

[붙임5] 2026년 소·중·대 산학협력프로젝트(캡스톤디자인) 결과보고서

2026년 전남대학교 소프트웨어중심대학사업 소·중·대 산학협력프로젝트(캡스톤디자인) 결과보고서

프로젝트명	강화학습을 이용한 협력 보드게임 AI					
Github url 주소	https://github.com/miffy9999/SpaceGent.git					
팀 명	Divergent			과제수행기간	2026. 3. 25. ~ 6. 23.	
지도교수	학 과	컴퓨터정보통신공학과		성 명	김명진	
프로젝트 수행인원 (※팀장은 첫줄에 기입)	학과(부·복수전공)	학년	학번	성명	연락처	E-Mail
	컴퓨터정보통신공학과	3	223400	박진석	010-6423-8996	jjinseok3639@gmail.com
	컴퓨터정보통신공학과	3	223364	정의준	010-9177-8996	apea8888@gmail.com
참여 기업	기업명	멘토명		직위	연락처(HP)	E-Mail
	모바일앱개발협동 조합	최원서		대표	010-4537-7808	dev@mobileap pdev.kr

위와 같이 2026년 전남대학교 소프트웨어중심대학사업
산학협력프로젝트(캡스톤디자인) 지원 프로그램 결과보고서를 제출합니다.

2026년 6월 21일

신청자명(대표학생) : 박진석
지도교수 : 김명진

박진석
(인)

전남대학교 소프트웨어중심대학사업단장 귀하

산학협력프로젝트(캡스톤디자인) 결과보고서(요약)

프로젝트명	강화학습을 이용한 협력 보드게임 AI		
수행기간	2026. 3. 25. ~ 6. 23.	소요예산	280,208원
소요예산 세부내역	-회의비: 177,000원 -SW활용비: 103,208원 -재료비: 0원		
참여인원	구분	인원수	성명(모두 기재)
	교수	1	김명진
	석박사과정	-	
	학부생	2	박진석, 정의준
	기업체	1	최원서
	계	4	
추진배경	협동 카드게임은 부분관측·제한통신 하에서 다수 에이전트가 공동 목표를 달성해야 하는 다중 에이전트 강화학습의 대표적 시험대이다. 대상 게임 The Crew는 완전협동·부분관측·극도로 제한된 통신·all-or-nothing 구조를 모두 갖추어, 완전정보 버전조차 NP-complete이며 통신이 제한되면 결정불가능에 속한다. 게다가 선행 연구들은 표준 학습 환경과 비교 기준이 부재하여 상호 비교가 어렵다. 이에 게임 전체를 한 번에 푸는 비현실적 목표 대신, 어려움의 본질만 보존한 최소 환경에서 강화학습의 학습 가능성과 병목을 규명할 필요가 있다.		
목표 및 내용	The Crew의 어려움의 핵심(은닉정보 하 협력·제한통신·조합적 과제 배분)만 남긴 최소 학습 커널을 Unity ML-Agents 기반으로 구현하고, 다중 에이전트 강화학습(MA-POCA)이 협력 정책을 어디까지 학습할 수 있는지 측정하는 것을 목표로 한다. 핵심 방법론으로 사람이 설계한 휴리스틱 참조 기준선, 탐색 기반 SO-ISMCTS 정책, 강화학습을 동일 환경에서 비교하는 3중 체계를 구축해 과제 난이도 문제와 학습 실패 문제를 분리하였다. 강화학습은 단일 과제 조건에서 미션 성공률 약 38%에 수렴해 랜덤 드래프트 휴리스틱 기준선(37.8%)과 유사한 수준이었다. 정체 원인을 규명하기 위한 분해 실험에서, 플레이를 고정하고 드래프트만 학습시킨 경우 51.2%로 자기손패 기준선(54.6%)에 근접한 반면, 드래프트를 고정하고 플레이만 학습시킨 경우 33.3%에 정체하여 잔여 병목이 협력 플레이 쪽에 있을 가능성을 진단하였다.		
기대효과	본 환경과 3중 비교 체계는 표준 환경과 비교 기준이 없던 The Crew 협력 연구의 공백을 메우는 재현 가능한 통합 환경을 제공한다. 또한 강화학습의 정체를 과제 난이도와 학습 실패로 분해해 진단함으로써, 후속 연구가 자원을 집중할 방향을 제시한다. 학습된 롤아웃·상대 모델링과 탐색의 혼합 등 후속 방법론의 정량적 비교 기준선으로도 기능할 수 있다.		

