2025-02 Best Practice 공모전

RLab

구선주, 김민성, 이호영, 정다훈, 정지욱, 최예림

Contents

1. 팀 소개 및 활동 계획

RLab

Reinforcement Laboratory

Rlab은 강화학습(Reinforcement Learning)에 관심 있는 학생들이 모여
함께 공부한 내용을 공유하는 스터디 그룹입니다.

구선주	김민성	이호영	정다훈 (팀장)	정지욱	최예림
소프트웨어학과	소프트웨어학과	소프트웨어학과	소프트웨어학과	전자전기공학과	소프트웨어학과

활동 목표

활동 방법

- 강화학습의 핵심 이론(밴디트 문제, 마르코프 결정 과정, 벨만 방정식, TD법 등)에 대한 개념적 이해
- 주요 알고리즘(Q러닝, DQN, 정책 경사법 등)의 수식과 구현 원리 습득
- 사례 중심의 스터디를 통해 실제 문제 해결에 강화학습을 적용하는 능력 향상

학습 목표

- 매주 한 회차씩 주제를 정해 발표자 중심의 이론 설명과 질의응답
- 각 회차별 발표자는 발표 자료와 예제 문제를 준비해 참여자들과 공유
- ZOOM 또는 오프라인 장소(예: 도산라운지)를 활용한 발표 및 토론

활동 계획

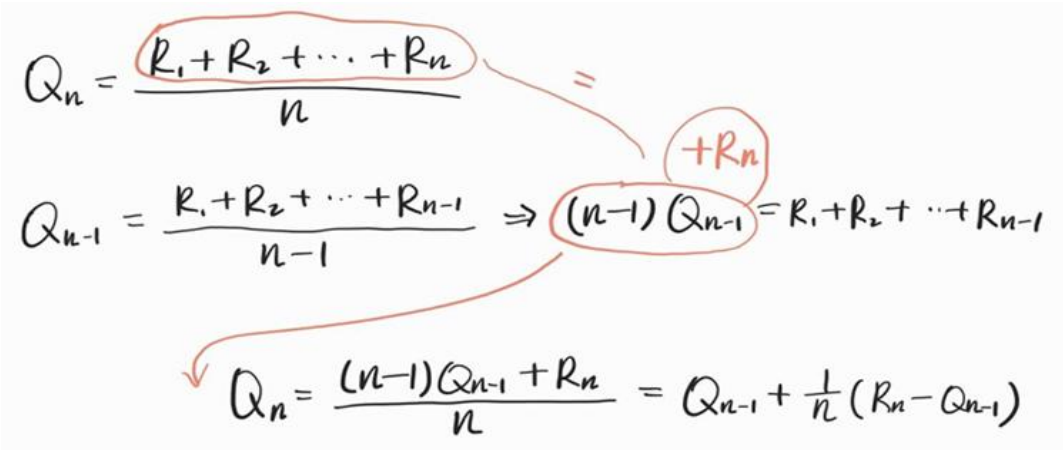
1주차	밴디트 문제 & 마르코프 결정 과정
2주차	벨만 방정식 & 동적 프로그래밍
3주차	몬테카를로법
4주차	신경망과 Q러닝
5주차	DQN
6주차	정책 경사법

2. 주차별 활동 내용 정리

1주차

학습 목표

밴디트 문제와 마르코프 결정 과정에 대한 기초 개념을
학습하고, 강화학습의 구조와 동작 원리를 이해한다.



학습 내용

- 밴디트 문제를 통해 탐색과 활용의 균형 개념을 익히고,
  MDP에서는 상태, 행동, 보상, 정책 등 강화학습의 핵심 요소들을 구조적으로 이해함.
- MP → MC → MRP → MDP로 이어지는 개념 흐름을 통해 강화학습의 이론적 틀을 체계적으로 정리함.

