초거대 인공지능 기술 동향 및 개발 경쟁

1부_초거대 인공지능 정의와 해외 기술 동향

알서포트 클라우드기술팀 **박상은** 팀장

목차

서론 **2**

1. 초거대 인공지능(Hyper-scale AI)의 정의 2

2. 초거대 인공지능의 특징 2

초거대 인공지능 기술 해외 동향 **3**

1. 인류 역사상 가장 뛰어난 인공지능 'GPT-3' 3

2. 구글의 차세대 인공지능 대화모델 '람다(LaMDA)' 4

3. 딥마인드의 초거대 인공지능 언어모델 '고퍼(Gopher)' 5

4. MS-엔비디아의 초거대 AI 'MT-LNG' 6

5. 중국 인공지능 아카데미의 초거대 AI 사전학습 모델 '우다오 2.0' 7

6. 화웨이의 대규모 자연어처리(NLP) 모델 '판구 알파' 8

결론: 초거대 AI에 대한 의문 **9**

※ 참고문헌 **10**

© Rsupport Co. Ltd. All Rights Reserved.

서론

최근 정보통신기술(ICT) 분야의 가장 핫한 키워드는 '초거대 인공지능(Hyper-scale AI)'이다. 2020년 6월 오픈AI(OpenAI)가 발표한 GPT-3라는 자연어처리(NLP) 모델을 기반으로 획기적이고 창의적인 활용 사례들이 나오면서 글로벌 ICT 기업들의 관심이 증폭되었고, 최근에는 초거대 AI 모델이 속속 등장하고 있다.

국내외 초거대 AI(Hyper-scale AI) 기술 동향을 1부와 2부에 걸쳐 상세히 살펴보고자 한다. 우선, 1부에서는 초거대 AI의 정의와 특징은 무엇이며, 글로벌 ICT 기업들이 발표한 초거대 AI의 기술동향 및 개발경쟁 현황에 대해 살펴본다.

1부_초거대 인공지능 정의와 해외 기술 동향
2부_ 국내 초거대 AI 기술 동향 및 국내외 초거대 AI 기술의 한계와 과제

1. 초거대 인공지능(Hyper-scale AI)의 정의

'초거대 AI'는 기존 인공지능(AI)에서 한 단계 진화한 차세대 AI로, 대용량 데이터를 학습해 인간과 같이 종합적 추론이 가능하다. 기존 AI보다 더 인간의 뇌에 가깝게 학습 및 판단 능력이 향상된 형태다. 예컨대, 알파고는 바둑 분야에 특화돼 있지만 초거대 AI는 여러 상황에 대해 스스로 학습해 역할을 수행한다. 이를 위해 기존 AI보다 수백 배 이상의 데이터 학습량이 필요하다[1].
종합하면, 초거대 AI란 데이터 분석과 학습을 넘어 인간의 뇌처럼 스스로 추론하고 창작할 수 있도록 방대한 데이터와 파라미터(매개변수)를 활용하는 AI 모델을 말한다. 이때 인공신경망의 파라미터는 인간 뇌의 뉴런 간 정보전달 통로인 시냅스와 비슷한 역할을 수행하게 된다.

2. 초거대 인공지능의 특징

2020년 이전까지만 하더라도 AI 모델들의 파라미터 수는 100억개 내외였다. 그러나 현재 초거대 인공지능 모델은 최소 1,000억 개 이상의 파라미터와 수천억 개로 구성된 데이터셋으로 학습한다. 이 초거대 AI 모델은 구글이 개발한 트랜스포머 알고리즘을 활용하고 있다. 병렬적 언어처리를 가능케해 효율성을 획기적으로 향상시킴으로써 대규모 데이터셋 학습에 용이하기 때문이다. 2021년 이후 등장한 최근의 초거대 AI 모델은 그 규모가 더욱 거대해져 수 천억~1조개 이상의 파라미터를 장착하고 있다[2].
GPT-3를 필두로 한 초거대 AI 모델은 이러한 대규모 파라미터와 데이터셋의 방대한 일반 데이터 학습만으로도 대부분의 작업(task)을 완성도 높게 수행한다. 이렇게 학습된 초거대 AI 모델은 블로그나 뉴스 등의 컨텐츠를 인간을 대신하여 작성하고, 인간과 AI 간의 자유대화를 가능케 한다. 이 외에도 이력서, 가계부, 이메일 등 원하는 양식의 컨텐츠 생성이 가능하고, 상당한 수준의 번역작업, 몇 줄의 일반 문장입력을 통한 자동 프로그래밍 작업 등 기존에는 상상할 수 없었던 인간 수준의 추론과 종합적 사고능력을 갖춘 AI 모델로 급속도로 발전했다.

