氏　　名　Nicolas Bougie

学位(専攻分野)　博士(情報学)

学 位 記 番 号　総研大甲第 2242 号

学位授与の日付　2021年3月 24日

学位授与の要件　複合科学研究科 情報学
学位規則第6条第1項該当

学 位 論 文 題 目　Efficient Reinforcement Learning through Improved Cognitive Capabilities

論 文 審 査 委 員　主　査　市瀬 龍太郎
情報学専攻 准教授
山田 誠二
情報学専攻 教授
杉山 磨人
情報学専攻 准教授
稲邑 哲也
情報学専攻 准教授
森山 甲一
名古屋工業大学 情報工学専攻 准教授

(Form 3)

Summary of Doctoral Thesis

Name in full Nicolas Bougie

Title Efficient Reinforcement Learning through Improved Cognitive Capabilities

A long-standing goal in reinforcement learning is to create agents that can solve complex decision-making tasks through trial and error interactions with the environment. While reinforcement learning has shown impressive advances in a plethora of simulated tasks including game-playing, board-games, or robot control, these results are shyly echoed in the real world. This is often due to huge amounts of interactions required to reach decent performance, which can be intractable in real-world settings. All these considerations lead to one fundamental question: how to improve sample efficiency in reinforcement learning agents enough that they can be practically applied to real-world tasks? We believe that developing methods that integrate human-like cognitive capabilities is the key to this answer. Cognitive capabilities are aspects of mental functioning, which include the capability for planning, memorizing, thinking abstractly, learning from experience, common sense reasoning, and making intrinsically motivated choices. These capabilities are one of the skill sets that distinguishes efficient learning in humans from data-consuming learning in artificial agents. Thus, the first chapter outlines this ideology in detail and concretizes these ideas into algorithms that are embodied with improved cognitive capabilities. Next, the second chapter provides the necessary background to understand the research topics presented in the following chapters. Next, the thesis is divided into three parts that focus on improving different aspects of cognitive capabilities.

In Part I, we start by tackling one of the key problems in reinforcement learning: how to overcome the lack of curiosity in reinforcement learning algorithms? One natural form of learning is to explore the environment and accumulate knowledge. However, in the real world, rewards are naturally sparse or poorly-defined, which is an important issue since hoping to stumble into a goal state by chance within an acceptable number of interactions is unlikely. While humans are accustomed to operating with rewards that are extremely sparse, reinforcement learning agents lack an essential cognitive capability: curiosity. That is, humans dedicate much time and energy to exploring and gathering information; and often the search for information is unrelated of a foreseeable extrinsic gain/reward, as if learning were

reinforcing in and of itself. This behavior emerges through our capability to make intrinsically motivated choices, our intrinsic desire to learn and understand. Inspired by curious behaviors in humans, we develop curiosity-driven agents that can learn to solve complex problems featuring extremely sparse rewards. However, in order to be useful for real-world tasks, several major limitations inherent in curiosity must be addressed.

First, we propose to generate an exploration bonus based on the agent's knowledge about its environment in order to encourage the gradual acquisition of new skills. Most prior work aim to model environmental dynamics, however, it tends to limit long-horizon performance due to model drift and does not directly consider the agent's knowledge - rather than visiting all possible states the agent should focus on task-relevant regions. Our approach (GoCu) escapes these drawbacks by operating directly on the agent's policy and encouraging the acquisition of task-relevant skills (Chapter 3).

Second, a potential limitation of our proposed method as well as prior approaches is the vanishing issue inherent in curiosity. As the agent becomes more familiar with its environment, curiosity may vanish quickly during training, leaving the agent with no incentive to further explore the environment and reducing its feedback to extrinsic reward only. Moreover, curiosity-driven agents tend to fall and stay trapped in poor local optima due to local sources of entropy in the environment (i.e. stochastic environments). On the other hand, we address both challenges by proposing a robust formulation of curiosity, PoBP, that uses the agent's learning progress on a multi-step horizon scale (Chapter 4).

Finally, while very good results have been achieved on some hard exploration tasks, these algorithms face a fundamental limitation: they do not explicitly encourage and promote deep exploration. That is, curiosity only captures the consequences of short-term decisions on the environment. Deep exploration that involves coordinated decisions over long time horizons is beyond the reach of most methods. We present a formulation of curiosity that can capture meaningful visual features and salient environmental dynamics at different scales; and then built an algorithmic framework (FaSo) that combines two curiosity components, explicitly promoting deep exploration (Chapter 5).

