Towards Robust and Adaptable Real-World Decision Making Systems

Yanchao Sun

Introduction

Empowered by high-performance computing techniques and large volumes of datasets, machine learning (ML) has achieved significant and rapid development in a wide range of applications in the past decade. In particular, reinforcement learning (RL) imitates the natural learning process of humans and can give birth to intelligent agents for automatic (sequential) decision making. By learning from interactions and maximizing long-term utilities, RL agents are shown to achieve superhuman performance in many scenarios such as video games and board games [6, 8, 7]. However, the success of these learning-based intelligent agents is mainly evaluated in stable and stationary simulated environments, while there are still many challenges impeding the application of these intelligent agents in real-world scenarios.

Why is the real world challenging? We live in the era of information explosion, while *noisy, false or even adversarial information* is mixed with useful information needed for making decisions. Recent work has revealed that a well-trained agent could greatly fail when the input is slightly but adversarially perturbed [3, 2]. That is, although an agent can learn to perform well in a clean environment, it may not make correct decisions in a noisy environment, making it risky to deploy these agents in high-stakes real world applications. On the other hand, the real world we live in is *ever-changing, where previous experience and methods are often challenged by new and unseen problems.* For example, the unprecedented COVID-19 pandemic has caused dramatic changes of our lives in the past few years; everyone is more or less forced to adapt to a changed environment and adopt a new lifestyle. However, adapting to a new environment is not easy for intelligent agents as they require a large amount of data samples and computation to learn — it takes days if not months for face recognition systems to get used to masks on faces. Therefore, the ability of quickly and reliably adapt to changes in environment is critical for intelligent agents to make real-time decisions.

Overview of My Research. Towards building more trustworthy and more general intelligent systems, my research aims to understand and improve the robustness and adaptability of decision-making agents. In particular, I focus on the challenging sequential decision making problem and the RL regime, and propose formulations and solutions centered on the following aspects.

- **Robustness.** The interactions between the agent and the environment are subject to perturbations. For instance, an autonomous driving system [4] may suffer from GPS signal interference or wind turbulence. Motivated by this practical challenge, my research addresses the following questions. How vulnerable existing intelligent agents are against possibly noisy or even adversarially perturbed inputs [10, 14]? Can we make sure that the agent works as expected no matter how the input is perturbed [5, 13]?

- **Efficient Adaptability.** A trained agent could later face a shift of task specification (e.g., navigating to a new location) or changes of the underlying environment (e.g., navigating in a new city). In this case, traditional methods may discard the old model and train a new agent from scratch which is usually expensive and inefficient. On the contrary, my research makes the agent adapt to the new setting with the main idea of "learning by analogy" [9]. More concretely, when the task or the environment gets changed, my goal is to let the agent efficiently adapt to the new environment by transferring pre-learned knowledge [11, 15, 12].

The above two aspects are essentially complementary: robustness focuses on the *external environmental shifts* where outside noise or adversaries perturb the agent-environment interactions, while adaptability emphasizes the *internal environmental shifts* that directly changes the structure of the underlying environments. To address these two challenges, I have been working on multiple projects spanning reinforcement learning, adversarial learning, representation learning, transfer learning, and their intersections. Below I will introduce the selected works of mine as well as my future research agenda.

Current Research

Focus 1: Improving the Robustness of Sequential Decision Making. Different from simple and stable simulators (e.g. stand-alone video games), real-world applications consist of both natural noise and artificial/adversarial noise that perturb the interactions between RL agents and the environment. Although there are many studies on the robustness of supervised learning models, the robustness of RL agents is under-explored, and is challenging due to the environmental uncertainty and unknown reward accumulated over long horizons. I have conducted research based on the following two cases.

