# 모방을 통한 자율 사물의 행동지능 구현

한국전자통신연구원 ▌최진철·배희철·박찬원

## 1. 서 론

공자는 지혜를 얻는 세 가지 방법으로 사색과 모방, 경험을 꼽았다. 이들 중 모방은 가장 쉬운 방법이고, 경험을 통해 지혜를 얻는 것은 가장 어려운 방법이라고 했다. 이러한 경험과 모방은 강화학습(Reinforcement Learning)과 모방학습(Imitation Learning) 같은 기계학습(Machine Learning) 기술의 모티브가 되었으며, 최근에는 로보틱스의 최적 제어나 자율주행 자동차의 행동 지능 도출을 위한 방법론으로 진화하고 있다[1].

강화학습은 특정 환경에서 학습을 수행하는 에이전트(Agent)가 반복적인 시행착오를 경험하면서 최대의 보상(Reward)을 얻을 수 있는 행동이나 의사결정이 무엇인지를 스스로 깨우치는 기계학습의 한 부류이다[2]. 지도 및 비지도 학습과는 달리 강화학습은 정적인 데이터 세트에 의존하지 않고 에이전트가 실제로 동작하며 얻은 경험으로부터 학습한다. 2족 보행, 점프, 달리기, 공중제비 등을 보여준 보스턴 다이내믹스(Boston Dynamics)의 휴머노이드 로봇 아틀라스(Atlas)[3], 딥마인드의 바둑 AI 알파고(AlphaGo)[4], 핵융합로 내부 초고온의 불안정한 플라즈마를 제어하는 알고리즘[5] 개발 등은 강화학습이 큰 축을 담당하고 있다.

모방학습은 지도학습의 한 형태로써 에이전트가 전문가의 행동과 의사결정을 모방하여 작업을 수행하는 정책(Policy)을 도출하는 기계학습 기법의 한 부류이다. 모방학습의 목적은 숙련된 전문가가 작업을 성공적으로 처리하기 위해 행동하는 정책 또는 전문가의 상태에 따른 행동 시퀀스 맵(Behavioral Sequence Map)을 도출하는 것이다. 즉, 복잡한 보상설계(Reward Shaping)나 정교한 작업계획(Task Planning)을 수립하는 대신 작업을 수행하는 시연을 학습하게 되는 것이다. 이러한 특성은 다른 유형의 기계학습에 비해 다양한 장점을 가진다[6]. 첫째로 특정 작업에 대한 정책을 도출하는 데 필요한 시간과 노력을 줄일 수 있다. 둘째로 보상만으로는 포착할 수 없는 요소를 전문가의 행동을 통해 배울 수 있다. 셋째, 인간 시연자의 전문 지식을 활용함으로써 에이전트가 작업을 보다 효율적이고 효과적으로 수행할 수 있다.

모방학습은 작업을 어떻게 수행되어야 하는지 잘 아는 숙련된 전문가가 있는 환경에서 유용하게 활용될 수 있다. 예를 들어 자율주행 분야에서 전문 드라이버가 먼저 특정 상황에서 자동차를 어떻게 조작해야 하는지 시연을 보이고, 이후 에이전트가 시시각각 변화하는 환경에 대처하는 드라이버의 행동을 모방하는 정책을 학습하는 것이다. 실제 자율주행 기술을 선도하고 있는 구글 웨이모(Waymo)와 영국 웨이브(Wayve) 등이 모방학습을 적극적으로 활용하고 있다[7,8].

또한 모방학습은 작업의 복잡한 특징들을 모두 담아낼 수 있는 보상 함수(Reward Function)를 명시하기 어려운 환경에도 유용하다. 예를 들어 휴머노이드 에이전트가 옆돌기(Cartwheel)와 공중제비(Backflip)를 하도록 보상함수를 설계하는 것은 매우 어려운 일이다. 이러한 문제에는 모방학습을 사용하여 보상함수를 명시적으로 지정하지 않고도 해당 작업을 잘 수행할 수 있는 행동 정책을 학습할 수 있다[3].

최근 모방학습 기술은 다른 기계학습 분야와 같이 새로운 접근방식과 컴퓨팅 능력의 향상으로 인해 빠르게 발전하고 있다. 그러나 고품질의 시연 데이터 취득, 다양한 감각 데이터 처리, 다중 에이전트 모방, 벤치마크 가상환경 구축, 일반화 문제 등 해결해야 다수의 도전과제에도 직면하고 있다. 이에 본 논문에서는 모방학습 기술의 구체적인 동작구조, 기술발전 트렌드, 당면과제, 성능분석을 위한 벤치마크 현황 등을 살펴보고 향후 발전 방향에 대해 고찰하고자 한다.

* 본 연구 논문은 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음[23ZR1100, 자율적으로 연결·제어·진화하는 초연결 지능화 기술 연구].

