

2. 멀티모달 AI 생성 모델(주요 생성형 AI 모델)

2.1. 트랜스포머(Transformer): 멀티모달 AI의 언어 천재, 문맥을 파악하여 맥락 있는 결과를 생성하다

① 왜 나오게 되었는가?

트랜스포머는 인공지능이 언어를 더 잘 이해하고 생성할 수 있도록 하기 위해 등장했다. 이전의 자연어 처리 모델은 RNN(순환 신경망, Recurrent Neural Network)이라는 방식을 사용했는데, 이는 마치 책을 한 문장씩 차례대로 읽어나가는 것과 비슷했다. 이 방식은 긴 문장을 이해하는 데 어려움이 있었고, 처리 속도도 느렸다. 예를 들어, “나는 어제 친구와 함께 영화를 보았다. 그것은 정말 재미있었다.”라는 문장에서 “그것”이 무엇을 가리키는지 파악하는 데 어려움이 있었다. 트랜스포머는 이러한 문제를 해결하고, 더 나아가 텍스트뿐만 아니라 이미지, 음성 등 다양한 형태의 데이터를 함께 처리할 수 있는 멀티모달 AI의 기반을 발전 및 고도화시킨다.

② 쉽게 설명하면?

트랜스포머는 문장 전체를 한 번에 이해하는 인공지능 모델이다. 이는 마치 사람이 문장을 읽을 때 전체적인 맥락을 파악하는 것과 비슷하다. 예를 들어, “배가 바다를 향해한다”와 “배가 고프다”라는 두 문장에서 ‘배’라는 단어의 의미가 다르다는 것을 문맥을 통해 이해하는 것과 같다. 트랜스포머는 이러한 방식으로 단어들 사이의 관계를 파악하고, 문장의 의미를 정확하게 이해한다. 또한, 여러 문장을 동시에 처리할 수 있어 매우 빠르게 작동한다.

③ 예시와 함께 좀 더 자세히 알아볼까?

트랜스포머의 핵심은 ‘어텐션 메커니즘(Attention Mechanism)’이라는 기술이다. 이는 마치 사람이 문장을 읽을 때 중요한 부분에 집중하는 것과 같다. 예를 들어, “나는 파란 하늘을 보며 행복을 느꼈다”라는 문장에서 ‘행복’이라는 감정은 ‘파란 하늘’과 연관이 있다. 트랜스포머는 이러한 관계를 파악하여 문장의 의미를 이해한다.

실생활에서의 예를 들어보자. 당신이 외국어 학습 앱을 사용하고 있다고 가정해보자. 이 앱이 트랜스포머 기술을 사용한다면, “I love eating apples”라는 문장을 번역할 때 단순히 단어 하나하나를 번역하는 것이 아니라, 문장 전체의 맥락을 고려하여 “나는 사과 먹는 것을 좋아한다”라고 자연스럽게 번역할 수 있다.

④ 왜 중요한가?

트랜스포머의 중요성은 크게 두 가지로 나눌 수 있다. 첫째, 언어 이해와 생성 능력을 크게 향상시켰다. 이는 번역, 요약, 질문 답변 등 다양한 언어 관련 작업의 성능을 획기적으로 개선한다. 둘째, 다양한 형태의 데이터를 함께 처리할 수 있는 기반을 마련한다. 이는 텍스트뿐만 아니라 이미지, 음성 등을 함께 이해하고 생성할 수 있는 멀티모달 AI의 발전을 이끈다.

예를 들어, 의료 분야에서 트랜스포머 기반 AI는 환자의 증상 설명(텍스트), X-ray 이미지, 과거 진료 기록 등을 종합적으로 분석하여 더 정확한 진단을 내리는 데 도움을 줄 수 있다. 이처럼 트랜스포머는 AI가 보다 인간적이고 종합적인 판단을 할 수 있게 해주는 중요한 기술이다.

⑤ 어디에 활용되는가?

트랜스포머는 우리 일상 곳곳에서 활용되고 있다. 가장 대표적인 예는 번역 서비스이다. 구글 번역¹¹⁾과 같은 번역 앱들은 트랜스포머 기술을 활용하여 더 자연스러운 번역을 제공한다. 또한, 스마트폰의 음성 비서나 챗봇도 트랜스포머 기술을 기반으로 작동한다. 이들은 사용자의 질문을 정확히 이해하고 적절한 답변을 생성한다.

더 나아가, 최근에는 달-이(DALL-E)나 미드저니(Midjourney)와 같은 이미지 생성 AI에도 트랜스포머 기술이 활용되고 있다. 이들은 텍스트 설명을 기반으로 이미지를 생성하는데, 예를 들어 “해변에서 서핑하는 강아지”라는 설명만으로도 그에 맞는 이미지를 만들어낼 수 있다.

교육 분야에서도 트랜스포머 기술이 활용되고 있다. 개인화된 학습 콘텐츠를 추천하거나, 학생의 답변을 자동으로 채점하고 피드백을 제공하는 등 다양한 방식으로 활용되고 있다.

