DEEP REINFORCEMENT LEARNING: A STATE-OF-THE-ART WALKTHROUGH

Aristotelis Lazaridis[1], Anestis Fachantidis[2], Ioannis Vlahavas[1]
[1]Intelligent Systems Lab, School of Informatics, Aristotle University of Thessaloniki
[2]Medoid AI

IJCAI 2021 MONTREAL

Abstract

- The power of Deep Reinforcement Learning (Deep RL) methods comes from the combination of an already established and strong field of Deep Learning, with the unique nature of Reinforcement Learning methods.
- It is, however, deemed necessary to provide a compact, accurate and comparable view of these methods and their results for the means of gaining valuable technical and practical insights.
- In this work we gather the essential methods related to Deep RL, extracting common property structures for three complementary core categories: a) Model-Free, b) Model-Based and c) Modular algorithms.
- We discuss the key differences of the various kinds of algorithms, indicate their potential and limitations, as well as provide insights to researchers regarding future directions of the field.

Introduction

- Reinforcement Learning (RL) is built upon the principles of learning via agent-environment interaction through actions and rewards, and achieved never-before-seen breakthroughs when combined with Deep Learning methods.
- We perform an analysis of several state-of-the-art Deep RL algorithms and divide them into three main categories, according to their anatomy and core functionality:
  i. **Model-free**: algorithms that do not rely on learning the state transition probability function for solving the optimization task in hand
  ii. **Model-based**: approximation of model dynamics in order to reach a solution
  iii. **Modular algorithms**: algorithms that are used as modules for functional Deep RL algorithms so as to improve original performance and increase implementation flexibility and stability
- We briefly describe and perform a compact comparative analysis of the state-of-the-art algorithms for each of these categories
- We present higher-level comparison figures of their performance on 57 games included in the Arcade Learning Environment (ALE) and the MuJoCo physics simulation platform

Paper Contribution

- Primarily, this comparative survey is to be used as a guide for becoming acquainted with state-of-the-art Deep RL algorithms, knowing their pros and cons, their relations, performance capabilities, as well as distinguishing the cases where it's more appropriate for methods to be used
- Such work allows the Deep RL community and researchers in general (especially early-stage researchers) to view specific traits of state-of-the-art algorithms in a categorized manner, as well as distinguish architectural designs developed for the means of advancing the fields of Deep Learning and Deep RL
- This kind of survey also aims at giving insights regarding novel algorithm implementations and variations of existing algorithms
- Lastly, readers can develop intuition regarding capabilities and potential of the algorithms and their characteristics

Model-Free Algorithms

- Algorithms with a model-free nature constitute the epitome of a direct learning process through experience
- An agent within an environment attempts to learn the optimal policy for solving a task by directly transforming the experience gathered as a result of performed actions, into a resulting policy
- Two main categories of the model-free family:
  i. **Deep Q-Learning**: approximation of the optimal action-value function $Q^*(s,a)$ through the use of deep neural networks
  ii. **Policy Gradient**: algorithms which make use of the Policy Gradient Theorem

