Prof. Bartosz Zieliński
Faculty of Mathematics and Computer Science
Jagiellonian University
bartosz.zielinski@uj.edu.pl

Kraków, December 29, 2025

Review of the doctoral dissertation by
Michał Nauman, MSc
entitled
"Sample-efficient actor–critic algorithms in reinforcement learning"

1. Research problem and its significance

In his dissertation, the doctoral candidate explores how reinforcement learning methods can learn an effective policy through minimal interaction with the environment. He introduces novel methods and conducts experiments to demonstrate their effectiveness.

2. Author's contribution

This dissertation focuses on the issue of sample efficiency in Reinforcement Learning (RL). It is a timely and important topic, since real-world interactions can be costly, time-consuming or risky. For example, when training autonomous vehicles or medical robots. The dissertation consists of six high quality publications. However, their descriptions in Section 4 are definitely too short, too technical, and understandable only for the experts in the field. I missed the figures, which could explain the motivation and solution idea in a way comprehensible for machine learning scientists without in-depth RL knowledge. Apart from this, however, the content is quite clear.

The contributions presented in the dissertation can be summarised as follows:

- **P1:** The doctoral candidate proposed Model-Based Many-Actions (MBMA), an approach that leverages dynamic models for many-actions sampling in the context of Stochastic Policy Gradients (SPG). MBMA addresses issues associated with existing implementations of many-actions SPG, and yields lower bias and comparable variance to SPG estimated from states in model-simulated rollouts.
- **P2:** He tested multiple regularization techniques from recent, state-of-the-art algorithms across several diverse tasks. He revealed that certain combinations of regularizations consistently demonstrate robust and superior performance. Moreover, he showed that a simple Soft Actor-Critic agent, appropriately regularized, finds a policy performing similarly to those obtained with model-based approaches.

- **P3:** He proposed Decoupled Policy Actor-Critic (DAC), a model-free algorithm that features two distinct actor networks: a pessimistic actor for temporal-difference learning and an optimistic actor for exploration. DAC requires significantly fewer computational resources and matches the performance of leading model-based methods.
- **P4:** He showed that the critic error can be approximated via a recursive, fixed-point model, similar to that of the Bellman value. Based on that, he proposed Validation Pessimism Learning (VPL), an algorithm which uses a small validation buffer to adjust the level of pessimism throughout the agent's training.
- **P5:** He proposed the BRO (Bigger, Regularized, Optimistic) algorithm. The key insight behind BRO is that strong regularization allows for effective scaling of the critic networks, which, paired with optimistic exploration, leads to superior performance.
- **P6:** He showed that value-based off-policy RL methods are predictable, i.e., it is possible to predict the performance of large-scale runs based on the performance of small-scale runs. He also demonstrated that data and compute requirements to attain a given performance level lie on a Pareto frontier, controlled by the updates-to-data (UTD) ratio. Finally, he determined the optimal allocation of a total resource budget across data and compute for a given performance.

The obtained results have been included in the proceedings of four very prestigious conferences, which took place between 2023 and 2025. Three publications at ICML (CORE A*) and one at AAAI (CORE A*), IJCAI (CORE A*), and NeurIPS (CORE A*). In each of them, the doctoral candidate is the first author, which is an extraordinary achievement.

3. Correctness

The dissertation serves as a guide to the doctoral candidate aforementioned publications, presented at six top-tier conferences. Their detailed descriptions in Chapter 4 are preceded by a very nice introduction in Chapter 1, general motivation in Chapter 2, and contributions listed in Chapter 3. The dissertation ends with conclusions and future work in Chapter 5.

Reading the dissertation did not lead to any specific questions, as the publications contain very detailed descriptions. However, I would love to see some videos that show how the proposed algorithms behave for specific tasks, such as those presented in Figure 3 of P5 (DeepMind Control, MetaWorld, and MyoSuite). I would also like to ask the doctoral candidate to present the simplified motivation and intuition for each of the papers so that it is understandable for the whole committee during the defense.

4. Candidate's knowledge

Based on the reading, I believe that the doctoral candidate possesses well-established knowledge in the area of reinforcement learning. The candidate correctly employs advanced

mathematical tools to introduce novel methods and is able to plan and conduct experiments to demonstrate their effectiveness.

A review of the literature from the papers of the doctoral candidate allows me to conclude that he has up-to-date knowledge of the dissertation subject and is capable of performing a critical review of sources in order to identify interesting research directions.

5. Summary

The reviewed dissertation presents solutions to important and original problems, constituting an example of how reinforcement learning methods can learn an effective policy through minimal interaction with the environment.

The remarks in Section 3 of this review are purely discussion-oriented and do not diminish the results achieved by the doctoral candidate, nor do they affect the overall positive assessment of the submitted dissertation.

Therefore, in view of the above, I conclude that the doctoral dissertation by Michał Nauman, MSc, entitled "Sample-efficient actor-critic algorithms in reinforcement learning", **meets the requirements** of the Act on Higher Education and Science (Journal of Laws 2024, item 1571). Consequently, I recommend its acceptance and the admission of Michał Nauman, MSc, to the public defense.

Furthermore, due to the exceptionally high quality of the work and its publication at six CORE A* conferences, I recommend that the doctoral dissertation be awarded **with distinction**.

Yours sincerely,

Podpisany elektronicznie przez Bartosz Michał Zieliński 29.12.2025 15:19:32 +01'00'