

# GPT-1 트랜스포머 구조의 파라미터 재현 가능성에 관한 연구

1<sup>st</sup> 정 우철

AIFEL, MODULABS (10th Online Research Course)

Ulsan, Republic of Korea

jwc9202@naver.com

**Abstract**—본 연구에서는 GPT-1 모델의 트랜스포머 블록에 포함된 매개변수들이 훈련 과정에서 어떻게 재현될 수 있는지 분석한다. 이를 위해 동일한 조건에서 여러 번 실험을 진행하고, 각 실험에서 얻어진 매개변수 간의 차이를 비교했다. 실험 결과, 완전한 매개변수 재현은 어려운 것으로 나타났지만, 모델의 성능은 여러 번의 훈련에서도 일관되게 유지되었다.

**Index Terms**—GPT-1, 트랜스포머, 매개변수, 파라미터, 재현성

## I. 서론

최근 몇 년간, 자연어 처리(NLP) 분야에서는 대형 언어 모델들이 큰 관심을 끌고 있으며, 그 중에서 OpenAI의 GPT-1(Generative Pretrained Transformer-1)은 트랜스포머 아키텍처를 기반으로 하는 최초의 대형 언어 모델로서 큰 의미를 가진다. 트랜스포머 모델은 그 구조적 특성상, 병렬 처리와 효율적인 학습을 가능하게 하여, 자연어 처리 및 생성 분야에서 뛰어난 성과를 보여주었다. GPT-1 모델은 매우 큰 파라미터 수를 가지고 있으며, 이는 학습 시 모델의 가중치가 어떻게 초기화되고, 학습 데이터에 어떻게 영향을 받는지에 대한 중요한 논의를 불러일으켰다.

본 연구에서는 GPT-1 모델의 파라미터가 매 학습 시마다 재현 가능한지에 대해 살펴보고자 한다. 이는 모델의 훈련 결과가 동일한 조건 하에서 일관되게 재현될 수 있는지, 즉 모델의 학습 및 추론 과정에서 발생하는 '랜덤성'이 어떻게 영향을 미치는지를 분석하는 것이다.

## II. APPROACH

### A. GPT-1 모델 아키텍처

GPT-1은 트랜스포머 아키텍처를 기반으로 하며, 셀프 어텐션(self-attention) 메커니즘을 통해 단어 간의 관계를 포착한다. 초기의 트랜스포머 모델들은 주로 인코더와 디코더 구조로 나뉘었지만, 이를 바탕으로 한 GPT 모델은 디코더와 동일한 구조만을 사용하도록 설계되었다. GPT-1은 다음과 같은 주요 요소들로 구성된다:

- **입력 임베딩**: 각 단어는 고차원의 벡터로 매핑되어, 이를 토대로 입력 문장이 임베딩된다.
- **트랜스포머 블록**: 여러 개의 트랜스포머 블록이 층을 이루며, 각 블록은 자기-어텐션 메커니즘과 피드포워드 신경망으로 구성된다.

- **언어 모델링 헤드**: 최종 출력은 단어의 확률 분포를 예측하는 데 사용된다.

GPT-1 모델은 총 117M(1억 1700만) 개의 파라미터를 가지고 있으며, 이러한 규모의 파라미터 집합은 학습 시 모델의 초기화 방법, 난수 생성 및 하이퍼파라미터 설정에 따라 학습 후 도달하는 값이 영향을 받는다.

### B. 파라미터의 재현 가능성

모델 훈련 시, 파라미터의 초기화와 학습 과정에서의 랜덤성은 모델 성능과 결과의 일관성에 영향을 미칠 수 있다. 이러한 랜덤성은 다음과 같은 여러 요소에서 발생한다:

- 1) **파라미터 초기화**: 대부분의 신경망 모델은 가중치를 랜덤하게 초기화한다. GPT-1은 Glorot 초기화 또는 He 초기화 기법을 사용하며, 이 과정에서 난수 생성기가 중요한 역할을 한다.
- 2) **미니배치 샘플링**: 학습 과정에서 사용하는 미니배치는 매번 랜덤하게 샘플링된다. 이는 각 학습 단계에서 모델이 처리하는 데이터 샘플이 달라지므로 결과에 영향을 미친다.
- 3) **학습률 스케줄링**: 학습률은 훈련이 진행됨에 따라 점차적으로 감소하는 방식으로 설정된다. 학습률 스케줄링이 정확하게 설정되지 않으면, 훈련 결과는 불안정해질 수 있다.
- 4) **옵티마이저의 동작**: GPT-1은 Adam 옵티마이저를 사용하며, 이 또한 미니배치 및 그라디언트 계산에서 랜덤성이 존재한다. 옵티마이저의 동작은 훈련의 안정성 및 수렴에 중요한 영향을 미친다.
- 5) **하드웨어 및 소프트웨어 환경**: 학습을 실행하는 하드웨어 및 소프트웨어 환경, 특히 GPU/TPU 사용, 병렬 처리 방식, 랜덤 초기화 및 시드 설정 방식 등이 훈련의 재현성에 영향을 미칠 수 있다.

## III. METHODS

### A. 파라미터 재현성 보장 방법

모델 훈련의 재현성을 보장하기 위해서는 여러 가지 방법을 사용할 수 있다.

첫째, 난수 생성기 시드를 고정하는 것이 중요하다. GPT-1 훈련 시, 파라미터 초기화와 데이터 샘플링에서 사용되는

난수 생성기의 시드를 고정하여 같은 초기 조건을 설정할 수 있다.

둘째, 학습 환경의 일관성을 유지하는 것이다. 동일한 하드웨어와 소프트웨어 환경에서 훈련을 실행하는 것이 중요하다. 하드웨어의 아키텍처나 소프트웨어의 버전 차이가 결과에 영향을 미칠 수 있다. 이를 해결하기 위해서는 동일한 라이브러리 버전과 설정을 사용하는 것이 바람직하다.

셋째, 모델 훈련의 세부 설정을 엄격히 정의하는 것이다. 학습률, 배치 크기, 옵티마이저 등의 하이퍼파라미터를 동일하게 설정하고, 동일한 학습 스케줄을 적용해야만 훈련 결과가 재현될 수 있다.

실험을 통해 GPT-1 모델의 재현 가능성을 분석하기 위해, 동일한 학습 환경에서 여러 번 모델을 훈련하고 결과를 비교하였다. 실험에서는 다음과 같은 절차를 따랐다:

- 1) 동일한 데이터셋을 사용하여 모델을 훈련하였다.
- 2) 학습 시 난수 생성기의 시드를 고정하였다.
- 3) 동일한 하드웨어 환경에서 모델을 훈련하였다.
- 4) 모델 훈련 후, 테스트 세트에 대한 성능을 평가하였다.

#### IV. RESULTS

실험 결과, 동일한 설정 하에서 훈련된 모델들은 일정 부분 유사한 성능을 보였으나, 약간의 차이가 존재하였다. 이는 주로 모델의 학습 과정에서 발생하는 수치적인 차이 때문이며, 완벽하게 동일한 결과를 얻는 것은 어려웠다. 특히, 모델의 초기화 및 옵티마이저의 미세한 차이가 훈련 결과에 영향을 미쳤다.

#### V. CONCLUSIONS

GPT-1 모델의 파라미터는 다양한 요소로 인해 완벽하게 재현되는 것이 어려운 것으로 나타났다. 파라미터 초기화, 미니배치 샘플링, 옵티마이저 동작, 하드웨어 및 소프트웨어 환경 등이 모두 훈련 결과에 영향을 미친다. 그러나, 동일한 조건에서 실험을 반복하면 일정 부분 일관된 결과를 얻을 수 있으며, 재현성을 높이기 위해서는 난수 시드 고정, 하드웨어 및 소프트웨어 환경의 일관성 유지, 하이퍼파라미터 설정의 정확한 정의 등이 중요하다.

향후 연구에서는 모델의 재현 가능성을 높이기 위한 기술적 접근 및 방법론에 대해 더 많은 논의가 필요하며, 이를 통해 대형 언어 모델의 훈련 과정에서 발생할 수 있는 랜덤성을 보다 효과적으로 제어할 수 있을 것이다.

#### ACKNOWLEDGMENT

이 연구는 아이펠 온라인 10기 리서치 교육과정의 일환으로 진행되었으며, 본문의 내용은 OpenAI 사의 ChatGPT 4o-mini의 도움을 받아 작성되었다.

#### REFERENCES