이처럼 트랜스포머는 우리 생활 전반에 걸쳐 AI 기술의 성능을 높이고, 더욱 지능적이고 자연스러운 서비스를 제공하는 데 기여하고 있다.

11) Google Translate Architecture illustrated (VivienLa, 2024)

2.2. BERT (Bidirectional Encoder Representations from Transformers): 맥락을 이해하는 언어 모델, 멀티모달 AI의 기반을 다지다

① 왜 나오게 되었는가?

BERT는 언어의 복잡성과 맥락을 더 잘 이해하기 위해 개발되었다. 기존의 자연어 처리 모델은 마치 외국어를 단어장으로만 공부하는 학생처럼, 단어의 순서나 개별적인 의미에만 집중했다. 하지만 실제 언어는 훨씬 복잡하다. 예를 들어, “배”라는 단어는 문맥에 따라 ‘과일’일 수도 있고, ‘운송 수단’일 수도 있으며, ‘신체 부위’를 가리킬 수도 있다. 이렇게 문맥에 따라 달라지는 단어의 의미를 정확히 파악하기 위해 BERT가 등장했다. BERT는 ‘Bidirectional Encoder Representations from Transformers’의 약자로, ‘양방향 인코더 표현을 위한 트랜스포머’라고 해석할 수 있다.

② 쉽게 설명하면?

BERT는 마치 뛰어난 독해 능력을 가진 학생처럼 문장 전체의 맥락을 고려하여 각 단어의 의미를 이해한다. 이 모델은 문장을 앞에서부터 읽는 것뿐만 아니라, 뒤에서부터도 읽어 전후 맥락을 모두 파악한다. 이는 마치 추리 소설을 읽을 때 앞뒤 문맥을 모두 고려하여 범인을 추리하는 것과 비슷하다. BERT는 이러한 방식으로 대량의 텍스트 데이터를 학습하여, 다양한 언어 관련 작업에서 뛰어난 성능을 보인다.

③ 예시와 함께 좀 더 자세히 알아볼까?

BERT의 작동 방식을 일상적인 예로 설명해보자. 당신이 친구에게 “나 어제 배 먹었어”라는 문자를 받았다고 가정해보자. 여기서 ‘배’가 과일인지, 아니면 배를 타고 어딘가를 다녀왔다는 의미인지는 문맥을 알아야 파악할 수 있다. BERT는 이전 대화 내용, 계절, 친구의 평소 습관 등을 모두 고려하여 ‘배’의 의미를 정확히 파악할 수 있다.

또 다른 예로, “The chef made the salad with fresh greens and crisp apples(그 요리는는 신선한 콩과 아삭아삭한 사과로 샐러드를 만들었다.)”라는 영어 문장에서 BERT는 ‘crisp(아삭아삭한)’이라는 단어가 ‘apples(사과)’를 수식한다는 것을 이해한다. 이는 BERT가 문장 구조와 단어 간의 관계를 정확히 파악할 수 있음을 보여준다.

④ 왜 중요한가?

BERT의 중요성은 크게 두 가지로 나눌 수 있다. 첫째, BERT는 언어의 복잡성을 더욱 정확하게 이해할 수 있게 해준다. 이는 기계 번역, 질문 답변 시스템, 감정 분석 등 다양한 언어 관련 기술의 성능을 크게 향상시켰다. 예를 들어, BERT를 활용한 번역 시스템은 “I’m blue”라는 문장을 단순히 “나는 파랗다”로 번역하지 않고, 문맥에 따라 “나는 우울하다”로 정확하게 번역할 수 있다.

둘째, BERT는 다른 AI 기술과 결합하여 더욱 강력한 멀티모달 AI의 기반을 제공한다. 예를 들어, 이미지와 텍스트를 함께 이해하는 AI 모델에서 BERT는 텍스트 부분을 담당하여 이미지와 텍스트 간의 관계를 더욱 정확하게 파악할 수 있게 한다.

⑤ 어디에 활용되는가?

BERT는 우리 일상 곳곳에서 활용되고 있다. 가장 대표적인 예는 검색 엔진이다. BERT를 활용한 검색 엔진은 사용자의 검색 의도를 더 정확히 파악하여 관련성 높은 결과를 제공한다. 예를 들어, “갤럭시 s20 방수”라고 검색했을 때, BERT는 이것이 갤럭시 S20의 방수 기능에 대한 질문임을 이해하고 관련 정보를 우선적으로 보여준다.

고객 서비스 분야에서도 BERT는 중요한 역할을 한다. BERT를 활용한 챗봇은 고객의 문의 내용을 더 정확히 이해하고, 적절한 답변을 제공할 수 있다. 예를 들어, “환불이 언제 되나요?”라는 질문에 대해 BERT는 이전 대화 내용을 고려하여 해당 고객의 구체적인 주문 정보와 환불 정책을 연결지어 답변할 수 있다.

2.3. GAN Model (Generative Adversarial Networks): 창과 방패의 대결, 경쟁 속에서 진화하는 AI

① 왜 나오게 되었는가?

