Unlocking Exploration: Self-Motivated Agents Thrive on Memory-Driven Curiosity

Hung Le
Applied AI Institute, Deakin University
Geelong, Australia
thai.le@deakin.edu.au

Hoang Nguyen
Applied AI Institute, Deakin University
Geelong, Australia
s223669184@deakin.edu.au

Dai Do
Applied AI Institute, Deakin University
Geelong, Australia
s223540177@deakin.edu.au

ABSTRACT

Despite remarkable successes in various domains such as robotics and games, Reinforcement Learning (RL) still struggles with exploration inefficiency. For example, in hard Atari games, state-of-the-art agents often require billions of trial actions, equivalent to years of practice, while a moderately skilled human player can achieve the same score in just a few hours of play. This contrast emerges from the difference in exploration strategies between humans, leveraging memory, intuition and experience, and current RL agents, primarily relying on random trials and errors. This tutorial reviews recent advances in enhancing RL exploration efficiency through intrinsic motivation or curiosity, allowing agents to navigate environments without external rewards. Unlike previous surveys, we analyze intrinsic motivation through a memory-centric perspective, drawing parallels between human and agent curiosity, and providing a memory-driven taxonomy of intrinsic motivation approaches.

The talk consists of three main parts. Part A provides a brief introduction to RL basics, delves into the historical context of the explore-exploit dilemma, and raises the challenge of exploration inefficiency. In Part B, we present a taxonomy of self-motivated agents leveraging deliberate, RAM-like, and replay memory models to compute surprise, novelty, and goal, respectively. Part C explores advanced topics, presenting recent methods using language models and causality for exploration. Whenever possible, case studies and hands-on coding demonstrations will be presented.

KEYWORDS

Reinforcement Learning; Exploration; Intrinsic Motivation; Memory

ACM Reference Format:
Hung Le, Hoang Nguyen, and Dai Do. 2024. Unlocking Exploration: Self-Motivated Agents Thrive on Memory-Driven Curiosity. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 4 pages.

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand.

OVERVIEW

Duration

Half day.

Target Audience

This tutorial is mainly designed for students and academics who work on Reinforcement Learning. It is also open to research engineers and industry practitioners who need to apply efficient reinforcement learning in their jobs. Basic familiarity with reinforcement learning is assumed, and an additional understanding of deep learning and neural networks would be beneficial. No special equipment is required, but attendees are encouraged to bring their laptops to experiment with models hands-on.

Engaging the audience The tutorial comprises three parts, each followed by a QA session interspersed with interactive demos and coding guidelines. The audience is encouraged to ask questions throughout the session.

Tutorial Speakers

Leading the tutorial is Dr. Hung Le, a Research Lecturer at the Applied AI Institute, Deakin University, Australia. Assisting are two tutorial supporters, Hoang Nguyen and Dai Do, both PhD students at Deakin University.
Address: 75 Pigdons Rd, Waurn Ponds VIC 3216, Australia
Email: `thai.le@deakin.edu.au`
Phone: +61 3 522 72425
Website: https://thaihungle.github.io/

Brief Bio Dr. Hung Le is a Research Lecturer at Deakin University, Australia, and is a senior member of the Applied Artificial Intelligence Institute (A2I2) where he currently supervises 5 PhD students in research areas focused on machine learning (ML) and reinforcement learning (RL). Specializing in deep reinforcement learning, he is dedicated to pioneering new agents equipped with artificial neural memory. His extensive work in this area includes multi-modal, adaptive and generative memory, efficient policy optimization, and memory-based reinforcement learning agents. With applications spanning health, dialogue agents, robotics, reinforcement learning, machine reasoning, and natural language processing, Dr. Le consistently publishes in premier ML/RL/AI conferences and journals, including ICLR, NeurIPS, ICML, AAAI, IJCAI, TMLR, KDD, NAACL, ECCV, and AAMAS. He earned his Bachelor of Engineering (Honors) from Hanoi University of

Science and Technology and completed his PhD in Computer Science at Deakin University in 2015 and 2020, respectively.

Related Talks

This tutorial is related to recent presentations, expanding on topics covered in the following talks:

- Conference tutorial: *"Memory-Based Reinforcement Learning"*. The 35th Australasian Joint Conference on Artificial Intelligence (AJCAI'22), December 2022, Perth, Australia. Audience size: 50.
- Industrial talk: *"Memory for Lean Reinforcement Learning"*. FPT Software AI Center, May 2022, virtual, Vietnam. Audience size: 100.
- Conference tutorial: *"Neural machine reasoning"*. The 30th International Joint Conference on Artificial Intelligence (IJCAI'21), June 2021, virtual, Canada. Audience size: 50.
- Conference tutorial: *"From deep learning to deep reasoning"*. The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'21), August 2021, virtual, Singapore. Audience size: 50.