## 2. 모방학습의 동작 구조

그림 1은 모방학습을 수행하는 데 필요한 일련의 작업 흐름을 보여준다[9,10]. 모방학습 프로세스는 전문가가 작업을 완수하기 위한 과정을 시연하고, 이때 발생하는 상태 변화와 행동 궤적을 기록하고 저장하는 것에서 시작된다. 작업 시연은 전문가 또는 전문가가 조작하는 로봇/사물에 부착된 센서/카메라나 외부 환경에 설치된 장치 등을 통해 포착된다. 노이즈가 많거나 신뢰성이 낮은 시연 데이터는 학습 성능을 떨어뜨릴 수 있으므로 정교하게 수집하는 것이 중요하다.

다음으로 수집 데이터로부터 시연자의 상태와 환경 변화 등을 설명해 줄 수 있는 특징(Feature)을 추출하는 과정이 뒤따른다. 특징은 시연으로부터 모방해야 하는 정책을 학습하기 위해 활용되는 중요한 정보이다. 특징 추출은 입력 데이터의 특성과 분석 목적에 따라 임의적으로 선택되거나 미가공 상태로 바로 학습에 활용될 수 있다. 최근에는 신경망을 이용하여 원시 데이터를 학습하며 스스로 특징을 추출하는 표현학습(Representation Learning)이 많이 활용되고 있다[11].

어떠한 특징을 학습할 것인지가 정해지면 시연 데이터를 학습하여 정책을 도출하는 과정이 뒤따르게 된다. 시연 데이터를 모방하는 가장 직접적인 방법은 각 상태 인스턴스를 입력 벡터로, 전문가의 행동을 레이블(Label)로 규정하고 지도학습(Supervised Learning)을 하는 것이다. 다양한 데이터 분포로부터 이러한 경향성을 통계적으로 설명할 수 있는 모델을 도출하는 회귀분석(Regression)이나 주어진 입력 벡터가 어떤 종류의 값인지 표식하는 분류(Classification) 기법이 활용되고 있다. 대표적인 모방학습 알고리즘 중 하나인 행동복제(Behavioral Cloning)[1]도 회귀분석 기반 알고리즘으로 구현할 수 있다. 이외에도 시연자와 교류하며 학습 에이전트의 행동 궤적에 피드백을 제공하여 행동 정책을 개선시키는 대화형 모방학습[12]이나 전문가의 시연 데이터로부터 보상함수를 역으로 학습하고, 추정된 보상을 이용해 최적의 정책을 찾는 역강화학습(Inverse Reinforcement Learning)기반 방법론[13]들이 활용될 수 있다.

마지막은 에이전트가 이전 단계에서 학습된 정책을 기반으로 작업을 수행하면서 성능이 개선되도록 정책을 보완하거나 대체하는 단계이다. 모방학습 수행 과정에는 시연 오류나 데이터 수집 방식의 불완전성으로 인한 부정확한 시연 측정, 잘못된 일반화(Generalization), 시연자와 학습자 간 물리구조 및 환경 차이, 시뮬레이션에서 학습한 모델을 현실에 적용할 때의 이질성(Sim2Real Gap) 문제 등이 발생할 수 있어 강건한 행동을 재생산하는 정책 도출이 어려운 경우가 많다. 초기에는 학습 데이터의 다변화와 학습 알고리즘의 성능 개선을 통해 어느 정도 해결하였지만, 이 방식만으로는 한계가 있어 최근에는 실제 현장 데이터를 기반으로 기존 학습모델을 개선하기 위한 용도로 강화학습, 능동학습(Active Learning), 전이학습(Transfer Learning), 최적화, 미세조정(Fine Tuning) 기술, 적대

