

LLM 기반 공간 의미 분석을 통한 객체 위치 추론 방법

신성한, 권용현, 이소현, 김유진, 김재호*

세종대학교

seonghan.sejong@gmail.com, yonghyun.sejong@gmail.com, sohyun.sejong@gmail.com,
kimyj.sejong@gmail.com, *kimjh@sejong.ac.kr

Object Location Inference Method Using LLM-Based Spatial Semantic Analysis

Seonghan Shin, Yonghyun Kwon, Sohyun Lee, Yujin Kim, Jaeho Kim*
Sejong Univ.

요약

최근 Embodied AI의 발전에 따라 물리적 환경과 공간에 대한 정보 인식 및 추론을 하기 위한 연구가 이루어지고 있다. 특히, 대규모 언어 모델을 이용한 공간 인식과 추론을 위한 연구가 수행되고 있다. 그러나 복잡한 공간 관계와 물리적 제약을 반영하여 객체 위치를 추론하는 기존의 방법론은 제한적인 성능을 보였다. 이러한 한계를 극복하기 위해 본 논문에서는 맥락 데이터베이스를 기반으로 대규모 언어 모델과 논리적 추론 모델을 결합한 객체 위치 추론 방법을 제안하고 이를 구현한다. 제안된 방법은 공간 구조 이해를 강화하고 물리적 위치 제약을 일관되게 반영함으로써 객체 추론의 정확도 및 안정성을 향상시키며, 이를 통해 에이전트의 공간 구조 및 역할 이해에서의 궁정적인 효과를 기대한다.

I. 서론

최근 Embodied AI의 발전에 따라, 에이전트가 물리적 환경에서 자율적으로 인식, 추론, 행동을 수행할 수 있도록 하는 연구가 활발히 이루어지고 있다. 자율 에이전트를 효과적으로 구현하기 위해서는 에이전트가 주변 환경을 분석하여 공간의 구조와 역할을 이해하고, 이를 바탕으로 객체의 위치를 추론하는 능력이 필수적이다[1].

최근 대규모 언어 모델(Large Language Model, LLM)이 자연어로 표현된 공간 정보를 처리하고, 이를 의미적으로 추론하는 데 높은 잠재력을 보이고 있다. 그러나 모델의 추론 능력 한계로 인해 복잡한 구조를 정확히 추론하는 데에는 어려움이 존재한다[2]. 이러한 한계를 극복하기 위한 시도로, 공간 관계를 각 속성 별로 분류하여 하나의 통합된 형식으로 변환하여 LLM을 미세 조정하는 SpaRP 프레임워크[3]가 제안되었다. 해당 프레임워크는 LLM의 공간 추론 성능을 향상시켰으나, 미세 조정에 필요한 높은 비용과 연산 자원으로 인해 실제 적용에는 제약이 따른다.

본 연구에서는 별도의 미세 조정 없이도 LLM의 공간 추론 능력을 향상시키기 위해, 선언적 규칙과 물리적 위치 제약 조건을 기반으로 논리 문제를 계산하는 Answer Set Programming(ASP)[4] 모듈을 LLM과 결합한 새로운 공간 추론 방법을 제안한다. 제안된 방법은 객체 위치 추론 문제에 적용되었으며, 일반 LLM 모델과 SpaRP 데이터셋[3]으로 미세 조정된 LLM 모델 대비 ASP[4]가 결합된 모델이 위치 추론의 정확도 및 안정성에 미치는 영향을 실험적으로 검증하였다.

II. 본론

2.1 시스템 개요

그림 1은 본 연구에서 제안하는 객체 위치 추론 시스템의 전체 구조이다. 본 시스템은 사용자가 텍스트 형태로 입력한 객체명을 바탕으로 방 및 가구 후보를 생성한 뒤, 이를 맥락 데이터베이스에 저장된 실제 가구와 생성된 가구 후보를 대응한다. 이후 물리적 위치 제약을 반영한 제약 기반 추론 과정을 통해 예측된 최종 위치를 우선순위로 제시한다.