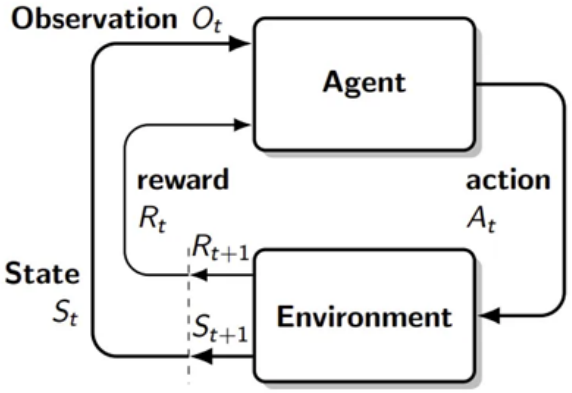
팀성찰

이번 활동을 통해 밴디트 문제와 MDP에 대한 실습을 통해 이론뿐만 아니라 실제 적용
가능성까지 함께 살펴보며 향후 심화 학습을 위한 기반을 마련할 수 있었다.

2주차

학습 목표

강화학습의 벨만 방정식과 동적 프로그래밍의 개념을 이해하고,
이를 통해 최적 정책 계산의 구조를 학습하여 복잡한 문제 해결 능력의 기초를 다진다.



학습 내용

- 벨만 방정식과 동적 프로그래밍(DP)을 중심으로 최적 정책 도출의 원리를 학습함
- 가치 반복과 정책 반복을 통해 상태 가치와 정책의 관계를 구조적으로
  이해하고 실습을 통해 개념을 구체화함

$$
\begin{aligned}
v(s) &= \mathbb{E}[G_t \mid S_t=s] \\
&= \mathbb{E}[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots \mid S_t=s] \\
&= \mathbb{E}[R_{t+1} + \gamma(R_{t+2} + \gamma R_{t+3} + \cdots) \mid S_t=s] \\
&= \mathbb{E}[R_{t+1} + \gamma G_{t+1} \mid S_t=s] \\
&= \mathbb{E}[R_{t+1} + \gamma v(S_{t+1}) \mid S_t=s] \quad \text{→ 재귀연산식}
\end{aligned}
$$

Immediate reward / State value

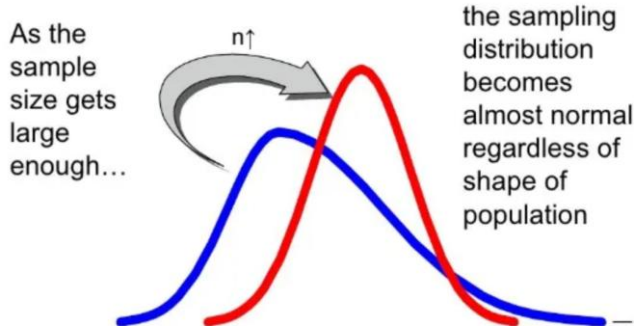
팀성찰

이번 활동을 통해 벨만 방정식과 동적 프로그래밍의 원리를 깊이 있게 이해하고,
수식과 구현을 함께 다루며 강화학습의 계산 구조에 대한 실질적인 감각을 키울 수 있었다.

3주차

학습 목표

몬테카를로법과 시간차 학습의 원리 및 차이점을 이해하고
각 기법이 강화학습에서 어떻게 사용되는지 학습한다.



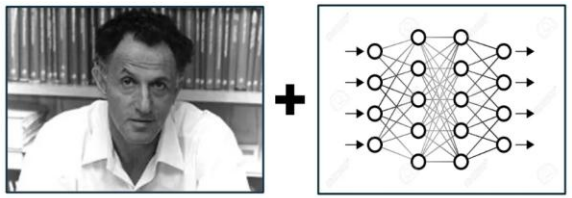
학습 내용

- **몬테카를로법(MC):** 에피소드가 종료된 후 얻은 전체 리턴을 기반으로 상태가치 또는 행동가치를 추정하며 반복을 통해 점차 정확도를 높이는 방식. → 완전한 에피소드가 필요
- **시간차 학습(TD):** 에피소드 종료를 기다리지 않고 현재 상태에서 얻은 보상과 다음 상태의 추정 가치를 기반으로 바로 업데이트를 수행함. → online/incremental learning이 가능
- 두 방법 모두 policy evaluation에 사용되며, 이후 일반화된 SARSA, Q-learning 등의 기법으로 확장됨.

