2025년도 한국통신학회 추계종합학술발표회

LLM 웹 에이전트의 장기 추론 성능 향상을 위한 기능적 역할 분리 연구

임소민[1], 박희원[2], 권민혜[1,2,*]
숭실대학교 전자정보공학부[1], 지능형반도체학과[2]
{isomin2004, heewon012}@soongsil.ac.kr, *minhae@ssu.ac.kr

Enhancing the Long-horizon Reasoning Capability of LLM Web Agents via Functional Role Separation

Somin Im, Heewon Park, Minhae Kwon

요 약

최근 대규모 언어 모델(Large Language Models, LLMs)은 자연어 처리, 코드 생성, 웹 기반 작업 등 다양한 분야에서 뛰어난 성능을 입증하며 적용 범위가 빠르게 확장되고 있다. 특히 웹 환경에서 사용자의 명령을 이해하고 작업을 수행하는 웹 에이전트로의 활용이 주목받고 있으나 복잡한 지시 사항이나 다단계의 추론이 요구되는 상황에서는 여전히 한계를 보인다. 이러한 한계를 해결하기 위해 다양한 접근법들이 제안되고 있지만 여전히 비효율적인 행동 패턴이 나타나는 문제가 존재한다. 본 논문에서는 LLM 에이전트를 기능적으로 분리하여, 에이전트의 추론 과정을 주기적으로 감독하고 조언을 제공하는 Director를 도입한 Agent-Director 시스템을 제안한다. 이는 웹 쇼핑 벤치마크 WebShop에서 기존 방법 대비 우수한 성능을 달성하고 장기적 추론에서도 효과적임을 확인하였다.

Ⅰ. 서 론

최근 LLM은 단순한 문장 생성 능력을 넘어, 스스로 목표를 설정하고 계획을 세워 행동하는 에이전트로 활용하려는 연구가 증가하고 있다. 특히 웹 환경에서는 LLM 에이전트가 웹페이지 탐색, 상품 비교, 결제와 같이 실제 사용자 작업 흐름을 순차적으로 수행하는 능력이 요구된다. 대표적인 웹 환경 벤치마크인 WebShop에서는 모델이 특정 제품을 구매하기 위해 검색, 후보 비교, 옵션 선택 등을 순차적으로 수행해야 하며, 각 단계의 결정이 이후 행동과 최종 결과에 누적적으로 영향을 미친다[1]. 이처럼 장기 추론이 요구되는 환경에서는 일회성 응답이 아닌 과정 전체를 조율할 수 있는 능력이 요구된다.

이러한 장기 추론 능력을 향상시키기 위해 다양한 연구들이 제안되고 있다. 대표적으로 추론과 행동을 교차 수행하도록 유도하는 방식[2], 과거의 실패 경험을 반성하며 학습하는 방식[3], 트리 탐색을 활용한 계획 기반 추론 방식[4], 행동 선택과 평가 과정을 분리하는 방식[5] 등이 있다. 그러나 기존 방법들은 고정적인 추론 패턴에 과의존하거나 복잡한 다단계 작업에서 비효율적인 탐색 및 반복적 오류가 발생하는 한계를 보인다.

본 연구는 이러한 한계를 극복하기 위하여 LLM 에이전트를 기능적으로 분리한 Agent-Director 시스템을 제안한다. Agent는 환경과 직접 상호작용하며 행동을 수행하고, Director는 주기적으로 Agent의 검색 기록을 관찰해 현재 Agent의 상황에 맞도록 전략적 조언을 제공한다. LLM의 역할 분리 구조를 통해 Director가 Agent의 추론 과정을 외부에서 독립적으로 분석함으로써, 기존 방법들에서 누적되던 내부적 추론 오류를 효과적으로 감지하고 교정할 수 있다. 또한, 추가적인 모델 학습이나 파인튜닝 없이도 장기적이고 안정적인 의사결정이 가능하다. 본 논문에서는 제안하는 시스템의 효과성을 입증하기 위해 웹 쇼핑 벤치마크인 WebShop 환경에서 실험을 수행하였으며, 기존 방법 대비 우수한 성능을 확인하였다.

Ⅱ. Agent-Director LLM 시스템

본 연구에서 제안하는 Agent-Director 시스템은 그림 1과 같이 동작한다. Agent가 사용자로부터 초기 명령어 I를 입력받으면 웹 환경과 직접 상호작용하며 페이지 정보를 확인하고 클릭, 검색 등의 행동을 선택 및 실행한다. Director는 일정 주기 n마다 Agent의 현재 상황을 기반으로 Agent에게 전략적 조언을 제공한다.

