



경남ICT 협회(AI연구회)

ChatGPT가 처음 5일 동안 100만 명의 사용자를 넘어 인터넷을 휩쓸었기 때문에 어떤 기계 학습 알고리즘이 내부에서 실행되고 있는지 궁금할 것입니다.

ChatGPT는 "RLHF(Reinforcement Learning from Human Feedback)"라는 특정 유형의 강화 학습을 사용하지만 높은 수준에서 LLM(Large Language Model)의 예입니다.

이 가이드에서는 주요 용어, 사용 사례 등을 포함하여 대규모 언어 모델에 대해 알아야 할 모든 것을 설명합니다.

- LLM(대형 언어 모델)이란 무엇입니까?
- LLM에 대해 알고 있는 7가지 주요 용어
- LLM에서 사용되는 주요 알고리즘
- 대규모 언어 모델 미세 조정
- 신속한 엔지니어링 기술 이해
- LLM의 한계

LLM(대형 언어 모델)이란 무엇입니까?

대규모 언어 모델은 대화 또는 기타 자연 언어 입력에 대해 인간과 유사한 응답을 생성하기 위해 방대한 양의 텍스트 데이터(입기: ChatGPT의 경우 전체 인터넷)에 대해 훈련된 인공지능의 하위 집합입니다.

이러한 자연어 응답을 생성하기 위해 LLM은 다층 신경망을 사용하여 복잡한 데이터를 처리, 분석 및 예측하는 심층 학습 모델을 사용합니다.

- LLM은 종종 인간의 텍스트와 구별할 수 없는 고품질의 일관된 텍스트를 생성하는 능력이 독특합니다.

이 최첨단 성능은 인간 언어의 뉘앙스를 학습할 수 있도록 일반적으로 최소 수십억 단어에 달하는 방대한 텍스트 코퍼스에서 LLM 을 교육함으로써 달성됩니다.

언급했듯이 가장 잘 알려진 LLM 중 하나는 OpenAI 에서 개발한 Generative Pretrained Transformer 3의 약자인 GPT-3입니다.

1,750 억 개의 매개변수가 있는 GPT-3는 현재까지 가장 크고 강력한 LLM 중 하나이며 번역, 요약 및 시 쓰기를 포함한 광범위한 자연어 작업을 처리할 수 있습니다.

ChatGPT: GPT 에서 1,750 억 개의 매개변수란 무엇입니까?

- GPT 는 "Generative Pre-trained Transformer"의 약자로, 인공지능 언어 모델 중 하나입니다. 매개변수란 모델을 구성하는 가중치(weight)와 편향(bias) 등의 파라미터(parameter)를 의미합니다. GPT 모델은 Transformer 구조를 사용하여 구성되며, 이 구조는 매개변수의 크기가 매우 큼니다.
- 2021년 6월 기준, GPT-3 모델의 매개변수는 175billion 개였습니다. 이는 이전 모델인 GPT-2 모델의 매개변수인 1.5billion 개보다 약 116배 큼니다. 이 많은 매개변수를 가진 모델은 매우 복잡한 언어 처리 작업을 수행할 수 있으며, 대규모 데이터셋을 이용한 사전학습(pre-training)을 통해 일반적인 자연어 이해 능력을 가지고 있습니다.
- GPT-3 모델의 1750 억 개의 매개변수를 이해하는 한 가지 예시는, 이 매개변수 수가 인간 뇌의 뉴런 수와 비슷하다는 것입니다. 인간 뇌의 뉴런 수는 약 1000 억 개에서 100 조 개 정도로 추정됩니다. 따라서, GPT-3 모델의 매개변수 수는 인간 뇌의 뉴런 수와 비슷하거나 그보다 큰 것으로 추정됩니다.
- 또 다른 예시로는, GPT-3 모델의 매개변수 수가 인터넷에서 사용되는 모든 웹사이트의 데이터 크기보다도 크다는 것입니다. 이는 매우 큰 규모의 데이터셋을 사용하여 훈련된 모델이라는 것을 보여줍니다.
- 이러한 큰 규모의 매개변수를 가진 모델은 매우 정교한 자연어 처리 작업을 수행할 수 있으며, 다양한 응용 분야에서 사용될 수 있습니다. (이상 ChatGPT Answer)

ChatGPT 는 GPT-3의 확장이며 블로그 게시물에서 다음과 같이 강조합니다.