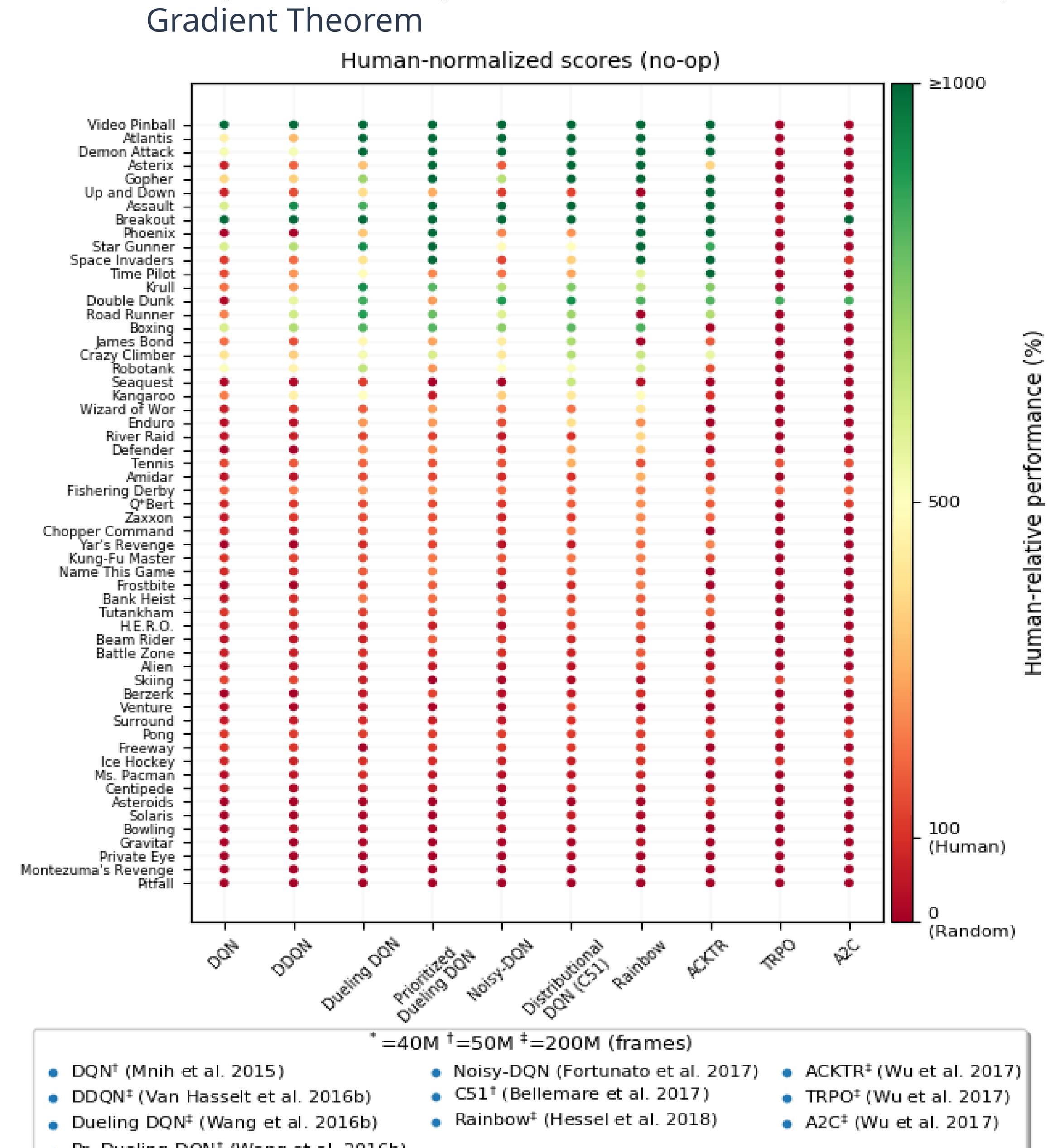

Figure 1 (above): Heatmap visualization of performance comparison between model-free algorithms in Atari games.