1. 프로젝트 개요

프로젝트명	강화학습을 이용한 협력 보드게임 AI
주제영역	☐ 생활 ☐ 업무 ☐ 공공/교통 ☐ 금융/핀테크 ☐ 의료 ☐ 교육 ☐ 유통/쇼핑 ☒ 엔터테인먼트
기술분야	☐ IoT ☐ 모바일 ☐ 데스크톱 SW ☒ 인공지능 ☐ 보안 ☐ 가상현실 ☐ 빅데이터 ☐ 자동제어기술 ☐ 블록체인 ☐ 영상처리 ☐ 기타()
성과목표	☒ 논문게재 및 포스터발표 ☐ 앱등록 ☐ 프로그램등록 ☐ 특허 ☐ 기술이전 ☐ 실용화 ☐ 공모전(<i>공모전명</i>) ☐ 기타()

2. 프로젝트 추진배경

협동 카드게임은 부분관측·제한통신 환경에서 다수의 에이전트가 공동 목표를 달성해야 하는 다중 에이전트 강화학습(MARL)의 대표적 시험대이다. 본 과제의 대상인 The Crew: 아홉 번째 행성을 찾아서는 4인 완전협동 트릭테이킹 게임으로, ① 팀 전원이 함께 이기거나 지는 완전 협동 구조, ② 자기 손패만 관측 가능한 부분관측, ③ 미션당 극도로 제한된 통신 토큰, ④ 과제 하나라도 실패하면 미션이 실패하는 all-or-nothing 구조를 모두 갖는다.

이 게임은 이론적으로도 어려운 부류에 속한다. 모든 패가 공개된 완전정보 버전조차 NP-complete이며, 통신 대역폭이 행동 대역폭보다 좁은 부분관측 협동게임은 결정불가능(undecidable)에 속한다는 결과가 알려져 있다. 즉 제한된 통신 토큰은 단순한 제약이 아니라 협력 학습을 본질적으로 어렵게 만드는 구조적 장벽이다. 한편 The Crew를 다룬 선행 연구들은 SAT 분석, 그리디 휴리스틱, MCTS, Transformer 기반 계획 등 다양하게 존재하나 표준 학습 환경과 비교 기준이 부재해 상호 비교가 어렵다. 따라서 게임 전체를 한 번에 푸는 비현실적 목표 대신, 어려움의 본질만 보존한 최소 환경에서 강화학습의 학습 가능성과 병목을 규명할 재현 가능한 통합 환경이 필요하여 본 과제를 선정하였다.

3. 프로젝트(주제) 목표 및 내용

본 과제의 최종 목표는 The Crew의 어려움의 핵심(은닉정보 하 협력·제한통신·조합적 과제 배분)만 남긴 최소 학습 커널을 Unity ML-Agents 기반으로 구현하고, 다중 에이전트 강화학습(MA-POCA)이 협력 정책을 어디까지 학습할 수 있으며 그 한계와 병목은 무엇인지 측정하는 것이다.

이를 위해 게임 구성 요소를 학습 대상과 제외 대상으로 명시적으로 분리하였다. 학습 루프에 포함한 핵심 커널은 ① 드래프트(과제 배분), ② 트릭테이킹(협력 플레이), ③ 통신 토큰(제한된 정보 공유) 세 가지이며, 50개 미션의 특수규칙·조난신호 등은 관측 벡터에 슬롯만 예약(헤드룸)하고 비활성화하여 확장 가능한 설계로 환원을 되돌릴 수 있게 하였다.

핵심 방법론은 강화학습 결과를 해석할 독립 기준을 마련하는 3중 비교 체계이다. ① 사람이 설계한 휴리스틱 참조 기준선, ② 탐색 기반 SO-ISMCTS 협력 베이스라인, ③ MA-POCA 강화학습을 동일 환경에서 측정하여 과제 난이도 문제와 학습 실패 문제를 분리하였다.

주요 결과로, 강화학습은 단일 과제 조건에서 미션 성공률 약 38%에 수렴해 랜덤 드래프트 휴리스틱 기준선(37.8%)과 유사한 수준이었다. 정체 원인을 규명하기 위한 분해 실험에서 플레이를 고정하고 드래프트만 학습시킨 경우 51.2%로 자기손패 기준선(54.6%)에 근접한 반면, 드래프트를 고정하고 플레이만 학습시킨 경우 33.3%에 정체하여, 잔여 병목이 협력 플레이 쪽에 있을 가능성을 진단하였다.

4. 시스템 구성 및 내용

전체 시스템은 ① 학습 환경 계층, ② 학습 알고리즘 계층, ③ 비교·측정 계층의 세 계층으로 구성된다.