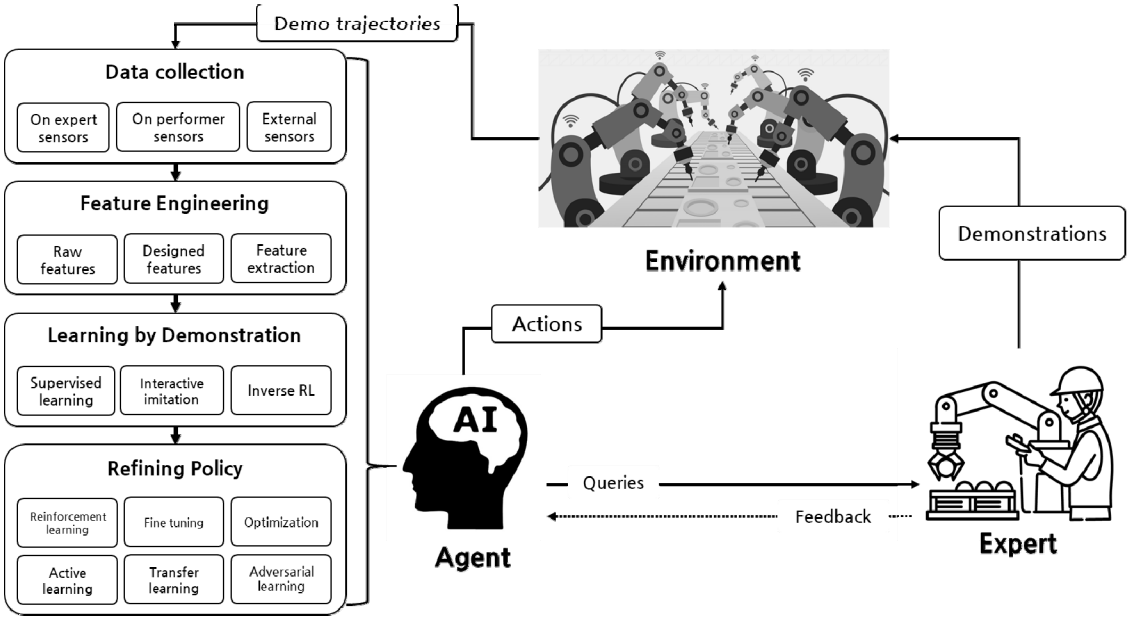



그림 1 모방학습 프로세스의 일반적인 작업흐름

적 학습(Adversarial learning) 등이 활발하게 응용되고 있다.

## 3. 모방학습 기술의 발전 트렌드

현재 모방학습은 다양한 분야에서 광범위하게 활용되고 있다. 본 절에서는 모방학습의 발전을 주도하고 있는 기술 트렌드에 대해 논의한다.

1) **감각정보 활용 다양화:** 전통적인 모방학습은 특정 기술에 대한 전문가 작업 시연을 학습하는 데 중점을 두었다. 이때 시연 데이터는 일반적으로 운동감각 티칭(Kinesthetic Teaching) 형태, 즉 특정 상태나 상황에서 전문가가 어떻게 행동하였는지에 대한 운동감각 데이터로 제공되었으며, 고품질의 시연 데이터 확보를 위해서 충분한 시간과 비용이 필요했다. 최근에는 이렇게 정형화된 시연 데이터를 활용하는 방식 대신 시각이나 촉각 정보처럼 레이블이 없거나(Unlabeled) 비정형(Unstructured) 정보를 활용하는 모방학습 연구가 활발하다. VIL(Visual Imitation Learning)[14]과 VINN (Visual Imitation through Nearest Neighbors)[15]은 로봇 매니퓰레이터(Manipulator)에 부착된 카메라를 통해 촬영한 영상을 학습하여, 물체 밀기나 파지 등의 행동 정책을 모방하는 정책을 도출한다. 최근에는 사람의 일상활동이나 요리하는 과정이 담긴 동영상을 모방학습하여 로봇 매니퓰레이터 두 대가 서랍이나 오븐을 열고, 채소나 칼, 냄비 등을 집는 등 10여 종의 업무를 처리할 수 있는 VRB(Vision-Robotics Bridge)[16]이 유명 컴퓨터 비전 학회인 CVPR 2023을 통해 공개되었다. 이외에도 Seq2Seq IL[17]과 BMP[18]는 기존 모방학습에서 해결하기 어려웠던 접촉이 많은(Contact-rich) 조작 작업 문제를 해결하기 위해 촉각 피드백 정보를 학습에 활용했다. 이와 같이 다양한 감각정보를 활용하는 모방학습은 명시적인 레이블이 없이 감각 데이터에서 관심 객체나 정보를 인지·추적하는 기술, 비정형 데이터를 학습하며 스스로 학습 데이터를 분류하는 자기 지도학습(Self-Supervised Learning), 감각 정보를 행동 정보로 재구성하는 모션 리타겟팅(Motion Retargetting) 기술 등이 함께 활용되는 경우가 많다.
2) **하이브리드 접근방식:** 모방학습의 주목할 만한 경향 중 하나는 여러 유형의 학습을 결합하여 성능을 향상시키는 하이브리드 접근방식이다. 예를 들어, 강화학습과 모방학습을 결합하면 에이전트가 새로운 더 나은 해결책을 찾기 위해 환경을 탐색하면서 인간의 시연을 통해 학습할 수 있다. 대표적인 사례로 딥마인드의 DQfD(Deep Q-Learning from Demonstrations)[19]를 꼽을 수 있다. DQfD은 먼저 시간 차이 손실(Temporal Difference Loss)과 지도 손실(Supervised loss)을 조합한 손실함수를 최적화하면서 소량의 시연 데이터를 사전학습(Pretrain)한다. 여기서 지도 손실은 모방하는 방법을 학습할 수 있게 해주며, 시간 차이 손실은 강화학습을 가능하게 해주는 요소로 동작한다. 에이전트는 사전학습 정책을 이용하여 환경과 상호작용하게 하는데 이때 수집되는 데이터와 시연 데이터를 최적화된 비율로 혼합하여 학습을 반복적으로 수행하며 정책을 업데이트한다. 이와 같은 하이브리드 접근방식은 다양한 학습방식의 장점을 활용하여 개별적인 한계를 극복하고 전반적인 성능을 향상하는 데 도움이 될 수 있다.
3) **적대적 모방학습(Adversarial Imitation Learning):** 적대적 모방학습은 에이전트가 실수를 하도록 속이려는 적대적 상대로부터 학습하는 기술이다. 이 기술은 학습된 정책의 적응력(Adaptability)과 회복력(Resilience)을 증가시킬 수 있어 비디오 게임이나 로보틱스, 컴퓨터 비전 등 다양한 분야에 적용되고 있다. 적대적 모방학습의 대표적인 알고리즘은 GAIL(Generative Adversarial Imitation Learning)[20]이다. GAIL은 모방학습을 분포 매칭(Distribution Matching) 문제로 규정하고, GAN 기법[21]을 활용하여 전문가 정책과 학습 정책에 의해 유도된 분포 간 제슨-섀넌 발산(Jensen-Shannon Divergence) 최소화를 목표로 한다. GAIL은 모방학습을 개선하기 위해 강화학습이나 다양한 발산(Diver- gence)을 도입하는 다양한 연구[22,23]에 영감을 주었다.
4) **다중 에이전트 모방학습(Multi-Agent Imitation Learning):** 다중 에이전트 모방학습은 인간의 시연이 아닌 다수 에이전트들의 행동으로부터 학습하는 기법이다. 해당 기법은 다양하고 넓은 범위의 행동을 행동할 수 있는 잠재력을 가지고 있어, 물류, 농업, 군사 등 다양한 군집 로봇 응용 분야에 관심을 받고 있다. 이러한 다중 에이전트 모방학습에서 각 에이전트는 서로의 행동에서 학습하거나, 유용한 정보를 사용하여 자신

