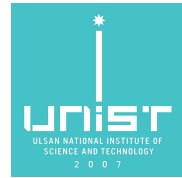


보도자료

온라인은 4.21(화) 배포 즉시
지면 4.22(수) 조간
부터 보도해 주시기 바랍니다



자료문의

대외협력팀: 장준용 팀장, 양윤정 담당
인공지능대학: 한승열 교수

052-217-1227
052-217-3455

UNIST 한승열 교수팀, 세계 3대 AI 학회 ICLR에 논문 3편 동시 채택 - 인공지능대학원 한승열 교수팀 피지컬 AI의 필수 기술인 강화학습 분야에서 성과 거둬 - 데이터 오류·장기 과업·협력 환경 변화 대응하는 강화 학습 기술 제시-

UNIST가 피지컬 AI 기술의 핵심인 강화학습 분야에서 국제적 연구 경쟁력을 입증했다.

UNIST는 인공지능대학원 한승열 교수팀의 연구 성과 논문 3편이 오는 23일 브라질 리우데자네이루에서 열리는 표현학습국제학회(ICLR)에 채택됐다고 21일 밝혔다. ICLR은 신경정보처리시스템학회(NeurIPS), 국제머신러닝학회(ICML)와 함께 세계 3대 AI 학회로 꼽히는 학회다.

최상위 AI 학회는 심사 기준이 까다로워 단일 연구실에서 3편의 논문이 동시에 채택된 것은 드문 사례다. 실제 올해 ICLR에는 전 세계에서 1만 9000여 편의 논문이 제출됐으며 이 중 약 27%인 5,300여 편만이 심사를 통과했다.

이번 성과는 모두 강화학습 분야에서 나왔다. 강화학습은 AI가 환경과 상호작용하며 시행착오를 통해 최적의 행동을 스스로 찾아내는 학습 방식으로, 로봇이나 자율주행차가 불확실하고 예측 불가능한 현실 물리 환경에 직접 부딪히며 상황을 인지하고 돌발 변수에 대처하도록 만들 수 있어 피지컬 AI 시대의 필수 기술로 꼽히고 있다.

한 교수팀의 첫 번째 성과는 산업 현장에서 바로 수집한 오프라인 데이터로도 인공지능을 효과적으로 학습시킬 수 있는 ‘자기 개선 스킬 학습법’(SISL)이다. 스킬은 복잡하고 긴 작업을 효율적으로 해결하기 위해 로봇 등의 연속적인 행동을 재사용 가능한 여러 세부 단위로 나눈 것이다. SISL은 더 유용한 스킬을 능동적으로 발견하고, 모델 내의 노이즈를 지속적으로 제거해 오류가 섞인 데이터로도 높은 작업 성공률을 유지할 수 있다.

복잡하고 긴 작업의 성공률을 높일 수 있는 기술도 내놨다. 강화학습은 과업이

복잡할 경우 목표를 여러 개의 하위 단위로 쪼개 수행하는데, 문제는 로봇이 물리적으로 도달할 수 없는 잘못된 하위 목표를 설정할 경우 전체 학습 효율이 떨어진다. 연구팀은 실패한 시도와 부분적으로 성공한 데이터를 분석해 도달 가능한 목표를 명확히 구분하는 ‘엄격한 하위 목표 실행(SSE)’ 학습기술을 개발해 이 문제를 해결했다.

군집비행과 같이 여러 AI 에이전트들이 협력해야 하는 상황에서 발생하는 최적화 문제를 해결할 수 있는 해법도 제시했다. 다중 에이전트 강화학습(MARL) 환경에서 훈련 중 상황이 변해 최적의 행동 기준이 달라질 때 AI가 차선책에만 머무르는 현상을 극복한 성과다. 연구팀은 하나의 최적 행동에만 의존하던 기존 알고리즘의 단점을 보완해, 여러 대안적 행동 가치를 동시에 기억하고 평가하는 ‘연속적 하위 가치 Q-러닝(S2Q)’ 기술을 개발했다.

