

# 자율 이동 로봇 연구에서의 강화학습 및 대형 언어 모델의 적용 가능성 탐색

채송화<sup>1</sup>, 성다훈<sup>2</sup>, 임유진<sup>3</sup>  
<sup>1</sup>숙명여자대학교 IT 공학과 석사과정  
<sup>2</sup>숙명여자대학교 IT 공학과 박사과정  
<sup>3</sup>숙명여자대학교 인공지능공학부 교수

watermelon97@sookmyung.ac.kr, ekgns324@sookmyung.ac.kr, yujin91@sookmyung.ac.kr

## A Study on the Application of Reinforcement Learning and Large Language Models to Autonomous Mobile Robots

SongHwa Chae<sup>1</sup>, Da-Hun Seong<sup>2</sup>, Yujin Lim<sup>3</sup>

<sup>1,2</sup>Dept. of Information Technology Engineering, Sookmyung Women's University

<sup>3</sup>Div. of Artificial Intelligence Engineering, Sookmyung Women's University

### 요 약

사전 설정을 통한 고정된 경로에 의존하는 AGV(Automated Guided Vehicle) 시스템은 동적 환경에 취약하여 다양한 환경에 유연하게 대응할 수 있는 AMR(Autonomous Mobile Robot)이 대안으로 주목받고 있다. AMR의 자율적인 판단 능력을 높이기 위한 제어 방법으로 강화학습이 제시되고 있으나, 해석 가능성과 적용성의 제약 조건이 존재한다. 본 연구는 이러한 문제를 보완할 방안으로 강화학습과 대형 언어 모델의 연계를 제안하며, 기술 융합 가능성과 직면한 과제를 검토하고자 한다.

### 1. 서론

최근 많은 산업 분야에서 효율적인 물류 관리 및 작업 자동화의 필요성이 급증함에 따라 동적인 환경 속에서도 환경에 따라 자율적인 판단을 기반으로 이동 및 경로 선택이 가능한 자율 이동 로봇(Autonomous Mobile Robot, AMR)에 대한 연구가 활발히 이루어지고 있다. 특히 비정형적 실내 공간에서 사용되는 AMR의 수요는 일반적인 제조업뿐만 아니라, 병원, 사무 환경 등 다양한 환경에서 증가하고 있다[1].

기존 사용되던 자동 유도 차량(Automated Guided Vehicle, AGV)은 자석, 마그네틱테이프 등과 같은 외부 유도 방식을 따라 작동되고 있어, 환경 변화가 발생하면 경로 변경에 제한적인 성능을 보인다. 따라서 AGV의 구조는 사전에 설정된 경로를 기반으로 수행되는 간단한 반복 작업에는 효율적이나, 동적인 특성으로 인해 모든 경로를 미리 설정할 수 없는 비정형적인 환경에서는 효율성이 떨어진다. 이 때문에 변동성이 높은 환경에서도 인식한 환경 정보를 기반으로 자율적인 판단을 내리고 경로를 계획할 수 있는

AMR 기반 기술로의 전환이 이루어지고 있다[2].

그러나 AMR이 더욱 효과적으로 운영되기 위해서는 장애물 회피, 다중 목표 판단, 실시간 경로 재계획 등 복합적인 제어 기술이 요구된다. 그러나 전통적인 제어 기법인 PID(Proportional-Integral-Derivative control)나 MPC(Model Predictive Control)는 비정형적 환경에 효과적으로 대응하는 데 한계가 있다. 그 때문에 환경 모델이 없이도 반복 학습을 통해 최적의 정책을 스스로 학습하여 환경 정보가 부족하거나, 동적으로 변화하는 환경에서도 효율적으로 적응할 수 있는 강화학습이 제어 기법의 대안으로서 주목받고 있다.

본 연구에서는 AGV와 AMR의 기술적 차이를 바탕으로 AMR 기반 자율 이동체의 강화학습 적용 가능성을 조사한다. 이어 강화학습의 주요 한계를 식별하고, 대형 언어 모델(Large Language Model, LLM)을 통해 한계를 보완하고 성능 개선 가능성을 탐색하여 향후 연구 방향을 논의하고자 한다.

### 2. AGV 및 AMR의 기능적 특징 비교

AGV는 자석, 마그네틱 테이프와 같이 외부적인 유

도 방식, 즉, 사전 설정된 경로를 따라 이동하는 운반 자동화 시스템이다. AGV 시스템은 중앙 집중형 제어 구조로, 환경 변화 대응에 제한적이라는 특징을 가지고 있으며 특히 반복적이고 예측 가능한 단순한 작업이 아닌 동적이거나 복잡한 환경에서는 효율적인 기기 운용에 한계를 보인다[3].