팀성찰

이번 학습을 통해 MC와 TD의 차이를 이론적으로 정리하고 실제 강화학습 알고리즘을 구성할 때 어떤 상황에 어떤 방법을 적용해야 할지 판단할 수 있는 기초를 다질 수 있었음.

4주차



$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left(r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t) \right)$$

Experience Replay + Target Network + Reward Clipping

학습 목표

Q-learning의 한계를 극복하기 위해 신경망을 결합한
DQN(Deep Q-Network)의 구조와 학습 방식을 이해하고,
강화학습이 고차원 상태 공간에서도 적용 가능한 원리를 습득한다.

학습 내용

- 기존 Q-learning을 심층 강화학습으로 확장한 DQN의 구조와 학습 절차를 체계적으로 학습함.
- 상태 공간이 클 때 Q-table 방식이 갖는 한계를 설명하고, 이를 해결하기 위한 함수 근사(FNN)를 통한 Q-value 예측 방식을 학습함.
- DQN의 핵심 요소인 Experience Replay(경험 재사용)와 Target Network를 도입하여 학습 안정성을 확보하는 구조를 이해하고, 이를 코드 예시와 함께 실습함.

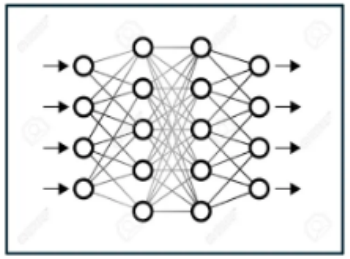
팀성찰

이번 학습을 통해 MC와 TD의 차이를 이론적으로 정리하고 실제 강화학습 알고리즘을 구성할 때 어떤 상황에 어떤 방법을 적용해야 할지 판단할 수 있는 기초를 다질 수 있었음.

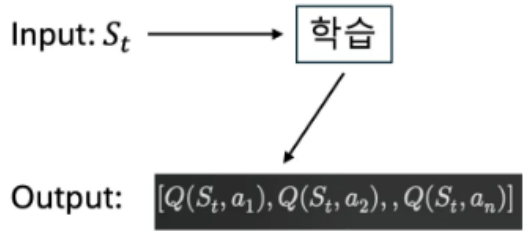
5주차

학습 목표

DQN 구조와 다양한 개선 기법을 학습함으로써, 더 안정적이고 효율적인 강화학습 모델 설계를 위한 기반을 다진다.



Q net



학습 내용

- Double DQN, Dueling DQN, Prioritized Replay, Fixed Q-target 등 DQN의 성능을 높이는 주요 기법들을 학습
- Atari 게임 사례와 코드 실습을 통해 적용 방법을 익힘

팀성찰

심화 기법 학습과 활발한 질의응답을 통해 이해도를 높였다. 시각화와 코드 실습으로 안정적인 학습 과정을 체감했으며, 모델 선택 기준을 고민하고 프로젝트 기반을 마련한 시간이었다.

6주차

학습 목표

정책 경사법의 수학적 원리와 학습 절차를 이해하고,
가치 기반 방법과의 차이 및 장단점을 파악하여 유연한 정책 최적화 기법을 익힌다.

학습 내용

- 정책 경사법의 핵심 개념과 수식을 학습함
- REINFORCE 알고리즘 및 baseline 기법을 통해 학습 안정화 방식을 익힘
- 실습을 통해 REINFORCE 구현과 정책 학습 과정을 시각적으로 확인함

	Φ_t
가장 간단한 정책 경사법	$G(\tau)$
REINFORCEMENT	G_t
베이스라인	$G_t - b(S_t)$
행위자-비평자	$R_t + \gamma V(S_{t+1}) - V(S_t)$
Q 함수	$Q(S_p, A_t)$
가치 함수 → 베이스라인	$Q(S_t, A_t) - V(S_t) = A(S_t, A_t)$

팀성찰

정책 경사법의 원리와 가치 기반 방식과의 차이를 구조적으로 이해하고, variance 문제와 baseline 전략을 함께 고민하며 정책 기반 학습의 핵심 개념을 익힌 의미 있는 시간이었다.