의 성능을 향상시킬 수 있다. GAIL 알고리즘을 다중 에이전트 환경으로 확장한 MAGAIL(Multi-Agent GAIL)[24]과 에이전트의 개별 정책과 잠재적 조직화 모델(Latent Coordination Model)을 동시에 모방학습하는 CMAIL(Coordinated Multi-Agent Imitation Learning)[25]이 대표적인 다중 에이전트 모방학습 연구 사례이다.

## 4. 행동 정책 학습 및 평가를 위한 가상환경 기반 벤치마크 도구

객체 인지 및 분류, 객체 추적, 자연어 처리 등과 같은 인공지능 분야는 알고리즘 학습과 추론 과정에서의 성능을 정량화하기 위해 표준화된 데이터 세트를 벤치마크로 사용하고 있다. 반면에 환경과 상호작용해야 하는 특징과 충분한 양의 학습 데이터가 필요한 기계학습의 속성으로 인해 로봇의 행동모델 평가를 위한 벤치마크는 대부분 가상환경 기반의 시뮬레이터 형태로 공개되고 있다. 신뢰성 높고 특정 요소에 편향되지 않는 벤치마크를 개발하는 것은 매우 어려운 과제이다. 특히 로봇을 조작하는 행동모델에 대한 평가는 환경과 상호작용을 통해 얻게 되는 동적 데이터에 기반하여 이루어져야 하기에 훨씬 더 도전적이다[26]. 본 절에서는 모방학습이나 강화학습을 통해 도출되는 정책을 학습하고 평가하기 위한 가상환경 기반의 벤치마크 도구에대해 소개한다.

오픈AI는 간단한 2D 에이전트 제어환경과 Mujoco 물리엔진 기반의 가상환경, 아타리(Atari)게임 환경 등을 연계해주는 Gym툴킷[27]을 공개했다. 이외에도 3D 엔진 개발 플랫폼인 유니티(Unity)의 머신러닝 에이전트(ML-Agents)[28], 오픈소스 로봇 시뮬레이터 가제보(Gazebo)의 Gym-Gazebo2[29] 등이 가상환경 속 에이전트의 행동모델 학습과 평가에 활용되고 있다.

강화학습에 딥러닝을 결합한 심층 강화학습 기술 연구 초기에는 Gym툴킷에 포함된 전통 제어(Classic control) 환경이나 Box2D, Mujoco 물리엔진 기반의 2D/3D 로봇 제어 환경 등이 많은 논문에서 벤치마크로 활용되었다. 그러나 이러한 기존의 벤치마크는 가상환경 내 객체의 위상(Topology)과 기하학적 변형 어려움, 현실의 동적 환경 모델 미적용, 여러 유형의 조작 작업(Task) 생성을 위한 지원 부족 등으로 이유로 관심이 점점 떨어지고 있다. 이에 최근에는 실세계의 다양한 로봇 에이전트와 환경 입출력 파라미터 및 객체 속성 커스터마이즈 지원, 다양한 객체 모델 지원, 보다 복잡하고 정교한 행동조작을 요구하는 작업 에피소드들을 제공하거나 저작 기능을 지원하는 가상환경 벤치마크들이 새롭게 공개되고 있다.