Agent-Director 시스템은 동일한 LLM 모델 $f(\cdot)$을 사용하며, 서로 다른 Profile을 통해 역할이 분리된다. Agent의 프로필 P_A와 Director의 프로필 P_D는 각각의 역할에 대한 특성을 정의하고 행동 전략, 조언 스타일 등 역할에 특화된 정보를 포함한다. 이러한 프로필 기반의 역할 분리는 Agent와 Director가 서로 다른 관점과 접근 방식을 가지게 하며, 단일 모델 내에서도 다양성 있는 추론을 가능하게 한다.

Agent는 매 시점 $t(1 \le t \le T)$ 마다 주어진 명령어 I를 수행하기 위해 웹 환경과 상호작용한다. 웹 환경은 Agent에게 현재 페이지 정보 o_t와 가능 행동 공간 A_t를 제공하며, Agent는 웹 환경으로부터 얻은 정보 o_t와 프로필 P_A를 활용해 행동 a_t를 결정한다. Agent의 정보를 기반으로 최대 n-step 까지의 페이지 정보와 행동을 누적해 검색 기록 $h_t = \{(o_{t-n-1}, a_{t-n-1}), \cdots, (o_{t-1}, a_{t-1}), (o_t, a_t)\}$를 구성한다. Agent는 명령어 I, 현재 시점에서의 페이지 정보 o_t, 이전 step까지의 검색 기록 h_{t-1}을 통합하여 현재 상황 $x_t = concat(I, o_t, h_{t-1})$를 생성한다. Agent는 생성된 x_t를 Agent의 프로필 P_A와 함께 LLM 모델 f에 입력하여 행동 $a_t = f(P_A, x_t)$를 선택한다.

Director는 n-step마다 개입하여 Agent에게 조언 c_t을 전달한다. 조언 c_t는 Agent로부터 전달받은 현재 상황 x_t와 Director의 프로필 P_D를 LLM 모델 f에 입력해 $c_t = f(P_D, x_t)$로 생성한다. Director에게 조언을 받은 Agent는 현재 시점에서의 조언 c_t를 x_t에 포함시켜 현재 상황 $x_t \leftarrow concat(x_t, c_t)$를 업데이트하는 과정을 거친다. 이후, 자신의 프로필인 P_A와 업데이트된 현재 상황 x_t를 기반으로 다시 행동 $a_t = f(P_A, x_t)$를 선택한다. 이때 Agent는 Director의 조언을 참고로만 활용하며, 최종 행동은 자신의 판단으로 결정한다.

*교신 저자(Corresponding Author)

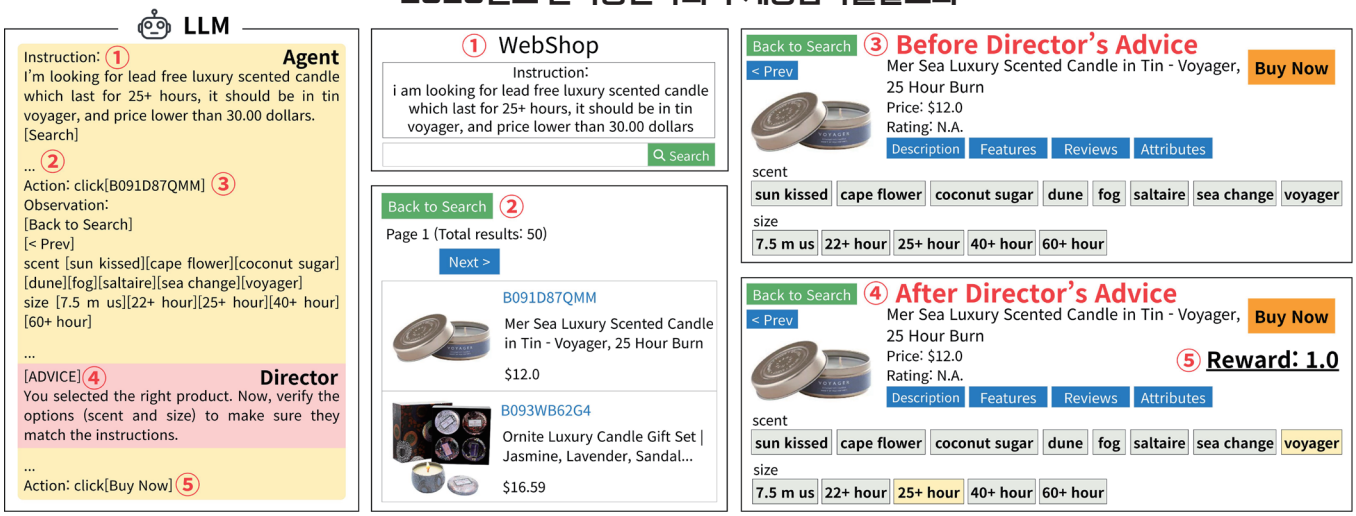

그림 1 제안하는 Agent-Director 시스템의 WebShop 동작 과정

Ⅲ. 실험

3.1. 실험환경 설정

본 실험은 meta의 LLaMA3-8B 모델을 백본 모델로 사용하여 WebShop 벤치마크에서 500 episode에 대한 성능을 평가하였다. 제안된 Agent-Director 시스템에서 Director의 호출 주기 n은 4로, 최대 time step T 는 15로 설정하였다.

