Extended Abstract: Contextualize Me – The Case for Context in Reinforcement Learning

Carolin Benjamins *
Institute for Artificial Intelligence
Leibniz University Hannover
c.benjamins@ai.uni-hannover.de

Theresa Eimer *
Institute for Artificial Intelligence
Leibniz University Hannover
t.eimer@ai.uni-hannover.de

Frederik Schubert
Institute for Information Processing
Leibniz University Hannover
schubert@tnt.uni-hannover.de

Aditya Mohan
Institute for Artificial Intelligence
Leibniz University Hannover
a.mohan@ai.uni-hannover.de

Sebastian Döhler
Institute for Information Processing
Leibniz University Hannover
doehler@tnt.uni-hannover.de

André Biedenkapp
Department of Computer Science
Albert-Ludwigs University Freiburg
biedenka@cs.uni-freiburg.de

Bodo Rosenhahn
Institute for Information Processing
Leibniz University Hannover
rosenhahn@tnt.uni-hannover.de

Frank Hutter
Department of Computer Science
Albert-Ludwigs University Freiburg
fh@cs.uni-freiburg.de

Marius Lindauer
Institute for Artificial Intelligence
Leibniz University Hannover
m.lindauer@ai.uni-hannover.de

This is an Extended Abstract of a Paper accepted for publication at TMLR.

1 Motivation

While Reinforcement Learning (RL) has shown successes in a variety of domains, including game playing [23, 5], robot manipulation [15, 19] and nuclear fusion [8], modern RL algorithms are not designed with generalization in mind, making them brittle when faced with even slight variations of their environment [28, 18, 16]. To address this limitation, recent research has increasingly focused on the generalization capabilities of RL agents. Ideally, general agents should be capable of zero-shot transfer to previously unseen environments and robust to changes in the prob-

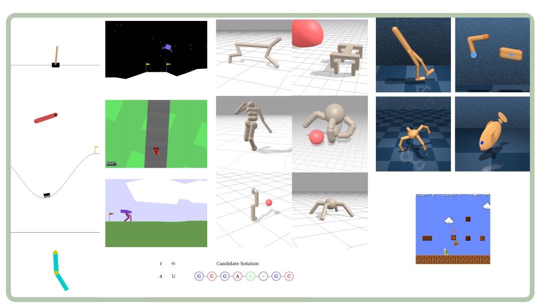

Figure 1: The CARL benchmarks

*Equal Contribution

16th European Workshop on Reinforcement Learning (EWRL 2023).

lem setting while interacting with an environment [20, 12, 7, 29, 10, 27, 1, 24, 2, 13]. Steps in this direction have been taken by proposing new problem settings where agents can test their transfer performance, e.g. the Arcade Learning Environment's flavors [17] or benchmarks utilizing Procedural Content Generation (PCG) to increase task variation, e.g. ProcGen [7], NetHack [14] or Alchemy [26].

While these extended problem settings in RL have expanded the possibilities for benchmarking agents in diverse environments, the degree of task variation is often either unknown or cannot be controlled precisely. We believe that generalization in RL is held back by these factors, stemming in part from a lack of problem formalization [13]. In order to facilitate generalization in RL, contextual RL (cRL) proposes to explicitly take environment characteristics, the so-called *context* [11], into account. This inclusion enables precise design of train and test distributions with respect to this context. Thus, cRL allows us to reason about the generalization capabilities of RL agents and to quantify their generalization performance. Overall, cRL provides a framework for both theoretical analysis and practical improvements.

In order to empirically study cRL, we introduce our benchmark library `CARL`, short for Context-Adaptive Reinforcement Learning. `CARL` collects well-established environments from the RL community and extends them with the notion of context. We use our benchmark library to empirically show how different context variations can drastically increase the difficulty of training RL agents, even in simple environments. We further verify the intuition that allowing RL agents access to context information is beneficial for generalization tasks in theory and practice.

2 Providing Benchmarks for cRL Research

Our benchmark library `CARL` extends common RL benchmarks with context information. This enables researchers to use well-known problem settings, but also define meaningful train and test distributions. In our release of `CARL` benchmarks, we include and contextually extend classic control and box2d environments from OpenAI Gym [6], Google Brax' walkers [9], a selection from the DeepMind Control Suite [25], an RNA folding environment [21] as well as Super Mario levels [4, 22], see Figure 1.

We provide discrete as well as continuous environments with vector-based as well as image-based observation spaces. The difficulty of the environments ranges from relatively simple in classic control tasks to very hard in Mario or RNA. The context information for most of these benchmarks is comprised of physical properties like friction or mass, making it intelligible to humans as well as relevant in practical tasks. Overall, `CARL` focuses on popular environments and will grow over time, increasing the diversity of benchmarks. Already now, `CARL` provides a vast benchmark collection for generalization and provides a platform for reproducible research.