표 1에 최근 공개된 로봇 매니퓰레이터의 행동모델 평가를 위한 9종의 가상환경 벤치마크를 비교하였다[30].

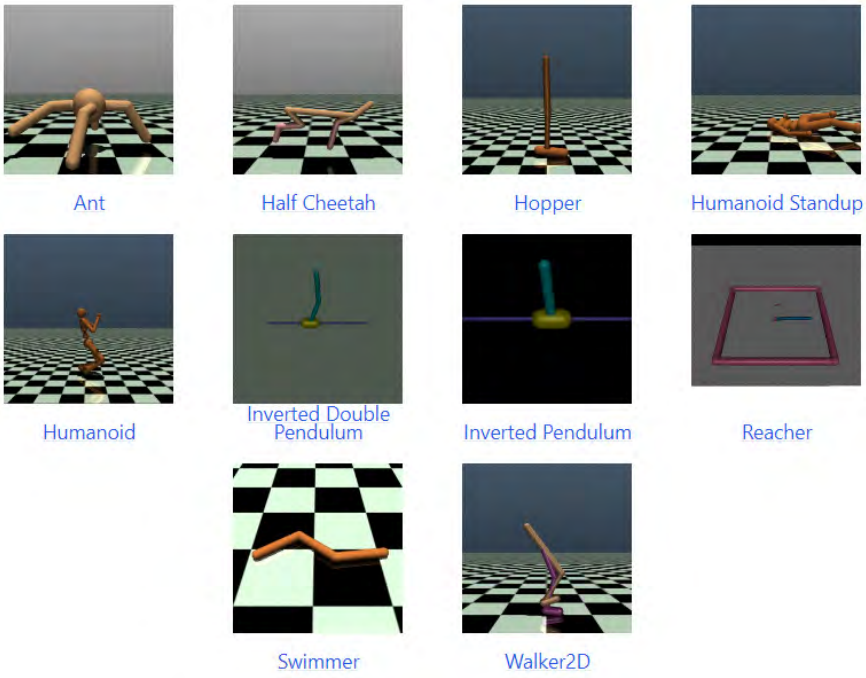
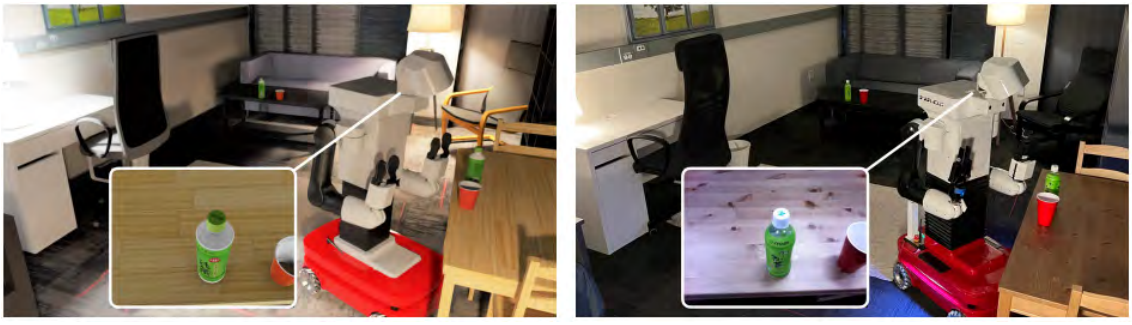



그림 3 행동모델 벤치마크를 위한 Mujoco 기반 가상환경 (출처: www.Gymlibrary.dev)

표 1 로봇 행동모델 학습 및 평가를 위한 차세대 가상환경 벤치마크 비교

| 벤치마크 | BEHAVIOR-1K [31] | Habitat 2.0 [32] | IsaacGym [33] | MetaWorld [34] | RLBench [35] | ManipulaTho r[36] | Robosuite [37] | TDW [38] | ManiSkill2 [30] |
|---|---|---|---|---|---|---|---|---|---|
| 개발주도 기관 | Stanford Univ. | Meta AI | NVIDIA | UCB/ Stanford Univ. | Imperial College London | Allen Institute for AI | Texas Univ./ Stanford Univ. | MIT/IBM | UCSD/ 칭화대 |
| Grasp구현 | Abstract | Abstract | Physical | Physical | Abstract | Abstract | Physical | Abstract | Physical |
| 데모 궤적 개수 | – | – | – | Procedural | Procedural | – | ~2000 | – | 〉30k |
| 다중 컨트롤러 지원 | – | Yes | No | No | Yes | No | Yes | No | Yes |
| Visual RL/ IL baselines | Limited | Full | No | No | No | Full | No | No | Full |
| 객체 모델 개수 | 5215 | YCB | – | 80 | 28 | 150 | 10 | 112 | 2144 |
| Scene 개수 | 50 | 105 | – | – | – | 30 | – | 105 | – |
| Ray tracing 지원 | OmniGibson | – | – | – | – | – | NVISII | Unity | Kuafu |
| 도메인 랜덤화 | – | No | Yes | No | Yes | No | Yes | No | Yes |
| Rigid body 시뮬레이션 | OmniGibson | Bullet | PhysX 5 | Mujoco | CoppeliaSim | Unity | Mujoco | Unity | SAPIEN |
| Soft body 시뮬레이션 | OmniGibson | – | – | – | – | – | – | – | Warp-MPM |



