자연어 기반 자율이동을 위한 Dual-LLM 접근법

김인곤, 신수용*
국립금오공과대학교
20246104@kumoh.ac.kr, *wdragon@kumoh.ac.kr

Dual-LLM Approach to Natural Language-Guided Autonomous Navigation

In Gon Kim, Soo Young Shin*
Kumoh National Institute of Technology.

요 약

본 연구는 로봇의 자율이동 과정에서 자연어 명령을 효과적으로 처리하기 위해 Dual-LLM 기반 접근법을 제안한다. 로컬 환경에서는 Gemma-3 모델을 활용하여 사용자의 명령을 저지연으로 해석하고, 클라우드 환경에서는 GPT-4 mini 모델을 이용해 고차원 추론 및 보조적 의사결정을 수행한다. 이를 통해 기존 ROS2 내비게이션 대비 성공률과 안정성이 향상됨을 시뮬레이션을 통해 확인하였다.

Ⅰ. 서 론

최근 로봇의 자율이동(autonomous navigation)은 물류, 서비스, 재난 대응 등 다양한 응용 분야에서 핵심 기술로 자리잡고 있다. 기존의 ROS2 내비게이션 스택(Nav2)은 지도 기반 경로 계획과 비용맵 기반 주행 제어를 통해 높은 수준의 자율주행 기능을 제공한다[5]. 그러나 이러한 방식은 정확한 좌표 입력과 사전 파라미터 튜닝에 크게 의존하며, 사용자의 자연어 명령을 직접적으로 반영하기 어렵다는 한계를 가진다.

한편, 대규모 언어모델(LLM: Large Language Model)은 자연어 이해와 추론 능력을 기반으로 로봇과 사용자의 상호작용을 혁신할 수 있는 잠재력을 보여주고 있다[1][2]. 사용자가 “A 구역으로 가”와 같이 모호한 명령을 내렸을 때, LLM은 이를 구조화된 명령으로 변환하거나, 불확실한 경우 사용자에게 재질문을 생성하여 명확성을 확보할 수 있다. 하지만 단일 LLM만으로는 두 가지 문제가 존재한다. 경량 모델은 응답 속도는 빠르지만 복잡한 추론 능력이 부족하고, 대규모 모델은 추론 성능은 뛰어나지만 응답 지연(latency)과 연산 비용이 크다는 점이다[4].

본 연구에서는 이러한 한계를 극복하기 위해 Dual-LLM 기반 자율이동 접근법을 제안한다. 구체적으로, 로컬 환경에는 Gemma-3(4b-qat) 모델을 배치하여 저지연 자연어 명령 해석을 수행하고[3], 클라우드 환경에는 GPT-4mini 모델을 활용하여 고차원 추론과 보조적 의사결정을 담당하도록 설계한다. 이를 통해 사용자는 직관적인 자연어 명령을 통해 로봇을 제어할 수 있으며, 시스템은 상황에 따라 동적으로 두 모델을 적절히 활용하여 효율성과 안정성을 동시에 확보한다.

Ⅱ. 본론

제안하는 Dual-LLM 기반 자율이동 시스템은 로컬 모델(Gemma)과 클라우드 모델(GPT)이 상호 보완적으로 동작하도록 설계되었다. Gemma는 Edge 단에서 Candidate Generator 역할을 수행하며, 사용자의 자연어 명령을 빠르게 처리하여 잠재적 실행 후보를 생성한다. 여기에는 NLU(Natural Language Understanding), 맵 정보 활용, 비전 인식 결과를 종합한 행동 계획이 포함되며, ASK/ACT/WAIT와 같은 기본 정책을 통해 즉각적인 실행 가능 여부를 판정한다.

반면 GPT는 Cloud 단에서 Policy Selector로 동작하며, Gemma가 생성한 후보를 입력으로 받아 보다 정교한 의사결정을 수행한다. GPT는 MoA(Mixture of Agents), ToT(Tree of Thoughts), Reflexion과 같은 고차원 추론 전략을 적용할 수 있으며, 긴 문맥(Long-context)과 멀티 타스크(Multi-task) 추론을 통해 단일 후보 이상의 최적 정책을 선택한다. 이러한 구조는 Gemma가 경량 모델로서 빠른 반응성을 제공하고, GPT가 복잡한 상황에서 심층 추론을 담당하는 이중 계층 구조를 형성한다[2][3]. 두 모델 간의 통신은 비동기적 요청-응답 구조로 이루어지며, Gemma가 먼저 후보 실행안을 생성한 뒤 필요 시 GPT API에 전달하여 보정된 정책을 수신한다. 최종적으로 결정된 결과는 ROS2 Nav2 스택으로 전달되어 /navigate_to_pose 실행, 주행 파라미터 수정, 혹은 사용자 재질문과 같은 구체적 동작으로 이어진다[5].

