1장

출처: **허정준.** *LLM을 활용한 실전 AI 애플리케이션 개발*. 책만, 2024.

1. LLM의 기초 뼈대 세우기

1.1 LLM이란?

- **LLM(Large Language Model)** 은 대규모 텍스트 데이터를 학습하여 자연어를 이해하고 생성하는 모델
- 텍스트의 패턴을 분석하고, 다음에 올 단어를 예측하는 방식으로 문장을 생성하는 구조

1.2 주요 발전 과정

- **2013년**: 구글이 Word2Vec 모델을 발표하여 단어 간의 의미를 벡터로 표현하는 방법을 제시
- **2017년**: 트랜스포머(Transformer) 아키텍처의 등장으로 자연어 처리 분야에 혁신 발생
- **2018년**: OpenAI의 GPT(Generative Pre-trained Transformer) 모델 등장, 대화형 AI와 자연어 생성의 중요한 기반이 된 사건

---

2. 딥 러닝과 언어 모델링

2.1 딥 러닝의 정의

- **딥 러닝(Deep Learning)** 은 인간의 뇌 구조를 모방한 신경망을 기반으로 데이터의 패턴을 학습하는 인공지능 기법
- 수백만 개의 뉴런과 계층적인 구조를 통해 대량의 데이터를 처리하고 비정형 데이터를 이해하는 방식

2.2 언어 모델의 작동 원리

- 언어 모델은 주어진 텍스트의 다음 단어를 예측하여 문장을 완성하거나 새로운 문장을 생성하는 기술
- 대표적인 언어 모델:
  - **GPT**: 입력된 텍스트의 패턴을 학습하고 다음 단어를 예측하여 문장을 생성하는 방식
  - **BERT**: 양방향 텍스트를 이해하고 문맥에 맞는 단어를 예측하는 모델

---

3. 전이 학습과 사전 학습

3.1 임베딩(Embedding) 정의

- **임베딩(Embedding)** 은 단어를 고정된 차원의 벡터로 변환하여 컴퓨터가 이해할 수 있는 형식으로 만드는 기술
- 단어 간의 유사도를 벡터 공간에서 나타내는 방식으로, 자연어 처리에서 중요한 역할을 수행

3.2 Word2Vec과 임베딩

- **Word2Vec** 모델은 단어를 벡터로 변환하여 의미적으로 유사한 단어들이 벡터 공간에서 가까운 거리를 갖게 함
- 이를 통해 단어 간의 의미 관계를 벡터 연산으로 다룰 수 있게 됨 (예: "왕 - 남자 + 여자 = 여왕")

3.3 전이 학습(Transfer Learning)

- **전이 학습**은 기존에 학습된 모델을 바탕으로 새로운 과제를 해결하는 방식
- **사전 학습(Pre-training)** 된 모델을 특정한 작업에 맞게 **미세 조정(Fine-tuning)** 하여 사용하는 기법

3.4 사전 학습의 중요성

- **사전 학습**된 모델은 대량의 데이터를 바탕으로 언어의 일반적인 구조와 패턴을 학습하는 단계
- 이후 **미세 조정**을 통해 새로운 데이터에 맞게 학습하여 다양한 자연어 처리 작업에 적용 가능

---

4. 트랜스포머 아키텍처

4.1 트랜스포머의 등장

- **트랜스포머(Transformer)** 는 2017년 구글에서 발표한 자연어 처리 모델로, 이전의 순차적 처리 방식(RNN, LSTM)을 대체
- **병렬 처리**가 가능해 속도와 성능 면에서 기존 모델들을 크게 뛰어넘는 성능을 발휘한 모델

4.2 어텐션 메커니즘(Attention Mechanism)

- **어텐션(Attention)** 은 모든 입력 단어들이 문맥을 고려하여 중요도를 계산하는 방식
- 이를 통해 모델은 문맥을 이해하고 중요한 단어에 더 집중하는 특징

4.3 트랜스포머의 구조

- **인코더-디코더 구조**로 이루어져 있으며, 각 층은 **자기 어텐션(Self-Attention)** 과 **피드 포워드 네트워크**로 구성
- 트랜스포머는 모든 입력을 동시에 처리하기 때문에 **병렬 처리**가 가능

---

5. LLM의 활용과 한계

5.1 LLM의 활용 사례

- **대화형 AI**: 자연어를 이해하고 대화하는 챗봇, 고객 서비스, 가상 비서 등에 활용
- **자동 번역**: LLM은 텍스트를 이해하고 다른 언어로 번역하는 데 뛰어난 성능을 보이는 모델
- **콘텐츠 생성**: 블로그 글, 기사, 소설 등 다양한 콘텐츠 생성에 활용

5.2 LLM의 한계

- **메모리 사용량**: 트랜스포머 모델은 대규모 연산과 메모리 사용을 요구, 특히 긴 문장을 처리할 때 성능 저하 발생
- **데이터 편향성**: 학습된 데이터에 편향이 있을 경우 생성되는 텍스트에도 편향이 반영될 수 있는 문제점
- **연산 비용**: 대규모 데이터를 처리하고 학습하는 데 높은 연산 비용이 소요되는 한계

---

6. 미래의 언어 모델과 트렌드

6.1 새로운 아키텍처 연구

- 트랜스포머의 성능을 넘어서는 새로운 아키텍처 연구가 활발하게 진행 중
- **효율성**과 **성능**을 모두 갖춘 새로운 모델 개발을 위한 노력이 이어지고 있음

6.2 LLM의 확장 가능성

- **멀티모달 모델**: 텍스트 외에도 이미지, 오디오 등 비정형 데이터를 동시에 처리하는 모델의 개발이 진행 중
- **응용 분야 확대**: 교육, 의료, 법률 등 다양한 산업에서 LLM을 활용한 혁신이 기대됨