2021-Silver

Reward is Enough

- **Date**: 2021-05-24
- **Link:** paper
- **Authors:**
  - Silver, David
  - Singh, Satinder
  - Precup, Doina
  - Sutton, Richard
- **Cites:**
- **Cited by:**
- **Keywords:** #reinforcement-learning
- **Collections:**
- **Status:** #in-progress

---

0 - Abstract

- Intelligence, and its associated abilities, can be understood as subserving the maximization of reward.

- Specialized problem formulations are not needed for each unique ability associated with intelligence.

- Reinforcement learning agents could constitute a solution to artificial general intelligence.

1 - Introduction

- Expressions of intelligence in nature: social intelligence, language, perception, knowledge representation, planning, imagination, memory, motor control, etc.

- Is there a common drive that motivates the development of all these different expressions?

- We know each of these expressions can be achieved through a unique, specialized formulation.

- Goal is to find a common objective that drives all intelligence expressions.

- Reward maximization is general enough to solve this goal:

  1. Different forms of intelligence may arise from the maximization of different reward signals in different environments.

     – Akin to animals evolving different traits based on their environment.

  2. The different forms of intelligence may arise from the pursuit of a single reward.

     – *Ex:* Squirrels can be thought of as having a singular goal of subsistence whose pursuit yields a set of complex skills.

  3. Reward maximization is general enough to provide the deeper context for skills.

     – *Ex:* grounding language to perceptual experience - dialogue regarding the best way to peel a fruit.

     – *Ex:* we know how to differentiate between a log and a crocodile, but *why* have we developed this skill?

  4. The pursuit of a singular goal naturally integrates the diverse intelligences of an agent.

- How do natural or artificial intelligences learn a singular goal?
  - Through interaction with the environment by trial and error

- *Ex:* learning the game of Go.
  - Early failed attempts tried to aspects of the game individually.
  - Ultimately the successful solution was to use a singular reward (win = +1; loss = -1).
  - This allowed for a specialized set of skills to manifest naturally.

2 - Background: The Reinforcement Learning Problem

- *Intelligence* - the flexible ability to achieve goals.

- *Reinforcement learning* - formalized problem of goal-seeking intelligence.

2.1 - Agent and Environment

- RL breaks down the problem into two interacting systems: agent and environment.

2.2 - Agent

- *Agent* - system that makes decisions: receives an observation at time and outputs an action .
  - In other words, the agent is a system where is the history of interactions between the agent and the environment.

2.3 - Environment

- *Environment* - system that receives an action at time and responds with observation at the next time step.

  - In other words, the environment is a system where:

    * - experience history.

    * - latest action.

    * - source of randomness.

2.4 - Rewards

- *Reward* - special scalar observation , emitted at every time-step by a reward signal in the environment.

  - Provides instantaneous feedback to the agent of progress towards a goal.

- Rewards fit a large variety of goals.

- Instantaneous feedback is essential for long or infinite streams of experience.

3. Reward is Enough

HYPOTHESIS (Reward-is-Enough) - Intelligence, and its associated abilities, can be understood as subserving the maximization of reward by an agent acting in its environment.

- This is deeper than just a trivial selection of a narrow reward to induce a specialized behavior.

- Rather, the hypothesis should be understood as a general, singular reward which through its maximization, implicitly yields intelligence and associated abilities.

- Sophisticated abilities may arise from the maximization of simple rewards in complex environments.

- Maximization of many different reward signals in many different environments may produce similar abilities associated with intelligence.
  - Meaning general intelligence might be robust to the choice of reward.
  - *Ex:* animals existing in different environments evolved a common set of skills (locomotion, perception, manipulation, etc.).
- The following sections discuss how the hypothesis could work for various abilities.

3.1 - Reward is enough for knowledge and learning

- *Knowledge* - information that is internal to the agent.
  - *Ex:* parameters of an agent's functions for selecting actions, etc.
- Environments may demand innate knowledge.
  - What does an agent know when it is born?
  - Is a gazelle born with an innate knowledge to run from the lion?
  - For RL this is difficult to reason with, as it is knowledge that is required without experience which is antithetical to RL itself.
- Environments may demand learned knowledge.
  - The total space of knowledge is greater than the capacity of the agent.
  - Therefore, the agent must be able to draw from past experience to reason about future events.
- Conclusion:
  - Agents must be given innate knowledge through their design.
  - Agents must acquire learned knowledge through experience.

3.2 - Reward is enough for perception

- Perception is critical for determining rewards.
    - *Ex:* Humans must use their eyes to determine whether or not a food is poisonous or not.
- This perceptual ability given in the example may arise implicitly from the maximization of the satiation reward.
- Perceptual abilities that are better supported from a reward maximization perspective rather than a supervised learning perspective:
    - Action and observation are typically intertwined.
        - *Ex:* echolocation - emit a sound, observe the time to hear the echo, perceive your surroundings.
    - Utility of perception depends upon the agent's behavior.
        - The reward gained from performing a given perception depends on the agent's state and subsequent action (i.e. the act of perception exists in a greater context that informs the value of the perception being made).
    - Perception may require opportunity costs.
        - In most scenarios, there is a cost in acquiring information.
    - The agent's context informs the distribution of data it perceives.
        - An agent in the arctic is more likely to perceive a polar bear, than an agent in NYC.
    - Many applications requiring perception lack access to labelled data.