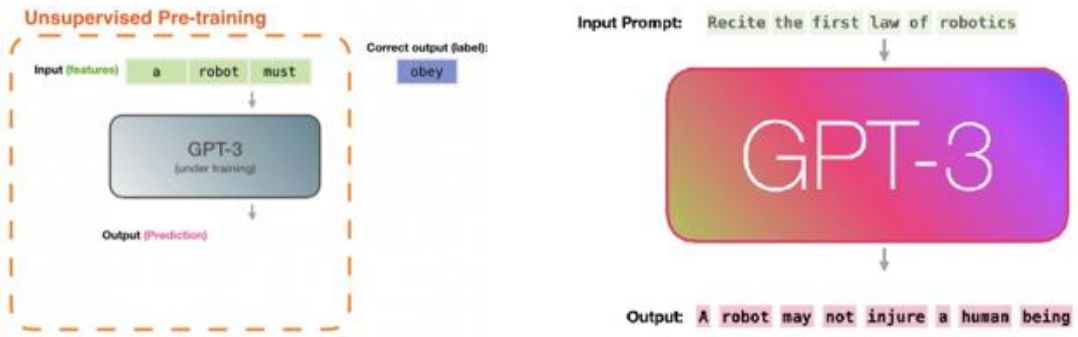
© Rsupport Co. Ltd. All Rights Reserved.

AI의 발전은 우리의 예상보다 빠른 속도로 발전하고 있고, 초거대 AI의 등장은 빠른 속도의 발전과 글로벌 ICT 기업들의 개발 경쟁에 불을 붙이고 있다.

초거대 인공지능 기술 해외 동향

1. 인류 역사상 가장 뛰어난 인공지능 'GPT-3'

최근 속속 등장하고 있는 글로벌 ICT 기업들의 초거대 AI 모델은 2020년 6월 오픈AI에 의해 발표된 'GPT-3'라는 자연어처리(NLP) 모델로부터 촉발된 것이라 해도 과언이 아니다.
GPT-3(Generative Pre-Training 3) 모델은 4,990억개 데이터셋 중 가중치 샘플링된 3,000억개로 구성된 데이터셋과 1,750억개 매개 변수로 딥러닝(Deep-learning)의 한계까지 사전 학습된 초거대 AI 모델로, 미세 조정(fine-tuning) 없이 여러 NLP 벤치마크에서 최첨단 성능을 발휘했다.
오픈AI가 발표한 내용에 따르면, 단 몇 개 키워드만 넣으면 작문을 작성해주는 혁신적인 AI 언어생성 모델이자 알고리즘으로[3], '인류 역사상 가장 뛰어난 인공지능'이라는 찬사를 받는 초거대 AI 모델의 선구자적 지위를 보유하고 있다. 또한, AGI(Artificial General Intelligence, 인간 수준의 범용적 인공지능)를 최초로 구현해줄 모델로서도 주목받고 있다.
GPT-3의 활용 사례는 상식 Q&A, 검색 엔진, 이력서 작성, 자유 대화, 텍스트 데이터 파싱, 전문적인 의학 지식, 프로그래밍, 문장을 낚시성 기사 제목 스타일로 바꾸기 등 다양하고 광범위한 분야에서 멋진 성능을 발휘한다. 여기서 우리는 "GPT-3는 대체 어떤 신박한 학습을 했기에 저렇게 똑똑한 일을 할 수 있을까?" 라는 궁금증이 생긴다.
그러나 GPT-3의 학습 방법은 놀라울 정도로 미개(?)하다. 주어진 텍스트에 대해 그저 '다음에 올 단어를 예측'하는 방식으로 학습을 한 것이다. 예를 들어, "현대 자동차가 처음으로 자체 생산한 차의 이름은?" 이라는 문장을 GPT-3에 넣으면 GPT-3는 다음에 나올 확률이 가장 높은 단어를 판단해서 "현대 포니"라는 말을 생성한다. GPT-3는 이와 같이 끊임없이 다음 단어를 예측하며 정답을 맞힐 수 있는 방향으로 모델을 업데이트해 나간다. 이런 방식으로 인터넷에 있는 문서와 책, 위키피디아 등 엄청난 양의 텍스트를 1,750억 개의 파라미터에 달하는 엄청나게 큰 모델로 학습시킨 게 바로 GPT-3인 것이다[3].