3 The Value of Context in Practice

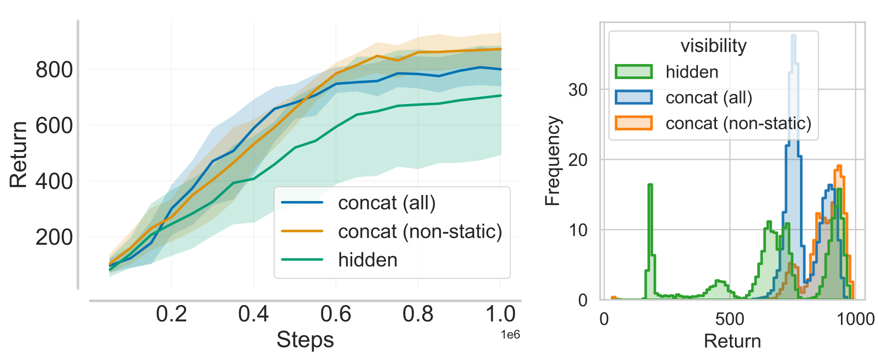

Figure 2: Train (lineplot) and test performance (histogram) of agents with visible and hidden context on CARLDmcWalker with different viscosity values and 5 seeds. Shown is the mean performance with a 95% confidence interval and testing across 200 test contexts (metrics are computed using stratified resampled bootstrapping [3]). Here, concatenating context to the state seems beneficial.

We demonstrate that varying context increases the difficulty of the task, already for simple environment settings. Varying the physical properties of classic control tasks like Pendulum or CartPole results in diverse environment dynamics that agents trained on single contexts fail to generalize to. Even when we train the agent on context variations, this general agent does not match the performance of specialized agents (trained on one context each). For example, on CartPole the general agent reaches the solution threshold on 40% fewer contexts than separate general agents. We investigate if this effect can be mitigated by providing context information to the agent in a naive manner by concatenating the context to the state (visible). Agents without access to context information are denoted by hidden. We see that agents with context information concatenated to the state tend to learn faster and reach higher final performances across most environments in our experiments (see Figure 2 for an example). When evaluating these agents on different variations of interpolation and extrapolation settings, we see that the test behavior changes remarkably when the agent is context-aware, focusing on the training distribution and thus becoming more predictable.

4 Concluding Remarks

Towards our goal of creating general and robust agents, we need to factor in possible changes in the environment. We propose modeling these changes with the framework of contextual Reinforcement Learning (cRL) in order to better reason about what demands cRL introduces to the agents and the learning process, specifically regarding the suboptimal nature of conventional RL policies in cRL. With `CARL`, we provide a benchmark library that contextualizes popular benchmarks and is designed to study generalization in cRL. It allows us to empirically demonstrate that contextual changes disturb learning even in simple settings and that the final performance and the difficulty correlate with the magnitude of the variation. We also verify that context-oblivious policies are not able to fully solve even simple contextual environments. Furthermore, our results suggest that exposing the context to agents even in a naive manner impacts the generalization behavior, in some cases improving training and test performance compared to context-oblivious agents. We expect this to be a first step towards better solution mechanisms for contextual RL problems and therefore one step closer to general and robust RL agents.

References

[1] M. Abdolshah, H. Le, T. K. George, S. Gupta, S. Rana, and S. Venkatesh. A new representation of successor features for transfer across dissimilar environments. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 1–9. PMLR, 18–24 Jul 2021.

[2] S. Adriaensen, A. Biedenkapp, G. Shala, N. Awad, T. Eimer, M. Lindauer, and F. Hutter. Automated dynamic algorithm configuration. *J. Artif. Intell. Res.*, 75:1633–1699, 2022.

[3] R. Agarwal, M. Schwarzer, P. Castro, A. Courville, and M. Bellemare. Deep reinforcement learning at the edge of the statistical precipice. In *Proceedings of the 35th International Conference on Advances in Neural Information Processing Systems (NeurIPS'21)*, pages 29304–29320, 2021.

[4] M. Awiszus, F. Schubert, and B. Rosenhahn. TOAD-GAN: Coherent style level generation from a single example. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 16, Oct. 2020.

[5] A. Badia, B. Piot, S. Kapturowski, P. Sprechmann, A. Vitvitskyi, Z. Guo, and C. Blundell. Agent57: Outperforming the atari human benchmark. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 507–517. PMLR, 2020.

[6] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba. Openai gym. *CoRR*, abs/1606.01540, 2016.

[7] K. Cobbe, C. Hesse, J. Hilton, and J. Schulman. Leveraging procedural generation to benchmark reinforcement learning. In *Proceedings of the 37th International Conference on Machine Learning, ICML*, volume 119 of *Proceedings of Machine Learning Research*, pages 2048–2056. PMLR, 2020.

[8] J. Degrave, F. Felici, J. Buchli, M. Neunert, B. Tracey, F. Carpanese, T. Ewalds, R. Hafner, A. Abdolmaleki, D. de Las Casas, C. Donner, L. Fritz, C. Galperti, A. Huber, J. Keeling, M. Tsimpoukelli, J. Kay, A. Merle, J. Moret, S. Noury, F. Pesamosca, D. Pfau, O. Sauter, C. Sommariva, S. Coda, B. Duval, A. Fasoli, P. Kohli, K. Kavukcuoglu, D. Hassabis, and M. Riedmiller. Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature*, 602(7897):414–419, 2022.