- ChatGPT 는 2022년 초에 교육을 마친 GPT-3.5 시리즈의 모델에서 미세 조정되었습니다. 여기에서 3.5 시리즈에 대해 자세히 알아볼 수 있습니다.

LLM에 대해 알고 있는 7가지 주요 용어

LLM에 대해 자세히 알아보기 전에 먼저 다음을 포함하여 알아야 할 몇 가지 주요 용어를 검토해 보겠습니다.

- 단어 임베딩: 단어의 의미를 숫자 형식으로 표현하여 AI 모델에 입력하고 처리할 수 있도록 LLM에서 사용되는 알고리즘입니다.
- 주의 메커니즘: AI가 출력을 생성할 때 텍스트의 감정 관련 단어와 같은 입력 텍스트의 특정 부분에 집중할 수 있도록 하는 LLM에서 사용되는 알고리즘입니다.
- 트랜스포머: 입력 데이터를 처리하기 위해 셀프 어텐션 메커니즘을 사용하는 LLM 연구에서 널리 사용되는 신경망 아키텍처 유형입니다.
- Fine-tuning LLMs: LLM을 더 작고 관련 있는 데이터 세트에서 교육하여 특정 작업 또는 도메인에 맞게 조정하는 프로세스입니다.
- Prompt engineering: LLM이 고품질의 일관된 출력을 생성하기 위한 입력 프롬프트의 숙련된 설계입니다.
- 편향(Bias): 교육 데이터 세트에 체계적이고 불공평한 선호도 또는 편견이 존재하며 LLM에서 학습하여 차별적인 결과를 얻을 수 있습니다.
- 해석 가능성: AI 시스템의 결과와 결정을 이해하고 설명할 수 있는 능력으로, 복잡성으로 인해 LLM에 대한 도전이자 지속적인 연구 영역입니다.

(Bias? 통계학과 머신러닝에서 사용되는 용어로, 모델에서 예측하는 값과 실제 값 사이의 차이를 나타내는 편향을 의미합니다. 머신러닝에서는 모델의 목적 함수를 최소화하는 데에 사용되는데, 모델이 복잡하고 데이터셋이 작은 경우, 모델은 특정한 경향성을 가질 수 있습니다. 이러한 경향성이 모델의 예측에 영향을 미치면, 모델이 편향되었다고 말합니다.)

예를 들어, 성별에 따른 연봉을 예측하는 모델에서, 모델이 훈련된 데이터셋이 남성의 연봉이 더 높은 경우, 모델은 여성의 연봉을 과소평가할 가능성이 높습니다. 이러한 경우 모델은 성별에 따라 편향되었다고 말할 수 있습니다.

편향을 감소시키기 위해서는 모델의 입력 데이터를 다양한 방식으로 변형하거나, 모델 자체의 구조를 변경해야 할 수 있습니다. 또한 데이터셋을 더 다양하고 균형있게 구성하여 편향을 줄일 수도 있습니다.

편향은 머신러닝에서 중요한 개념 중 하나이며, 모델의 성능을 개선하는 데에 중요한 역할을 합니다. 따라서 모델을 훈련하거나 평가할 때, 편향을 고려하는 것이 중요합니다.)

LLM에서 사용되는 주요 알고리즘

자연어 처리, 특히 LLM(대형 언어 모델) 분야는 이러한 AI 모델이 가능한 한 인간 언어에 가깝게 처리, 이해 및 출력할 수 있도록 하는 다양한 알고리즘에 의해 구동됩니다.

단어 임베딩, 어텐션 메커니즘 및 변환기를 포함하여 위에서 언급한 LLM에서 사용되는 몇 가지 주요 알고리즘을 좀 더 자세히 간략하게 살펴보겠습니다.