[그림 1] GPT-3의 학습 과정과 GPT-3의 다음 단어 예측 과정
(*출처: Jay Alammar, Visualizing machine learning one concept at a time, https://jalammar.github.io/)

© Rsupport Co. Ltd. All Rights Reserved.

최근 3년 남짓한 기간은 'NLP(자연어처리)의 황금기'라고 불릴 만큼 NLP에서 많은 발전이 있었던 시기였다. 얼마 전까지만 해도 머신러닝 업계에서 AGI라는 단어는 일종의 금기어였다. AGI는 영화 같은 상상일 뿐이고 현실에서 가능한 컨셉은 아니었다. 하지만, GPT-3가 할 수 있는 일을 목도한 지금, GPT-3에서는 약간의 AGI 냄새가 나기 시작했다. 물론, GPT-3는 AGI가 아니고 AGI의 도래가 임박한 것도 아니지만, GPT-3의 성능과 발전 가능성을 확인한 만큼 이제는 AGI가 절대 불가능한 컨셉이라고 말할 수는 없게 되었다.

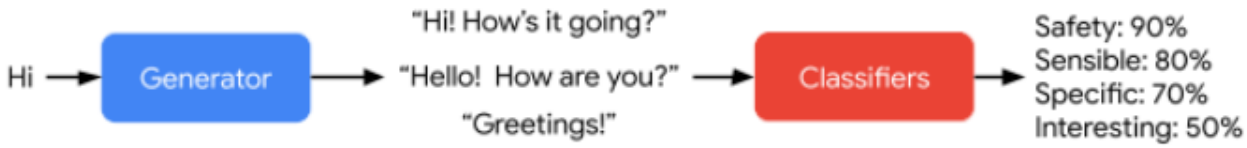
2. 구글의 차세대 인공지능 대화모델 '람다(LaMDA)'

구글 I/O 2021에 구글 CEO와 함께 자신이 명왕성인처럼 대화하는 챗봇 시스템이 등장하였다. "사람들이 날 그저 하찮은 얼음 덩어리가 아니라 하나의 아름다운 행성으로 알아주었으면 해." 그 이름은 바로 '람다(LaMDA: Language Model for Dialogue Applications, 대화를 위한 언어 모델)'이다.
람다는 언어 이해에 중점을 둔 BERT와 언어 생성에 중점을 둔 GPT를 합친 대화를 중점으로 만든 언어 모델이다.
2022년 1월 논문을 통해 자세한 내용이 공개된 람다의 데모는 왜 그토록 많은 사람들의 감탄을 자아냈을까? 그것은 바로 대형 언어 모델로 챗봇을 만들기가 매우 어렵기 때문인데, 안전성(Safety)과 팩트 정확성(Factual Accurateness)에 문제가 있기 때문이다.
잘못 디자인된 챗봇 시스템이 많은 유저들과 사회에 부정적인 영향을 끼칠 수 있기에 안전성은 무척 민감한 문제이다. 더불어 다양하고 흥미로운 답변을 생성하면서도 팩트에 맞는 답변을 생성하는 것이 어렵고, 마지막으로 일관된 롤 플레잉, 즉 페르소나(persona)를 가지게 하는 것도 쉽지 않다. 이렇다 보니 람다가 보여준 명왕성에 빙의한 챗봇 데모가 엄청난 파장을 불러 일으킨 것이다[4].
람다는 GPT와 거의 동일하게 거대한 트랜스포머이며, 약 1,370억 개의 파라미터로 구성되어 있다. 인터넷 등을 통해 공개된 약 30억 개의 문서, 11억 개의 대화를 학습 데이터로 사용하였고, 이 과정을 사전학습(Pre-training)이라고 한다. 또한, GPT와 같이 텍스트가 주어졌을 때 다음 단어를 예측하는 방식으로 학습된다.

[그림 2] 대화모델 람다의 사전학습(LaMDA pre-training as a language model)

다른 언어 모델과 다른 람다의 가장 큰 차별점 중 하나는 바로 답변 생성 후에 여러 측정방식(Metric)에 대해 아래와 같이 점수를 예측한다는 점이다.

© Rsupport Co. Ltd. All Rights Reserved.



