Research Evaluation Document

Name: Srinjoy Roy Student Code: T026
Submitted To: Research Fellow Advisory Committee (RFAC)

This document has been prepared for evaluating the progress made over the past eighteen months spent in the Institute for Advancing Intelligence (IAI), TCG-CREST.

1 Courseworks

Semester	Name of Courses	Marks Obtained (%)
First (IAI)	Artificial Intelligence & Machine Learning 1	93.0
	Cryptology & Security 1	91.0
	Automata & Formal Languages	89.0
	Discrete Mathematics	87.0
Second (IAI)	Artificial Intelligence & Machine Learning 2	87.0
	Design & Analysis of Algorithms	95.0
	Quantum Information & Cryptology	85.0
	Research Methodology	78.0
First (CMI)	Mathematical Logic	80.93

Table 1: Attended courses and the corresponding marks obtained (in percentage).

2 Work done so far

Title: Opportunistic Actor-Critic with Clipped Triple Q-learning.
Authors: Srinjoy Roy (IAI, TCG-CREST), Saptam Bakshi (IAI, TCG-CREST), Br. Tamal Maharaj (RKMVERI Belur).

Summary: Reinforcement Learning (RL) considers the paradigm of an agent interacting with its environment, and having the aim of learning reward-maximizing behavior. Actor-Critic methods, a type of model-free RL, have achieved state-of-the-art performance in many real-world continuous control domains. Despite their success, the wide-scale deployment of these models is still a far cry. Soft Actor-Critic (SAC) and Twin Delayed Deep Deterministic Policy Gradient (TD3), two of the best model-free deep RL algorithms in the past years, are based on actor-critic framework. SAC effectively addressed the problems of sample complexity and convergence brittleness to hyper-parameters. It outperformed all state-of-the-art algorithms including TD3 in harder tasks, whereas TD3 produced moderate results in all environments.

However, SAC suffers from inefficient exploration owing to the Gaussian nature of its policy which causes borderline performance in simpler tasks. The goal of this research is to introduce Opportunistic Actor-Critic (OPAC), an ensemble model-free deep RL algorithm that combines the central features of TD3 and SAC. It employs Clipped Triple Q-learning to fine-tune its value estimates. We have systematically evaluated OPAC on MuJoCo environments where the results show that OPAC performs consistently well in easy and challenging tasks. It also shows that OPAC outperforms TD3 and SAC in terms of average reward accumulated over time.

Submitted To: To be submitted very soon.

3 Future Work

Primary Topics: Differential Evolution (DE), Neural Architecture Search (NAS), and Reinforcement Learning (RL).

Summary: Differential Evolution (DE) is one of the best known stochastic real-parameter optimization algorithm. It falls under the class of gradient-free optimization algorithms and essentially follows an Evolutionary Algorithm (EA) style approach. The main difference between DE and EA is in their workflow. The classical EA pipeline is "Initialization → Crossover → Mutation → Selection" while DE follows "Initialization → Mutation → Crossover → Selection". Unlike EA, DE and it's variants perform a "difference vector" based mutation which is why it's named so.

Neural Architecture Search (NAS) is a sub-field of Automated Machine Learning (AutoML) that deals with automatically finding out best neural architectures for a specific set of tasks. The existing and most successful neural architectures in deep learning have been manually hand-engineered and fine-tuned by experts. But these methods are exhaustive, costly, and highly prone to human errors. This is exactly where the motivation of performing NAS lies - to have better, faster, and cost-effective ways of finding out neural architectures.

My Ph.D. is being supervised by Dr. Swagatam Das from ISI, Kolkata. Under his guidance, I am presently going through seminal papers on DE and NAS and implementing them in Python. This is to facilitate hands-on experience about the basic framework of DE (and it's successors), NAS benchmarks, and handling complex data-sets. Our objective is to find gaps in the literature so that we may use the tools at our disposal to solve interesting problems. In other words, we are aiming to come up with efficient state-of-the-art NAS methods using DE and/or RL.

4 Papers Read

[1] Noor H. Awad, Neeratyoy Mallik, and Frank Hutter. Differential evolution for neural architecture search. *CoRR*, abs/2012.06400, 2020.

[2] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. Openai gym, 2016.

[3] Kamil Ciosek, Quan Vuong, Robert Loftin, and Katja Hofmann. Better exploration with optimistic actor-critic, 2019.

[4] Swagatam Das and Ponnuthurai Nagaratnam Suganthan. Differential evolution: A survey of the state-of-the-art. *IEEE Transactions on Evolutionary Computation*, 15(1):4–31, 2011.

[5] Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. Neural architecture search: A survey. *J. Mach. Learn. Res.*, 20(1):1997–2017, jan 2019.

[6] Scott Fujimoto, Herke van Hoof, and David Meger. Addressing function approximation error in actor-critic methods. *CoRR*, abs/1802.09477, 2018.

[7] Shixiang Gu, Ethan Holly, Timothy P. Lillicrap, and Sergey Levine. Deep reinforcement learning for robotic manipulation. *CoRR*, abs/1610.00633, 2016.

[8] Tuomas Haarnoja, Aurick Zhou, Kristian Hartikainen, George Tucker, Sehoon Ha, Jie Tan, Vikash Kumar, Henry Zhu, Abhishek Gupta, Pieter Abbeel, and Sergey Levine. Soft actor-critic algorithms and applications. *CoRR*, abs/1812.05905, 2018.

[9] Jiafei Lyu, Xiaoteng Ma, Jiangpeng Yan, and Xiu Li. Efficient continuous control with double actors and regularized critics. *CoRR*, abs/2106.03050, 2021.

[10] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, February 2015.

[11] E. Todorov, T. Erez, and Y. Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5026–5033, 2012.

[12] Chris Ying, Aaron Klein, Esteban Real, Eric Christiansen, Kevin Murphy, and Frank Hutter. Nas-bench-101: Towards reproducible neural architecture search. *CoRR*, abs/1902.09635, 2019.

[13] Barret Zoph and Quoc V. Le. Neural architecture search with reinforcement learning. *CoRR*, abs/1611.01578, 2016.