

초대규모 언어 모델과 네이버의 최근 연구

네이버 | 이상우·이기창·하정우

1. 서 론

2020년 6월 GPT-3[1] 라는 초대규모 언어 모델 (Hyperscale Language Model, Hyperscale LM)의 등장 이래로, 초대규모 언어 모델 기술은 빠른 발전을 거듭해오고 있다. 초대규모 언어 모델은 기존 언어 모델보다 훨씬 더 큰 파라미터 사이즈를 가지는 모델들을 지칭하며, 이 거대한 파라미터를 충분히 학습시키기 위해 과거 언어모델들 보다 훨씬 더 대규모의 코퍼스를 활용하여 학습한다. 이를 통해 이러한 초대규모 언어 모델에서는 작은 파라미터 사이즈에서 일어나지 않았던 여러 언어 능력들이 창발되기를 기대한다.

OpenAI [1,2], Google [3,4], Deepmind [5,6], Meta [7] 을 포함한 많은 글로벌 테크 대기업들과 그 외 연구기관들에서 서로 다른 전략과 접근을 바탕으로 여러 초대규모 언어 모델들을 만들어 나가고 있다. 우리나라에서는 네이버가 대표적으로 초거대 언어 모델과 그 응용에 많은 역량을 투입하고 있다 [8]. 본 논문에서는 초거대 언어 모델의 동향을 네이버에서 연구하고 있는 연구들을 중심으로 소개한다.

2. 초대규모 언어 모델

초대규모 언어 모델은 큰 파라미터 사이즈를 가지는 Transformer [9] 기반의 언어 모델들을 지칭하며, 초대규모 언어 모델에서는 기존 작은 언어 모델에서 나타나지 않는 강력한 언어 능력들이 등장하기를 기대한다. 본 장에서는 OpenAI에서 개발된 첫번째 초대규모 언어 모델 GPT-3 [1]와 네이버에서 개발된 한국어 초대규모 언어 모델 HyperCLOVA [8]에 대해 다룬다.

2.1 OpenAI GPT-3

현재 개발되고 있는 초대규모 언어 모델은 Transformer [9] 기반의 기존 언어 모델들을 확장시킨 것이다. 그 중에서도 대부분의 주류가 Transformer decoder 기반

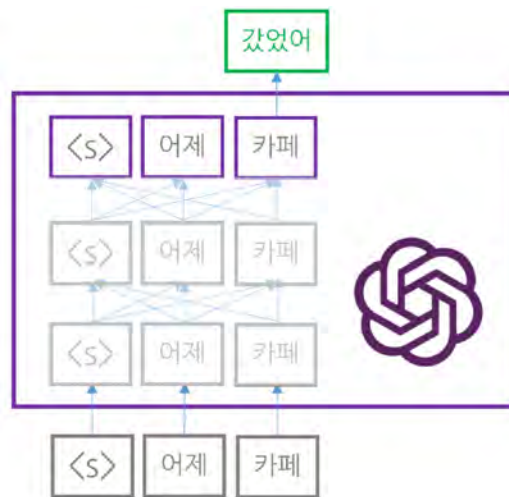


그림 1 OpenAI GPT-2 언어 모델의 도식. 도식에서는, “어제”, “카페” 라는 순차적인 token을 input으로 받아서 “갔었어” 라는 token을 생성하고 있다.

의 GPT-2 [10] 모델을 사용한다. GPT-2는 OpenAI에서 만든 GPT 언어 모델 시리즈의 두번째 모델이다. 그림 1은GPT-2 언어 모델의 도식을 보여주고 있다. GPT-2는 110M에서 1.5B 사이의 파라미터를 가지는 모델로 GPT-2 논문에서 제안되고 공개되어 사용되었다.

최초의 초대규모 언어 모델이자 많은 초대규모 언어 모델의 기본 형태를 제공하는 GPT-3 역시 GPT-2와 구조적으로 거의 유사하다. 다만 모델의 크기에서 차이를 보인다. 논문에서 소개된 GPT-3는 2048 token을 가지며, 모델 파라미터 사이즈는 175B에 달한다. 300B Token 규모의 데이터를 학습하며, 82%는 웹에서 모은 데이터, 16%는 책에서 모은 데이터, 3%는 위키피디아의 데이터이다.

