

블랙박스 최적화를 위한 협력적 이중 LLM 프레임워크에 관한 연구

권오승, 이진규, 이인규
고려대학교 전기전자공학과

worat@korea.ac.kr, jlee0070@korea.ac.kr, inkyu@korea.ac.kr

Cooperative Dual-LLM Framework for Black-box Optimization

Ohseung Kwon, Jinkyu Lee and Inkyu Lee

School of Electrical Engineering, Korea University

요약

본 논문에서는 블랙박스 (Black box) 환경에서의 문맥 내 예시 (In-context example)를 활용한 대규모 언어 모델 (Large Language Model, LLM) 기반 최적화 방법을 제안한다. 복잡한 수학적 모델을 사용하지 않고, 수집된 데이터와 사전 학습된 언어 모델을 결합하는 최적화 기법을 설계한다. 이중 LLM 에이전트를 사용하여 수집된 데이터를 통해 텍스트 기반 전략을 제시하고 이를 기반으로 새로운 행동을 탐색하는 최적화 구조를 제안한다. 실험결과를 통해 기존 LLM 최적화 기법 및 유전 알고리즘 대비 우수한 수렴 성능을 보여줌을 확인한다.

I. 서론

대규모 언어 모델 (Large Language Model, LLM)은 방대한 데이터를 기반으로 학습된 언어 모델로서, 자연어 이해뿐 아니라 추론 능력을 기반으로 다양한 분야에서 우수한 성능을 입증하고 있다. 최근 무선 통신 분야에서도 LLM 을 활용하여 복잡한 최적화 문제를 해결하려는 시도가 활발히 이루어지고 있다. 실제 무선 네트워크 환경은 간섭, 채널 불확실성 등으로 인해 정확한 수학적 모델링이 어려운 경우가 많으며, 시스템의 관측된 입력-출력데이터 만을 활용하여 해를 탐색하는 블랙박스 (Black-box) 최적화 방식이 요구된다. 이러한 환경에서 문맥 내 예시 (In-context example)을 통해 관측 정보로부터 해를 추론하는 LLM 기반 최적화 기법이 지속적으로 연구되고 있다 [1].

본 논문에서는 블랙박스 환경에서의 무선 네트워크 최적화를 위한 새로운 LLM Optimizer 구조를 제안한다. 제안 기법은 두 개의 에이전트가 각각 전략 제안과 행동 생성을 담당하도록 역할을 분리함으로써, 단일 에이전트 기반 최적화에서 발생할 수 있는 탐색 방향의 불안정성과 비효율을 완화하고자 한다. 특히, 전략 제안단계에서 탐색의 거시적 방향과 후보 해의 구조를 정제하고, 행동 단계에서 이를 바탕으로 구체적인 해를 생성하도록 설계하여 수렴 속도와 안정성을 동시에 향상시키는 것을 목표로 한다. 실험결과를 통해 제안 기법은 기존 LLM 기반 최적화 기법 대비 안정적인 수렴 특성을 보이며, 동일한 반복 횟수에서 안정적인 목적함수 값을 달성함을 확인하였다.

II. 시스템 모델

본 논문에서는 다음과 같이 정의된 최적화 문제를 푼다고 가정한다.

$$\max_{\mathbf{x}_{min} \leq \mathbf{x} \leq \mathbf{x}_{max}} f(\mathbf{x}) \quad (1)$$

여기서, \mathbf{x} 는 D 차원의 행동 벡터(Action Vector)이며, 각 요소는 $\mathbf{x} \in [\mathbf{x}_{min}, \mathbf{x}_{max}]^D$ 의 제약 조건을 가진다. $f(\cdot)$ 는 보상 함수이며, $f(\cdot)$ 를 최대화하는 최적해 \mathbf{x}^* 을 도출하고자 한다. 기존의 경사 하강법 (Gradient Descent)을 계산할 수 있는 환경이 아닌 보상 함수의

미분 가능성을 계산할 수 없는 블랙박스 환경을 가정한다.