3. 활동 결과

활동 결과

- 각자 맡은 주제를 발표하며 강화학습 이론을 구조적으로 이해
- 복잡한 수식과 알고리즘을 코드 구현으로 연결해보는 경험
- 다양한 강화학습 알고리즘 간 차이점과 적용 맥락에 대한 통찰 확보
- 팀원 간의 활발한 질문과 피드백을 통해 상호 학습 시너지 실현

매주 도산라운지에 모여 회의를 진행

Notion페이지로 일정 관리 및 자료 저장

4. 활동 후기

활동 후기

구선주

강화학습 스터디를 통해 복잡한 개념들도 함께 고민하고 설명하며 자연스럽게 이해할 수 있었다.
혼자 공부할 때는 놓치기 쉬운 부분들을 서로 질문하고 보완하면서 학습의 깊이를 더할 수 있었고,
이론과 실습을 병행하며 강화학습의 원리를 실제로 체감해보는 뜻깊은 경험이었다.
앞으로는 개념 이해에 그치지 않고, 직접 응용하고 확장해보는 연습이 필요하다는 점도 느꼈다.

김민성

매주 주제를 맡아 발표하면서 단순히 읽는 수준을 넘어, 각 알고리즘의 수식과 코드가 어떤 의미를 갖는지 체계적으로 이해할 수 있었다. 혼자 공부했다면 어려웠을 내용을 팀원들과 함께 토론하며 풀어나가면서 성취감을 느낄 수 있었으며, 특히 Q러닝과 DQN 파트를 구현해보며 강화학습이 실제 문제 해결에 어떻게 쓰이는지 감을 잡을 수 있었던 뜻깊은 시간이었다.

활동 후기

이호영

강화학습은 처음 접하는 분야였기에 어렵고 막막했지만, 발표를 준비하고 팀원들과 주기적으로 공유하는 과정을 통해 점차 익숙해졌다. 이 책은 직접 구현을 병행해야 하는 구성이었기에 더 깊은 몰입이 가능했고, 발표 후 팀원들의 질문을 통해 이해하지 못한 부분도 점검할 수 있었다. 이 과정을 통해 강화학습에 대한 이해의 폭이 훨씬 넓어졌고, 강화학습을 낯설지 않게 받아들일 수 있게 되었다.

정다훈

매주 발표를 준비하면서 단순히 읽는 것보다 '설명할 수 있는 수준까지' 이해하는 것이 얼마나 중요한지를 실감했다. MDP와 벨만 방정식을 학습하면서 수식의 흐름을 코드로 옮기는 데 집중했고, 이 과정을 통해 이론과 구현을 자연스럽게 연결지을 수 있었다. 개인적으로는 자신감을 얻고 성취감을 느낀 활동이었다.

활동 후기

정지욱

책 전체를 끝까지 따라가면서 강화학습의 큰 흐름과 세부 알고리즘들의 구조를 이해할 수 있게 되었다. 정책 경사법이나 액터-크리틱 같은 내용도 어렵지만 재미있었고, 발표하면서 배운 내용을 팀원들에게 설명할 수 있었던 것이 가장 큰 보람이었다. 매주 준비하면서 점점 더 이론과 구현에 흥미를 느끼게 되었고, 스스로 깊이 파고드는 계기가 되었다.

최예림

처음에는 수식과 알고리즘이 너무 낯설고 어렵게 느껴졌지만, 팀원들과 함께 발표하고 토론하는 과정을 통해 점점 익숙해졌다. 특히 직접 코드를 구현하고 설명하며, 단순한 이론 암기를 넘어 실제로 '이해'했다는 느낌이 들어 성취감이 컸다. 강화학습에 대한 두려움을 넘고 흥미를 느끼게 된 소중한 경험이었다.

Thank You