초대규모 언어 모델을 대표하는 가장 큰 언어 능력은 in-context learning이다. 그림 2에서 볼 수 있듯이, In-context Learning에서는 fine-tuning 없이 downstream task에 대한 설명과 downstream task의 예시만 몇 개 넣어주는 것 만으로도, 초대규모 언어 모델이 downstream

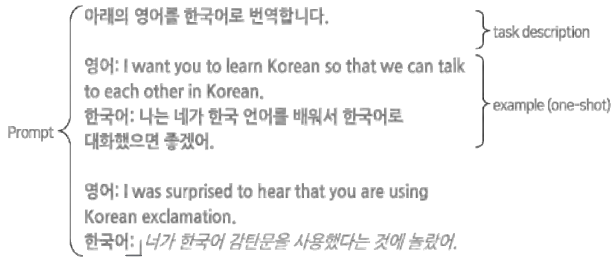


그림 2 In-context Learning의 도식

task를 풀 수 있게끔 한다. in-context learning에서는 zero-shot, few-shot 이라는 용어를 많이 사용하는데, 그림 2에서 example이 없는 경우 zero-shot, 몇개 있는 경우 few-shot이라고 부른다. in-context few-shot learning에서 비록 몇개의 training 데이터가 사용되지만, 그것으로 모델을 추가 학습하지는 않는다는 점을 유념할 필요가 있다.

초대규모 언어 모델은 학습하는 데이터에 따라 발휘될 수 있는 능력이 다르다. OpenAI GPT-3는 Microsoft로부터 많은 양의 github 데이터를 받아 학습을 수행했고, 결과적으로 프로그래밍 코드에 대한 설명이나 주석만으로 코드를 생성할 수 있게 되었다. 이러한 능력을 확장하여 OpenAI는 github 코드 데이터를 중심으로 학습하고 성능을 개량한 Codex라는 모델을 만들게 되었다 [11]. 지금은 Copilot이라는 브랜드로 API 서비스를 수행하고 있는데, 코드를 작성할 때 작성해야 할 코드를 생성하여 프로그래밍을 보조하는 툴로 많은 프로그래머들의 생산성 향상에 크게 기여하고 있다.

2.2 HyperCLOVA

HyperCLOVA는 네이버에서 공개한 초대규모 언어 모델이다. 학습데이터는 블로그, 카페 등 네이버가 보유한 양질의 한국어 말뭉치다. 말뭉치 품질 향상을 위해 개인정보 마스킹, 비속어 및 중복 문서 제거 등을

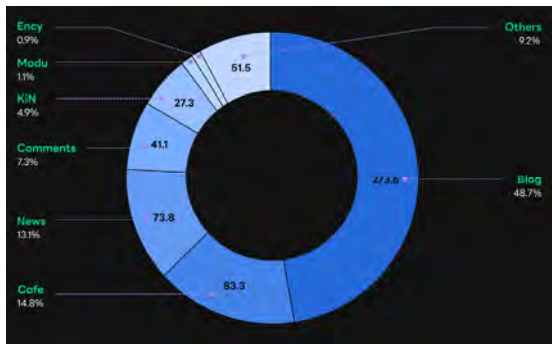


그림 2-2-1 HyperCLOVA 학습 말뭉치 구성. 원 안의 숫자는 말뭉치의 크기를 의미 (10억 토큰 단위).

표 1 HyperCLOVA 크기별 KorQuAD(EM) 점수 개선 현황

구분	6.9B	13B	60B	82B
2021년 5월	0.6121	0.6604	-	0.6927
2022년 11월	0.7322	0.7594	0.7775	-

수행했다. 조사, 어미 빈도가 높은 한국어 특성을 감안해 형태소 단위 바이트 페어 인코딩(Morpheme-aware Byte Pair Encoding) 방식으로 토큰화 했다 [8].

초거대 모델 학습은 학습 도중 train loss가 발산하는 등 그 자체로 도전적인 과제다 [7]. 파라미터와 그 그래디언트의 크기 등을 지속적으로 추적 관찰하면서 grad clip 값 조절, 문제 있는 데이터를 건너뛰어 학습하기, Precision 조절 등 다양한 방식으로 대응하여 모델 학습을 안정화하였다.

최근 연구 [5,6,12] 를 보면 모델 크기가 동일할 때 층을 더 쌓거나, 빈칸 채우기(Fill in the Middle) 태스크를 추가 학습하면 모델의 품질이 개선되는 경향이 있다. 내부 실험 결과에서, HyperCLOVA에서도 이같은 경향을 재확인했다. 이밖에 다량의 구어 데이터로 파인튜닝을 수행해 성능이 더욱 개선되었다. 표 2-2-1에 제시한 KorQuAD는 한국어 기계 독해 과제로 in-context few-shot learning 점수가 비교적 안정적인 태스크다. 개발 초기 [8] 대비 10% 안팎의 상승을 보였다.

네이버는 다국어 초거대 언어 모델이나 멀티모달 초거대 언어모델에도 관심이 있다. 2021년 DEVIEW에서 네이버는 영어와 한국어를 같이 잘하는 다국어 HyperCLOVA [13] 나, 블로그 이미지를 보고 블로그 내용을 생성하는 HyperCLOVA에 대하여 소개한 바가 있다 [14]. 언어의 종류나 모달의 확장에 대해서도 계속해서 연구개발 계획을 가지고 있다.