III. LLM 기반 최적화 기법

본 절에서는 (1)의 최적해를 도출하기 위한 LLM 기반 최적화 기법을 제안한다. 제안하는 시스템은 그림 1 과 같이 전략을 제안하는 Strategy-LLM (S-LLM)과 행동을 생성하는 Action-LLM (A-LLM)의 독립적인 2 개의 LLM 에이전트와 메모리 베파, 샘플링 연산자, 프롬프트 생성기로 구성된다.

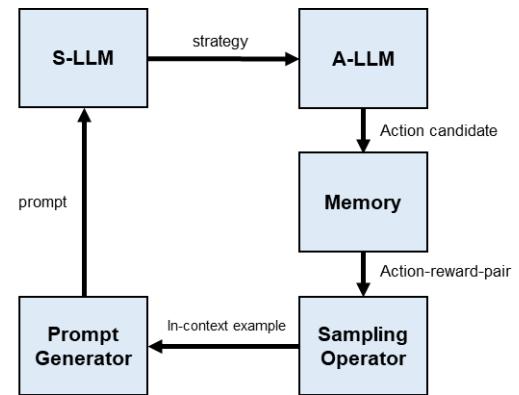


그림 1. 전략 추출 (S-LLM)과 행동 제안 (A-LLM)을 결합한 제안 최적화 구조

전체적인 최적화 과정은 상위 성능 샘플을 토대로 텍스트 기반 전략을 제안한 뒤, 해당 전략을 기반으로 새로운 행동 벡터를 생성하는 구조로 설계된다. 첫 번째 에이전트인 S-LLM 은 현재 모집단 (Population)에서 성능이 우수한 상위 k 개의 샘플을 입력으로 받아, 보상을 최대화하기 위한 핵심 규칙을 단일 문장 형태의 전략으로 요약 및 출력한다. 이후 A-LLM 은 S-LLM 이 생성한 전략과 탐색 범위를 규정하는 제약 조건을 활용하여, 다음 세대에서 평가될 후보 해 (Candidate solution)를 생성한다. 이때 A-LLM 은 수치 데이터로부터 직접 패턴을 학습하기보다 전략 기반 지침을 통해 탐색 방향을 유도함으로써 탐색 효율을 높인다. 생성된 후보 해는 목적함수 평가

를 거친 후, 모집단 내 성능이 가장 낮은 개체를 대체하는 방식으로 모집단이 갱신된다. 이러한 절차를 미리 설정된 세대 (generation) 수만큼 반복되며, 반복에 따라 모집단이 점진적으로 개선되어 최적해로 수렴한다. 샘플링 단계에서는 탐색 안정성을 위해, 이전 세대에서의 최우수 해를 보존하는 엘리티스트 (elitist) 샘플링을 고려한다.

IV. 실험결과

제안하는 알고리즘의 성능을 검증하기 위해, 무선 통신의 대표적인 최적화 문제 중 하나인 간섭 채널(interference channel)에서의 전력 제어(power control) 환경을 고려하였다. 본 실험에서는 다음과 같은 스펙트럼 효율(spectral efficiency, SE)의 합을 보상함수로 정의하였다.

$$f(\mathbf{x}) = \sum_{d=1}^D \log_2 \left(1 + \frac{P_{tx} |h_{dd}|^2 x_d}{P_{tx} \sum_{d' \neq d} |h_{d'd}|^2 x_{d'+1}} \right) \quad (2)$$

여기서, 행동 벡터는 $\mathbf{x} = [x_1, \dots, x_D]^T$ 로 정의되며 $x_d \in [0,1]$ ($d = 1, \dots, D$)는 송신기 d 의 전력 분배 비율을 나타낸다. $h_{d'd}$ 는 송신기 d' 로부터 수신기 d 로의 채널계수를 의미하며 Rayleigh fading을 따른다고 가정한다. $P_{tx} = 10\text{ W}$ 는 송신 전력을 나타내며, 분모의 상수항 1은 잡음 전력을 의미한다.

사용한 언어 모델은 Llama 3.2:3B [2]이며, 모집단 크기(population size), 세대별 생성 개체 수(offspring size), 그리고 ICL 예시 개수는 모두 4로 고정하였다. 또한 송수신 쌍의 수는 $D = 3$ 으로 설정하였다. 본 논문에서는 다음과 같은 기법들을 제안하는 기법과 비교하였다.

