✓	✓
ViZDoom	3D	Stylized Maze, Synthetic	✓	✗
DeepMind Lab	3D	Stylized Maze, Synthetic	✓	✓
Malmo	3D	Synthetic	✓	✓
Starcraft	3D	Synthetic	✓	✓

- Recent breakthroughs in RL research [1] have been powered in part by advances in deep learning and the availability of diverse simulation environments to train RL agents.
- However, they lack multiple modalities limiting sensorimotor data to only one single stream; they lack naturalistic appeal in terms of their layout, appearance and objects

Key Idea: Lifelong Learning in Embodied Agents

- **Curriculum learning** - virtual environments are easy to modify in terms of complexity thereby making it easier to train agents in progressive fashion
- **Short-term and long-term goals** - these are equivalent respectively with skills and composition of skills
- **Mimic agents in real-world scenarios** - embodied learning tries to encapsulate the real world dynamics in a simulated environment as faithfully as possible, which creates more "real life"" like setup
- **Cause-and-effect learning** - rich, multi-modal data streams can help agents to understand the causal relationships of various events and opportunities associated with each object, through interactions as afforded by the objects and the environment.

Virtual Embodiment Environments: Brief Summary

House 3D [2]

+ Realistic and extensible environment built on top of the SUNCG dataset.
+ Supports rendering photo-realistic 3D visuals with support for diverse 3D objects and layouts
+ Can be customized to some extent

– Does not allow defining varying degrees of difficulty in tasks

HoME [3]

+ Provides natural language descriptions of objects and audio rendering
+ Supports rigid body dynamics and external forces like gravity
+ Supports "adding" or "removing" objects from scenes

MINOS [4]

+ Specifically designed for multi-sensory navigation models
+ Supports material variation (for texture and colors) and object clutter variation
+ Supports navigation task and goal specification

– Not as rich as Home and House3D

VirtualHome [5]

+ Crowd-sourced a dataset of "programs" for performing different activities in a house which can be seen as a composition of skills

– Does not allow for creating variations of a scene

Recommendations for a Lifelong Learning testbed

- **Support a multitude of tasks of different difficulty**
  - The lowest level tasks should require only one skill to solve tasks
  - Complex tasks should require a composition of skills learned in the previous levels.
  - Tasks should be structured, probably in a hierarchy, such that task complexity would increase as the learning agent moves up. Solving a task in the $i+1^{th}$ level should require the agent to solve the task at level i and also learn to compose previously acquired skills.
- **Facilitate modifications and incremental progress**
  - Easy addition or removal of objects from a scene or a variety of scenes.
  - Facilitate the incremental expansion of the data that the agent sees over time.
  - By providing the flexibility to scale up the size of the environment, the quantity of objects with which the agent can interact, etc.
- **Evaluate on resistance to catastrophic forgetting**
  - The environment should continue to challenge it with previously seen tasks
  - Hierarchy and compositional tasks
- **Short-term and long-term planning**
  - Tasks should be generated in the environment in a way that requires the agent to do both short-term and long-term planning.
  - Dealing with many goals that span different time scales
  - Test an agent's capacity to learn different types of knowledge and to generalize across time scales.
- **Inbuilt Provisions for Training and Evaluation**
  - Metrics for evaluating forward and backward transfer

References

- David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484, 2016.
- Y. Wu, Y. Wu, G. Gkioxari, and Y. Tian. Building Generalizable Agents with a Realistic and Rich 3D Environment. *ArXiv e-prints*, January 2018.
- S. Brodeur, E. Perez, A. Anand, F. Golemo, L. Celotti, F. Strub, J. Rouat, H. Larochelle, and A. Courville. HoME: a Household Multimodal Environment. *ArXiv e-prints*, November 2017.
- M. Savva, A. X. Chang, A. Dosovitskiy, T. Funkhouser, and V. Koltun. MINOS: Multimodal Indoor Simulator for Navigation in Complex Environments. *ArXiv e-prints*, December 2017.
- Xavier Puig, Kevin Ra, Marko Boben, Jiaman Li, Tingwu Wang, Sanja Fidler, and Antonio Torralba. Virtualhome: Simulating household activities via programs. In *CVPR*, 2018.