이상현, 황재박, 조용현 연구원이 각 연구의 제1저자로 참여했다.

연구수행은 과학기술정보통신부 정보통신기획평가원의 ‘자율드론상용화를 위한 과제 지향 강화학습 핵심기술 개발’, ‘AI 스타 펠로우십 프로그램’, ‘인공지능대학원 지원 사업’과 한국연구재단의 ‘LLM 기반 다중 에이전트 강화학습을 통한 대규모 자율군집제어 End-to-End 기술 개발 사업’의 지원을 받아 이뤄졌다.

한승열 교수는 “이번 연구는 제한된 데이터와 불확실한 환경에서도 강화학습을 안정적으로 적용할 수 있는 가능성을 보여준 것”이라며 “자율주행, 로봇, 스마트 제조 등 다양한 분야로의 확장이 기대된다”고 말했다.

붙임: 연구결과개요 3편, 연구자 이력사항. (끝)

연구 결과 개요 (SISL)

논문명	SELF-IMPROVING SKILL LEARNING FOR ROBUST SKILL-BASED META-REINFORCEMENT LEARNING
저자정보	한승열(교신저자), 이상현(제1저자), 배상준, 박이삭

1. 연구배경

강화학습은 시행착오를 통해 보상을 최대화하는 방향으로 행동을 스스로 배우는 인공지능 학습 방식이다. 메타 강화학습은 이러한 학습 경험을 바탕으로, 처음 보는 새로운 문제에도 빠르게 적응할 수 있도록 만든 확장된 강화학습 방법이다. 하지만 실제 환경의 문제는 여러 단계를 거쳐야 하는 경우가 많아, 단순한 행동 학습만으로는 한계가 있다. 이를 해결하기 위해 '행동 묶음'인 스킬(skill)을 미리 학습해 사용하는 방식이 도입됐지만, 학습에 쓰이는 데이터에 오류나 불완전한 행동이 섞여 있을 경우 잘못된 행동까지 함께 학습되는 문제가 있었다. 특히 이러한 오류는 긴 작업 과정에서 누적되며 전체 성능을 떨어뜨리는 요인으로 작용했다.

2. 연구내용

연구팀은 데이터의 잡음을 줄이면서 스킬을 점진적으로 개선하는 '자기 개선형 스킬 학습(SISL)' 방법을 제안했다.

이 방법은 이미 배운 행동을 활용해 문제를 해결하는 과정과, 기존 데이터 주변에서 더 나은 행동을 새로 탐색하는 과정을 분리해 동시에 수행하는 구조다. 탐색 과정에서 성과가 좋은 행동만 골라 다시 학습에 반영하면서, 학습 데이터 자체를 점점 정제해 나간다. 또한 기존 데이터에도 다시 점수를 매겨 실제 문제 해결에 기여하는 행동에 더 높은 비중을 두도록 설계했다.

이 방법의 효과는 여러 장기 과제 환경에서 확인됐다. 로봇이 여러 동작을 순서대로 수행하는 주방(Kitchen) 환경이나, 미로를 찾아가는 Maze2D 환경 등에서 기존 방법은 데이터에 잡음이 많아질수록 성능이 크게 떨어졌다. 반면 제안된 방법은 잡음이 큰 조건에서도 안정적으로 높은 성능을 유지했다.

특히 일부 미로 환경에서는 잡음이 큰 데이터에서도 거의 모든 과제를 성공적으로 해결했으며, 기존 방법들이 성능 저하를 보이는 상황에서도 일관된 학습 결과를 보였다. 이는 학습 과정에서 불필요한 데이터를 걸러내고 유용한 행동 중심으로 스킬을 개선한 결과다.

3. 기대효과

이 기술은 데이터가 완벽하지 않은 현실 환경에서도 안정적으로 학습할 수 있는 강화학습 방법을 제시한다는 점에서 의미가 있다.