학습 환경 계층은 Unity ML-Agents 상에 구현한 최소 커널로, 드래프트·트릭데이킹·통신 토큰을 처리한다. 관측 벡터는 총 313차원으로 내 손패(40), 바닥/과제 풀 슬롯(40), 선택상(5), 팀 태스크(162), 플레이어별 남은 손패·트릭 승리 수(8), 페이즈 플래그(2), 통신(24), 특수규칙 예약(32)으로 구성된다. 행동 공간은 [10, 2, 4]이며 페이즈별 마스킹으로 드래프트 선택과 트릭 플레이를 분리하고, follow-suit 위반 카드는 자동 마스킹된다.

학습 알고리즘 계층은 MA-POCA(Multi-Agent POrthumous Credit Assignment)로, 4개 에이전트가 하나의 그룹 보상을 공유하고 중앙집중 critic이 각 에이전트의 기여도를 추정해 신용을 할당한다. 동일한 신경망이 페이즈 마스킹을 통해 드래프트와 플레이를 모두 출력하는 단일 네트워크 구조를 사용한다. 보상은 과제 완수 +1/N, 실패 -1(즉시 종료), 전부 완수 +1.0으로 설정하였다.

비교·측정 계층은 게임 규칙을 파이썬으로 재현한 휴리스틱 참조 기준선, SO-ISMCTS 탐색 정책, 강화학습 결과를 동일 환경에서 측정·대조하는 3중 체계로 구성되며, SSH 터널링이 차단된 원격 서버 환경에서 tfevents를 직접 파싱하는 분석 파이프라인을 포함한다.

5. 프로젝트 결과물에 대한 기술

본 과제의 최종 결과물은 재현 가능한 통합 학습 환경과 3중 비교 체계, 그리고 단일 과제 협력 강화학습 베이스라인이다.

첫째, Unity ML-Agents 기반의 최소 학습 커널과 순수 파이썬 규칙 재현 환경을 구축하고, 휴리스틱 참조 사다리·SO-ISMCTS·MA-POCA를 동일 환경에서 비교하는 3중 체계를 완성하였다. 휴리스틱 참조 기준선은 단일 과제 조건에서 무작위 24.9%, 협력+랜덤 드래프트 37.8%, 협력+자기 손패 드래프트 54.6%, 완전정보 최적 드래프트 91.9%로 측정되어 강화학습 결과를 해석할 독립 척도를 제공한다.

둘째, SO-ISMCTS 협력 베이스라인은 단일 owner-held 과제에서 46.5%로 규칙기반(23.5%) 대비 약 2배, 목표 카드가 동료 손에 갇히는 협력 시나리오에서는 능동적 피딩(active feeding) 적용 후 미션 성공 34.6%·완수율 49.9%로 규칙기반 대비 큰 격차를 보였다.

셋째, 강화학습 베이스라인은 미션 성공률 약 38%에 수렴하였고, 통신 추가·관측 확장·커리큘럼 임계값 하향 등 정제 원인 후보를 데이터로 차례로 기각하였다. 이어 2×2 분해 실험으로 드래프트는 격리 시 기준선에 근접하나(51.2%) 플레이는 그렇지 못함(33.3%)을 확인하여, 잔여 병목이 협력 플레이 쪽에 있다는 예비 진단을 얻었다. 다만 분해 실험은 단일 시드·미완주 상태의 예비 결과임을 명시한다.

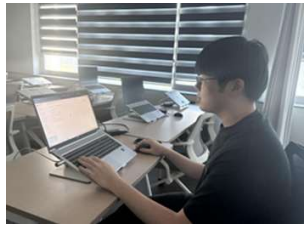
6. 프로젝트 진행내용

1) 참여인원 및 담당 역할

연번	소속학과	성명	수행역할 분담내용
1	컴퓨터정보통신공학과	박진석	원격 서버에 학습 환경 구축 및 RL 학습 관리. 휴리스틱을 활용한 task 완수율 측정. 선행연구 조사 및 분석.
2	컴퓨터정보통신공학과	정의준	The Crew 를 복을 참조한 Unity RL 학습 환경 모델링. SO-ISMCTS를 활용한 task 완수율 측정. Github 협업 환경 구축.

