거대 언어 모델 에이전트의 기대 기반 메타인지 기술

황규석 [1], 배유석 [2], 황중원 [2]
[1] 서울과학기술대학교 컴퓨터공학과 학부생
[2] 한국전자통신연구원

kyoosuk99@seoultech.ac.kr, baeys@etri.re.kr, jwhwang@etri.re.kr

Expectation Based Metacognition for LLM Agent

Kyoo Suk Hwang[1], Yu Seok Bae[2], Joong-won Hwang[3]
[1]Dept. of Computer Engineering, Seoul National University of Science and Technology
[2]Visual Intelligence Lab, Electronics and Telecommunications Research Institute

요 약

최근 거대언어모델(Large Language Model)에 기반한 에이전트 모델의 연구가 활발하게 이루어지고 있다. 본 논문은 이러한 LLM 기반이 자신의 행동의 실패를 인지하고 계획을 다시 수립할 수 있도록 하는 메타인지적 기능에 대한 연구를 다룬다. 본 연구에서 LLM 에이전트는 계획 단계에서 행동의 결과에 대한 기대를 출력하고, 이후 관측 결과와 행동이 일치하는지 여부에 따라 계획을 수정한다. 이는 별도의 검증 데이터셋을 요구하지 않으며 LLM 의 확장성을 극대화할 수 있는 방법이다. ScienceQA, Mathvista, MMMU 세 가지 벤치마크에서 본 논문에서 제안하는 방법이 메타인지를 수행하지 않는 방법에 비해 효과가 있음을 확인할 수 있었다.

1. 서론

최근 거대 언어 모델(Large Language Model, LLM)의 강력한 일반화 능력을 사용한 에이전트 모델들에 대한 연구가 활발하게 이루어지고 있다. 특히 최근 ReAct[1]는 프롬프트 형태로 추론과 행동을 교차 생성하여 상호보완하는 생각사슬(Chain-of-Thought, CoT)을 유도하여 큰 발전을 이루었다. ReAct는 LLM이 단순히 언어적 추론을 수행하는 수준을 넘어, 외부 환경과 상호작용하며 일종의 에이전틱 AI로 작동할 수 있음을 보여준 초기 사례라는 점에서 의미가 있다. 그러나 이러한 방식은 스스로의 실패나 한계를 메타적으로 인식하고 반성적으로 활용하지 못하기 때문에, 경험을 축적하며 점진적으로 개선하는 능력에는 한계가 따른다. 이와 달리 Reflexion[2]은 행동의 실패를 인지하는 메타인지적 자기반성을 도입하여, 실패 경험을 구조화된 피드백으로 전환하고 이를 장기 기억으로 저장함으로써, 반복적 시도 속에서 자기강화적 학습과 성능 개선을 가능하게 한다.

그러나 Reflexion의 경우, LLM 에이전트의 실패 여부를 평가하기 위하여 미리 정해진 태스크별 평가 모듈(Evaluator)을 도입하는데, 이 평가 모듈은 미리 정해진 태스크에 대하여 정답 정보(Ground Truth)를 가진 평가 데이터셋을 통하여 학습되거나 휴리스틱하게 구현된다. 이에 따라 정답 정보를 가진 데이터셋을 모으는 비용이 들 뿐만 아니라, 평가 모듈이 평가할 수 있는 태스크의 종류가 한정되어 있어 LLM의 강력한 확장성을 온전히 활용하지 못하는 원인이 된다.

이러한 한계를 극복하기 위하여 본 논문에서는 정답 정보 기반이 아닌 기대 기반의 평가 모듈을 제안한다. 즉 LLM 에이전트가 행동을 정하는 계획 단계에서 행동의 결과가 어떨지 기대를 하고, 이 기대를 잘 만족시켰는지 평가하는 방식으로 수행하여, 단순 행동의 수행뿐 아니라, 수행 능력을 고려한 계획의 타당성까지 포함한 평가를 수행하는 방식이다. 이는 평가의 주체가 LLM이기 때문에 강력한 확장성이 있다는 장점을 갖고 있다. 최종적으로는 이러한 평가 모듈을 활용한 멀티 에이전트 파이프라인을 구성하여 가장 효율적으로 에이전트를 선택 및 활용하여 문제를 풀 수 있는 방법을 설계하였다.

2. 방법

본 논문의 제안 방법에서 LLM 에이전트는 주어진 목표를 달성하기 위해 다단계의 행동을 수행한다. 각 단계는 계획-기대-행동-관찰-계획수정을 포함하는데,

이러한 동작은 프롬프트를 통한 생각사슬로 에이전트에게 지시된다. 그림 1은 전체 파이프라인을 보여주며, 자세한 프롬프트는 https://github.com/kstone-285/Smolcorrection에서 확인 가능하다.

