

# Abstract

Moritz Lange

This doctoral dissertation, titled `BEYOND TRIAL AND ERROR: IMPROVING REINFORCEMENT LEARNING THROUGH REPRESENTATION LEARNING AND REASONING` and authored by Moritz Johannes Lange, addresses the question of how to reduce the reliance on trial and error search in reinforcement learning (RL), in order to improve its performance, safety and trustworthiness. Most current RL agents rely on trial and error search in two ways: (I) During training they explore their environment to learn the value of taking different actions in different states, and (II) during planning they explore outcomes of actions with forward models to reach a desired goal.

Meaningful representations of environment states can enable agents to better predict the out-comes of their actions, while humans – who need to be able to trust agents – can better understand an agent’s decision basis. Furthermore, if the agent is capable of reasoning about the actions re-quired to reach a goal, its behavior becomes more efficient while its decisions become explainable. Both aspects improve on the trial and error nature of most current RL approaches, because ac-tions become more informed, up to a point where the agent can identify suitable actions through deduction instead of exploration.

This dissertation presents five studies that together advance the understanding of how to use representation learning and reasoning towards that goal. The first two studies mainly address representation learning. They empirically show under which circumstances representation learning can improve efficiency during training, and explore suitable types of representations: Re-pre-sentations are best learned by modular encoders, should convey information about environment dynamics, and interpretable representations can improve performance as well as trustworthiness. Presented examples include location and heading representations for a navigation task, and object-based state representations for reasoning about classical mechanics.

The latter three studies mainly address reasoning. In particular physical reasoning, as agents often operate in physical environments, and causal reasoning because predictions based on causal understanding are more robust than those based solely on statistical correlations. These studies contain the first survey of the field of physical reasoning in machine learning, and identify a lack of RL-based physical reasoning benchmarks. They additionally propose a denoising diffusion model that can generate trajectories of multiple interacting objects – or agents – conditioned on initial states and goal states. Lastly, they identify pitfalls of modern gradient-based causal discovery approaches, which can fall victim to distributional biases in their training data and then deduce an incorrect causal model. They also provide insights into how approaches can avoid those pitfalls, which will be important for efficient and scalable causal reasoning in future RL approaches.

Together, these studies contribute to the development of reasoning agents that require less trial and error search during learning and potentially none during planning, which makes them more efficient, safer and more trustworthy for the humans that work with them.

# Kurzfassung

Moritz Lange

Diese Dissertation mit dem Titel `BEYOND TRIAL AND ERROR: IMPROVING REINFORCEMENT LEARNING THROUGH REPRESENTATION LEARNING AND REASONING` (deutsche Übersetzung: `JENSEITS VON VERSUCH UND IRRTUM: BESSERES VERSTÄRKUNGSLERNEN DURCH REPRÄSENTATIONSLERNEN UND SCHLUSSFOLGERN`), verfasst von Moritz Johannes Lange, untersucht wie eine Verringerung von Versuch und Irrtum im Verstärkungslernen (RL) die Performanz, Sicherheit und Verlässlichkeit verbessern kann. Die meisten modernen RL-Agenten verwenden das Prinzip von Versuch und Irrtum einerseits während des Lernens zum Erkunden und Bewerten ihrer Umgebungszustände und andererseits auch während der Planung mit Vorwärtsmodellen bei der Suche nach Aktionen, die einen gewünschten Zielzustand herbeiführen.

Aussagekräftige Repräsentationen (Kodierungen) von Zuständen erlauben Agenten die Konsequenzen ihrer Aktionen besser vorherzusehen und Menschen können die Entscheidungsgrundlage solcher Agenten besser nachvollziehen. Wenn Agenten außerdem von einem Zielzustand ausgehend die erforderlichen Aktionen schlussfolgern können, werden ihre Entscheidungen erklärbarer und effizienter. Beides reduziert die Notwendigkeit von Versuch und Irrtum, weil der Agent sich die korrekte Aktion herleiten kann, anstatt ausprobieren zu müssen.

Diese Dissertation stellt fünf Studien vor, mit neuen Erkenntnissen darüber wie Repräsentationslernen und Schlussfolgern hierzu genutzt werden können. Die ersten beiden Studien befassen sich mit Repräsentationslernen. Sie zeigen wie geeignetes Repräsentationslernen Agenten effizienter lernen lassen kann und präsentieren Beispiele: Modulare Ansätze sowie Repräsentationen, die Informationen über Umgebungsdynamiken enthalten und interpretierbar sind. Präzenterte Beispiele dafür sind Ort und Blickrichtung für Navigationsaufgaben und Objekt-basierte Repräsentationen. Die drei anderen Studien befassen sich mit algorithmischem Schlussfolgern bezüglich Physik und Kausalität. Ersteres ist wichtig da Umgebungen oft physikalisch sind, letzteres ermöglicht robustere Vorhersagen als Modelle die Korrelationen lernen. Diese Studien umfassen die erste Übersichtsarbeit zu algorithmischem Schlussfolgern über physikalische Prozesse, mit dem Ergebnis, dass es kaum physikalische Referenzumgebungen gibt, die auf RL basieren. Sie präsentieren auch ein Diffusionsmodell das Trajektorien mehrerer interagierender Objekte – oder Agenten – generiert, bei vorgegebenen Anfangs- oder Endzuständen. Weiterhin identifizieren sie problematische Einflüsse von Datenverteilungen auf gradientenbasierte kausale Lernmethoden, die dazu führen können, dass falsche kausale Zusammenhänge gelernt werden. Es werden Möglichkeiten aufgezeigt, solche Methoden unabhängig von diesen Einflüssen zu machen.

Zusammengenommen tragen diese Studien zur Entwicklung von Agenten bei, die durch ihre Befähigung zum Schlussfolgern weniger Versuch und Irrtum während des Lernens benötigen, und während des Planens möglicherweise ganz darauf verzichten können. RL-Agenten werden so effizienter, sicherer und verlässlicher für die Menschen, die mit ihnen arbeiten.