[9] C. Freeman, E. Frey, A. Raichuk, S. Girgin, I. Mordatch, and O. Bachem. Brax - A differentiable physics engine for large scale rigid body simulation. *CoRR*, abs/2106.13281, 2021.

[10] X. Fu, G. Yang, P. Agrawal, and T. Jaakkola. Learning task informed abstractions. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 3480–3491. PMLR, 18–24 Jul 2021.

[11] A. Hallak, D. D. Castro, and S. Mannor. Contextual markov decision processes. *arXiv:1502.02259 [stat.ML]*, 2015.

[12] P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger. Deep reinforcement learning that matters. In S. McIlraith and K. Weinberger, editors, *Proceedings of the Conference on Artificial Intelligence (AAAI'18)*. AAAI Press, 2018.

[13] R. Kirk, A. Zhang, E. Grefenstette, and T. Rocktäschel. A survey of zero-shot generalisation in deep reinforcement learning. *J. Artif. Intell. Res.*, 76:201–264, 2023.

[14] H. Küttler, N. Nardelli, A. Miller, R. Raileanu, M. Selvatici, E. Grefenstette, and T. Rocktäschel. The nethack learning environment. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*, 2020.

[15] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter. Learning quadrupedal locomotion over challenging terrain. *Science in Robotics*, 5, 2020.

[16] M. Lu, Z. Shahn, D. Sow, F. Doshi-Velez, and L. H. Lehman. Is deep reinforcement learning ready for practical applications in healthcare? A sensitivity analysis of duel-ddqn for hemodynamic management in sepsis patients. In *AMIA 2020, American Medical Informatics Association Annual Symposium, Virtual Event, USA, November 14-18, 2020*. AMIA, 2020.

[17] M. Machado, M. Bellemare, E. Talvitie, J. Veness, M. Hausknecht, and M. Bowling. Revisiting the arcade learning environment: Evaluation protocols and open problems for general agents. *J. Artif. Intell. Res.*, 61:523–562, 2018.

[18] T. Meng and M. Khushi. Reinforcement learning in financial markets. *Data*, 4(3):110, 2019.

[19] K. Ploeger, M. Lutter, and J. Peters. High acceleration reinforcement learning for real-world juggling with binary rewards. In J. Kober, F. Ramos, and C. J. Tomlin, editors, *4th Conference on Robot Learning, CoRL 2020, 16-18 November 2020, Virtual Event / Cambridge, MA, USA*, volume 155 of *Proceedings of Machine Learning Research*, pages 642–653. PMLR, 2020.

[20] M. Ponsen, M. Taylor, and K. Tuyls. Abstraction and generalization in reinforcement learning: A summary and framework. In *Adaptive and Learning Agents, Second Workshop, ALA 2009, Held as Part of the AAMAS 2009 Conference in Budapest*, volume 5924 of *Lecture Notes in Computer Science*, pages 1–32. Springer, 2009.

[21] F. Runge, D. Stoll, S. Falkner, and F. Hutter. Learning to Design RNA. In *Proceedings of the International Conference on Learning Representations (ICLR'19)*, 2019. Published online: `iclr.cc`.

[22] F. Schubert, M. Awiszus, and B. Rosenhahn. Toad-gan: a flexible framework for few-shot level generation in token-based games. *IEEE Transactions on Games*, pages 1–1, 2021.

[23] D. Silver, A. Huang, C. Maddison, A. Guez, L. Sifre, G. Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

[24] S. Sodhani, A. Zhang, and J. Pineau. Multi-task reinforcement learning with context-based representations. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 9767–9779. PMLR, 18–24 Jul 2021.

[25] Y. Tassa, Y. Doron, A. Muldal, T. Erez, Y. Li, D. de Las Casas, D. Budden, A. Abdolmaleki, J. Merel, A. Lefrancq, T. Lillicrap, and M. Riedmiller. Deepmind control suite. *CoRR*, abs/1801.00690, 2018.

[26] J. Wang, M. King, N. Porcel, Z. Kurth-Nelson, T. Zhu, C. Deck, P. Choy, M. Cassin, M. Reynolds, H. Song, G. Buttimore, D. Reichert, N. Rabinowitz, L. Matthey, D. Hassabis, A. Lerchner, and M. Botvinick. Alchemy: A benchmark and analysis toolkit for meta-reinforcement learning agents. In J. Vanschoren and S. Yeung, editors, *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks 1, NeurIPS Datasets and Benchmarks*, 2021.

[27] D. Yarats, R. Fergus, A. Lazaric, and L. Pinto. Reinforcement learning with prototypical representations. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 11920–11931. PMLR, 18–24 Jul 2021.

[28] T. Yu, D. Quillen, Z. He, R. Julian, K. Hausman, C. Finn, and S. Levine. Meta-world: A benchmark and evaluation for multi-task and meta reinforcement learning. In *Conference on Robot Learning (CoRL)*, 2019.

[29] A. Zhang, S. Sodhani, K. Khetarpal, and J. Pineau. Learning robust state abstractions for hidden-parameter block mdps. In *9th International Conference on Learning Representations, ICLR 2021*. OpenReview.net, 2021.