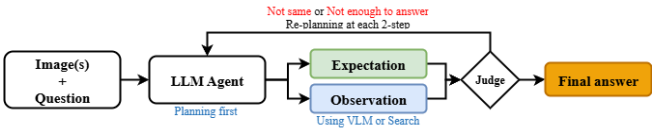
2.1 에이전트 행동의 범위

일반적으로 LLM 기반 파운데이션 모델은 다양한 도메인과 과제를 어우르는 범용성에서는 뛰어나지만, 특정 작업에서는 특화 에이전트보다 성능이 떨어지는 문제가 있으며, 언어 외 모달리티를 처리하거나 외부 지식을 활용하기 어렵다는 문제가 있다. 이를 해결하기 위하여 통상적으로 도구적 기능을 연결하고, LLM 에이전트가 이를 호출하여 사용하는 방식이 널리 사용된다. 본 논문에서는 검색 기능, 검출 기능, 계수 기능, 비교 기능, 문자열 인식 기능, 차트 해독 기능을 하는 도구를 연결시켰으며, 이 중 검색기능을 제외한 기능들은 시각언어모델(Vision Language Model, VLM)을 활용하는 방식을 취하였다. 이는 행동 결과에 대한 관찰을 언어적인 형태로 나타내기 위함이다.

2.2 기대-결과 일치 여부 판정과 계획 수정

최종 목표를 달성하기 위하여 LLM 에이전트는 여러 단계의 행동을 계획하게 된다. 이 때, 제안 방법에서는 프롬프트와 지시문을 통하여 LLM 에이전트가 계획된 행동에 대한 기대를 출력하도록 유도하였다. 이를 위하여 맥락정보로 이전의 계획-행동과 행동 결과에 대한 관찰이 입력된다.

이렇게 출력된 기대는 VLM 이나 검색도구를 통하여 검색된 결과에 대한 관찰과 쌍으로 비교되게 된다. 이 비교는 역시 LLM 에 의하여 이루어지는데, 이는 관찰과 기대가 모두 언어의 형태로 전환되어 있기 때문에 가능하다. 최종적으로 이 쌍이 동일한 의미를 지니지 않았다면, 계획이나 수행과정에 실패가 있었다고 생각하고, 새로운 행동을 탐색한다. 실험적으로 이러한 의미의 일관성을 통한 메타인지가 제대로 수행되었는지를 직접 물어보는 방식의 메타인지보다 높은 성능을 보인다는 점이 확인되었다.



(그림 1) 멀티 에이전트 파이프라인.

3. 실험설계 및 결과

본 논문에서 제안하는 방법을 시험하기 위하여 다단계 문제를 해결하는 벤치마크로 ScienceQA, Mathvista, MMMU 를 사용하였다. LLM 에이전트의 모델은 Qwen2.5-Coder 을 사용하였으며, 행동으로 호출되는 VLM 으로 Qwen2.5-VL 을 사용하였다. 비교군으로는 단순 VLM 과 메타인지 없이 행동하는 LLM 에이전트를 선택하였다.

표 1 의 결과는 제안 방법의 효과가 모델 규모에 따라 다름을 보여준다. 이는 소규모 모델이 복잡한 프롬프트를 제대로 인지하고 활용하지 못하는 반면, 대규모 모델(32B)은 더 많은 지식을 기반으로 이를 활용해 적절한 도구를 선택하며 Mathvista, MMMU 와 같은 복잡한 문제 해결에 성공했기 때문이다.

그러나 ScienceQA 처럼 단순한 문제에는 과도하여 성능 저하를 일으켰다. 이는 문제의 복잡도를 판단하여 선택적으로 적용하는 전략이 필요함을 보여준다.

<표 1> 제안 방법의 정량적 성능 비교(Accuracy)

(3B-AWQ)	VLM only	Single Agent	Multi Agent(ours)
ScienceQA	**46.80**	44.74	40.09
Mathvista	**60.40**	43.71	37.25
MMMU	**33.11**	32.64	31.33
(7B-AWQ)	VLM only	Single Agent	Multi Agent(ours)
ScienceQA	**71.34**	62.33	46.33
Mathvista	**68.00**	50.33	50.75
MMMU	**35.47**	34.40	33.20
(32B-AWQ)	VLM only	Single Agent	Multi Agent(ours)
ScienceQA	**91.87**	89.25	87.50
Mathvista	71.40	74.24	**76.49**
MMMU	53.22	56.13	**59.58**

4. 결론

본 논문에서는 LLM이 자신의 행동 결과에 대한 기대를 출력하고 실제 행동의 결과와 일치여부를 확인함으로써 계획과 행동이 제대로 수행되었는지 확인할 수 있는 메타인지 기능을 구현하였다. 이를 활용하여 계획수정을 하는 LLM 에이전트는 모델의 규모가 커질수록 그렇지 않은 에이전트에 비해 복잡한 벤치마크에서 높은 성능을 보여주었다. 향후 추가적인 연구를 통해 Reflexion과 같이 장기 계획에 메타인지 내용을 반영하는 방법을 연구할 예정이다.

사사

본 연구는 과학기술정보통신부(MSIT) 지원으로 정보통신기획평가원(IITP)의 연구개발사업(No. RS-2022-II220124, 스스로 학습 역량을 인지하고 활용하여 적정한 결과를 제공하는 인공지능 기술 개발)의 지원을 받아 수행되었습니다.

참고문헌

[1] Yao, Shunyu, et al. "React: Synergizing reasoning and acting in language models." International Conference on Learning Representations (ICLR). 2023.

[2] Shinn, Noah, et al. "Reflexion: Language agents with verbal reinforcement learning." Advances in Neural Information Processing Systems 36 (2023): 8634-8652.