In Part II, we focus on giving agents the capability for leveraging external guidance in order to enable common sense reasoning. In the real world, reinforcement learning requires large amounts of data due to reward sparsity, but also because most systems learn tasks from scratch, i.e., without any prior knowledge. While this is convenient in simulated domains where interactions are virtually unlimited, this assumption rarely holds in the real

world - training an end-to-end reinforcement learning system with no prior assumptions about the domain often induces millions or billions of interactions to reach reasonable performance. Learning without prior knowledge seems to be an approach that is rarely taken in humans - they use their common sense reasoning to extract initial biases as well as strategies on how to approach a problem. Precisely, common sense reasoning refers to the cognitive capability to make presumptions about the type and essence of ordinary situations based on prior experiences and domain knowledge (also called common sense priors). Research in cognitive systems has shown that many of these challenges can be addressed by exploiting domain knowledge and the coupling between domain knowledge and learning. Thus, we propose generic mechanisms that employ human guidance to transfer human knowledge into reinforcement learning. Our core idea is to develop novel types of guidance that require minimal human effort and provide additional meaningful information to the agent.

First, we introduce human-like planning and domain knowledge to enhance information given to the agent. The central concept is to augment the agent's input with high-level information that are easy to interpret for the agent and applicable to many situations. We then derive a framework for the integration of human-like planning based on the high-level information and simple visual recognition, improving sample efficiency and facilitating various forms of common sense reasoning (Chapter 6).

Second, we give a reinforcement learning agent the capability of leveraging existing human expertise. Rather than integrating human guidance and domain knowledge designed expressly to solve the task being learned, we want to reduce human effort by leveraging existing datasets related to the task being learned. Moreover, most methods in the reinforcement learning literature lack interpretability which may limit their applicability in some real-world domains. Therefore, we present methods for extracting rules from existing datasets; and then build an interpretable agent whose internal representation is based on these rules (Chapter 7).

Finally, one limitation of our methods as well as most prior work is that the agent and the teacher cannot actively share insights. Hence, the agent cannot cope with the changes in the environment or query the teacher when it struggles. To overcome these challenges, we introduce the concept of active goal-driven demonstrations to actively query the demonstrator only in hard-to-learn and uncertain regions of the state space. The depicted method relies on a novel form of human guidance, goal-driven demonstrations, that are easier and more intuitive to provide for a demonstrator than full demonstrations while precisely matching the agent's needs (Chapter 8)

In Part III, developmental studies have shown that interactive capabilities in humans emerge incrementally through the aggregation of multiple internal and external feedback signals. Hence, one natural question that arises is: how to simultaneously learn from these two classes of supervision (i.e. curiosity and human guidance) to achieve human-like sample-efficient learning? To answer this question, we propose a hierarchical reinforcement learning framework that exploits the hierarchical structure of the task to integrate different modes of supervision at different levels. Our key design principle is to introduce (non-expert) human guidance at the high-level for long-term planning and common sense reasoning. At the low-level, we make use of curiosity to improve sample efficiency and drive the learning of subpolicies, particularly in tasks featuring sparse rewards. We further show how curiosity relates to automatic sub-goal discovery. (Chapter 9).

In the final chapter, we first summarize our contributions to the state-of-the-art in the domain of reinforcement learning. Our specific focus was on sample-efficient learning to expand the availability of reinforcement learning to real-world domains. Motivated by this ambition, we devoted our research to make reinforcement learning agents much less reliant on large amounts of interactions by improving their cognitive capabilities. Through evaluations on multiple benchmark tasks (e.g. Atari games, Minigrid), we show that curiosity is vital to rapidly acquire new skills and improve exploration efficiency in extremely sparse reward environments, outperforming prior approaches in both sample-efficiency and performance. Besides, even in the absence of any extrinsic reward signal, curiosity provides enough indirect supervision for learning useful behaviors and skills. Furthermore, we demonstrate that introducing this cognitive capability enables our agents to master tasks featuring real-world characteristics (e.g. stochasticity, poorly-defined rewards, temporally-extended exploration patterns) that traditional methods fail to solve.

We also show that the proposed forms of human guidance improve various forms of common sense reasoning, enhancing data efficiency and final performance. They reduce the number of required interactions and human effort enough to scale RL to practical tasks such as robot control, trading, 3D navigation, or game-playing. Moreover, the depicted forms of human guidance can be used in tasks that are too challenging for even humans to perform well and be provided by a non-expert.

These contributions together demonstrate that improving cognitive capabilities in reinforcement learning agents is essential to drastically enhance sample efficiency and performance. These capabilities inspired by learning in humans make artificial agents suitable for real-world domains.

Finally, we conclude the final section with a discussion of the main findings and we also make some anticipations on the future development in this field. Especially, leveraging both internal motivation and external guidance allowed us to reduce the number of interactions by several orders of magnitude and to outperform human experts in a wide range of domains such as game-playing, trading, 3D navigation, or robot control. Hence, the proposed algorithms have improved efficiency in reinforcement learning agents enough that they can be practically applied to real-world tasks. In the future, these algorithms could be used to develop sample-efficient and learning machines, with substantial benefits for society as a whole.