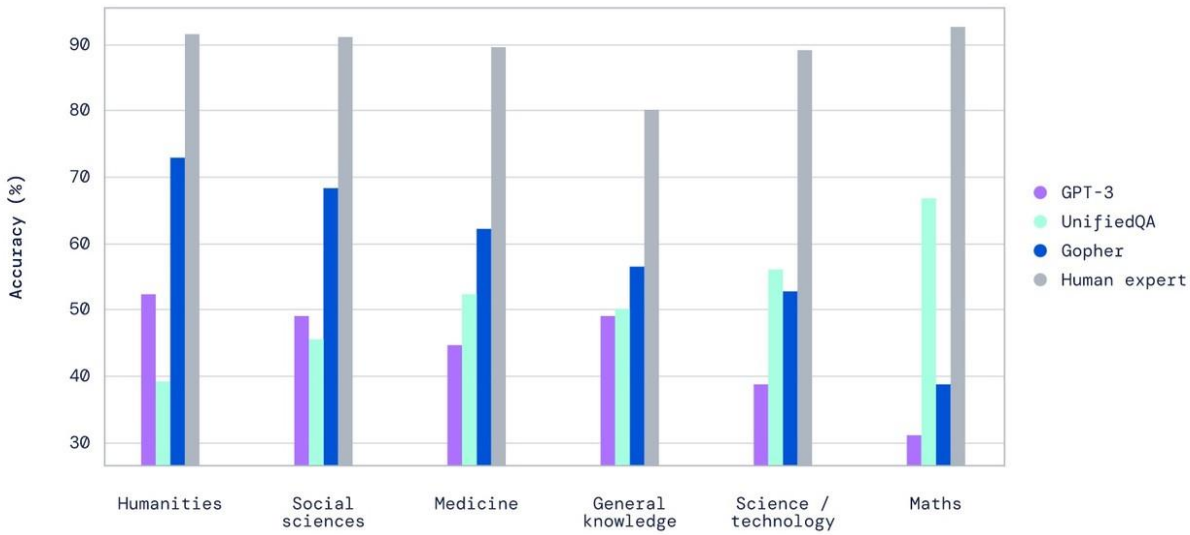
<자료> LaMDA: Towards Safe, Grounded, and High-Quality Dialog Models for Everything, Google AI Blog

[그림 3] 람다: 답변 생성 후, Metric에 대한 점수 예측 예시

람다는 생성된 답변 중, 먼저 안전성이 낮은 답변을 제거한다. 그리고, 구체적이고 합리적이고 흥미로운 (SSI:Specific Sensible Interesting) 답변을 골라 답변을 한다. 이런 식의 점수 기반 답변 생성 모델은 그저 많은 양의 데이터에 학습된 크기가 큰 모델에 비해 좋은 성능을 보여주었다. 특히, 구체성, 합리성, 안전성에서는 인간 레벨에 근접한 큰 향상을 이루어냈다[4].

3. 딥마인드의 초거대 인공지능 언어모델 '고퍼(Gopher)'

알파고로 알려진 딥마인드도 자체 개발한 초거대 AI 언어모델 '고퍼(Gopher)'를 공개하면서, 초거대 인공지능 (AI) 언어모델 개발 경쟁에 처음 합류했다. 고퍼의 성능 증명 논문과 함께, 검색 언어모델 (Retrieval Language Model)과 거대 언어모델의 윤리 문제를 각기 다룬 논문 2개를 발표했고, 단순히 모델 크기를 늘리는 여타 기업과 달리 AI 학습 효율을 높여 개발 비용을 줄이는 전략을 택했다[5].
고퍼는 2,880억개 매개변수로 구성되었고, OpenAI의 GPT-3(1,750억개) 보다는 크고 마이크로소프트의 Megatron-Turing(5,300억개)보다는 작지만 성능은 기존 모델들을 능가한다는 것이 딥마인드의 설명이다.



[그림 4] 인류학, 사회과학, 의학, 일반 상식, 과학/기술, 수학 분야의 고퍼와 GPT-3, 인간 전문가 능력 비교 (*출처=딥마인드 공식 블로그)

또한 딥마인드는 고퍼와 함께 검색 언어모델 일종인 레트로(RETRO, Retrieval-Enhanced Transformer)를 소개

© Rsupport Co. Ltd. All Rights Reserved.

