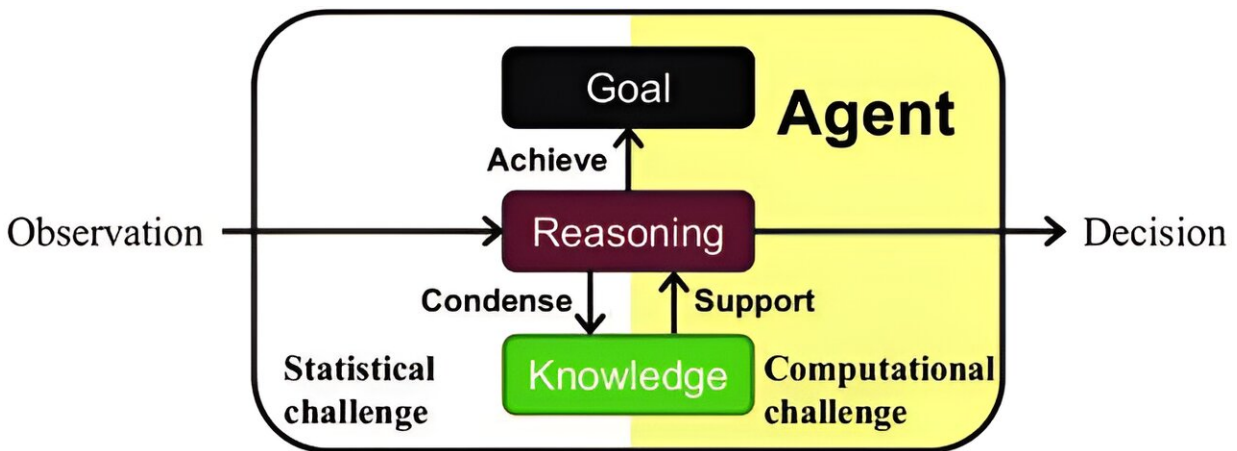# A new approach to intelligent decision-making in reinforcement learning

August 1 2023

An agent can be understood in terms of 3 parts working together: the knowledge part, the reasoning part, and the goal. The knowledge part retains everything the agent has learned from its entire history. The reasoning part is a computational process that processes newly arrived information based on the learned knowledge. It helps condense information into various kinds of knowledge and concludes the actions the agent should take. The goal is something that orients the reasoning process and is aligned with the agent's lifelong interests. Credit: Chenyang Wu et al.

A new paper published in *Intelligent Computing* presents the primary challenges of reinforcement learning for intelligent decision-making in complex and dynamic environments.

Reinforcement learning is a type of machine learning in which an agent learns to make decisions by interacting with an environment and receives rewards or penalties.

The agent's goal is to maximize long-term rewards by determining the best actions to take in different situations. However, researchers Chenyang Wu and Zongzhang Zhang of Nanjing University are convinced that reinforcement learning methods that rely solely on rewards and penalties will not succeed at producing intelligent abilities such as learning, perception, social interaction, language, generalization and imitation.

In their paper, Wu and Zhang identified what they see as the shortcomings of current reinforcement learning methods. A major issue is the amount of information that needs to be collected through trial and error.

Unlike humans who can use their past experiences to reason and make better choices, current reinforcement learning methods heavily rely on agents that try things out repeatedly on a large scale to learn how to perform tasks. When dealing with problems that involve many different factors influencing the outcome, it is necessary for agents to try out a huge number of examples to figure out the best approach.

If the problem increases slightly in complexity, the number of examples needed grows fast, making it impractical for the agent to operate efficiently. To make matters worse, even if the agent had all the information needed to determine the best strategy, figuring it out would still be very hard and time-consuming. This makes the learning process slow and inefficient.

Both statistical inefficiencies and computational inefficiencies hinder the practicality of achieving general reinforcement learning from

scratch. Current methods lack the efficiency required to unlock the full potential of reinforcement learning in developing diverse abilities without extensive computational resources.

Wu and Zhang argue that statistical and computational challenges can be overcome by accessing high-value information in observations. Such information can enable strategy improvements through observation alone, without the need for direct interaction. Imagine how long it would take for an agent to learn to play Go by playing Go—in other words, through trial and error.

Then imagine how much faster an agent could learn by reading Go manuals—in other words, by using high-value information. Clearly, the ability to learn from information-rich observations is crucial for efficiently solving complex real-world tasks.

High-value information possesses two distinct characteristics that set it apart. First, it is not independent and identically distributed, implying that it involves complex interactions and dependencies, distinguishing it from past observations. To fully comprehend high-value information, one must consider its relationship with past information and acknowledge its historical context.

The second feature of high-value information is its relevance to computationally aware agents. Agents with unlimited computational resources may overlook high-level strategies and rely solely on basic-level rules to derive optimal approaches. These agents disregard higher-level abstractions, which may introduce inaccuracies, and prioritize computational efficiency over accuracy.

Only agents that are aware of computational trade-offs and capable of appreciating the value of computationally beneficial information can effectively leverage the benefits of high-value information.

In order for reinforcement learning to make efficient use of high-value information, agents must be designed in new ways. In accordance with their formalization of intelligent decision-making as "bounded optimal lifelong reinforcement learning," Wu and Zhang identified three fundamental problems in agent design:

1. Overcoming the non-independent and identically distributed nature of the information stream and obtaining knowledge on the fly. This requires connecting the past to the future and transforming the continuous flow of information into useful knowledge for future use.

   However, limited computational resources make it impossible to remember and process the entire interaction history. Therefore, a structured knowledge representation and online learning algorithm are necessary to organize information incrementally and overcome these constraints.

2. Supporting efficient reasoning given bounded resources. Firstly, universal knowledge that facilitates understanding, predicting, evaluating and acting is no longer enough under computing constraint. To address this challenge, efficient reasoning demands a structured knowledge representation that exploits the problem structure and helps the agent reason in a problem-specific way, which is essential for computing efficiency.

   A second aspect of the reasoning process is sequential decision-making. This plays a pivotal role in guiding agents to determine their actions, process information and develop effective learning strategies. Consequently, meta-level reasoning becomes necessary to maximize the utilization of computing resources. Thirdly, successful reasoning requires agents to effectively combine their internal abilities with the information gleaned

from external observations.

3. Determining the goal of reasoning to ensure that the agent seeks long-term returns and avoids being driven solely by short-term interests. This is known as the exploration–exploitation dilemma. It involves finding a balance between exploring the environment to gather new knowledge and exploiting the best strategies based on existing information.

   This dilemma becomes more complicated when considering the computational perspective, as the agent has limited resources and must balance between exploring an alternative way of computation and exploiting the best existing approach. Because exploring everything in a complex environment is impractical, the agent relies on its existing knowledge to generalize to unknown situations. Resolving this dilemma requires aligning the reasoning goal with the agent's long-term interests. There is still much to understand, especially from the computational perspective.