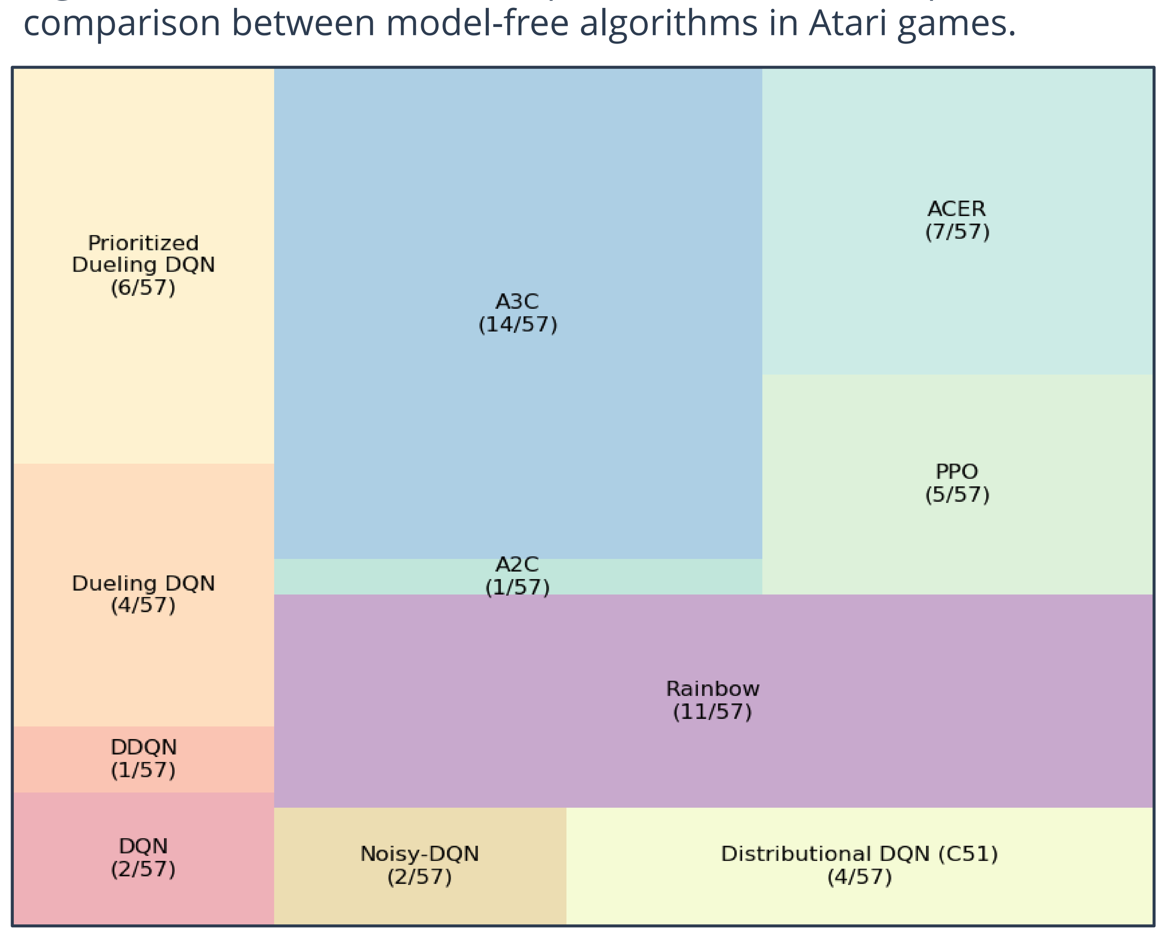

Figure 2 (below): TreeMap visualization representing the number of wins for each model-free algorithm in Atari games.

Model-Based Algorithms

- Model-based learning methods attempt to learn the underlying environment stochastic dynamics using the experience gained to accurately predict future states
- Then, planning methods are used to solve the given task
- Two main categories of the model-based family:
  i. **Manifest state-based**: algorithms that use only manifested (i.e. directly observable) states to create the model of the environment
  ii. **Latent state-based**: algorithms that find latent (i.e. non-directly observable) states to create the model of the environment