Why this topic?

Powered by the high-capacity representation of deep learning and advanced computing infrastructure, current reinforcement learning agents demonstrate mastery in learning intricate policies that map from complex state spaces to vast action spaces [32]. However, they necessitate hundreds of millions or even billions of environmental steps to kickstart the learning process, resulting in prolonged exploration periods [2, 12]. This is feasible only in simulation scenarios, proving challenging for real-world applications such as robotics or industrial planning. It is crucial to optimize the exploration process to enable the adoption of current RL techniques in real-world settings, providing intrinsic mechanisms to motivate agents to exhibit reasonable behaviour at the earliest opportunity.

Viewing intrinsic exploration through the lens of memory, akin to human cognition, is important for understanding the efficiency of self-motivated RL agents. Human-like memory systems enable agents to retain valuable experiences, learn from past interactions, and expedite the adaptation process, significantly influencing many aspects of RL [20–22, 24, 25]. Analyzing intrinsic motivation from a memory perspective not only aligns RL approaches with human-like exploration but also opens vast avenues for further investigation, pushing the boundaries not only of RL but also of AI at large.

DETAILED OUTLINE

The tutorial is designed for 3 hours + 20-minute break. The content is organized into three parts in which Part A covers background and problem introduction, Part B reviews well-established exploration approaches using human-like memory models, and Part C presents advanced topics on intrinsic motivation touching on emerging technologies such as large language models and causality where implicit memory mechanisms are used.

Part A: Reinforcement Learning Fundamentals and Exploration Inefficiency (30 minutes)

The session opens with an overview and speaker introductions. It then explores fundamental reinforcement learning concepts, emphasizing the exploration-exploitation tradeoff, and addresses exploration challenges in deep reinforcement learning. The session concludes with a QA and brief demonstration for audience engagement. Details are given below:

- Welcome and Introduction (5 minutes)
  - Overview of the tutorial
  - Brief speaker introductions
- Reinforcement Learning Basics (10 minutes)
  - Key components and frameworks
  - Classic exploration [13, 29]
- Exploring Challenges in Deep RL (10 minutes)
  - Hard exploration problems
  - Simple exploring solutions [11, 15]
- QA and Demo (5 minutes)

Part B: Surprise and Novelty (110 minutes, including a 20-minute break)

The session starts with an introduction of principles and frameworks for intrinsic motivation, encompassing reward shaping and the taxonomy of memory systems in driving agent exploration. It then delves into slow and careful (system I) memory architectures for modelling surprise-based curiosity. Following a short break, the session introduces novelty-based intrinsic motivation through diverse memory systems, characterized as fast and readily accessible (system II). The session wraps up with replay memory techniques that resemble associative memory, followed by an interactive QA and demo section. The detailed outline is as follows:

- Principles and Frameworks (10 minutes)
  - Reward shaping and the role of memory
  - A taxonomy of memory-driven intrinsic exploration
- Deliberate Memory for Surprise-driven Exploration (25 minutes)
  - Forward dynamics prediction [1, 27, 33]
  - Advanced dynamics-based surprises [5, 10, 17, 18]
  - Ensemble and disagreement [28, 37]
- Break (20 minutes)
- RAM-like Memory for Novelty-based Exploration (25 minutes)
  - Count-based memory [4, 36]
  - Episodic memory [30, 34]
  - Hybrid memory [2, 3, 19]
- Replay Memory (20 minutes)
  - Performance-based replay [9, 12]
  - Entropy-based replay [14, 26]
- QA and Demo (10 minutes)

Part C: Advanced Topics (60 minutes)

The session kicks off with recent exploration methods using language-based knowledge, including pre-trained large language models. It then shifts to an emerging line or research direction where causal discovery guides exploration. Concluding the session, a closing remarks recaps key learning and touches on other intrinsic motivation approaches, followed by a QA session and a concluding demo. Below is the detailed outline:

- Language-guided exploration (20 minutes)
  - Language-assisted RL [6, 35]
  - LLM-based exploration [8, 14]
- Causal discovery for exploration (20 minutes)
  - Statistical approaches [23, 31]
  - Deep learning approaches [7, 16]
- Closing Remarks (10 minutes)
- QA and Demo (10 minutes)

REFERENCES

[1] Joshua Achiam and Shankar Sastry. 2017. Surprise-based intrinsic motivation for deep reinforcement learning. *arXiv preprint arXiv:1703.01732* (2017).