Word Embedding

단어 임베딩은 단어의 의미를 숫자 형식으로 표현하는 데 사용되므로 LLM에서 사용되는 기본 알고리즘으로, AI 모델에서 처리할 수 있습니다. 이는 유사한 의미를 가진 단어가 서로 더 가깝게 위치하는 고차원 공간에서 단어를 벡터에 매핑하여 달성됩니다.

ChatGPT: Word Embedding?

- Word Embedding은 자연어 처리(Natural Language Processing) 분야에서 단어를 벡터로 표현하는 기법입니다. 이를 통해 컴퓨터가 단어를 이해하고 처리할 수 있게 됩니다.
- 기존에는 단어를 One-hot Encoding 방식으로 벡터화 했습니다. 이 방식은 단어의 등장 여부를 나타내는 0과 1로 이루어진 벡터를 만드는 것입니다. 예를 들어, 5,000개의 단어를 다루는 문제에서 "cat"은 $[0, 0, 0, \dots, 1, 0, 0, \dots, 0]$ 와 같은 벡터로 표현됩니다. 그러나 이 방식은 단어의 의미나 문맥을 고려하지 않기 때문에 한계가 있습니다.
- Word Embedding은 이러한 문제를 해결하기 위해 등장한 기법입니다. Word Embedding은 각 단어를 고정된 크기의 실수 벡터로 변환하는 것입니다. 예를 들어, 100차원의 벡터로 변환한다면 "cat"은 $[0.23, 0.65, -0.12, \dots, 0.34, -0.78, 0.91, \dots]$ 와 같은 벡터로 표현됩니다. 이 방식은 단어의 의미나 문맥을 고려할 수 있도록 하여, 단어 간 유사도를 계산하거나, 문서 분류, 감정 분석 등 다양한 자연어 처리 작업에서 활용될 수 있습니다.
- Word Embedding을 학습하는 방법으로는 CBOW(Continuous Bag-of-Words)와 Skip-gram 등이 있으며, 이러한 방법들은 대표적인 Word Embedding 알고리즘인 Word2Vec에 적용됩니다. 또한, Word Embedding은 딥러닝 모델인 RNN, CNN, Transformer 등과 함께 사용되어 자연어 처리의 성능을 높이는 데에도 활용됩니다. (이상 ChatGPT Answer)

Attention Mechanisms

어텐션 메커니즘은 LLM의 또 다른 중요한 알고리즘으로, 시가 출력을 생성할 때 입력 텍스트의 특정 부분에 집중할 수 있도록 합니다. 이를 통해 LLM은 주어진 입력의 맥락이나 감정을 고려하여 보다 일관되고 정확한 응답을 얻을 수 있습니다.

ChatGPT: Attention Mechanisms?

- Attention Mechanism은 딥러닝 모델에서 주로 사용되는 메커니즘으로, 입력 데이터 중에서 특정 부분에 더 집중하여 처리하는 기법입니다. 특히 자연어 처리 분야에서 RNN, Transformer 등과 함께 많이 사용됩니다.
- Attention Mechanism은 입력 데이터의 각 부분(예: 입력 문장의 각 단어)에 대한 가중치를 계산하여, 더 중요한 부분에 더 큰 가중치를 부여하는 방식으로 동작합니다. 이를 통해 모델은 입력 데이터의 중요한 부분에 더 집중하여 처리하며, 결과적으로 더 정확한 예측을 할 수 있습니다.
- 예를 들어, 번역기 모델에서는 입력 문장과 출력 문장 간의 관련성을 파악하여 번역을 수행합니다. Attention Mechanism은 이 때 입력 문장의 각 단어에 대한 가중치를 계산하여, 번역할 때 더 중요한 단어에 더 집중합니다. 이를 통해 모델은 문맥에 따라 적절한 번역 결과를 생성할 수 있습니다.
- Attention Mechanism은 다양한 방식으로 구현될 수 있으며, 대표적인 방법으로는 Bahdanau Attention과 Dot Product Attention 등이 있습니다. Bahdanau Attention은 입력 데이터의 모든 부분에 대한 가중치를 계산하는 방식이며, Dot Product Attention은 입력 데이터 간의 내적(Dot Product)을 계산하여 가중치를 부여하는 방식입니다.
- Attention Mechanism은 자연어 처리 분야 뿐만 아니라, 이미지 처리 분야에서도 활용되며, 이미지에서 주요 부분에 더 집중하여 처리하는 방식으로 적용됩니다. (이상 ChatGPT Answer)