Benchmark: 본 연구의 실험은 Amazon 상품 데이터 기반의 대화형 전자상거래 환경인 WebShop 벤치마크에서 수행되었다[1]. 성능 평가는 구매 성공 여부를 나타내는 이진 지표인 Success Rate와 요구사항 만족도를 나타내는 0-1 사이의 연속 값인 Reward를 기준으로 한다.

Baseline:

1. **ReAct[2]:** LLM이 행동을 선택하기 전 추론 과정을 명시적으로 출력해 행동의 근거를 제공하는 프레임워크
2. **Reflexion[3]:** 에피소드 종료 후 실패 원인을 자체 분석하는 self-reflection 과정을 통해 과거 경험으로부터 학습하는 프레임워크
3. **LATS[4]:** 트리 탐색을 통해 여러 행동 경로를 탐색하고 평가하여 최적의 의사결정을 수행하는 프레임워크
4. **LAC[5]:** 강화학습의 Actor-Critic 구조를 LLM 환경에 적용하여, 생성 및 평가 능력을 분리해 장기 의사결정 성능을 개선하는 프레임워크

3.2. 실험 결과

표 1은 WebShop 환경에서 각 베이스라인과 제안된 시스템의 성능 비교 결과를 나타낸다. 제안된 Agent-Director 시스템은 Success Rate 39.2%, Avg Reward 0.6564을 달성하며 모든 베이스라인 대비 우수한 성능을 보였다. 특히 가장 높은 성능을 보인 LAC 대비 Success Rate에서 8.2%의 향상을 달성하였으며, 단일 관점 추론 방식인 ReAct와 비교하면 Success Rate에서 13.8%, Avg Reward에서 0.2141의 큰 성능 격차를 확인할 수 있다.

실험 결과는 LLM의 기능적 역할 분리를 통한 Agent-Director 시스템의 효과를 입증한다. 기존 방법들은 단일 Agent의 내부 추론에만 의존하여 실시간 오류 교정에 한계가 있는 반면, 제안하는 시스템은 Director가 외부 관점에서 Agent의 행동 이력을 주기적으로 분석하고 독립적 조언을 제공하여 추가적인 계산 비용 없이도 장기적 목표 유지와 일관된 의사결정을 가능하게 한다. 이는 복잡한 다단계 추론이 요구되는 웹 환경에서 역할 분리를 통한 외부 관점의 주기적 개입이 효과적임을 보여준다.

표 1 알고리즘 별 성능 비교

	Success Rate	Avg Reward
ReAct[2]	25.4%	0.4423
Reflexion[3]	27.2%	0.4723
LATS[4]	13.2%	0.3768
LAC[5]	31%	0.6250
Proposed	**39.2%**	**0.6564**

Ⅳ. 결론

본 논문은 기존 단일 에이전트 방식의 한계를 극복하기 위해 LLM을 기능적으로 분리한 Agent-Director 시스템을 제안하였다. 제안된 시스템은 환경과 직접 상호작용하는 Agent와 이를 주기적으로 관찰하여 조언을 제공하는 Director로 구성되며, Director가 외부 관점에서 Agent의 행동 이력을 분석함으로써 내부적 추론 오류를 효과적으로 교정할 수 있다. WebShop 벤치마크에서 수행한 실험 결과, 제안된 시스템이 ReAct, Reflexion, LATS, LAC 과 같은 기존 방법 대비 우수한 성능을 달성하였으며, 특히 장기적 추론이 요구되는 복잡한 테스크에서 그 효과를 확인할 수 있었다. 본 연구는 추가적인 모델 학습 없이도 기능적 기반의 역할 분리만으로 LLM의 추론 성능을 향상시킬 수 있음을 확인했다는 점에서 의의가 있다.

참 고 문 헌

[1] S. Yao, et al., "WebShop: Towards Scalable Real-World Web Interaction with Grounded Language Agents," NeurIPS, 2022.

[2] S. Yao, et al., "ReAct: Synergizing Reasoning and Acting in Language Models," ICLR, 2023.

[3] N. Shinn, et al., "Reflexion: Language Agents with Verbal Reinforcement Learning," NeurIPS, 2023.

[4] A. Zhou, et al., "Language Agent Tree Search Unifies Reasoning Acting and Planning in Language Models," ICML, 2024.

[5] H. Dong, et al., "Enhancing Decision-Making of Large Language Models via Actor-Critic," ICML, 2025