(a) BEHAVIOR-1K의 가상환경 (b) 실세계 로봇 동작 이미지

그림 4 실사 수준의 높은 그래픽 품질을 보여주는 BEHAVIOR-1K 가상환경 벤치마크 (출처: 참고문헌 [31])

표 1에 나타낸 벤치마크들은 벤치마크 관심 요소에 따라 기능이나 구현방식에 차별성이 존재한다. 예를 들어 각 벤치마크의 모든 시뮬레이션 백 엔드(Back-end)는 물리적으로 로봇 매니퓰레이터 그리퍼(Gripper)의 파지(Grasp)와 관련된 정교한 데이터 수집과 제어 기능을 지원할 수 있다. 그러나 벤치마크는 고수준의 작업처리 현황에만 관심을 두고 그리퍼의 상태를 개/폐(Open/Close)나 0과 1 사이의 실수 등으로 추상화된 정보로만 제공하기도 한다. 표 1에서 다중 컨트롤러 지원은 벤치마크가 여러 컨트롤러 구현을 지원하고, 각 작업에 대해 특정 인터페이스를 선택할 수 있는 기능 제공 여부를 의미하며, 도메인 무작위화(Domain Randomization)는 강건한(Robust) 모델 학습을 위해 시뮬레이션 속 객체와 환경 속성을 무작위로 변형할 수 있는 기능의 지원 여부를 의미한다. BEHAVIOR-1K 벤치마크의 경우 엔비디아(Nvidia)의 옴니버스(Omniverse) 엔진과 PhysX5[39]기반의 옴니깁슨(OmniGibson)을 3D 가시화 및 물리엔진으로 사용하여 실사 수준의 높은 그래픽 품질을 보여주었다(그림 4 참조).

## 5. 모방학습 기술의 과제

모방학습은 로보틱스에게 인간의 행동양식을 가르칠 수 있는 큰 잠재력을 가지고 있다. 이 방식은 시행착오를 거듭하며 최적화하는 여타 기술과 비교하여 시간과 비용, 컴퓨팅 자원을 절약할 수 있으며, 정교한 조작 능력의 습득까지 가능하게 줄 실마리를 제공하고 있다. 그럼에도 불구하고 모방학습은 아직 해결해야 할 과제들이 많다[40].

모방학습의 첫 번째 과제는 전문가 시연 품질 문제이다. 시연의 품질이 떨어지거나 작업처리의 중요한 요소를 포착하지 못한다면 학습 알고리즘이 유용한 정보를 추출하지 못할 수 있다. 이는 학습된 정책의 성능 저하로 이어지게 된다. 이러한 문제를 극복하기 위해서는 전문가 시연을 정교하게 포착하기 위한 장비와 소프트웨어를 도입하거나 다양한 환경과 광범위한 시나리오 상에서 수행된 시연 데이터로 확장시키는 방법이 활용될 수 있다. 또한, 능동학습(Active Learning)을 도입하여 추가 데이터가 필요한 영역에서 전문가에게 부족한 시연을 요청하여 학습에 반영할 수도 있어야 할 것이다.

또 다른 과제는 분포이동(Distribution Shift) 문제이다. 이 문제는 전문가 시연이 에이전트와 다른 환경이나 조건에서 수집되는 경우 발생할 수 있다. 학습 알고리즘이 새로운 데이터 분포에 적응할 수 없으면, 작업 시나리오에 일반화되기 어려워 성능이 저하될 수밖에 없다. 이러한 문제에는 전문가 시연 분포와 학습 데이터 분포 간의 차이를 명시적으로 표현하여 데이터 분포에 강건한 정책을 학습하는 도메인 적용(Domain Adaptation) 기술이 활용될 수 있다.

세 번째로는 차원의 저주(Curse of Dimensionality) 문제이다. 이는 변수의 수가 매우 많은 고차원(High Dimension) 데이터를 다룰 때 겪을 수 있다. 예를 들어 다수의 이미지나 영상을 입력받아 다관절 로봇 매니퓰레이터의 출력 행동을 매핑시키는 학습 알고리즘의 경우 고려해야 하는 경우의 수가 많아 실행이 매우 까다로워진다. 그래서 최근에는 고차원 데이터 처리에 효과적인 심층 신경망을 학습 모델로 주로 활용하고 있으며, 일부는 상태 공간의 차원을 줄이기 위해 차원 축소 기법을 적용하기도 한다.