이에 반해, AMR은 SLAM(Simultaneous Localization and Mapping) 기반의 자율 주행 기술을 바탕으로 운용되어 외부 인프라에 대한 정보가 없어도 주변을 인식하고 경로를 스스로 계획할 수 있다. LiDAR, IMU(Inertial Measurement Unit), 카메라 등 다양한 센서를 사용하여 실시간으로 환경 정보를 수집하고, 수집된 정보를 기반으로 장애물을 회피하며 목적지까지 자율적인 판단하에 이동할 수 있다. 이러한 기술적 특징으로 인해 복잡하고 예측 불가능한 환경에서도 안정적으로 운용될 수 있어 다양한 산업 분야 적용 가능성이 확대되고 있다[4].

AGV의 단순하고 안정적인 구조는 정적 환경 속 반복 작업 수행에는 탁월하나, 변화에 대한 적응 및 대응력이 약한데 반하여, AMR의 경우 높은 환경 인식을 통해 다양한 시나리오 속 유연성과 운영 효율성을 입증하고 있다[5]. 이러한 기술적 차이를 고려하였을 때, 다양한 실제 환경을 인식하고, 환경 변화에 따라 능동적으로 적응 및 대응할 수 있는 기술의 지속적인 연구의 필요성이 대두되고 있다. 이에 따라 본 논문에서는 AMR의 자율성을 향상시키기 위한 제어 기법으로 강화학습의 적용 가능성을 고찰하고, 더 나아가 강화학습의 한계를 보완할 수 있는 방안으로 LLM과의 융합 가능성에 대해 논의한다.

### 3. AMR 기반 강화학습 연구 및 한계

AMR 제어의 기존 기법으로는 PID 및 MPC와 같은 전통적 제어 기법이 사용됐다. 이 전통적 기법들은 환경 조건이 정적일 때 안정적 제어가 가능하나, AMR에 탐색되는 객체가 자주 이동하거나 이동 경로가 복잡한 환경 조건에서는 성능적 한계가 발생한다. 특히 환경 변화에 따라 실시간으로 제어 입력을 계산해야 할 경우, AMR의 운용 효율성은 급격하게 저하될 수 있다[6]. 즉, 계산량이 높지 않고 빠르게 판단할 수 있는 장점에 반해 복잡한 환경에서는 효율적으로 운행되기 어렵다는 것이 기존 기법의 한계임이 지속해서 논의됐다.

반면, 강화학습은 환경과의 상호작용을 통해 에이전트가 보상을 최대화하고, 정책을 학습하여 최적의 선택을 끌어내는 방식을 채택하고 있어, 사전 모델이나 명시적인 지도가 없이도 상황에 맞는 전략을 스스로 도출할 수 있다. AMR은 주행 경로가 고정되지 않

고 사람의 움직임뿐만 아니라 다양한 객체들의 변화를 적용하여 이동해야 하므로, 강화학습은 AMR의 제어 기법으로 주목받고 있다.

[7]은 시뮬레이션 기반의 강화학습을 활용하여 환경 정보 없이도 이동 경로를 학습할 수 있는 시스템을 제안하였다. [8] 또한 주어진 환경 정보 없이 시각 센서만으로 환경을 인식하고, 인식 결과에 따라 장애물을 피하며 자율적으로 목적지에 도달할 수 있는 정책 학습 구조를 제시했다.

최근의 연구들은 다양한 환경 속에서 강화학습 기반의 AMR 제어 기법 성능을 검증하고 있다. [9]는 혼잡한 환경에서 강화학습 기반 내비게이션 정책이 AMR 이동 경로를 안정적으로 학습하여 동적으로 이동하는 장애물을 회피함에 탁월함을 입증하였다. [10]은 강화학습의 보상 설계를 통해 다중 목표가 주어지는 시나리오에서도 효과적인 경로 최적화를 달성하였다. 또한 [11]은 전통적인 경로 계획 알고리즘과 강화학습 기반 알고리즘을 비교한 실험을 통해 환경 정보가 없는 동적 환경에서의 충돌 회피 성능이 우수함을 증명하였다.

강화학습은 다양한 시나리오에서 다양한 목표 지점을 탐색하거나 충돌 횟수를 감소시키는 등의 안전한 주행 성능을 입증하며 AMR 제어의 유망한 접근법으로 평가받고 있으나 강화학습에는 여전히 몇 가지 한계가 존재한다. 예를 들어, 시뮬레이션과 실제 환경 간의 간극이 존재하는 Sim-to-real Gap 문제는 지속해서 제기되고 있다. 뿐만 아니라 강화학습 정책은 일반적으로 해석이 어려운 블랙박스 모델로, 정책의 의도나 오류 발생 원인을 이해하고 조정하는 것이 어렵다는 단점이 명확히 존재한다. 이 외에도 전통 제어 기법보다 학습에 요구되는 시간 및 연산 자원 요구가 매우 높다.