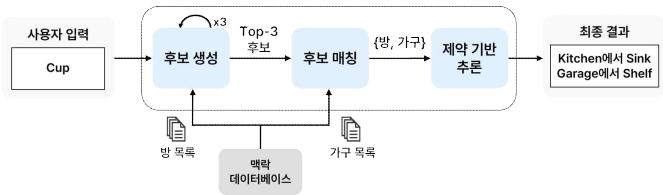


그림 1. 실험의 구조도

2.2 후보 생성 모듈

본 모듈은 사용자가 입력한 객체명을 GPT-4o-mini[5]에 질의하여, 해당 객체가 존재할 가능성이 높은 방 후보 3개를 맥락 데이터베이스의 방 목록에서 추출한다. 동시에, GPT-4o-mini[5]를 사용하여 객체가 배치될 가능성이 높은 가구 후보 3개를 새롭게 생성한다. 이 과정은 총 세 차례 반복 수행되며, 페드백 구조를 통해 생성된 방 및 가구 후보의 정확도를 점진적으로 향상시킨다. 해당 모델은 입력 토큰 최적화를 통해 응답 시간과 호출 비용을 효과적으로 절감하여 세 차례 반복 수행 구조에서도 효율적인 후보 생성이 가능하다.

2.3 후보 매칭 모듈

본 모듈은 후보 생성 모듈에서 전달된 3개의 방 후보와 가구 후보를 입력받아, 각 방에 포함된 가구 목록과 가구 후보 간의 의미적 유사도를 평가하고 가장 유사한 항목을 식별한다. 구체적으로, 각 방 후보에 포함된 가구 목록을 맥락 데이터베이스에서 조회하여 방별 가구 스키마 리스트를 구성한다. 이후 SBERT(Sentence-BERT)[6]를 이용해 가구 후보와 맥락 데이터베이스에서 추출된 가구 이름을 임베딩 벡터로 변환한 뒤, 코사인 유사도를 계산하여 가구 후보와 의미적으로 가장 유사한 가구 항목을 식별한다. 단순 임베딩과 달리 대조 학습으로 미세 조정된 SBERT[6]는, 가구 후보와 방별 스키마 내 가구 이름의 코사인 유사도를 직접 계산함으로써 짧은 텍스트인 가구명에도 높은 분별 성능을 보인다.

2.4 제약 기반 추론 모듈

본 모듈은 후보 매칭 모듈을 통해 대응된 {방, 가구} 쌍을 ASP[4]를 이용한 크기 호환성, 상호 배타성, 접근성 등의 도메인 제약을 적용한다. 이후, 이러한 모든 제약을 충족하는 위치 조합을 ASP 추론 엔진[4]을 통해 생성하고, 각 조합 위반 여부를 판단하여 최종 위치 후보를 도출한다. ASP[4]는 선언적 규칙 기반의 추론 방식을 사용하여, 해당 추론 방식과 제약 조건을 토대로 위치나 프롬프트 길이에 제약 없이 일관되게 적용할 수 있다. 이를 통해 LLM 단독 사용 시 발생하기 쉬운 제약 미반영 문제를 방지하고, 결과의 논리적 일관성을 보장할 수 있다.

III. 실험 구현 결과 및 성능

3.1 데이터셋

본 실험에서는 맥락 데이터베이스 구축을 위해, 맵 스키마에 16개 방과 각 방마다 중복 없는 5개 가구로 구성된 16개의 방 스키마를 정의하였다. 이때 각 방에는 해당 공간에 일반적으로 배치되는 4개의 가구와 더불어, 실험의 제약 조건을 반영하기 위해 배치 가능성이 낮은 가구 1개를 의도적으로 포함했다. 성능 평가를 위해 총 30개의 평가 데이터셋을 제작하였으며, 각 데이터셋은 제약 조건을 갖는 특정 가구와 해당 가구가 실제로 배치된 방 및 방 내 정확한 위치를 정답으로 설정하여 성능을 검증하였다.

3.2 실험 방법 및 성능 지표

본 실험에서는 기본 GPT-4o-mini[5], SpaRP 형식[3]으로 미세 조정된 GPT-4o-mini-FT, 그리고 본 연구에서 제안한 모델인 GPT-4o-mini + ASP를 비교 대상으로 설정하였다. GPT-4o-mini[5]는 입력된 객체명을 기반으로 생성된 방 후보와 16개 방의 전체 가구 목록을 프롬프트에 포함하여 평가를 수행하였다. GPT-4o-mini-FT는 동일한 프롬프트 구성에 SpaRP 형식[3]으로 변환된 평가 데이터셋을 반영하여 일관되게 평가하였다. GPT-4o-mini + ASP는 방 목록과 평가 데이터셋만을 입력으로 받아 방 및 가구 후보를 생성하고, SBERT[6] 기반 의미 기반 유사도 대응과 A SP[4] 제약을 적용하여 평가를 진행하였다. 성능 평가는 Top-k Hit Rate 와 1회 추론 시 사용된 프롬프트 토큰 수를 기준으로 이루어졌다.