네 번째로는 경험하지 않은 상태나 학습하지 않은 데이터에도 충분한 성능을 보여줄 수 있는 일반화(Generalization)의 문제이다. 전문가 시연의 품질에 영향받는 모방학습의 특성상 시연이 불완전하거나 다양한 상황을 포함하지 못하면 새로운 시나리오로 일반화하는데 어려움을 겪을 수밖에 없다. 모방학습의 일반화를 위해 도메인 지식을 융합하거나 미세조정 기술을 적용하고 있으나 아직 개선의 여지가 많다.

## 6. 결 론

모방학습 기술은 인간의 전문 지식을 통해 작업을 처리하는 행동 정책을 배울 수 있는 훌륭한 접근방식으로써 로보틱스와 자율주행, 제조 및 의료, 금융 등 다양한 응용 분야에서 상당한 가능성을 보여주고 있다. 반면에 전문가 시연 품질 문제, 분포이동 문제, 차원의 저주 문제, 일반화 문제, 확장성(Scalability) 등 아직 극복해야 할 몇 가지 난관이 존재하고 있다.

현재 많은 연구자들이 모방학습 기술의 여러 난제를 극복하고자 노력하고 있으며, 머지않아 에이전트가 인간의 지식으로부터 자신의 지식을 획득할 수 있는 경로를 스스로 구축함으로써 복잡하고 정교한 작업을 독립적으로 실행할 수 있는 지능형 시스템의 출현이 예상된다. 이처럼 지속적으로 발전하는 모방학습 기술은 향후 인공지능 분야의 중요한 연구 테마로 자리 잡을 것으로 기대된다.

## 참고문헌


[1] T. Osa, et al., “An Algorithmic Perspective on Imitation Learning,” arXiv preprint, abs/1811.06711, 2018.

[2] R. Sutton and A. Barto, “Reinforcement Learning: An Introduction,” MIT Press Cambridge, Second Ed., 2018.

[3] X. Peng et al., “DeepMimic: Example-Guided Deep Reinforcement Learning of Physics-Based Character Skills,” arXiv preprint, arXiv:1804.02717, 2018.

[4] D. Silver et al., “Mastering the game of Go with Deep Neural Networks and Tree Search,” Nature, vol. 529, no. 7587, pp. 484-489, 2016.

[5] J. Degrave et al., “Magnetic Control of Tokamak Plasmas Through Deep Reinforcement Learning,” Nature, vol. 602, pp. 414-419, 2022.

[6] P. Abbeel and A. Y. Ng, “Apprenticeship learning via inverse reinforcement learning,” in Proc of ICML, 2004.

[7] E. Bronstein et al., “Hierarchical Model-Based Imitation Learning for Planning in Autonomous Driving,” arXiv preprint, arXiv:2210.09539, 2022.

[8] A. Hu et al., “Model-Based Imitation Learning for Urban Driving,” arXiv preprint, arXiv:2210.07729,