이러한 강화학습의 한계를 보완하기 위해 최근, 자연어 기반으로 조건을 설정하고 정책을 해석하여 모델 해석 가능성을 높일 수 있는 방안으로 대형 언어 모델 LLM을 강화학습에 결합하려는 시도가 이루어지고 있다.

### 4. LLM을 통한 보완 가능성

강화학습의 제약을 극복하기 위한 새로운 접근으로 LLM의 활용 가능성이 주목받고 있다. LLM은 자연어를 기반으로 직관적인 명령어를 설정하고, 정책을 맥락적으로 해석할 수 있으며 다양한 센서 및 환경 정보 등의 서로 다른 데이터들을 함께 처리하는 멀티모달 통합을 가능하게 함으로써 강화학습의 한계로 여겨지는 해석 불투명성을 완화할 수 있는 보완적 수단으로 부상하고 있다. 이러한 기술은 자율 이동체를

포함한 다양한 지능형 시스템의 제어 성능과 설명 가능성 향상에 이바지할 수 있어 강화학습과 LLM 결합은 꾸준한 논의가 이루어지고 있다.

예를 들어, [12]는 시뮬레이션과 현실 간의 교통 동역학 차이를 바로잡아 강화학습의 현실 적용성을 높였다. LLM을 통해 도로 상황 및 날씨 등의 맥락 정보를 바탕으로 동역학 변화를 추론하여 강화학습 단독 사용보다 시뮬레이터와 현실 사이의 괴리를 개선할 수 있음을 시사한다.

[13]은 LLM이 보상 설계와 도메인 랜덤 파라미터를 생성하게 하여 Sim-to-real 과정에서의 인간 개입을 최소화하면서도 동시에 강화학습의 성능 향상을 달성했다.

[14]는 강화학습 정책의 의사결정을 사람에게 설명하기 위해 LLM을 사용하였다. 행동 시퀀스를 요약한 후, LLM에 프롬프트로 제공하여 자연어 해석을 생성함으로써 강화학습의 해석 불가능성을 LLM으로 보완하였다. 이는 LLM을 통해 인간이 이해할 수 없었던 강화학습의 블랙박스 영역을 인간이 해석하고 이해할 수 있는 설명 가능성을 입증한다.

[15]에서는 강화학습의 한계를 극복하는 것에 그치지 않고, 강화학습과 LLM을 활용하여 로봇의 상황 대처 유연성을 극대화하는 방향성을 제시하기도 한다. 로봇이 작업을 수행하는 과정에서 얻는 다양한 피드백을 LLM이 자연어로 해석하고, 이를 바탕으로 계획을 조정하거나 행동을 수정하여 변화하는 상황에 유연하게 대처하도록 하였다. 이 연구는 강화학습과 LLM의 결합이 로봇의 자율성과 대응성을 높이는 것에 어떤 기여를 할 수 있는지를 시사한다.

이러한 연구들은 강화학습의 구조적 한계를 보완하고, 고차원적 판단이 요구되는 자율 시스템 내에서의 의사결정을 기존보다 유연하고 직관적으로 발전시키기 위한 수단으로써의 LLM의 활용 가능성을 입증한다. 이러한 관점에서 보았을 때, AMR 연구에 있어 강화학습과 LLM의 결합은 필수적인 연구 방향으로 판단된다.

## 5. 결론 및 향후 연구 방향

본 연구는 기존 AGV와 AMR의 기술적 차이를 바탕으로 비정형적이고 동적인 환경에서의 AMR 활용 가능성을 고찰하였다. AGV는 고정 경로와 외부 유도 인프라에 의존하는 구조로 인해 환경 변화에 취약하고, 변동성이 높은 환경에서의 적응성에 한계를 가진다. 반면, AMR은 센서 융합, SLAM 기반 자율 주행 등의 기술을 통해 다양한 환경에서 실시간 판단과 주행을 할 수 있다는 점에서 기존 AGV보다 실용성 측면에서 효율적이다.

AMR의 자율성 향상을 위한 제어 기법으로 강화학습은 환경 정보 없이도 목표 지향적인 행동을 학습할 수 있어 자율 이동체의 성능을 향상시키는 유망한 접근으로 평가된다. 특히 객체가 자주 이동하는 등의 동적인 실내 공간이나, 다중 목표를 지닌 환경, 변화 대응이 자주 요구되는 경우 강화학습은 기존 PID, MPC와 같은 전통적인 제어 방식에 비해 우수한 적응성과 유연성을 보이지만 Sim-to-real Gap, 정책의 불투명성과 같은 제약도 여전히 존재한다.

이러한 강화학습의 제한점을 보완하기 위한 대안으로 최근 들어 강화학습과 LLM의 융합에 대한 관심이 부상하고 있다. 자연어에 특화되어 있는 LLM은 사용자에게 강화학습의 의사결정 구조를 설명하고 시뮬레이션과 현실 사이의 전이 문제를 해결하는 등 강화학습의 효율성과 성능을 높이는 핵심 기술로 고려되고 있다.