특히 로봇 제어나 자율주행처럼 여러 단계를 거쳐야 하는 문제에서, 일부 잘못된 데이터가 전체 성능을 망가뜨리는 문제를 줄이고 보다 신뢰성 있는 의사결정을 가능하게 할 것으로 기대된다.

연구 결과 개요 (SSE)

논문명	STRICT SUBGOAL EXECUTION: RELIABLE LONG-HORIZON PLANNING IN HIERARCHICAL REINFORCEMENT LEARNING
저자정보	한승열(교신저자), 황재박(제1저자), 이상현, 김정모

1. 연구배경

강화학습은 시행착오를 통해 보상을 최대화하는 방향으로 행동을 배우는 인공지능 기술이다. 목표 지점을 정해두고 그곳에 도달하도록 학습하는 방식도 널리 쓰이는데, 실제 환경에서는 목표까지 여러 단계를 거쳐야 하는 경우가 많아 학습이 어려워진다.

이를 해결하기 위해 문제를 여러 중간 목표(서브골)로 나눠 단계적으로 해결하는 방식이 사용된다. 하지만 기존 방법은 “중간 목표에 실제로 도달했는지”를 엄격하게 따지지 않고 학습을 진행하는 경우가 많아, 도달하기 어려운 목표를 계속 선택하거나 불필요하게 많은 단계를 거치는 문제가 있었다. 이런 비효율은 긴 작업일수록 더 크게 나타났다.

2. 연구내용

연구팀은 중간 목표를 확실하게 달성한 경우에만 다음 단계로 넘어가도록 하는 ‘Strict Subgoal Execution(SSE)’ 방법을 제안했다.

이 방법은 목표 공간을 그래프로 나누고, 현재 상태에서 도달 가능한 중간 목표만 선택하도록 학습을 진행한다. 특히 시도 과정에서 실패한 지점과 일부만 성공한 지점을 구분해 기록하는 방식(Frontier Experience Replay)을 도입해, 실제로 도달 가능한 범위를 경계처럼 구분한다. 이를 통해 도달하기 어려운 목표는 자연스럽게 배제되고, 신뢰할 수 있는 경로 중심으로 학습이 이루어진다.

또한 실패가 자주 발생하는 경로는 비용을 높게 설정해 회피하도록 만들고, 탐색 전용 정책을 따로 뒤 아직 충분히 탐색되지 않은 영역도 함께 살펴보도록 했다. 이 구조를 통해 불필요한 시도를 줄이면서도 전체 탐색 범위는 유지할 수 있도록 설계했다.

이 방법의 효과는 다양한 장기 과제 환경에서 확인됐다. 미로 탐색(AntMaze), 열쇠를 순서대로 획득해야 하는 KeyChest, 장애물을 피해 목표를 찾아가는 Reacher 과제 등에서 기존 방법들은 중간 목표를 잘못 선택해 학습이 지연되거나 실패하는 경우가 많았다. 반면 제안된 방법은 더 적은 단계의 의사결정으로 목표에 도달했으며, 복잡한 과제에서도 높은 성공률과 빠른 학습 속도를 보였다.

특히 여러 중간 목표를 순서대로 해결해야 하는 과제에서도, 별도의 순서 정보를 주지

않아도 필요한 단계를 스스로 찾아내며 안정적으로 목표를 달성하는 모습을 보였다.

3. 기대효과

이 기술은 복잡한 작업을 여러 단계로 나눠 해결해야 하는 상황에서, 불필요한 시행착오를 줄이고 더 효율적인 의사결정을 가능하게 한다는 점에서 의미가 있다.

로봇이 여러 동작을 순서대로 수행해야 하는 작업이나, 자율주행처럼 경로 선택이 중요한 문제에서 더 적은 판단으로 안정적으로 목표에 도달할 수 있는 기반 기술로 활용될 수 있다.