- LLMO-E [1]: 기존 ICL 기반 LLM 최적화 기법으로, 행동 생성(추론) 모델만을 사용하는 단일 에이전트 구조.
- 유전 알고리즘 (Genetic Algorithm, GA)[3]: 기존 진화 알고리즘 기반 최적화 기법.
- 무작위 탐색 (Random): 무작위로 행동 벡터 후보를 샘플링하는 기법
- 지역 최적해 (local optimal)[4]: weighted minimum mean-squared error (WMMSE) 기반의 전력 제어 최적 기법

그림 2는 반복 횟수에 따른 평균 스펙트럼 효율 성능을 나타낸다. 시뮬레이션 결과, 제안하는 기법은 약 20 회의 반복만에 Local optimal 성능에 근접하며, 비교 기법들 중 가장 우수한 수렴 속도와 안정성을 보였다. 반면, 유전 알고리즘은 초기 단계에서 빠른 성능 향상을 보이나, 최적해에 도달하지 못한 채 국소 해에 머무르는 조기 수렴 성능을 확인할 수 있다. 또한, 단일 에이전트 기반의 LLMO-E는 반복이 진행됨에 따라 점진적인 성능 개선을 보였으나, 제안 기법 대비 수렴 속도가 저조함을 확인하였다. 이러한 결과는 제안하는 이중 LLM 구조의 최적화과정을 전략 추출과 행동생성 단계로 구분함으로써 각 에이전트의 추론 부담을 분산시켰고, 블랙박스 환경에서도 효율적으로 최적해를 도출할 수 있음을 보여준다.

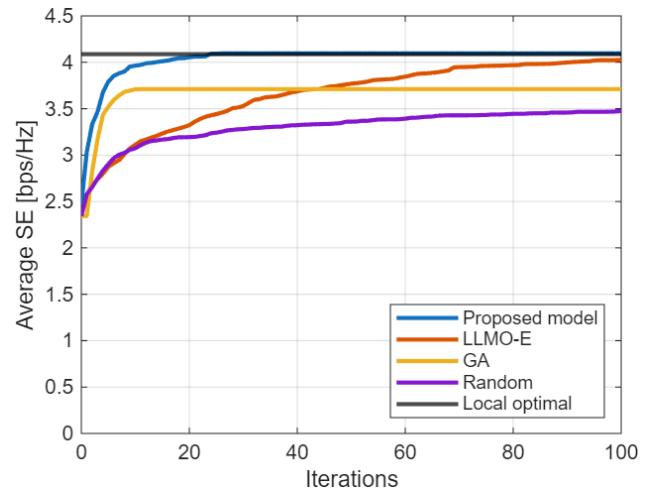


그림 2. Iteration에 따른 SE 성능

V. 결론

본 논문에서는 최적화 성능 극대화를 위해 전략 추출과 행동 생성을 분리한 새로운 이중 LLM 기법을 제안하였다. 제안하는 구조는 두 에이전트 간의 협력적 역할 분담을 통해 탐색과 활용의 균형을 효과적으로 조절함으로써, 기존 단일 모델 대비 우수한 수렴 속도와 최적화 성능을 입증하였다. 향후 연구에서는 본 프레임워크를 분산형 구조로 확장할 계획이다. 에이전트 간 방대한 수치 데이터를 직접 교환하는 대신, 텍스트 기반의 전략 정보만을 송수신함으로써 통신 오버헤드를 획기적으로 절감하는 고효율 제어 구조를 설계하고자 한다.

ACKNOWLEDGMENT

본 연구는 한국연구재단의 지원을 받아 수행된 연구임.
(RS-2022-NR070834)

참고 문헌

- [1] H. Lee, W. Zhou, M. Debbah, and I. Lee, "On the convergence of large language model optimizer for black-box network management," submitted to IEEE Trans. Commun., 2025.
- [2] H. Touvron et al., "LLaMA: Open and efficient foundation language models," Feb. 2023. [Online] Available: <https://arxiv.org/abs/2302.13971>.
- [3] J. H. Holland, Adaptation in Natural and Artificial Systems. Ann Arbor: The University of Michigan Press, 1975.
- [4] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," IEEE Trans. Signal Process., vol. 59, pp. 4331–4340, Sep. 2011.