2) 회의 및 SW멘토링 진행

번호	일시/장소	회의/멘토링 내용(상세히 작성)	관련 사진
1	2026. 3. 31. (12:00~13:30) / 공대7호관 229호	- 딥 씨 크루게임을 환경에 모델링 계획 수립 - 버전 및 라이브러리 버전 설정클라우드 활용 방안 탐색. - 멘토링 시 질문할 내용 및 프로젝트 내용 정리	
2	2026. 4. 9. (12:00~13:30) / 공대7호관 229호	- Git LFS를 사용하여 보드게임 모델링을 위한 에셋을 관리하고 README에 설치 프로세스 업로드 하여 공유. - .gitignore 파일 설정 및 ML-Agents 인터페이스 및 프로젝트 구조 README에 반영하여 공유. - GameManager 추가 및 버그 수정. 통신 토큰 메커니즘 및 카드 UI 스프라이트 연결. 소나 토큰 및 게임 UI (HUD) 추가. - ML-Agents 학습 시 필요한 연산량 및 메모리 계산. 이에 따른 컴퓨팅 자원 및 소요 시간 계산	
3	2026. 4. 28. (12:00~13:30) / 공대7호관 229호	- 게임 플레이 안정화 - 교내 GPU 서버 연결 테스트 및 환경 구축 - 팀 역할 분담 및 규칙 수정 반영하여 README 업데이트	

4	2026. 6. 10. (15:30~17:00) / 공대7호관 229호	- task 배정 순간 owner_score(드래프트 보상)로 뽑힌 owner가 4명 중 얼마나 좋은가를 평가해 그룹 보상 ± 0.3(DraftShapeScale). 드래프트 액션에 즉시 크레딧하도록 수정 - 드래프트 ϵ-greedy 커리큘럼 추가, shaping 제거 반영. - 트릭 승리 카드(winningCardsInHand) 별도 누적하도록 결정	
5	2026. 6. 12. (17:00~18:30) / 공대7호관 229호	- 인수인계·발표·재현을 위해 로컬 메모리/세션 분석 - MCTS 성능 측정의 feeding 우선순위 복구	

7. 프로젝트 세부일정 및 내용

No.	작업 내용	3월				4월				5월				6월				담당자	비고
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4		
1	계획서 작성																	박진석	
2	선행연구 조사 및 분석																	박진석	
3	최소 학습 커널 설계 및 환경 모델링																	박진석	
4	Unity ML-Agents 학습 환경 구축																	정의준	
5	휴리스틱 참조 기준선 측정																	박진석	
6	SO-ISMCTS 구현 및 협력 베이스라인 측정																	정의준	
7	강화학습 학습 및 분해 실험																	박진석	

8. 결과물에 대한 향후 활용계획

본 과제에서 구축한 통합 학습 환경과 3중 비교 체계, 예비 단계의 분해 실험 결과는 후속 연구로 자연스럽게 이어진다.

첫째, 예비 결과의 확정이다. 단일 시드·미완주 상태인 두 분해 실험을 4시드·20M 스텝으로 완주하고, 특히 탐색 강제 스케줄을 받지 못했던 플레이 학습 실험에 동일 처방을 적용해 재측정함으로써 잔여 병목이 협력 플레이 쪽에 있다는 잠정 해석을 엄밀히 검증한다.

둘째, 환원했던 게임 요소의 점진적 복원이다. 관측 벡터에 예약해 둔 특수규칙 32칸을 단계적으로 활성화하여 신경망 재구성 없이 전체 게임 수준으로 학습 가능성을 끌어올린다.

셋째, OpenSpiel 기반 공개 연구 환경 배포이다. 현재 Unity·파이썬으로 이원화된 환경을 OpenSpiel 표준 인터페이스에 맞추어 리팩터링·공개하여, 후속 연구자가 별도 구현 없이 여러 알고리즘을 비교할 수 있는 재현 가능한 벤치마크로 제공한다.

넷째, 학위논문으로의 연결이다. 분해 실험의 완주와 엄밀화 항목을 보완하여, 캡스톤에서 설계·예비 검증한 가설을 학위논문에서 확정하는 연속적 연구 흐름을 구성한다. 이로써 표준 환경 부재라는 선행 연구의 공백을 메우고, 학습된 롤아웃·상대 모델링과 탐색의 혼합 등 후속 방법론의 정량적 비교 기준선으로 활용될 수 있다.

9. 참고자료 및 문헌

[1] Reiber 외, The Crew: The Quest for Planet Nine is NP-Complete (2021)

[2] Coulombe & Lynch, Characterizing the Decidability of Finite State Automata Team Games with Communication (2022)

[3] An Analysis of the Cooperative Card Game The Crew (2021)

[4] Chan, Performance of a Multi-Agent Greedy Algorithm in a Cooperative Game with Imperfect Information (2021)

[5] Litjens, An MCTS-based AI Playing The Crew (2022)

[6] Rebstock 외, Transformer Based Planning in the Observation Space with Applications to Trick Taking Card Games (2024)

[7] The Crew: a Strategy Analysis (Leiden, 2025)