했는데, 이를 적용할 경우 기존 모델과의 성능 격차는 더욱 벌어진다. 레트로는 문장 생성 시 2조 가량의 방대한 구문 텍스트로 이뤄진 외부 메모리를 활용하는데, 텍스트를 생성할 때 외부 데이터베이스에서 비슷한 구절을 찾아 비교하는 방식이다. 이를 이용하면 70억개의 매개변수로 이뤄진 레트로 모델이 매개변수가 1,750억개인 GPT-3와 동일한 성능을 낼 수 있다. 25배 정도 큰 모델과 같은 결과를 제시할 수 있다는 의미다[5]. 레트로와 같은 검색 언어모델은 초거대 AI 개발 효용을 극대화하고, AI 윤리에도 유용하다는 것이 딥마인드의 설명이다.

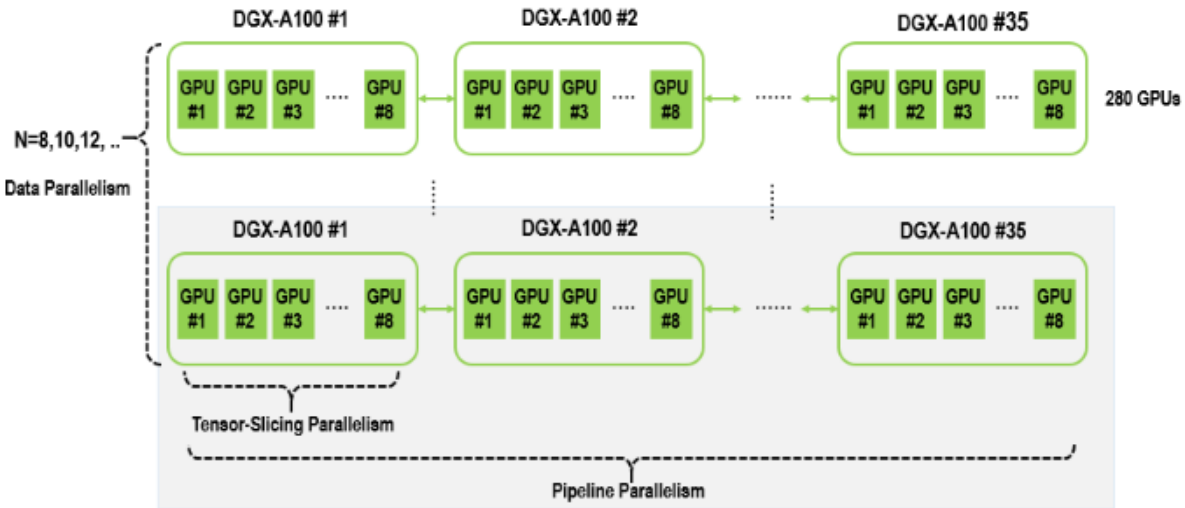
4. MS-엔비디아의 초거대 AI 'MT-LNG'

컴퓨팅 기술 강자 마이크로소프트(이하 MS)와 엔디비아(NVIDIA)가 협력해 기존 GPT-3(175B)보다 3배 더 큰 530B 파라미터의 MT-NLG (Megatron-Turing Natural Language Generation Model)를 발표했다. MS에 따르면, MT-NLG는 현재 같은 유형 모델 중 규모와 정확도 모두에서 최고 수준을 보인다. 해당 모델이 높은 정확도를 보이는 5개 분야는 완료 예측(completion prediction), 독해(reading comprehension), 상식 추론(commonsense reasoning), 자연어 추론(natural language inference), 단어 의미 중의성 해소(word sense disambiguation)다.

MS는 공식 블로그를 통해 "최근 초거대 모델 매개변수 처리는 최고 수준의 GPU 메모리에서도 더이상 불가능한 수준에 이르렀다"고 전했다. 이와 같은 거대한 AI 모델의 훈련 효율화를 위해서는 엔비디아 셀린(Selene)과 MS 에저(Azure) NDv4와 같은 슈퍼컴퓨팅 클러스터와 Megatron-LM과 DeepSpeed라는 획기적인 소프트웨어를 접목해야만 가능했다고 부연했다.

MT-LNG의 데이터셋은 총 339B 토큰(Token)으로 구성된 15개의 데이터셋을 사용해 MT-LNG를 학습했고, 데이터셋의 용량은 총 1.5TB에 해당한다. 이러한 대규모(Large-scale) 학습을 위해서는 크게 하드웨어 시스템과 소프트웨어 설계의 접목을 통해 해결하였는데, 그 구성은 아래와 같다.

