

강화학습 개요

정 태 수

고려대학교 산업경영공학부
tcheong@korea.ac.kr



도입

• 자전거 배우기

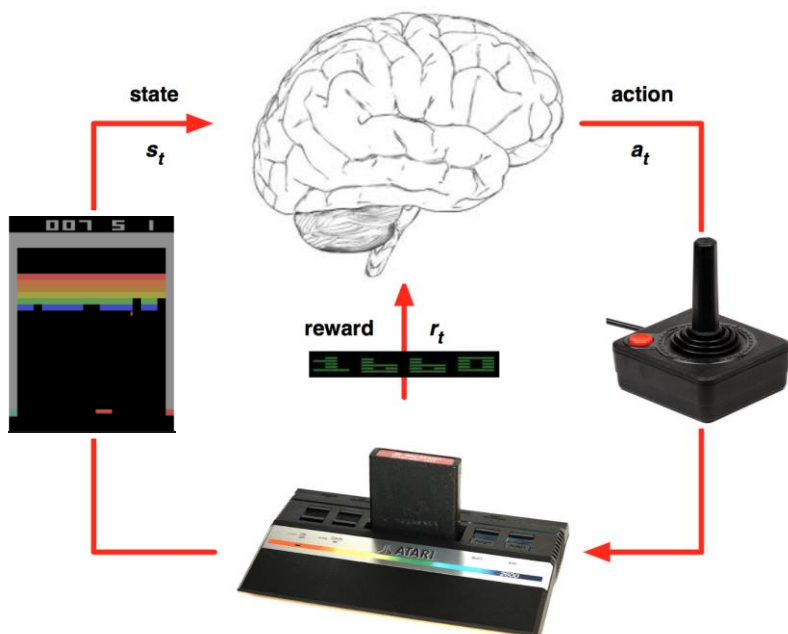
- 자전거 타는 법을 배울 때 구체적으로 타는 방법을 알지 못하고 배운다
- 타다 넘어지고 일어서다 보면 어떻게 해야 균형을 잡으면서 자전거를 운전할지를 스스로 학습하게 된다