[2] Adrià Puigdomènech Badia, Bilal Piot, Steven Kapturowski, Pablo Sprechmann, Alex Vitvitskyi, Zhaohan Daniel Guo, and Charles Blundell. 2020. Agent57: Outperforming the atari human benchmark. In *International Conference on Machine Learning*. PMLR, 507–517.

[3] Adrià Puigdomènech Badia, Pablo Sprechmann, Alex Vitvitskyi, Daniel Guo, Bilal Piot, Steven Kapturowski, Olivier Tieleman, Martin Arjovsky, Alexander Pritzel, Andrew Bolt, et al. 2019. Never Give Up: Learning Directed Exploration Strategies. In *International Conference on Learning Representations*.

[4] Marc Bellemare, Sriram Srinivasan, Georg Ostrovski, Tom Schaul, David Saxton, and Remi Munos. 2016. Unifying count-based exploration and intrinsic motivation. *Advances in neural information processing systems* 29 (2016).

[5] Yuri Burda, Harrison Edwards, Amos Storkey, and Oleg Klimov. 2018. Exploration by random network distillation. In *International Conference on Learning Representations*.

[6] Devendra Singh Chaplot, Kanthashree Mysore Sathyendra, Rama Kumar Pasumarthi, Dheeraj Rajagopal, and Ruslan Salakhutdinov. 2018. Gated-attention architectures for task-oriented language grounding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

[7] Oriol Corcoll and Raul Vicente. 2020. Disentangling causal effects for hierarchical reinforcement learning. *arXiv preprint arXiv:2010.01351* (2020).

[8] Yuqing Du, Olivia Watkins, Zihan Wang, Cédric Colas, Trevor Darrell, Pieter Abbeel, Abhishek Gupta, and Jacob Andreas. 2023. Guiding pretraining in reinforcement learning with large language models. *arXiv preprint arXiv:2302.06692* (2023).

[9] Adrien Ecoffet, Joost Huizinga, Joel Lehman, Kenneth O Stanley, and Jeff Clune. 2021. First return, then explore. *Nature* 590, 7847 (2021), 580–586.

[10] Aleksandr Ermolov and Nicu Sebe. 2020. Latent world models for intrinsically motivated exploration. *Advances in Neural Information Processing Systems* 33 (2020), 5565–5575.

[11] Meire Fortunato, Mohammad Gheshlaghi Azar, Bilal Piot, Jacob Menick, Ian Osband, Alex Graves, Vlad Mnih, Remi Munos, Demis Hassabis, Olivier Pietquin, et al. 2017. Noisy networks for exploration. *arXiv preprint arXiv:1706.10295* (2017).

[12] Quentin Gallouédec and Emmanuel Dellandréa. 2023. Cell-free latent go-explore. In *International Conference on Machine Learning*. PMLR, 10571–10586.

[13] Aurélien Garivier and Eric Moulines. 2011. On upper-confidence bound policies for switching bandit problems. In *International Conference on Algorithmic Learning Theory*. Springer, 174–188.

[14] Yijie Guo, Jongwook Choi, Marcin Moczulski, Shengyu Feng, Samy Bengio, Mohammad Norouzi, and Honglak Lee. 2020. Memory based trajectory-conditioned policies for learning from sparse rewards. *Advances in Neural Information Processing Systems* 33 (2020), 4333–4345.

[15] Jakob Hollenstein, Sayantan Auddy, Matteo Saveriano, Erwan Renaudo, and Justus Piater. 2022. Action Noise in Off-Policy Deep Reinforcement Learning: Impact on Exploration and Performance. *Transactions on Machine Learning Research* (2022).

[16] Xing Hu, Rui Zhang, Ke Tang, Jiaming Guo, Qi Yi, Ruizhi Chen, Zidong Du, Ling Li, Qi Guo, Yunji Chen, et al. 2022. Causality-driven Hierarchical Structure Discovery for Reinforcement Learning. *Advances in Neural Information Processing Systems* 35 (2022), 20064–20076.

[17] Hyoungseok Kim, Jaekyeom Kim, Yeonwoo Jeong, Sergey Levine, and Hyun Oh Song. 2019. EMI: Exploration with Mutual Information. In *International Conference on Machine Learning*. PMLR, 3360–3369.

[18] Kuno Kim, Megumi Sano, Julian De Freitas, Nick Haber, and Daniel Yamins. 2020. Active world model learning with progress curiosity. In *International conference on machine learning*. PMLR, 5306–5315.