Transformers

트랜스포머는 LLM 연구에서 널리 사용되는 신경망 아키텍처 유형입니다. 이러한 네트워크는 self-attention 메커니즘을 사용하여 입력 데이터를 처리하여 인간 언어의 장기적인 종속성을 효과적으로 캡처할 수 있습니다.

이러한 알고리즘은 LLM이 자연어 입력을 처리 및 이해하고 가능한 한 인간과 유사한 출력을 생성할 수 있도록 하므로 LLM의 성능에 매우 중요합니다.

ChatGPT: Transformers ?

- Transformers 는 딥러닝 모델 중 하나로, 주로 자연어 처리 분야에서 사용되는 모델입니다. 기존의 RNN, LSTM 등의 모델에서 발생하는 장기 의존성 문제를 해결하고, 대규모 데이터를 다루는 데에 효과적입니다.
- Transformers 는 Attention Mechanism 을 기반으로 하며, 입력 시퀀스와 출력 시퀀스 간의 관계를 모델링하는 데에 사용됩니다. 이를 위해 Self-Attention 이라는 메커니즘을 사용하여 입력 시퀀스의 모든 단어 간의 관계를 파악합니다. 이후, Encoder 와 Decoder 라는 두 개의 모듈을 사용하여 입력 시퀀스와 출력 시퀀스 간의 관계를 모델링합니다.
- Encoder 는 입력 시퀀스를 입력 받아, Self-Attention 과 Feed-Forward Network 를 사용하여 입력 시퀀스를 처리합니다. 이후, Decoder 는 Encoder 에서 출력된 정보와 이전 시점의 출력 결과를 사용하여 출력 시퀀스를 생성합니다. Decoder 에서는 다시 Self-Attention 과 Encoder-Decoder Attention 을 사용하여 입력 시퀀스와 출력 시퀀스 간의 관계를 모델링합니다.
- Transformer 모델은 이러한 방식으로 대규모 데이터를 다룰 수 있으며, 번역, 요약, 챗봇 등 다양한 자연어 처리 작업에서 높은 성능을 보입니다. 특히, 최근의 자연어 처리 분야에서는 Transformer 를 기반으로 한 모델들이 SOTA(State of the Art)를 경신하고 있습니다. 대표적인 Transformer 모델로는 BERT, GPT 등이 있습니다. (이상 ChatGPT Answer)

대규모 언어 모델 미세 조정

대규모 언어 모델 미세 조정은 특정 작업 또는 도메인에 대한 범용 모델을 조정하는 처리를 의미합니다.

이는 예를 들어 시가 특정 도메인의 패턴과 뉘앙스를 학습할 수 있도록 일련의 프롬프트와 이상적인 응답을 제공하여 당면한 작업과 관련된 더 작은 데이터 세트에서 LLM 을 교육함으로써 달성됩니다.

예를 들어 미세 조정된 LLM 은 다음에 대해 학습할 수 있습니다.

- 의료 진단을 지원하는 의료 기록 데이터 세트
- 법률 자문을 제공하는 법률 문서 데이터 세트입니다.
- SEC 파일링 또는 애널리스트 보고서와 같은 재무 데이터 세트

이 맞춤형 접근 방식은 종종 ChatGPT 와 같은 범용 LLM 을 사용하는 것과 비교하여 특정 작업에서 우수한 성능을 제공합니다.

OpenAI는 GPT-3 미세 조정 문서에서 다음과 같이 작성합니다.