3.3 실험 결과

표 1은 기본 GPT-4o-mini[5] 모델과, SpaRP 형식[3]으로 미세 조정된 GPT-4o-mini-FT, 그리고 ASP 제약[4]을 적용한 제안 모델의 Top-k Hit Rate 성능을 비교한 결과이다. SpaRP 형식[3]으로 미세 조정된 GPT-4o-mini-FT 모델 대비, 제안 모델은 Top-3 Hit Rate을 53.3%에서 60.0%로 6.7%p 개선하였으며, Top-5 Hit Rate에서도 60.0%를 유지하여 ASP 제약 도입의 유효성을 확인하였다.

표 1. 제약 조건별 성능 평가

모델	Hit@1	Hit@3	Hit@5
GPT-4o-mini Only	26.7	40	40
GPT-4o-mini-FT(SpaRP)	33.3	53.3	60
GPT-4o-mini+ASP(Ours)	26.7	60	60

표 2. 프롬프트 토큰 사용량

모델	토큰 사용량
GPT-4o-mini Only	900 - 910
GPT-4o-mini-FT(SpaRP)	900 - 910
GPT-4o-mini+ASP(Ours)	420 - 450

표 2는 세 모델의 1회 추론당 프롬프트 사용량을 비교한 결과이다. 기본 GPT-4o-mini[5] 모델과 SpaRP 형식[3]으로 미세 조정된 GPT-4o-mini-FT 모델은 900 - 910 토큰을 소모한 반면, 제안 모델은 불필요한 가구 정보를 프롬프트에서 제거함으로써 420-450 토큰으로 추론 비용을 약 49.5% 절감하였다. 이는 토큰 길이 제한의 영향을 최소화하며, 맥락 데이터베이스가 확장되어도 일정한 토큰 사용량을 유지하여 시스템 확장에 따른 비용 증가를 방지한다.

IV. 결론

본 연구에서는 후보 생성 모듈, 후보 매칭 모듈, 제약 기반 추론 모듈과 맥락 데이터베이스를 결합하여 공간 의미를 분석하고, 이를 기반으로 객체 위치를 추론하는 새로운 접근법을 제안하였다. 제안한 방법은 공간 정보를 직접 입력받는 기존 LLM 기반 방식보다 객체 위치 추론 정확도에서 유의미한 향상을 보였으며, 특히 Top-3 기준 성능에서 기존 연구 대비 우수한 결과를 나타냈다. 또한 토큰 사용량을 크게 줄임으로써 추론 비용을 효과적으로 절감할 수 있음을 실험을 통해 확인하였다.

본 시스템은 공간이 다양한 제약 조건을 일관되게 반영할 수 있어, 에이전트가 공간의 구조와 역할을 이해하고 객체 위치 추론 정확도를 향상하는데 기여할 것으로 기대한다. 향후 연구에서는 객체를 이미지나 오디오 같은 멀티모달 입력 형태로 제공하여 추론하는 기능을 구현하고자 한다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 정보통신방송 혁신인재양성(메타버스융합대학원)사업(IITP-2025-RS-2023-002545 29)의 연구 결과로 수행되었고, 산업통상자원부 및 산업기술평가관리원(KEIT)(RS-2022-00154678) 연구비 지원받았으며, 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학ICT연구센터(ITRC)의 지원을 받아 수행된 연구임(IITP-2025-RS-2021-II211816)

참 고 문 헌

- [1] Raychaudhuri, S., & Chang, A. X. "Semantic Mapping in Indoor Embodied AI--A Comprehensive Survey and Future Directions.", arXiv preprint arXiv:2501.05750. 2025.
- [2] Yan, H., Hu, X., Wan, X., Huang, C., Zou, K., & Xu, S. "Inherent limitations of LLMs regarding spatial information.", arXiv preprint arXiv:2312.03042. 2023.
- [3] Rizvi, M. I. H., Zhu, X., & Gurevych, I. "Sparc and sparp: Spatial reasoning characterization and path generation for understanding spatial reasoning capability of large language models.", arXiv preprint arXiv:2406.04566. 2024.
- [4] Lifschitz, V. *Answer set programming* (Vol. 3). Cham: Springer. 2019.
- [5] Hurst, A., Lerer, A., Goucher, A. P., Perelman, A., Ramesh, A., Clark, A., ... & Kivlichan, I. "Gpt-4o system card.", arXiv preprint arXiv:2410.21276. 2024.
- [6] Reimers, N., & Gurevych, I. "Sentence-bert: Sentence embeddings using siamese bert-networks.", arXiv preprint arXiv:1908.10084. 2019.