• *Robustness under Execution-Time Perturbation.* Perturbations can exist during the execution of a trained agent such that the agent fails to take correct actions. Is there a way to measure the robustness of an agent, and improve its robustness under any perturbations? As the old saying goes, "if you know yourself and your enemy, you will never lose a battle". We first systematically investigate the worst-case cumulative reward of any given agent with bounded observation perturbations [14]. We design an efficient and asymptotically optimal approach to learn such worst-case attacks, and it helps evaluate the robustness of a given agent. This work, recognized by a best paper award at a NeurIPS workshop, significantly outperforms prior adversarial attack methods for RL agents, and it reveals that existing RL models are extremely vulnerable to observation perturbations. Therefore, it is crucial to train robust RL agents for high-stakes applications. Then, we propose a novel and efficient robust training algorithm that achieves state-of-the-art robustness in multiple domains [5]. Different from prior robust training methods that are either driven by heuristics and not robust under strong attacks, or driven by strong attackers but too expensive in practice, our robust training algorithm efficiently learns a lower bound of the worst-case policy value, and improves this worst-case value by policy optimization. Moreover, we look beyond the commonly considered observation attacks with bounded ℓ_p perturbations. We formulate the communication attacks in multi-agent systems, and propose the first certifiable defense framework that is guaranteed to make robust decisions when a proportion of communication messages are (arbitrarily) perturbed [13].

• *Robustness for Training.* Perturbations can also happen during training time. It has been shown that an adversary can control the trained model to have certain (bad) behaviors by perturbing a small fraction of training data, which is known as poisoning attacks [1] and has been well-studied in supervised learning. However, an RL agent can be trained in an online manner where the training data is uncertain and unknown beforehand. It is unclear yet how robust an agent is against training-time perturbations. To tackle this problem, we establish the first unified formulation and practical algorithm of poisoning attacks for online on-policy RL learners [10]. The proposed method quantifies the robustness of RL agents against poisoning attacks, which facilitates more future study on robustness during training.

Focus 2: Improving the Adaptability of Reinforcement Learning. Most existing RL algorithms are developed for a pre-specified stationary environment. However, real-world environments and demands are subject to constant changes. Are RL agents able to efficiently adapt to new environments? For example, if we have an auto-driving car that works well in rural areas during the daytime, can it be fine-tuned to drive in urban areas during the nighttime? What if its sensory configurations get upgraded and result in a different observation space? Addressing these questions can lead to more versatile and general artificial intelligence, for which I have proposed a series of solutions.

• *Adapting by Knowledge Transfer.* A key to achieving adaptability is knowledge transfer from the known to the unknown, while it is not easy to let agents automatically decide what to transfer and how to transfer, and avoid negatively affecting future performance. We improve the adaptability of agents by theoretically grounded knowledge transfer on various levels. On the first level, we focus on adapting to unseen states within a single environment, and we develop an algorithm to "learn by analogy" with guaranteed efficiency [9]. On the second level, we transfer knowledge from task to task, by proposing a provably efficient algorithm that utilizes the modular similarities across tasks [11]. Stepping towards another level, we let the agent adapt to drastically different observation spaces by transferring similarities of latent dynamics. We propose a theory-inspired transfer algorithm which, for the first time, achieves transfer learning from a vector-input environment to a pixel-input environment [15].

• *Adapting by Representation Pretraining.* Pretraining is an increasingly prevalent direction to improve agents' adaptability. By mapping high-dimensional inputs into a lower-dimensional representation space while keeping the most informative features, one can greatly reduce the learning burden of downstream tasks. Our recent work formulates the pretraining pipeline for RL tasks, and introduces a pretraining approach that leverages both perceptual and control-relevant information in the pretrained representation [12]. The pretrained model can quickly adapt to multiple downstream tasks across various environmental structures.

Future Agenda

Towards Better Robustness and Adaptability: Pushing the Boundaries. Following my previous work, I aim to provide more general and theoretically grounded solutions to improve the robustness and adaptability of RL agents. More specifically, my ongoing and future work includes (1) characterizing the optimality of robustness of RL agents, and the trade-off between natural performance and robustness from a game-theoretical perspective, (2) developing RL approaches with certifiable robustness in both the training phase and the deployment phase, (3) pretraining generic foundation models from multiple modality that can be applied to a broad range of decision making problems, and (4) designing lifelong learning agents to fulfill the needs of multitasking.