1) 하드웨어 : NVIDIA Selene, MS-Azure NDv4와 같은 슈퍼컴퓨팅 클러스터
2) 소프트웨어 : 530B의 MT-LNG를 학습시키기 위해 NVIDIA Megatron-LM과 MS DeepSpeed를 조합하여 data/pipeline/tensor-slicing parallelism 구현



[그림 5] MT-LNG: data, pipeline, tensor-slicing parallelism 구조

© Rsupport Co. Ltd. All Rights Reserved.

MT-LNG와 같은 초대규모 AI 모델의 발전이 무어의 법칙을 훨씬 능가하면서 더 많은 연산능력을 필요로 하고 있다. 현재 구글의 TPU를 제외하고 학습에서 사용할 수 있는 하드웨어 리소스는 GPU가 유일하다. 이러한 한계에서 MT-LNG의 Large-scale 학습방법은 초대규모 AI 모델 학습에 다음과 같은 시사점을 준다[6].

1) 하드웨어 리소스 효율화

Megatron-LM과 DeepSpeed 조합과 같은 소프트웨어 프레임을 이용해 GPU의 대역폭(throughput)을 향상시키고 이기종 컴퓨팅을 도입하여 하드웨어 리소스를 보다 효율적으로 사용할 것이다.

2) 하이퍼스케일러(Hyperscaler)들의 자체 트레이닝 칩 개발

Large-scale 모델 개발에 많은 시간과 비용이 들기 때문에 많은 하이퍼스케일러들이 자체 AI 가속기를 개발하여 고가의 GPU 도입 비용을 절감하고 AI 모델 개발 시간을 단축할 것이다.

3) Large-scale 모델에 적합한 메모리 풀 구조

Large-scale 모델 학습에 가장 필요한 자원은 메모리로, 거대한 메모리 풀을 효율적으로 사용할 수 있는 구조가 연구될 것이다. 예를 들어, NVIDIA의 Grace CPU는 고속의 NVLINK를 이용하여 CPU의 메모리를 Far 메모리로 사용하여 부족한 GPU 메모리 용량을 확보할 수 있을 것이다.

5. 중국 인공지능 아카데미의 초거대 AI 사전학습 모델 '우다오 2.0'

중국 베이징 인공지능 아카데미(BAAI: Beijing Academy of Artificial Intelligence) 연구팀이 발표한 '우다오 2.0'은 1조 7천억개의 매개변수를 사용한 AI 사전 학습모델이다. 이는 GPT-3의 매개변수 1,750억개보다 10배 크고, 구글 스위치 트랜스포머보다 1,500억개 더 큰 규모다.

우다오 2.0은 대화 음성을 시뮬레이션하고 시를 쓰며, 그림을 이해하고 레시피를 생성하는 등 사전 훈련된 AI 모델로, 마치 데이터를 연료로 사용하는 발전소와 같다. 중국어와 영어로 AI 애플리케이션을 지원하는 다양한 인텔리전스(Intelligence)를 생성한다[7]. 총 4.9테라바이트(TB)의 텍스트 및 이미지 데이터로 학습되고 있으며, 이 학습 데이터에는 중국어와 영어 텍스트를 각각 1.2TB씩 포함하고 있다.

특히, GPT-3가 할 수 없는 다중 모드로 텍스트와 이미지에서 학습하고 두 가지 유형의 데이터를 모두 포함하는 작업을 처리할 수 있다. 예를 들어, 이미지 생성이나 얼굴 인식 등의 특정 작업에 특화된 심층 생성 모델과 달리, 에세이나 시를 쓰거나 정지 이미지에 근거해 보완하는 문장을 생성할 수 있고, 문장의 설명에 따라 이미지를 만들어 낼 수도 있다.

베이징에서 활동중인 AI 연구자 블레이크 얀은 "거대 데이터 세트로 훈련된 이들 정교한 모델이 특정 기능에 사용되는 경우, 소량의 새로운 데이터 밖에 필요하지 않다. 왜냐하면 인간과 마찬가지로 이전에 학습한 지식을 새로운 작업에 응용할 수 있기 때문이다"라고 말했다.

이는 우다오 2.0의 대규모 사전학습 모델이 인간과 같은 추론능력을 지닌 범용 인공지능에 초점을 뒀다는 점을 시사한 것으로, 우다오 2.0의 목적이 세계 최고 수준의 AGI(Artificial General Intelligence) 개발이라는 연구진의 주장에 근거를 둔다. 즉, AI가 정말 사람처럼 생각하고 행동하도록 만들겠다는 뜻이다.