연구 결과 개요 (S2Q)

논문명	RETAINING SUBOPTIMAL ACTIONS TO FOLLOW SHIFTING OPTIMA IN MULTI-AGENT REINFORCEMENT LEARNING
저자정보	한승열(교신저자), 조용현 (제1저자), 이선우

1. 연구배경

여러 개의 인공지능이 함께 협력해 문제를 해결하는 '멀티에이전트 강화학습'은 자율주행, 로봇 제어 등 다양한 분야에 쓰인다. 이때 각 인공지능의 행동 가치를 하나로 묶어 학습하는 방식이 널리 활용된다.

하지만 기존 방법은 매 순간 가장 좋다고 판단된 '하나의 행동'에 집중해 학습하는 구조다. 문제는 학습이 진행되면서 어떤 행동이 더 좋은지에 대한 판단이 바뀔 수 있다는 점이다. 처음에는 최선이라고 여겼던 선택보다, 이후 더 나은 행동이 드러나는 경우가 생긴다.

이때 기존 방식은 이미 선택했던 행동에 계속 의존하면서 다른 가능성을 충분히 남겨 두지 못한다. 결국 더 좋은 선택이 나타나도 이를 다시 찾아내지 못하고, 덜 좋은 전략에 머무르는 한계가 있었다.

2. 연구내용

연구팀은 이런 문제를 해결하기 위해, 가장 좋은 행동 하나만 고르는 대신 '차선의 선택지'까지 함께 기억하는 학습 방법을 제안했다.

이 방법은 먼저 가장 좋은 행동을 찾고, 그 다음에는 이미 선택된 행동을 제외한 상태에서 두 번째로 좋은 행동을 찾는다. 같은 방식으로 세 번째, 네 번째 후보까지 순서대로 확보한다. 말하자면 '1등을 뽑고 나면 2등, 3등도 따로 챙겨두는 방식'이다.

또 실제 행동을 선택할 때도 하나만 고집하지 않는다. 여러 후보 중에서 상황에 따라 확률적으로 선택하도록 해, 특정 행동에만 매달리지 않고 다양한 선택지를 계속 탐색하도록 했다.

실험에서는 스타크래프트 기반 협력 과제와 축구 시뮬레이션 환경에서 성능을 비교했다. 그 결과 기존 방법은 학습 중 최적 행동이 바뀌는 상황에서 적응이 늦거나 성능이 떨어진 반면, 제안된 방법은 미리 확보해둔 '차선의 행동'을 활용해 빠르게 새로운 최적 선택으로 전환했다.

특히 여러 인공지능이 동시에 협력해야 하는 복잡한 환경에서도, 다양한 후보를 유지하는 구조 덕분에 더 높은 성공률과 빠른 학습 속도를 보였다.

3. 기대효과

이 기술은 하나의 정답만 고집하는 대신, 여러 유망한 선택지를 함께 유지하면서 상황 변화에 유연하게 대응할 수 있다는 점에서 의미가 있다.

환경이 계속 바뀌는 자율주행, 로봇 협업, 군집 시스템 같은 분야에서, 더 빠르고 안정적으로 의사결정을 내릴 수 있는 인공지능 기술로 이어질 것으로 기대된다.

한승열 [교신저자] 교수 이력사항

1. 인적사항

- 소 속 : 울산과학기술원(UNIST)
- 전 화 : 052-217-3455
- 이메일 : syhan@unist.ac.kr



2. 학력

- 2013년: 한국과학기술원 전기및전자공학부 학사
- 2016년: 한국과학기술원 전기및전자공학부 석사
- 2021년: 한국과학기술원 전기및전자공학부 박사

3. 경력사항

- 2021.09 - 2021.10 : 한국과학기술원 Postdoctoral researcher
- 2021.11 - 현재 : 울산과학기술원 조교수

4. 전문 분야 정보

- Practical Reinforcement Learning for Real-world Applications
- Entropy-based Frameworks for Sample Efficient Exploration in RL
- Reinforcement Learning for High-dimensional Action Spaces

5. 연구지원정보

6. 기타