


2025-1 단러닝클럽 활동보고서 (비교과형)

팀 이름	RLab	활동 회차	_____3_____회
팀 유형	<input checked="" type="checkbox"/> 퓨리클럽 (튜터링: O / <input checked="" type="checkbox"/> X)	<input type="checkbox"/> 두런클럽	<input type="checkbox"/> 글로벌클럽 (튜터링 : O / X)
활동 정보			
팀장 이름	정다훈	활동일자	2025.04.28
활동장소	퇴계기념도서관 도산라운지	활동시간	12:00 – 14:00
참석자	김민성, 구선주, 정다훈, 최예림, 이호영		
결석자	정지욱	총 참여인원	5
활동내용			
주제	강화학습에 대해 알아본다.		
목표	에피소드 기반의 강화학습 기법인 몬테카를로법(Monte Carlo Method)과 시간차 학습(TD: Temporal Difference)을 비교하여 이해하고, on-policy 및 off-policy 정책 학습 방식과 탐험/활용 전략에 대한 개념을 체계적으로 정립한다.		
학습 내용	<p>이번 회차에서는 몬테카를로법(Monte Carlo Method)의 정의 및 작동 방식을 중심으로 학습하였다. 이 기법은 전체 에피소드를 기반으로 보상을 추정하는 방식으로, 환경 모델 없이도 경험을 통해 정책을 평가하고 개선할 수 있는 장점이 있다.</p> <p>이어 TD 학습(Temporal Difference Learning)과의 비교를 통해 에피소드 종료 전에도 학습이 가능한 TD의 효율성과 빠른 수렴 특성을 확인하였다. 또한 on-policy와 off-policy 정책 학습의 개념을 통해, SARSA와 Q-learning 등의 차이를 이해하고, 이를 구현하는 데 있어 필요한 정책 평가와 개선 전략도 함께 학습하였다.</p> <p>마지막으로 ϵ-greedy 정책을 활용한 탐험(Exploration)과 활용(Exploitation)의 균형 조정 전략도 실습을 통해 확인하였다.</p>		
팀성찰	<p>이번 활동은 강화학습의 실제 적용 가능성을 엿볼 수 있는 의미 있는 세션이었다. 특히 몬테카를로 방식과 TD 방식의 차이점에 대해 팀원들이 예제를 중심으로 토론하며 각 방식의 강점과 한계를 구체적으로 비교할 수 있었다.</p> <p>on-policy와 off-policy 개념도 서로 질문을 주고받으며 모호했던 부분을 명확히 할 수 있었고, ϵ-greedy 정책과 같은 실질적 전략에 대해 구현 예시를 함께 보며 실습 중심의 이해를 확장시켰다.</p> <p>전체적으로 강화학습 알고리즘들의 구조와 차이점을 구조적으로 바라보는 기회를 가졌으며, 이후 심화된 가치 기반 및 정책 기반 학습으로 이어질 기반을 다지는 시간이었다.</p>		
활동증빙	활동사진	활동자료 사진	
		<ul style="list-style-type: none"> 지원받은 도서를 활용하여 학습한 경우 도서 활용 증빙 사진을 첨부하세요. 	

개별성찰	구선주	이번 스터디에서는 TD(Temporal Difference) 학습, on-policy와 off-policy의 차이, 그리고 ϵ -greedy 정책에 대해 학습하였다. TD 학습은 실제로 다음 상태에서 얻은 보상을 기반으로 가치 함수를 업데이트하는 방법으로 샘플 하나로도 학습이 가능하다는 점에서 Monte Carlo 방식과 구별되었다. 또한 SARSA(on-policy)와 Q-learning(off-policy)의 차이와 정책 평가 방식, ϵ -greedy 정책을 통한 탐험과 활용의 균형 유지 방법에 대해 실제 예시를 통해 깊이 이해할 수 있었다.
	김민성	이번 스터디에서는 TD(Temporal Difference) 학습, on-policy와 off-policy의 차이, 그리고 ϵ -greedy 정책에 대해 학습하였다. TD 학습은 실제로 다음 상태에서 얻은 보상을 기반으로 가치 함수를 업데이트하는 방법으로 샘플 하나로도 학습이 가능하다는 점에서 Monte Carlo 방식과 구별되었다. TD는 예측과 실제 결과를 비교해 오차를 즉시 수정한다는 특징이 있어, 빠르고 효율적인 학습이 가능하다는 점을 이해했다. on-policy와 off-policy는 정책과 가치 함수 업데이트의 관계를 구분하는 개념으로 on-policy는 현재 따르는 정책을 그대로 평가하고 개선하는 반면, off-policy는 다른 행동 정책을 따라가면서 목표 정책을 학습할 수 있다. 이 차이는 탐험과 수렴 속도에 영향을 미치며 대표적으로 SARSA(on-policy)와 Q-learning(off-policy) 알고리즘이 존재한다는 점을 배웠다. 또한 ϵ -greedy 정책은 탐험(exploration)과 활용(exploitation)을 균형 있게 수행하기 위한 방법이다. 대부분 최선의 행동을 선택하지만 ϵ 확률로 무작위 행동을 선택함으로써 새로운 상태나 보상을 탐색할 수 있도록 한다. 이 방식을 통해 최적 정책을 찾기 위한 충분한 탐험이 보장될 수 있다는 점을 이해했다.
	이호영	이번 학습을 통해 몬테카를로 방법이 에피소드 전체를 활용해 보상을 추정하는 방식이라는 점을 알게 되었고, ϵ -greedy 전략이 탐험과 이용 사이의 균형을 조절하는 방법이라는 것도 명확히 이해할 수 있었다. 또한 온정책과 오프정책의 차이를 정확하게 구분하면서 강화학습 알고리즘들의 학습 구조를 좀 더 체계적으로 바라볼 수 있게 되었다.
	정다훈	이번 스터디를 통해 몬테카를로법에 대한 기본 개념을 이해할 수 있었다. 강화학습 모델에서는 항상 모델을 정확하게 알 수 없기 때문에 경험적 추정으로 정책을 개선하는 방법이 필수적이라는 것을 알게 되었다. on-policy와 off-policy 정책 학습의 구조를 정리하며, SARSA와 Q-learning 알고리즘의 차이를 명확히 인식하게 되었다. 이러한 내용을 통해 강화학습에 대한 전체적 이해도를 높일 수 있었다.
	정지욱	이번 학습을 통해 몬테카를로 방법이 에피소드를 전체 활용해 보상을 추정하는 방식이라는 점을 알게 되었고, ϵ -greedy 전략이 탐험과 이용 사이의 균형을 조절하는 방법이라는 것도 명확히 이해할 수 있었다. 또한 온정책과 오프정책의 차이를 정확히 구분하면서 강화학습 알고리즘들의 학습 구조를 좀 더 체계적으로 바라볼 수 있게 되었다.
	최예림	이번 스터디를 통해 몬테카를로법에 대한 기본 개념을 이해할 수 있었다. 강화학습 모델에서는 항상 모델을 정확하게 알 수 없기 때문에 경험적 추정을 통해 개선 하는 방법이 필수적이라는 것을 알게 되었다. on-policy와 off-policy에 대한 개념을 적립할 수 있었다. 해당 정책들을 사용한 SARSA와 Q-Learning 알고리즘에 대해서도 새롭게 알게 되어 강화학습에 대한 전반적인 이해를 높일 수 있었다.

※ 작성 후 반드시 PDF파일로 저장하여 영웅스토리에 업로드 하세요.