도입

• 벽돌깨기 게임

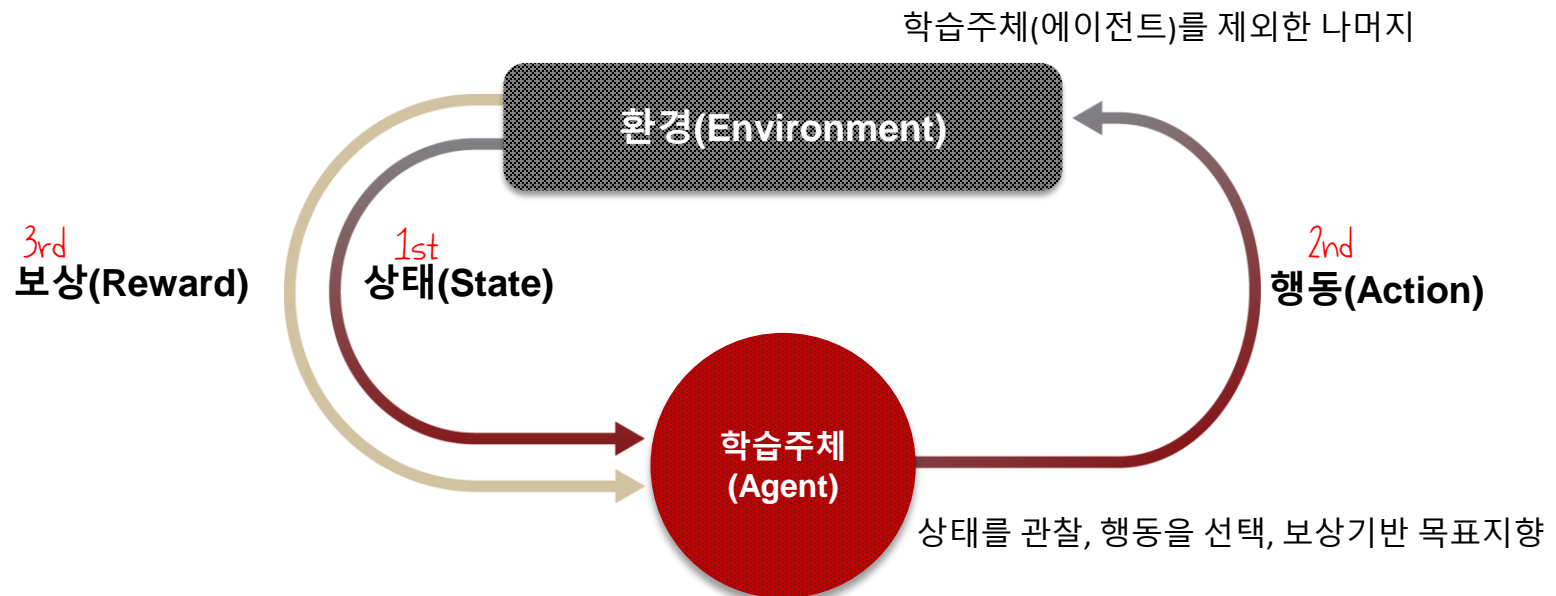


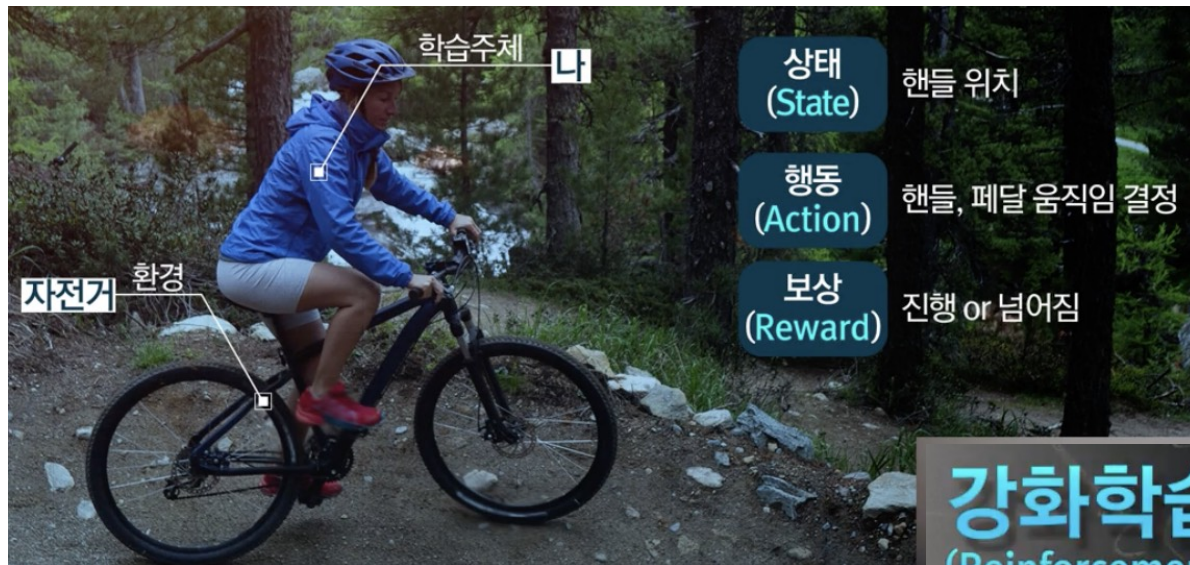


강화학습

• 강화학습 (Reinforcement Learning)

- 주어진 어떤 **상황**(state)에서 **보상**(reward)를 최대화할 수 있는 **행동**(action)에 대해 학습하는 것
- 학습 주체(agent)가 상황에 가장 적합한 행동을 찾기까지는 수많은 **시행착오** (trial & error)가 필요
 - 학습주체는 문제의 구조(환경에 대한 정보)를 모르는 상태에서 학습
- 현재 선택한 행동이 미래의 순차적 보상에 영향 (delayed reward)
 - 즉, 현재의 의사결정이 미래에 영향을 미침





강화학습 (Reinforcement Learning)

상태 (State) → 행동 (Action) → 보상 (Reward)

일련의 과정을 통해 학습

학습 주체가 상황에 가장 적합한 행동을 찾기까지 수많은 시행착오 필요



지도학습, 비지도학습과의 차이점

• 지도학습

- 레이블(label)이라는 정답 데이터가 주어진 상황에서 학습이 이루어지는 경우를 말하며, 이러한 잘 정의된 데이터로부터 즉각적인 피드백을 받으며 학습
- 주로 현재의 데이터들을 바탕으로 미래를 예측하거나, 정답이 있는 결과를 맞추는데 목적이 있음

• 비지도학습

- 정답에 해당하는 레이블(label)이 존재하지 않음으로 즉각적인 피드백을 받을 수 없으며, 주로 데이터 자체에 내재되어 있는 성질을 찾아내는데 활용

• 강화학습

- 잘 정의된 (labeled) 데이터, 즉 정답이 존재하는 데이터,가 아닌 환경과의 상호작용을 통해 얻은 보상(reward)로 부터 학습
 - 환경과의 상호작용을 통해 데이터를 취득