Form8・Separate Sheet
（様式８・別紙１）

Results of the doctoral thesis screening

博士論文審査結果

Name in Full
氏　　名　Nicolas Bougie

Title
論文題目　Efficient Reinforcement Learning through Improved Cognitive Capabilities

強化学習は，試行錯誤による環境との相互作用を通して，複雑な意思決定タスクを解決することができる学習手法である．強化学習は，ゲーム，ロボット制御など多くのシミュレーションタスクで目覚ましい成果を挙げてきたが，現実世界では，膨大な試行錯誤を繰り返すことができないため，実用化上の大きな障壁となっている．そこで，本博士論文では，強化学習における環境との相互作用をどのようにすれば，試行錯誤を減らし効率化できるかという問題に取り組んでいる．そのために，人間のような認知能力を統合することを検討した．認知能力とは，計画性，記憶力，抽象的な思考，経験からの学習，常識的な推論，内発的に動機づけられた選択を行う能力などの精神機能のことである．本博士論文では，強化学習を行うエージェントの認知能力を改善することで，より少数の試行錯誤で学習可能とするための学習手法の作成を試みた．

本論文は，３部構成となっており，全１０章からなる．第１章「Introduction」では，強化学習の効率性の問題に対して，研究の動機を述べると共に，本論文の貢献について説明している．

第２章「Background」では，本論文の理解に必要となる強化学習などの研究の背景について説明している．

続く，第３章から第５章までは，第１部「Learning to Act via Curiosity-Driven Exploration」としてまとめられ，好奇心に基づく学習に関する研究について述べられている．第３章「Skill-Based Curiosity for Intrinsically Motivated Reinforcement Learning」では，エージェントが持つ環境に対する知識に基づいて，好奇心となる内的報酬を生成する手法を提案し，その有効性を実験的に示した．第４章「Exploration via Progress-Driven Intrinsic Rewards」では，好奇心がすぐに消失してしまう問題に対して，新たな好奇心の定式化手法を提案し，その有効性を実験的に示した．第５章「Fast and Slow Curiosity for High-Level Exploration」では，好奇心による手法に長期的な視点が欠如している問題点に対して，２つの好奇心の要素を組み合わせた手法を提案し，その有効性を実験的に示した．

その後の第６章から第８章までは，第２部「Bridging the Gap Between Reinforcement Learning and Human Guidance」としてまとめられ，常識的な推論を行うために，外部からのガイダンスを活用する強化学習に関する研究について述べられている．第６章「Combining Deep Reinforcement Learning with Prior Knowledge and Reasoning」では，人間のような事前知識を利用できる学習手法を提案し，その有効性を実験的に示した．第７章「Towards Interpretable Reinforcement Learning with State Abstraction Driven

by External Knowledge」では，外部の既存データから規則を抽出して内部表現として利用することで，解釈可能なエージェントを構築する手法を提案し，その有効性を実験的に示した．第８章「Active Goal-Driven Learning」では，能動的に教師に問い合わせを行う機構を備えたエージェントを提案し，その有効性を実験的に示した．

続く，第３部「Leveraging Human Guidance and Curiosity for Sample-Efficient Learning」では，第１部，第２部の成果を受け，人間のガイダンスと好奇心の両方を用いた学習に関する研究について述べられている．第９章「Hierarchical Learning from Human Preferences and Curiosity」では，人間のガイダンスと好奇心という異なる２つの観点を階層的にまとめた強化学習手法を提案し，その有効性を実験的に示した．

最後の第１０章「Conclusion」では，博士論文の総括を行うと共に，展望を述べ，結論をまとめた．

公開発表会では，博士論文の章立てに従って発表が行われた．その後に行われた論文審査会及び口述試験では，審査委員からの質疑に対して的確に回答がなされた．

上記のように，本博士論文は，強化学習において，環境との相互作用をどのようにすれば，効率化できるかという問題に対し，その解決方法を示した点で，この研究分野の発展に貢献するものである．また，本論文で示した考え方により，人間のような認知能力をエージェントに組み込む手法の一端が明らかとなり，人間と協調可能な安全・安心な知的システムのための基盤技術開発という観点からも意義が認められる．さらに，博士論文の内容は，４本の査読付きジャーナル論文，３本の査読付き国際会議論文で発表されており，その内１本は最優秀論文賞を受賞するなど，社会からも評価されている．以上より，本論文は博士論文として，十分な水準であると審査委員全員一致で認められた．