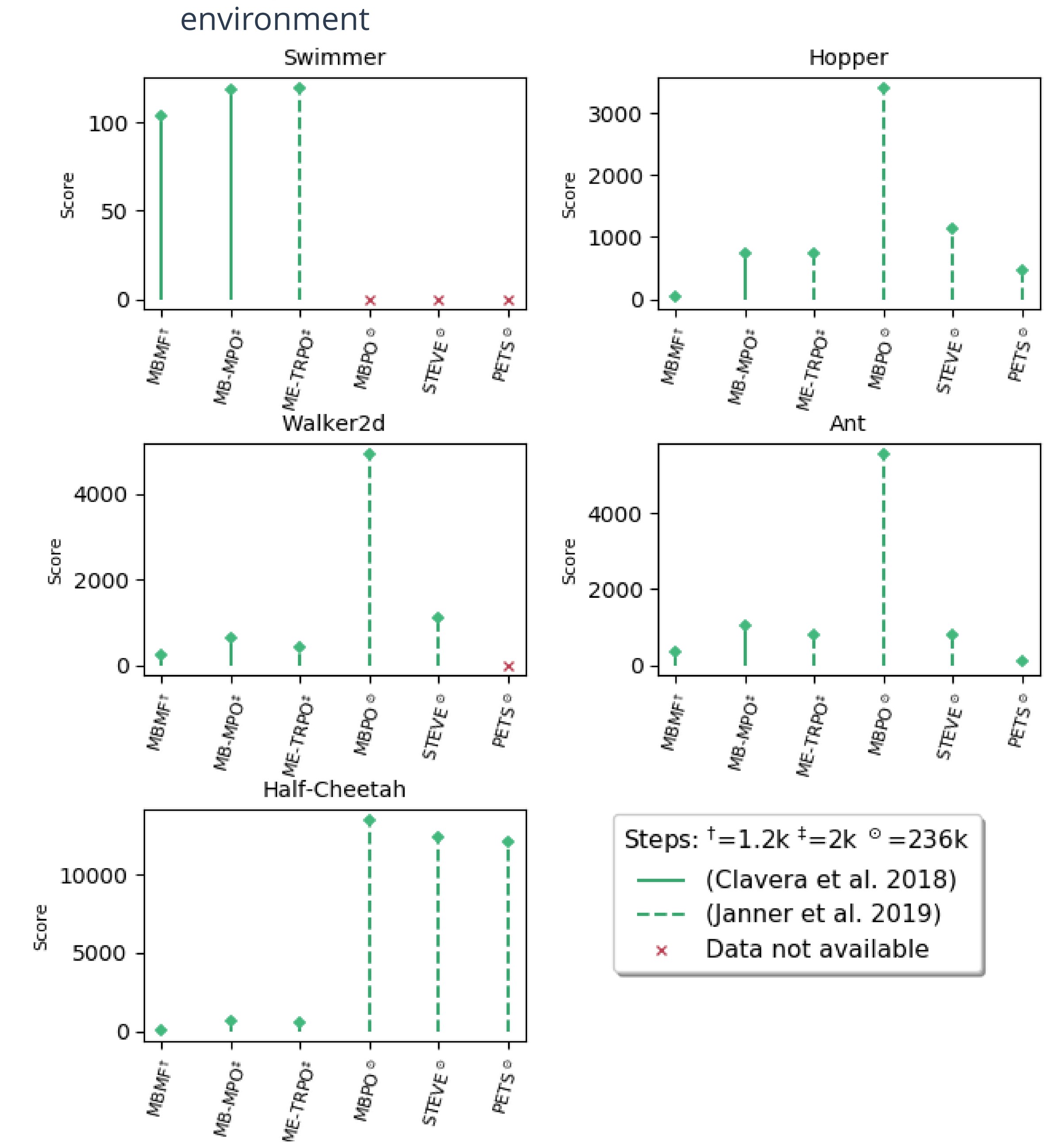

Figure 3: Performance of model-based algorithms on various MuJoCo tasks.

Discussion

- Policy Gradient methods are more sample-efficient than DQN-based methods, but they require a stronger theoretical background to extend existing architectures
- Model-based algorithms are far more sample efficient than model-free methods, but the lack of a core architecture makes it difficult to get involved with this category in general
- Novel combinations of modular frameworks and core algorithms can be implemented, while multiple frameworks can be used at once as well

Full Paper

Lazaridis, A., Fachantidis, A., & Vlahavas, I. (2020). Deep Reinforcement Learning: A State-of-the-Art Walkthrough. *Journal of Artificial Intelligence Research*, *69*, 1421-1471.

Modular Algorithms

- This family of algorithms is composed of frameworks designed to host model-free or model-based algorithms
- These frameworks aim to improve the injected algorithms in particular aspects, as in improving the strategy used for exploration, and/or equip them with various new features
- The following types of modular algorithms are distinguished:
  i. **Distributed frameworks**: ability of an algorithm to work in a distributed fashion
  ii. **Exploration frameworks**: efficient exploration strategies
  iii. **Unsupervised frameworks**: when extrinsic rewards are nonexistent, the agent crafts an intrinsic reward function
  iv. **Hierarchical frameworks**: intrinsic sub-goals and behaviors are generated, which solve given environments
  v. **Generalization frameworks**: the ability of an agent to function efficiently in new, unseen tasks and environments