강화학습의 예시

IBM Research

www.ibmwatson.com

IBM

How Watson Learns Superhuman Jeopardy! Strategies



Gerald Tesauro, IBM Research
Joint work with: David Gondek, Jonathan Lenchner,
James Fan, John Prager



© 2011 IBM Corporation

Playing Atari with Deep Reinforcement Learning

Volodymyr Mnih Koray Kavukcuoglu David Silver Alex Graves Ioannis Antonoglou

Daan Wierstra Martin Riedmiller

DeepMind Technologies





강화학습의 예시

• 자율주행차

- AWS DeepRacer
 - 강화학습을 실제로 체험해볼 수 있는 자율주행 모형차
 - 실제 차량의 18분의 1의 크기



• 에너지 관리

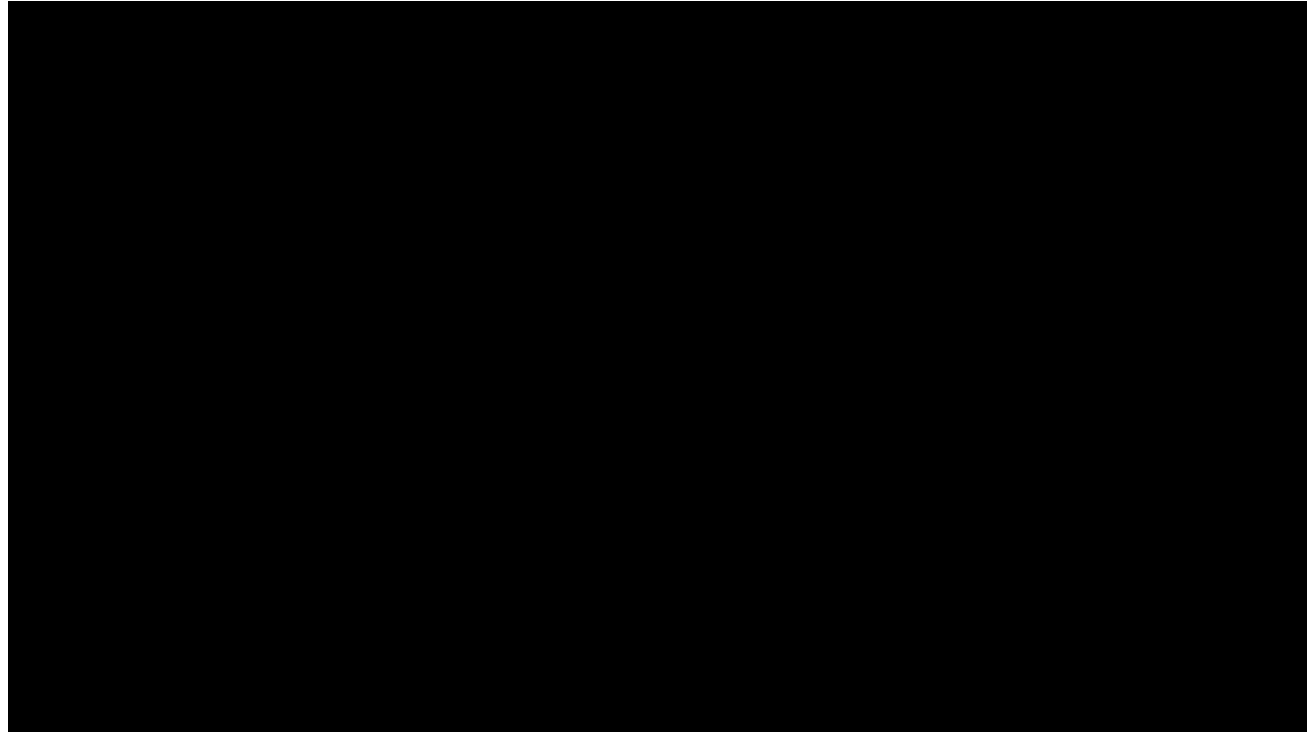
- 딥마인드의 구글 데이터 센터 에너지 관리
 - 관리자의 개입없이 데이터센터 에너지 관리가 AI 에이전트에 의해 자동으로 제어되고 있으며 40%의 에너지 사용량 절감이 있었다고 보고





강화학습의 예시

• 로봇 제어



https://www.youtube.com/watch?v=ydh_AdWZfIA

• 기타

- 주식 트레이딩 (Pit.ai, <https://techcrunch.com/2017/03/21/aihedgefund/>)
- 온라인 추천시스템
- 자연어처리 등

A grid of 3D wireframe rectangles, some tilted and some upright, arranged in a pattern that frames the central text.

THANK YOU

tcheong@korea.ac.kr