GAN 모델은 보다 현실적이고 다양한 데이터를 만들어내기 위해 등장했다. 이전의 AI 모델들은 마치 복사기처럼 주어진 데이터를 그대로 따라 만들기만 했다. 예를 들어, 고양이 사진을 만들려고 하면 기존 고양이 사진들을 평균 내어 흐릿하고 어색한 이미지를 만들어냈다. 이는 마치 여러 사람의 얼굴을 합성해 평균적인 얼굴을 만들면 누구의 얼굴과도 닮지 않은 모호한 이미지가 나오는 것과 비슷하다. GAN은 이러한 한계를 극복하고 더욱 생생하고 다양한 데이터를 만들어내기 위해 개발되었다.

② 쉽게 설명하면?

GAN은 ‘적대적 생성 신경망’이라는 뜻으로, 서로 경쟁하며 발전하는 두 개의 AI 시스템으로 구성된다. 하나는 ‘생성자’로 가짜 데이터를 만들어내고, 다른 하나는 ‘판별자’로 이 데이터가 진짜인지 가짜인지 구별한다. 이는 마치 화폐 위조범(생성자)과 경찰(판별자)의 끊임없는 숨바꼭질과 같다. 위조범은 점점 더 정교한 위조 지폐를 만들고, 경찰은 더욱 세밀하게 위조 지폐를 감별하는 능력을 키운다. 이 과정에서 두 시스템 모두 계속해서 발전하게 되고, 결과적으로 매우 현실적인 가짜 데이터를 만들어낼 수 있게 된다.

③ 예시와 함께 좀 더 자세히 알아볼까?

실생활에서 GAN의 원리를 이해할 수 있는 재미있는 예를 들어보자. 여러분이 부모님께 거짓말을 하지 않고 늦게 들어오는 방법을 찾으려 한다고 가정해보자. 여러분(생성자)은 계속해서 새로운 핑계를 만들어내고, 부

모님(판별자)은 이 핑계가 진실인지 거짓인지 판단하려 노력한다. 처음에는 “학교에서 청소당번이었어요”라는 단순한 핑계로 시작했지만, 부모님이 이를 의심하자 “선생님께서 급한 일이 생겨서 제가 대신 청소를 했어요”라는 식으로 점점 더 정교한 핑계를 만들어낸다. 이 과정에서 여러분의 핑계는 점점 더 그럴듯해지고, 부모님의 의심도 점점 더 섬세해진다.

GAN도 이와 비슷하게 작동한다. 예를 들어, 고양이 이미지를 생성하는 GAN에서 생성자는 처음에는 어설픈 고양이 그림을 만들어낸다. 하지만 판별자가 이를 가짜로 판별하면, 생성자는 더 세밀한 특징(예: 고양이의 수염, 귀 모양 등)을 추가하며 점점 더 실제 같은 고양이 이미지를 만들어낸다. 이 과정을 통해 GAN은 결국 실제 고양이 사진과 구별하기 어려운 수준의 이미지를 생성할 수 있게 된다.

④ 왜 중요한가?

GAN의 중요성은 그것이 가진 무한한 창조성에 있다. 이는 마치 인간의 상상력을 AI에 불어넣은 것과 같다. GAN은 단순히 주어진 데이터를 복제하는 것이 아니라, 완전히 새로운 것을 만들어낼 수 있다. 예를 들어, 패션 디자인 분야에서 GAN은 기존의 디자인을 학습한 후 전혀 새로운 옷 디자인을 제안할 수 있다. 이는 디자이너들에게 새로운 영감을 줄 수 있다.

또한, GAN은 데이터 부족 문제를 해결하는 데 큰 도움이 된다. 예를 들어, 의료 분야에서 희귀 질병의 X-ray 영상이 부족할 때, GAN을 사용하여 추가적인 가상의 X-ray 영상을 생성할 수 있다. 이렇게 생성된 데이터는 의료 AI의 학습 데이터로 활용되어 진단의 정확도를 높이는 데 기여할 수 있다.

⑤ 어디에 활용되는가?

GAN은 우리 일상 곳곳에서 활용되고 있다. 가장 대표적인 예는 스노우와 같은 스마트폰의 카메라 앱이다. 요즘 스마트폰으로 찍은 사진이 점점 더 좋아지는 이유 중 하나는 GAN 기술 덕분이다. GAN은 저화질의 사진을 고화질로 변환하거나, 어두운 곳에서 찍은 사진을 밝게 만드는 데 사용된다.

또 다른 예로, 온라인 쇼핑몰에서 가상 피팅 서비스를 들 수 있다. GAN을 활용하면 실제로 옷을 입어보지 않고도 자신의 사진 위에 원하는 옷을 입혀볼 수 있다. 이는 소비자들이 더 쉽고 재미있게 쇼핑을 할 수 있게 해준다.

엔터테인먼트 분야에서도 GAN의 활용도가 높다. 예를 들어, 영화나 게임에서 CG로 만든 캐릭터나 배경을 더욱 현실감 있게 만드는 데 GAN이 사용된다. 심지어 돌아가신 배우의 모습을 GAN으로 재현하여 영화에 출연시키는 것도 가능해졌다.