- GPT-3는 개방형 인터넷의 방대한 양의 텍스트에 대해 사전 훈련되었습니다. 몇 가지 예와 함께 프롬프트가 제공되면 종종 수행하려는 작업을 직관적으로 파악하고 그럴듯한 완료를 생성할 수 있습니다. 이를 종종 "퓨샷 학습"이라고 합니다.
- 미세 조정은 프롬프트에 맞출 수 있는 것보다 더 많은 예를 훈련하여 소수 학습을 개선하여 다양한 작업에서 더 나은 결과를 얻을 수 있도록 합니다. 모델이 미세 조정되면 더 이상 프롬프트에 예제를 제공할 필요가 없습니다.

LLM을 미세 조정하면 원래 교육 데이터에 존재할 수 있는 편향을 줄이는 데 도움이 될 수 있습니다. 특히, 보다 집중된 데이터 세트를 사용함으로써 LLM은 다양한 입력 세트에 대해 훈련될 수 있으므로 차별적 출력의 가능성을 줄일 수 있습니다.

즉, LLM 미세 조정에는 몇 가지 제한 사항이 있다는 점에 유의하는 것이 중요합니다. 예를 들어, AI 모델의 성능은 미세 조정 데이터 세트의 품질과 크기에 의해 제한될 수 있습니다. 또한 LLM을 미세 조정하는 것은 종종 상당한 도메인 전문 지식이 필요한 모든 프롬프트와 응답을 준비해야 하므로 시간이 많이 걸리고 자원 집약적인 프로세스가 될 수 있습니다.

이러한 문제에도 불구하고 미세 조정된 LLM은 특정 작업 및 영역에 대해 향상된 성능과 감소된 편향을 제공하므로 자연 언어 처리 분야에서 중요한 발전입니다.

성능을 개선하기 위해 LLMs를 미세 조정하는 것 외에도 LLM의 세계에 더 깊이 들어가면 고품질 출력을 생성하는 핵심 요소가 "신속한 엔지니어링 기술"임을 알게 될 것입니다.

신속한 엔지니어링 기술 이해

신속한 엔지니어링 기술은 고품질의 일관된 출력을 생성하기 위한 LLM(Large Language Models)에 대한 입력의 숙련된 설계를 말합니다. 입력 프롬프트의 품질이 생성된 텍스트의 품질에 큰 영향을 미칠 수 있으므로 이는 LLM 작업의 중요한 측면입니다.

신속한 엔지니어링에는 LLM에 대한 입력을 신중하게 제작하여 특정 방향으로 응답을 유도하는 작업이 포함됩니다. 여기에는 AI 시스템이 텍스트를 생성하도록 특정 주제나 컨텍스트를 제공하거나 출력에 통합할 특정 단어나 구를 제공하는 것이 포함될 수 있습니다.

효과적인 신속한 엔지니어링을 위해서는 LLM의 기능과 한계에 대한 깊은 이해와 매력적인 입력을 만드는 방법에 대한 예술적 감각이 필요합니다. 또한 프롬프트를 조금만 변경해도 출력이 크게 변경될 수 있으므로 세부 사항에 대한 예리한 눈이 필요합니다.

프롬프트 엔지니어링의 주요 측면 중 하나는 LLM이 일관된 텍스트를 생성할 수 있도록 충분한 컨텍스트를 제공하는 것입니다. 여기에는 모델이 컨텍스트를 이해하고 관련 응답을 생성하는 데 도움이 되는 특정 방식으로 배경 정보를 제공하거나 입력을 프레이밍하는 작업이 포함될 수 있습니다.

다음은 LLM이 지속적으로 개선됨에 따라 프롬프트 엔지니어링이 어떻게 사라질 수 있는지(또는 적어도 변경될 수 있는지)에 대한 흥미로운 스레드입니다.

LLM의 한계

LLM(대형 언어 모델)은 의심할 여지없이 흥미로운 인공지능 분야이지만 이러한 알고리즘에는 이해하고 고려하기에 좋은 몇 가지 주요 제한 사항이 있습니다.

