



Universiteit
Leiden
The Netherlands

Enhancing autonomy and efficiency in goal-conditioned reinforcement learning

Yang, Z.

Citation

Yang, Z. (2025, February 26). *Enhancing autonomy and efficiency in goal-conditioned reinforcement learning*. SIKS Dissertation Series. Retrieved from <https://hdl.handle.net/1887/4196074>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4196074>

Note: To cite this publication please use the final published version (if applicable).

Enhancing Autonomy and Efficiency in Goal-Conditioned Reinforcement Learning

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op woensdag 26 februari 2025
klokke 16:00 uur

door

Yang, Zhao 杨昭

geboren te Ningxian, China

in 1996

Promotor:

Prof.dr. A. Plaat

Co-promotores:

Dr. T. M. Moerland

Dr. M. Preuss

Promotiecommissie:

Prof.dr. J. Batenburg

Prof.dr. M. Bonsangue

Dr. V. François-Lavet

Dr. J. Liu

Prof.dr. M. Spaan

Vrije Universiteit Amsterdam
Lingnan University Hong Kong
Delft University of Technology



Universiteit
Leiden



Copyright © 2024 Zhao Yang. All rights reserved.

This PhD project was conducted in the Reinforcement Learning Group, Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands.

SIKS Dissertation Series No. 2025-10. The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Research presented in this thesis was supported by the China Scholarship Council (CSC). CSC No. 202007720053.

ISBN: 978-94-6506-997-5

Printed by: Ridderprint

To my grandmother.

Contents

1 Introduction	1
1.1 Research questions	4
1.2 Outline of this thesis	6
1.3 List of publications	8
2 Preliminaries	9
2.1 Reinforcement learning	9
2.2 Goal-conditioned reinforcement learning	11
2.2.1 Define the goal space	12
2.2.2 Select a goal	13
2.2.3 Reach the goal	14
2.2.4 Post-explore	15
3 World Models Increase Autonomy in Reinforcement Learning	17
3.1 Introduction	19
3.2 Related Work	21
3.3 Preliminaries	23
3.3.1 Reset-free RL	23
3.3.2 Model-based RL setup	24
3.4 Method	25
3.4.1 Back-and-Forth Go-Explore	25
3.4.2 Learning to Achieve Relevant Goals in Imagination	26
3.4.3 Implementation Details	27
3.5 Experiments	27
3.5.1 Results	29
3.5.2 Analysis	30

Contents

3.5.3 Ablations	32
3.6 Conclusion and Future Work	33
3.7 Experimental Details	34
3.7.1 Environments	34
3.7.2 Baseline Implementations	36
3.7.3 Hyperparameters	36
3.7.4 Results Clarification	38
3.7.5 Resource Usage	38
3.8 More Visualizations on Replay Buffer	38
3.9 Detailed Ablations	38
3.10 MBRL on Sawyer Door	40
3.11 More Analysis on Fetch Environments	43
3.12 Analysis on R3L	44
4 Continuous Episodic Control	45
4.1 Introduction	47
4.2 Background	49
4.3 Related Work	50
4.4 Continuous Episodic Control (CEC)	51
4.5 Experiments	55
4.5.1 Toy Examples	55
4.5.2 Continuous Navigation Tasks	58
4.5.3 Robotics Control Task	59
4.6 Conclusion and Future Work	60
5 Two-Memory Reinforcement Learning	63
5.1 Introduction	65
5.2 Background	67
5.2.1 Markov Decision Process	67
5.2.2 Deep Q-Learning	67
5.2.3 Episodic Control	68
5.3 Related Work	68
5.3.1 Episodic Control	69
5.3.2 Episodic Memory for Learning	69
5.3.3 Experience Replay	69
5.4 Two-Memory Reinforcement Learning (2M)	70
5.4.1 A Motivating Example	71

5.4.2 Switching	73
5.4.3 Learning	74
5.5 Experiments	74
5.5.1 Proof of Concept Under Tabular RL Setting	75
5.5.2 Results on MinAtar Games	76
5.5.3 Ablation Study	78
5.6 Conclusion and Future Work	81
6 First Go, then Post-Explore	83
6.1 Introduction	85
6.2 Related Work	86
6.3 Background	87
6.4 Methods	89
6.4.1 IMCEP	89
6.4.2 Post-exploration	91
6.5 Experiments	92
6.5.1 Experimental Setup	92
6.5.2 Hyper-parameter Settings	92
6.5.3 Results	94
6.6 Conclusion and Future Work	96
7 Conclusion	99
7.1 Answers to research questions	99
7.2 Limitations	102
7.2.1 Limitations of model-based reset-free methods	102
7.2.2 Limitations of episodic control methods	103
7.2.3 Limitations of post-exploration	105
7.3 Reflections on goal-conditioned reinforcement learning	106
7.4 Future research	107
7.5 Conclusion	108
Acknowledgements	123
Summary	125
Samenvatting	127
Titles in the SIKS dissertation series since 2016	129

Contents

Curriculum Vitae	147
-------------------------	-----