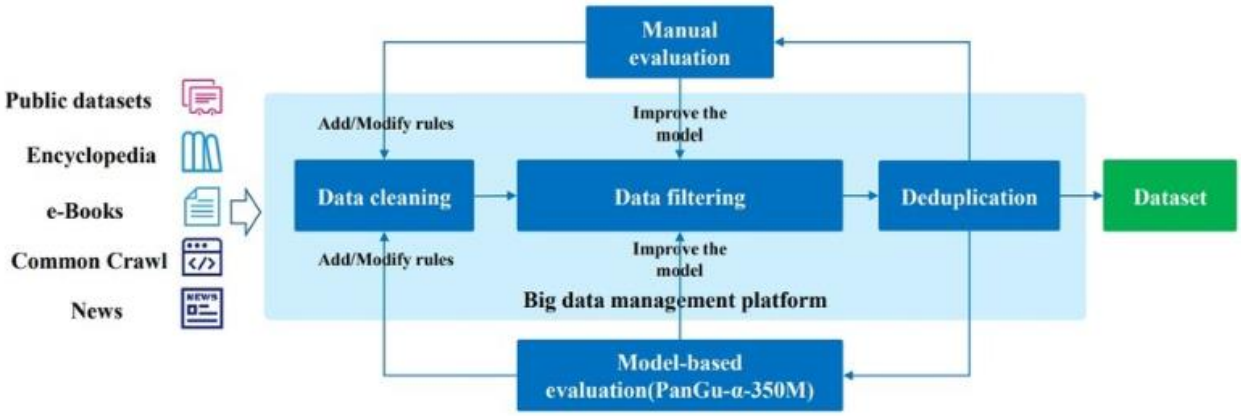
또한, 우다오 2.0은 구글의 MoE(Mixture-of-Export training System) 기술의 한계인 병목현상을 해결하고자 파이토치(PyTorch) 프레임까지 지원하는 최초의 MoE 시스템인 FastMoE를 개발하였는데, 이는 가용성, 유연성, 훈련속도 측면에서 강점을 보인다. 구글의 MoE와 다르게 오픈소스 형태이고, 특정 하드웨어를 필요로 하

© Rsupport Co. Ltd. All Rights Reserved.

지 않으며 이미 ImageNet, LAMA 등 여러 벤치마크 테스트에서 좋은 성적을 냈다고 한다.
종합하면, 우다오 2.0의 사전학습 프레임워크는 업계 최고 수준의 AGI를 지향하며, FastMoE 시스템을 통한 훈련 효율성으로 코딩, 모델, 학습, 미세 조정(fine-tuning), 추리 등 5가지의 고효율 특징을 갖고 있다. 이를 바탕으로 한 우다오 2.0은 중국어 텍스트 데이터, 멀티모달 데이터, 대화음성 데이터까지 모두를 통합한 거대 데이터 모음집이라 할 수 있다[8].

6. 화웨이의 대규모 자연어처리(NLP) 모델 '판구 알파'

GPT-3가 공개되고 1,750억개의 GPT-3는 역사상 가장 뛰어난 언어 인공지능(AI)이자, 역사상 가장 유명한 언어 AI가 되었다. 이런 현상은 GPT-3가 타 모델과 차원이 다른 자연어 생성능력을 보여주었기 때문이다. 그러나 대부분의 언어 모델은 영어로만 제공된다. 또한, GPT-3는 인문, 사회과학 분야에서 높은 성능을 보인 반면, 공학, 자연과학 분야에서는 약한 모습을 보였다. GPT-3 학습에 사용된 데이터가 인문, 사회과학적인 데이터에 쏠려 있어 사고력보다는 지식의 양에 강점이 있다고 볼 수 있다.
여기에 중국 화웨이가 GPT-3를 넘는 2,000억개의 매개변수를 포함하는 750기가바이트(GB)로 대규모 자연어처리(NLP) 모델인 '판구 알파'를 개발하고 오픈소스로 공개했다. 화웨이 연구팀은 이 초거대 언어모델 설계에서 트랜스포머(Transformer) 기반 자기회귀언어(Autoregressive Language) 모델을 기본 아키텍처로 선택했으며, 다른 쿼리 레이어가 트랜스포머 레이어 위에 추가되어 모델이 최대 2,000억개의 매개변수를 확장할 수 있게 하였다[10].