향후 연구는 강화학습과 LLM의 단순한 융합에 그치지 않고 자율 이동 로봇이 보다 복합적인 과업과 비정형적 사용자 요구를 실제 환경에서 어떻게 유연하게 처리할 수 있는지를 실증적으로 입증하는 방향으로 확장되어야 한다. 이러한 접근은 단순한 성능 개선을 넘어 실질적인 서비스 적용 가능성과 시스템 안정성 확보에 기여함으로써 자율 이동 기술의 진화를 가속화하는 중요한 전환점이 될 수 있을 것으로 판단된다.

## 사사문구

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원·학·석사연계 ICT 핵심인재양성지원(IITP-2025-RS-2022-00156299, 50)과 정부(중소벤처기업부)의 재원으로 기술개발사업 지원(RS-2024-00510727)을 받아 수행된 연구임

## 참고문헌

- [1] A. Loganathan and N. S. Ahmad, "A Systematic Review on Recent Advances in Autonomous Mobile Robot Navigation," *Engineering Science and Technology, an International Journal*, 40, 101343, 2023.
- [2] G. Fragapane, R. De Koster, F. Sgarbossa, and J. O. Strandhagen, "Planning and Control of Autonomous Mobile Robots for Intralogistics: Literature Review and Research Agenda," *European Journal of Operational Research*, 294(2), 405–426, 2021.
- [3] M. Aizat, N. Qistina, and W. Rahiman, "A Comprehensive Review of Recent Advances in Automated Guided Vehicle Technologies: Dynamic Obstacle Avoidance in Complex Environment Toward Autonomous Capability," *IEEE Transactions on Instrumentation and Measurement*, 73, 1–25, 2023.

- [4] R. Keith and H. M. La, "Review of Autonomous Mobile Robots for the Warehouse Environment," *arXiv preprint*, arXiv:2406.08333, 2024.
- [5] R. T. Silva, M. Brilhante, H. Sobreira, D. Matos, and P. Costa, "AGVs vs AMRs: A Comparative Study of Fleet Performance and Flexibility," *2024 7th Iberian Robotics Conference (ROBOT)*, Porto, Portugal, 2024, pp. 1–6.
- [6] D. Q. Mayne, "Model Predictive Control: Recent Developments and Future Promise," *Automatica*, 50(12), 2967–2986, 2014.
- [7] L. Tai, G. Paolo, and M. Liu, "Virtual-to-Real Deep Reinforcement Learning: Continuous Control of Mobile Robots for Mapless Navigation," *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vancouver, Canada, 2017, pp. 31–36.
- [8] G. Kahn, A. Villaflor, B. Ding, P. Abbeel, and S. Levine, "Self-Supervised Deep Reinforcement Learning with Generalized Computation Graphs for Robot Navigation," *2018 IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, Australia, 2018, pp. 5129–5136.
- [9] S. Liu, P. Chang, W. Liang, N. Chakraborty, and K. Driggs-Campbell, "Decentralized Structural-RNN for Robot Crowd Navigation with Deep Reinforcement Learning," *2021 IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, China, 2021, pp. 3517–3524.
- [10] S. Huang, X. Wu, and G. Huang, "Deep Reinforcement Learning-Based Multi-Objective Path Planning on the Off-Road Terrain Environment for Ground Vehicles," *arXiv preprint*, arXiv:2305.13783, 2023.
- [11] D. Arce, J. Solano, and C. Beltrán, "A Comparison Study Between Traditional and Deep-Reinforcement-Learning-Based Algorithms for Indoor Autonomous Navigation in Dynamic Scenarios," *Sensors*, 23(24), 9672, 2023.
- [12] L. Da, M. Gao, H. Mei, and H. Wei, "Prompt to Transfer: Sim-to-Real Transfer for Traffic Signal Control with Prompt Learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, Vancouver, Canada, 2024, Vol. 38(1), pp. 82–90.
- [13] Y. J. Ma, W. Liang, H. J. Wang, S. Wang, Y. Zhu, *et al.*, "Dreureka: Language Model Guided Sim-to-Real Transfer," *arXiv preprint*, arXiv:2406.01967, 2024.
- [14] X. Zhang, Y. Guo, S. Stepputtis, K. Sycara, and J. Campbell, "Explaining Agent Behavior with Large Language Models," *arXiv preprint*, arXiv:2309.10346, 2023.
- [15] W. Huang, F. Xia, T. Xiao, H. Chan, J. Liang, *et al.*, "Inner Monologue: Embodied Reasoning Through Planning with Language Models," *arXiv preprint*, arXiv:2207.05608, 2022.