- LLM의 주요 한계 중 하나는 편향에 대한 민감성입니다.

논의한 바와 같이 LLM은 막대한 양의 텍스트 데이터에 대해 교육을 받습니다. Twitter를 스크롤하면서 알 수 있듯이 이 입력 데이터에는 데이터에 존재하는 상당한 양의 편향이 포함될 수 있습니다. 입력 데이터의 편향의 결과는 AI의 차별적인 출력이며 기존의 사회적 불평등을 강화합니다.

- LLM의 또 다른 한계는 해석 능력이 부족하다는 것입니다.

LLM은 매우 복잡한 알고리즘이며 일반적으로 딥 러닝은 종종 "블랙 박스"라고 불리기 때문에 모델이 특정 출력에 어떻게 그리고 왜 도달했는지 정확히 알기 어렵습니다.

이로 인해 LLM의 결과를 신뢰하기 어렵고 고 부담 의사 결정 시나리오에서의 사용에 대한 의문이 제기될 수 있습니다.

마지막으로 LLM을 교육하고 실행하는 데 필요한 순전히 크기와 계산 능력이 상당한 제한이 될 수 있습니다. LLM 및 더 광범위하게는 딥 러닝에는 엄청난 양의 데이터와 계산 리소스가 필요하므로 개발 및 유지 관리 비용이 상당히 많이 들고 환경에 해를 끼칠 가능성은 말할 것도 없습니다.

전반적으로 LLM은 AI에서 인상적인 발전이지만 고려해야 할 중요한 제한 사항도 있습니다. LLM이 일상 생활에서 점점 더 중요한 역할을 함에 따라 연구원과 개발자는 이러한 한계를 해결하여 잠재력을 최대한 발휘해야 합니다.

요약 : 대규모 언어 모델(LLM)

논의한 바와 같이 LLM(Large Language Models)은 자연어 입력에 대해 인간과 유사한 응답을 생성하기 위해 대규모 텍스트 데이터 모음에서 훈련된 인공 지능 유형입니다.

- LLM에 대해 알아야 할 주요 용어: 워드 임베딩, 어텐션 메커니즘, 변환기, 미세 조정, 신속한 엔지니어링, 편향, 해석 가능성
- 주요 알고리즘에는 워드 임베딩, 어텐션 메커니즘, 변환기가 포함됩니다.
- LLM 미세 조정: LLM을 더 작고 관련 있는 데이터 세트에서 교육하여 특정 작업 또는 도메인에 맞게 조정하는 것을 말합니다.
- 신속한 엔지니어링: LLM이 고품질의 일관된 출력을 생성하기 위한 숙련된 입력 설계입니다.
- 편향: 이는 LLM에서 학습할 수 있고 차별적인 결과를 초래할 수 있는 데이터 세트의 체계적이고 불공평한 선호 또는 편견의 존재를 나타냅니다.
- 해석 가능성: AI 시스템의 결과와 결정을 이해하고 설명할 수 있는 능력으로, 복잡성으로 인해 LLM이 해결해야 할 과제입니다.

결론적으로 LLM(Large Language Models)은 인공 지능 분야에서 흥미로운 발전이며 ChatGPT가 입소문이 나면서 앞으로 몇 년 동안 일상 생활에서의 사용이 증가할 것 같습니다. 해결해야 할 문제와 윤리적 고려 사항이 분명히 있지만 LLM의 잠재적 용도는 방대하고 다양합니다.

LLM이 계속 발전하고 발전함에 따라 광범위한 산업 및 응용 분야에서 점점 더 중요한 역할을 할 가능성이 높습니다. 따라서 자연어 애호가이든 AI에만 관심이 있든 관계없이 LLM은 눈여겨볼 가치가 있습니다. 그리고 특정분야에 ChatGPT와 같은 자신의 AI 모델을 구축하는 것 또한 가능할 것입니다([myGPT Plan](#)은 다음편을 기대해주세요)

출처: <https://www.mlq.ai/what-is-a-large-language-model-llm/>

(번역 및 추가 및 보완 설명: 박종영)