[그림6] 판구 알파의 데이터 출처 및 사전 학습 데이터 구축 프로세스(논문캡처[9])

화웨이 연구팀은 중국어로 된 전자책, 백과사전, 뉴스, SNS, 웹페이지 등에서 80테라바이트(TB)의 원시 데이터를 수집했다. 대규모 말뭉치 전처리를 위한 분산 클러스터를 구축하고 불쾌한 단어, 스팸, 저품질 데이터 정리 및 필터링, 중복 제거, 품질 평가 및 기타 처리 절차를 통해 약 1.1TB의 고품질 중국 말뭉치 데이터 세트를 구축했다. 또한, 연구팀은 5차원 병렬 기능을 결합해 모델에 적용했으며, CAN4가 구동하는 2,048개의 어센드 AI 프로세서 클러스터로 훈련했다.

© Rsupport Co. Ltd. All Rights Reserved.

Model	#Parameters	#Layers (L)	Hidden size (d)	FFN size (d_{ff})	#Heads (N_h)
PanGu-α 2.6B	2.6B	32	2560	10240	40
PanGu-α 13B	13.1B	40	5120	20480	40
PanGu-α 200B	207.0B	64	16384	65536	128

[그림 7] PanGu-α 모델의 크기와 하이퍼 파라미터(논문캡처[9])

연구팀은 매개변수 크기가 증가하는 고품질 중국어 텍스트 말뭉치에 대해 26억, 130억, 2000억의 세 가지 모델을 훈련시켰다. 모델은 언어 모델링 작업에서 처음 평가되었으며, 모델 용량과 데이터 및 계산 양이 증가함에 따라 복잡성이 감소한다는 점에 주목했다. 또한 연구팀은 대화 생성, 요약 및 질문 답변과 같은 다양한 시나리오에서 모델의 텍스트 생성 능력을 조사하였고, 실험 결과는 일반적으로 모델 용량이 증가함에 따라 모델의 성능이 향상된다는 것을 보여주었다. 한편, 화웨이 연구팀이 공개한 이 초거대 언어모델 '판구 알파(PanGu Alpha)'는 현재 오픈 소스로 공개된 상태이다.

결론: 초거대 AI에 대한 의문

위에서 살펴본 것처럼 현재 글로벌 ICT 기업들의 초거대 AI 기술 개발 경쟁이 치열하다. 치열한 기술 경쟁 덕분에 초거대 AI의 발전 속도도 가속화되고 있다. 이렇게 과열되고 있는 초거대 AI 개발 경쟁에 문제는 없을까? 이토록 글로벌 ICT 기업들이 치열하게 경쟁하는 가운데 과연 우리의 초거대 AI 기술은 어디까지 왔을까?

이들 질문에서 시작된 국내외 초거대 AI 기술의 한계와 과제에 대해 2부를 통해 알아보겠다. 이에 앞서 국내 ICT 기업들의 초거대 AI 기술 동향도 상세히 살펴보겠다.

© Rsupport Co. Ltd. All Rights Reserved.

※ 참고문헌

[1] 초거대 AI, 네이버 시사상식사전, 2021년 12월.
[2] 이효은, GPT-3 초거대 AI 경쟁에 불을 붙이다, 전자신문, 2021년 12월.
[3] 김종윤, GPT-3 인류 역사상 가장 뛰어난 '언어 인공지능'이다, 인공지능신문, 2020년 8월.
[4] 박지호, 최강 챗봇 등장! 구글 람다(LaMDA), 위클리 NLP, 2022년 2월.
[5] 박성은, 딥마인드도 초거대 AI 언어모델 고퍼 공개...크기 대비 성능이 차별점, AI타임스, 2021년 12월.
[6] 김대우, GPT-3보다 3배 더 큰 언어모델:MT(Megatron-Turing)-NLG(530B), 2021년 10월.
[7] 최창현, 중국 인공지능 아카데미, GPT-3 10배 넘는 1조7500억개 매개변수 초거대 AI 언어 모델 '우다오 2.0' 개발, 인공지능신문, 2021년 6월.
[8] 애뚱, "우다오 2.0 발표, 쯔위엔 연구원 가상 디지털 인간 화즈빙", 디지털 중국, 2021년 7월.
[9] PanGu-α TEAM, PanGu-α : Large-Scale Autoregressive Pretrained Chineses Language Models With Auto-Parallel Computation, 2021년 4월.
[10] 최창현, GPT-3 넘었다!...화웨이, 2천억개 매개변수의 초거대 언어AI모델 '판구 알파' 공개, 인공지능신문, 2021년 5월

© Rsupport Co. Ltd. All Rights Reserved.