[19] Hung Le, Kien Do, Dung Nguyen, and Svetha Venkatesh. 2024. Beyond Surprise: Improving Exploration Through Surprise Novelty. *International Conference on Autonomous Agents and Multi-Agent Systems (To appear)* (2024).

[20] Hung Le, Thommen Karimpanal George, Majid Abdolshah, Dung Nguyen, Kien Do, Sunil Gupta, and Svetha Venkatesh. 2022. Learning to Constrain Policy Optimization with Virtual Trust Region. *Advances in Neural Information Processing Systems* 35 (2022), 12775–12786.

[21] Hung Le, Thommen Karimpanal George, Majid Abdolshah, Truyen Tran, and Svetha Venkatesh. 2021. Model-based episodic memory induces dynamic hybrid controls. *Advances in Neural Information Processing Systems* 34 (2021), 30313–30325.

[22] Hung Le and Svetha Venkatesh. 2022. Neurocoder: General-purpose computation using stored neural programs. In *International Conference on Machine Learning*. PMLR, 12204–12221.

[23] Phillip Lippe, Sara Magliacane, Sindy Löwe, Yuki M Asano, Taco Cohen, and Efstratios Gavves. 2023. Causal Representation Learning for Instantaneous and Temporal Effects in Interactive Systems. In *The Eleventh International Conference on Learning Representations*. https://openreview.net/forum?id=itZ6ggvMnzS

[24] Dung Nguyen, Phuoc Nguyen, Hung Le, Kien Do, Svetha Venkatesh, and Truyen Tran. 2022. Learning Theory of Mind via Dynamic Traits Attribution. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*. 954–962.

[25] Dung Nguyen, Phuoc Nguyen, Hung Le, Kien Do, Svetha Venkatesh, and Truyen Tran. 2023. Memory-augmented theory of mind network. In *AAAI Conference on Artificial Intelligence*.

[26] Junhyuk Oh, Yijie Guo, Satinder Singh, and Honglak Lee. 2018. Self-imitation learning. In *International Conference on Machine Learning*. PMLR, 3878–3887.

[27] Deepak Pathak, Pulkit Agrawal, Alexei A Efros, and Trevor Darrell. 2017. Curiosity-driven exploration by self-supervised prediction. In *International conference on machine learning*. PMLR, 2778–2787.

[28] Deepak Pathak, Dhiraj Gandhi, and Abhinav Gupta. 2019. Self-supervised exploration via disagreement. In *International conference on machine learning*. PMLR, 5062–5071.

[29] Daniel J Russo, Benjamin Van Roy, Abbas Kazerouni, Ian Osband, Zheng Wen, et al. 2018. A tutorial on thompson sampling. *Foundations and Trends® in Machine Learning* 11, 1 (2018), 1–96.

[30] Nikolay Savinov, Anton Raichuk, Damien Vincent, Raphael Marinier, Marc Pollefeys, Timothy Lillicrap, and Sylvain Gelly. 2018. Episodic Curiosity through Reachability. In *International Conference on Learning Representations*.

[31] Maximilian Seitzer, Bernhard Schölkopf, and Georg Martius. 2021. Causal influence detection for improving efficiency in reinforcement learning. *Advances in Neural Information Processing Systems* 34 (2021), 22905–22918.

[32] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. 2017. Mastering the game of go without human knowledge. *nature* 550, 7676 (2017), 354–359.

[33] Bradly C Stadie, Sergey Levine, and Pieter Abbeel. 2015. Incentivizing Exploration In Reinforcement Learning With Deep Predictive Models. *arXiv e-prints* (2015), arXiv–1507.

[34] Christopher Stanton and Jeff Clune. 2018. Deep curiosity search: Intra-life exploration can improve performance on challenging deep reinforcement learning problems. *arXiv preprint arXiv:1806.00553* (2018).

[35] Allison Tam, Neil Rabinowitz, Andrew Lampinen, Nicholas A Roy, Stephanie Chan, DJ Strouse, Jane Wang, Andrea Banino, and Felix Hill. 2022. Semantic exploration from language abstractions and pretrained representations. *Advances in Neural Information Processing Systems* 35 (2022), 25377–25389.

[36] Haoran Tang, Rein Houthooft, Davis Foote, Adam Stooke, OpenAI Xi Chen, Yan Duan, John Schulman, Filip DeTurck, and Pieter Abbeel. 2017. # exploration: A study of count-based exploration for deep reinforcement learning. *Advances in neural information processing systems* 30 (2017).

[37] Yao Yao, Li Xiao, Zhicheng An, Wanpeng Zhang, and Dijun Luo. 2021. Sample efficient reinforcement learning via model-ensemble exploration and exploitation. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 4202–4208.