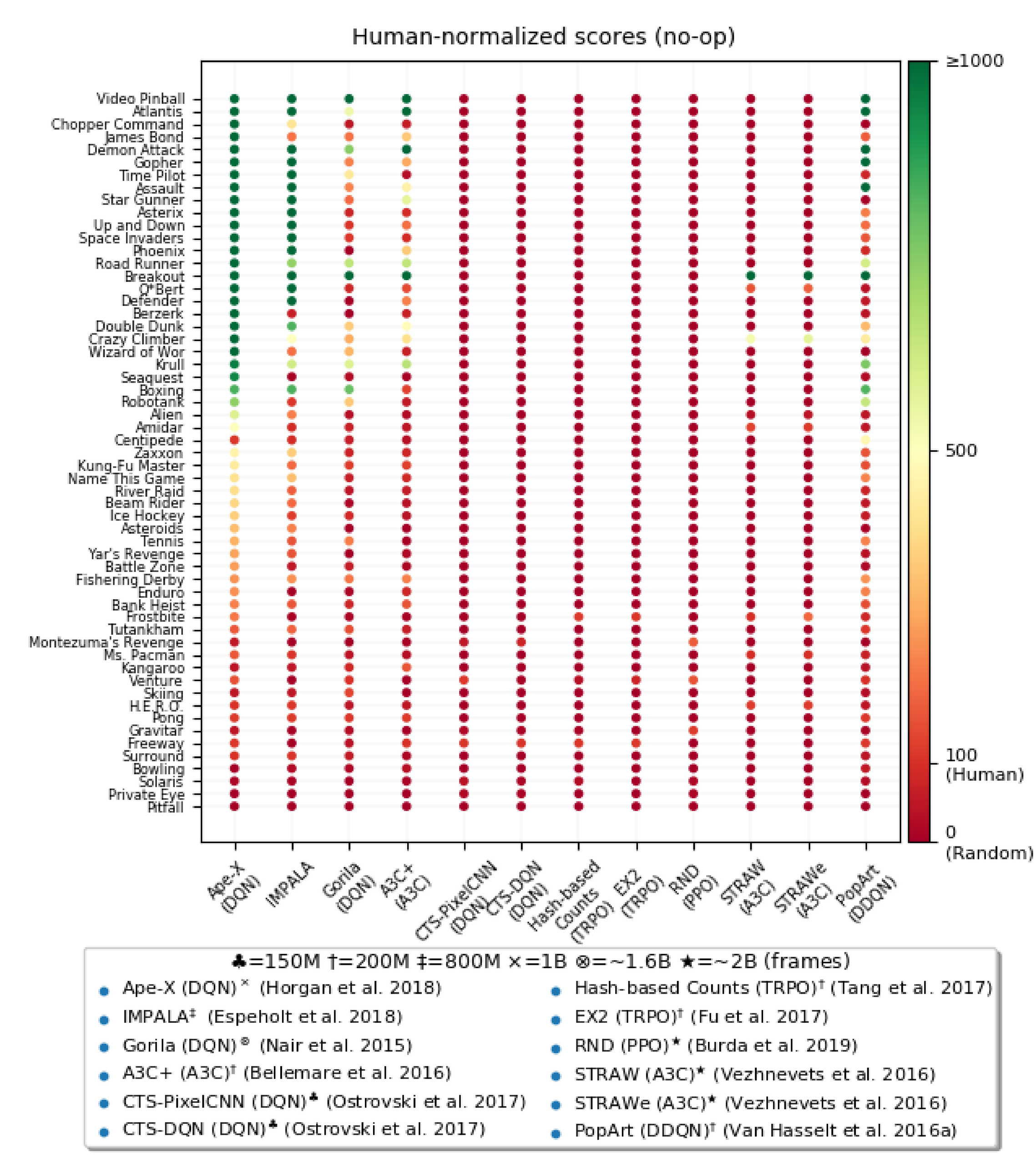

Figure 4: Heatmap visualization of performance comparison between Deep RL algorithms equipped with modular frameworks.

Conclusion

- Deep RL has expanded tremendously as a field and continues to grow rapidly, achieving super-human performance on various environments, and more specifically, video-games
- Onset of an era where competitiveness in gaming shifts from being human-controlled to bot-controlled
- Rise of new challenges in Deep RL and AI in general, towards building new foundations and protocols for AI, selecting the corresponding countermeasures, and/or fusing the realm of AI with our own reality