Towards Real-life Applications: Research for Social Good. I will also broaden my research and apply my expertise to more real-life research problems. In particular, I am interested in autonomous driving, home robots and healthcare systems, which can benefit the society and bring convenience to people with disabilities and diseases. RL approaches, in combination with other artificial intelligence techniques, have huge potential to build more intelligent systems by interacting with people and the environment. Robustness and adaptability are crucial in these high-stakes applications, to make sure that the trained agents can not only work for seen scenarios, but also guarantee safe behaviors under unexpected inputs and quickly produce good behaviors when changes happen.

In summary, my research enhances the robustness and adaptability of intelligent agents, which can lead to more trustworthy and general artificial intelligence (AI) for real-world decision making problems. I enjoy both exploring new research areas and diving deep into a certain topic. I would like to collaborate with people from different backgrounds and push the boundaries of AI techniques. I believe that the development of AI can lead to a brighter future of human society.

References

[1] B. Biggio, B. Nelson, and P. Laskov. Poisoning attacks against support vector machines. In *Proceedings of the 29th International Conference on International Conference on Machine Learning*, page 1467–1474, 2012.

[2] A. Gleave, M. Dennis, C. Wild, N. Kant, S. Levine, and S. Russell. Adversarial policies: Attacking deep reinforcement learning. In *International Conference on Learning Representations*, 2020.

[3] S. Huang, N. Papernot, I. Goodfellow, Y. Duan, and P. Abbeel. Adversarial attacks on neural network policies. *arXiv preprint arXiv:1702.02284*, 2017.

[4] B. R. Kiran, I. Sobh, V. Talpaert, P. Mannion, A. A. Al Sallab, S. Yogamani, and P. Pérez. Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[5] Y. Liang*, Y. Sun*, R. Zheng, and F. Huang. Efficient adversarial training without attacking: Worst-case-aware robust reinforcement learning. In *Proceedings of the 36th Conference on Neural Information Processing Systems*, 2022.

[6] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

[7] J. Schrittwieser, I. Antonoglou, T. Hubert, K. Simonyan, L. Sifre, S. Schmitt, A. Guez, E. Lockhart, D. Hassabis, T. Graepel, T. Lillicrap, and D. Silver. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588(7839):604–609, 2020.

[8] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, et al. Mastering the game of go without human knowledge. *nature*, 550(7676): 354–359, 2017.

[9] Y. Sun and F. Huang. Can agents learn by analogy? an inferable model for pac reinforcement learning.

In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '20, page 1332–1340, 2020.

[10] Y. Sun, D. Huo, and F. Huang. Vulnerability-aware poisoning mechanism for online rl with unknown dynamics. In *International Conference on Learning Representations*, 2021.

[11] Y. Sun, X. Yin, and F. Huang. Temple: Learning template of transitions for sample efficient multi-task rl. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 9765–9773, 2021.

[12] Y. Sun, S. Ma, R. Madaan, R. Bonatti, F. Huang, and A. Kapoor. Smart: Self-supervised multi-task pretraining with control transformers. *preprint*, 2022.

[13] Y. Sun, R. Zheng, P. Hassanzadeh, Y. Liang, S. Feizi, S. Ganesh, and F. Huang. Certifiably robust policy learning against adversarial communication in multi-agent systems. *arXiv preprint arXiv:2206.10158*, 2022.

[14] Y. Sun, R. Zheng, Y. Liang, and F. Huang. Who is the strongest enemy? towards optimal and efficient evasion attacks in deep RL. In *International Conference on Learning Representations*, 2022.

[15] Y. Sun, R. Zheng, X. Wang, A. E. Cohen, and F. Huang. Transfer RL across observation feature spaces via model-based regularization. In *International Conference on Learning Representations*, 2022.