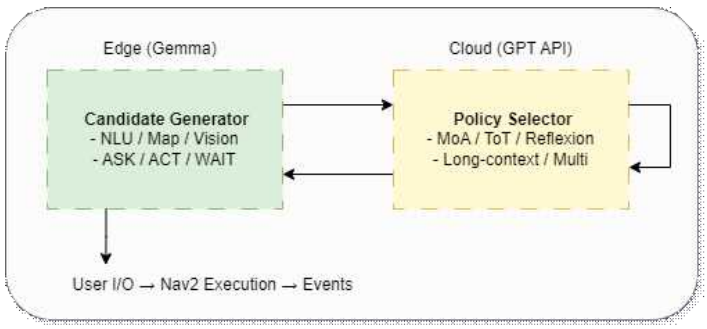

그림 1. Dual-LLM 통신 구조 (Gemma와 GPT 간 역할 분담 및 데이터 흐름)

그림 1. 은 제안하는 Dual-LLM 구조에서 Gemma와 GPT 간의 통신 및 역할 분담을 보여준다. 좌측의 Candidate Generator(Gemma)는 사용자 입력과 센서 데이터를 기반으로 실행 후보를 생성하고, 우측의 Policy Selector(GPT)는 이를 고차원적으로 평가하여 최종 정책을 결정한다. 이

러한 이중 구조는 단일 모델 대비 높은 신뢰성과 적응성을 제공하며, 특히 자연어 기반 자율이동에서 모호한 명령을 해석하거나 복잡한 주행 조건을 처리할 때 효과적이다.

Ⅲ. 실험 및 결과

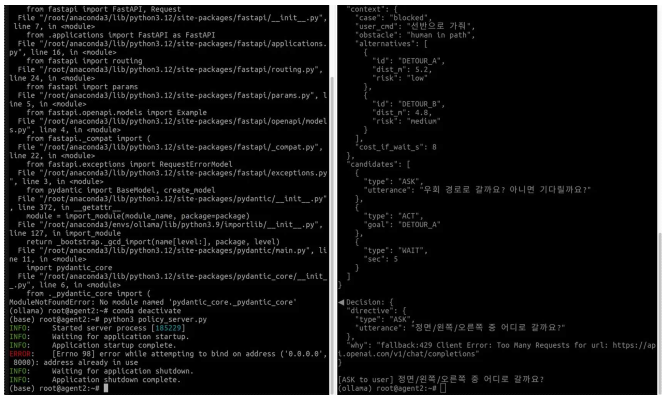

그림 2 Dual-LLM 시스템 실행 결과 로그 (좌: 서버 실행, 우: 후보안 생성 및 최종 질문)

제안한 Dual-LLM 기반 자율이동 시스템의 성능을 검증하기 위해 ROS2 Jazzy와 Gazebo Harmonic 환경에서 시뮬레이션 실험을 수행하였다. 실험 환경은 단순 창고 형태의 맵으로 구성하였으며, 사용자는 “선반으로 가줘”와 같은 자연어 명령을 입력하였다.

그림 2. 는 실제 실행 과정 중 시스템의 후보 행동 생성과 최종 의사결정 로그를 보여준다. 좌측은 서버 실행 및 FastAPI 기반 통신 과정, 우측은 “선반으로 가줘” 명령에 대해 탐지된 장애물(사람)을 인식한 뒤, **우회 경로(DET0UR_A, DETOUR_B) 및 대기(Wait)**의 후보안을 생성하고, 이들에 대해 “우회 경로로 갈까요? 아니면 기다릴까요?”라는 질문을 사용자에게 제안하는 모습을 나타낸다.

이와 같은 로그는 제안한 프레임워크가 단순히 좌표를 따라가는 것이 아니라, 자연어 기반 후보 생성(Gemma) + 정책 선택(GPT) 과정을 통해 상황에 맞는 행동을 도출함을 보여준다. 특히 Gemma는 ASK/ACT/WAIT 형태로 빠르게 후보안을 만들고, GPT는 이를 종합적으로 평가해 최종적으로 사용자에게 질문을 생성하거나 특정 경로를 선택하도록 한다.

Ⅳ. 결론

본 논문에서는 로봇의 자율이동 과정에서 자연어 명령을 처리하기 위해 Dual-LLM 기반 접근법을 제안하였다. 로컬 단에서는 Gemma-3 모델을 통해 저지연 명령 해석과 후보 행동 생성을 수행하고, 클라우드 단에서는 GPT-4mini를 통해 고차원 추론과 정책 선택을 담당하도록 설계하였다. 이를 통해 기존 ROS2 내비게이션 대비 명령 해석의 정확성과 주행 성공률을 향상시킬 수 있음을 ROS2 Jazzy + Gazebo Harmonic 환경에서의 실험을 통해 확인하였다.

제안된 구조는 단일 LLM 접근법의 한계를 극복하고, 실시간성(로컬)과 심층 추론(클라우드)을 결합한 새로운 형태의 자율이동 프레임워크로서 의의가 있다. 향후 연구에서는 멀티모달 센서 데이터를 통합하여 시각·언어 기반 의사결정을 확장하고, 실제 로봇 플랫폼에 적용하여 실증 실험을 진행할 예정이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원 - 학·석사연계ICT핵심인재양성 지원(IITP-2025-RS-2022-00156394, 50%)과 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-지역지능화혁신인재양성사업의 지원을 받아 수행된 연구임 (IITP-2025-RS-2020-II201612, 50%)..

참 고 문 헌

[1] Vemprala, S., et al. "ChatGPT for Robotics: Design Principles and Model Abilities." arXiv preprint arXiv:2306.17582, 2023.

[2] OpenAI. "GPT-4 Technical Report." arXiv preprint arXiv:2303.08774, 2023.

[3] Google. "Gemma: Open Models for Responsible AI." arXiv preprint arXiv:2403.08295, 2024.

[4] Huang, W., et al. "Language Models as Zero-Shot Planners: Extracting Actionable Knowledge for Embodied Agents." ICML, 2022.

[5] Macenski, S., Foote, T., Gerkey, B., Lalancette, C., & Woodall, W. "Robot Operating System 2: Design, architecture, and uses in the wild." Science Robotics, 7(66), eabm6074, 2022.