2022.
[ 9 ] A. Hussein et al., “Imitation Learning: A Survey of Learning Methods,” ACM Comput. Surv. (CSUR), vol. 50, no. 2, pp. 1-35, 2017.
[10] J. Delgado et al., “Robotics in construction: A critical review of the reinforcement learning and imitation learning paradigms,” Adv. Eng. Inf., vol. 54 pp. 101787, 2022.
[11] X. Chen et al., “An Empirical Investigation of Representation Learning for Imitation,” arXiv preprint, arXiv:2205.07886, 2022.
[12] X. Chen et al., “Interactive Imitation Learning in Robotics: A Survey,” arXiv:2211.00600, 2022.
[13] 이상광 외 “역강화학습 기술 동향,” 전자통신동향분석, vol. 34, no.6, 2019.
[14] S. Young et al., “Visual Imitation Made Easy,” arXiv preprint, arXiv:2008.04899, 2018.
[15] J. Pari et al., “The Surprising Effectiveness of Representation Learning for Visual Imitation,” arXiv preprint, arXiv:2112.01511, 2021.
[16] S. Bahl et al., “Affordances from Human Videos as a Versatile Representation for Robotics,” In Proc of CVPR, pp.13778-13790, 2023.
[17] W. Yang et al., “Seq2Seq Imitation Learning for Tactile Feedback-based Manipulation,” arXiv preprint, arXiv: 2303.02646, 2023.
[18] S. Stepputtis et al., “A System for Imitation Learning of Contact-Rich Bimanual Manipulation Policies,” arXiv preprint, arXiv:2208.00596, 2022.
[19] T. Hester et al., “Deep Q-learning from Demonstrations,” in Proc of AAAI, no. 394, pp. 3223-3230, 2018.
[20] J. Ho et al., “Generative Adversarial Imitation Learning,” arXiv preprint, arXiv:1606.03476, 2016.
[21] I. Goodfellow et al., “Generative adversarial nets.,” In Proc of NeurIPS, pp. 2672-2680, 2014.
[22] I. Kostrikov et al., “Discriminator-Actor-Critic: Addressing Sample Inefficiency and Reward Bias in Adversarial Imitation Learning,” arXiv preprint, arXiv: 1809.02925, 2018.
[23] X. Peng et al., “Variational Discriminator Bottleneck: Improving Imitation Learning, Inverse RL, and GANs by Constraining Information Flow,” arXiv preprint, arXiv:1810.00821, 2018.
[24] J. Song et al., “Multi-Agent Generative Adversarial Imitation Learning,” arXiv preprint, arXiv:1807.09936, 2018.
[25] H. Le et al., “Coordinated Multi-Agent Imitation Learning,” arXiv preprint, arXiv:1703.03121, 2017.
[26] B. Yang et al., “Generative Adversarial Imitation Learning,” IEEE Robot. Autom. Lett., vol. 5, no. 2, pp. 2094-2099, Apr. 2020.
[27] G. Brockman et al., “OpenAI Gym,” arXiv preprint, arxiv:1 606.01540, 2016.
[28] J. Arthur et al., “Unity: A General Platform for Intelligent Agents,” arXiv preprint, arxiv:1809.02627, 2018.
[29] N. G. Lopez et al., “Gym-Gazebo2, a Toolkit for Reinforcement Learning Using ROS 2 and Gazebo,” arXiv preprint, arxiv:1903.06278, 2019.
[30] J. Gu et al., “ManiSkill2: A Unified Benchmark for Generalizable Manipulation Skills,” arXiv preprint, arXiv:2302.04659, 2023.
[31] C. Lee et al., “BEHAVIOR-1K: A Benchmark for Embodied AI with 1,000 Everyday Activities and Realistic Simulation,” in Proc. of CoRL 2022.
[32] A. Szot et al., “Habitat 2.0: Training home assistants to rearrange their habitat,” Advances in Neural Information Processing Systems, 34: 251-266, 2021.
[33] V. Makoviychuk et al., “Isaac gym: High performance gpu-based physics simulation for robot learning,” arXiv preprint, arXiv:2108.10470, 2021.
[34] T. Yu et al., “Meta-World: A Benchmark and Evaluation for Multi-Task and Meta Reinforcement Learning,“ arXiv preprint, arXiv:1910.10897, 2019.
[35] S. James et al., “RLBench: The Robot Learning Benchmark & Learning Environment,” arXiv preprint, arXiv:1909.12271, 2019.
[36] K. Ehsani et al., “Manipulathor: A framework for visual object manipulation,” in Proc. of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4495-4504, 2021.
[37] Y. Zhu, et al., “Robosuite: A Modular Simulation Framework and Benchmark for Robot Learning,” arXiv preprint, arXiv:2009.12293, 2020.
[38] Y. Zhu, et al., “Threedworld: A platform for interactive multi-modal physical simulation,” arXiv preprint, arXiv:2007.04954, 2020.
[39] Nvidia, Corp. Physx. https://developer.nvidia.com/physx-sdk, 2022. Accessed: 2022-06-10.
[40] C. Priya et al., “Open Source Developments Novel Perspectives in Imitation Learning: Trends, Challenges, Future Directions,” Journal of Open Source Developments, vol.10, No.1, pp.26-35. 2023.

## 약 력


### 최 진 철

2005 아주대학교 전자공학부 졸업(학사)
2007 아주대학교 전자공학과 졸업(석사)
2012 아주대학교 전자공학과 졸업(박사)
2008 동핀란드 대학교 방문연구원
2012~현재 한국전자통신연구원 선임연구원
관심분야 : 모방학습, 강화학습, 역강화학습, 디지털트윈
Email : spiders22v@etri.re.kr



### 배 희 철

2005 부산대학교 산업공학과 졸업(석사)
2005 싱가포르국립대학교 산업시스템공학과 연구원
2014 부산대학교 산업공학과 졸업(박사)
2009~현재 한국전자통신연구원 책임연구원
관심분야 : 인공지능, 최적화, 기계학습, 강화학습
Email : hessed@etri.re.kr



### 박 찬 원

1993 광운대학교 컴퓨터공학과 졸업(학사)
1996 광운대학교 전자계산기공학과 졸업(석사)
2016 충남대학교 전자공학과 졸업(박사)
1996~1999 KAIST IDEC (연구원)
1999~현재 한국전자통신연구원(실장)
관심분야 : 사물인터넷, 머신러닝, 디지털트윈
Email : cwp@etri.re.kr