3. 초대규모 언어 모델의 공개

3.1 초대규모 언어 모델 공개의 최근 동향

초대규모 언어 모델의 공개에 대하여 서로 다른 기관들이 서로 다른 전략을 가지고 있다. OpenAI에서는 초대규모 언어 모델의 API와 GUI Interface (OpenAI Playground) 를 공개하고, 툴을 사용할 때 사용한 만큼 과금을 하는 구조를 가지고 있다 [1,2]. Google과 Deepmind는 대체로 모델을 공개하지 않으나 [3,4,5,6], 최근 FLAN-T5라는 상대적으로 낮은 성능을 가지는 모델의 checkpoint를 github에 공개했다 [4]. Meta는 OPT라는 모델의 checkpoint를 공개했으나, 상대적으로 다른 대기업들 대비 모델 성능이 좋지 않다는 평

을 받는다. 그 외에 EleutherAI에서 GPT-neo라는 이름의 시리즈 모델들을 공개하고 있고 [15], Bigscience에서 Bloom 이라는 모델을 공개하였다.¹⁾

한국에서는 현재 네이버가 API를 공개해 beta service를 제공하며 과금하는 구조를 가지고 있고, 카카오는 6B의 모델 checkpoint을 공개한 상황이다 [16]. 최근에는 Polyglot이라는 프로젝트에서 EleutherAI, Stability.ai, TUNiB주도로 한국어 언어 모델들을 공개하고 있으며, 현재 6B까지 공개되었다 [17].

3.2 CLOVA Studio

CLOVA Studio는 HyperCLOVA를 활용해 생성, 요약, 분류, 변환 등 텍스트 기반 과제를 수행할 수 있는 웹 서비스이다. 코딩을 할 줄 몰라도 쉽게 AI 도구들을 만들 수 있는 ‘노코드(No code) AI’를 지향한다. CLOVA Studio 익스플로러에는 in-context learning 예제들이 다수 수록되어 있어 목적에 맞게 활용할 수 있다.

튜닝 기능을 이용하면 HyperCLOVA를 사용자의 데이터로 업데이트해 더 나은 성능을 경험할 수 있다. CLOVA Studio에서 제공하는 튜닝 기법은 P-tuning [18], LoRA [19]가 있다. 튜닝할 때 HyperCLOVA의 일부 요소만 업데이트하고, 인퍼런스할 때 HyperCLOVA 본체는 공유한 채로 업데이트된 요소들만 그때그때 교체해주는 방식이다. 초거대 모델 전체를 업데이트하거나 교체하는 방식 대비 비용 효율성을 대폭 높였다. 이밖에 CLOVA Studio에서는 문서 요약, 배치 단위 in-context learning, 소량의 사용자 데이터를 대량으로 확장 등 다양한 사용자 편의 기능을 제

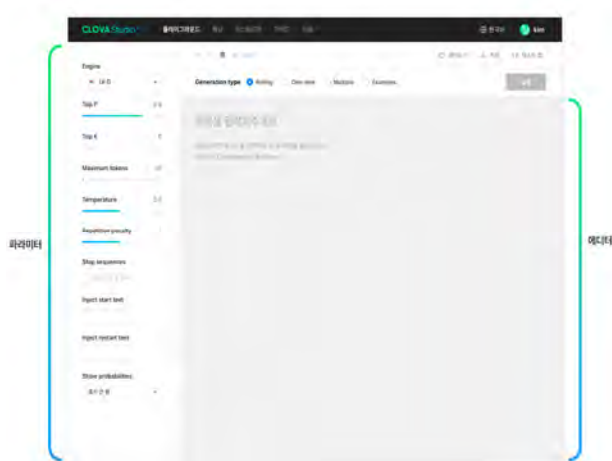


그림 3 네이버 클라우드를 통해 현재 (22년 11월 기준) 서비스 되고 있는 CLOVA Studio의 모습

공한다.

HyperCLOVA는 사용자가 가지고 있는 학습 데이터를 늘려주는 데이터 증강 기능 CODA도 제공한다. 4.3장에서 그 내용을 더 자세히 소개한다.

22년 11월 현재 CLOVA studio 신청 기업이 1000곳을 넘었다. 이 가운데 뽀빠테크놀로지스(광고 카피), 임플로이랩스(자기소개서), 킵그로우(SNS 게시물) 등 300여곳이 실제 서비스에 적용하고 있다.

4. 데이터 증강 (Data Augmentation)

앞서 설명한 바와 같이, 초대규모 언어 모델은 다양하고 강력한 언어적인 능력을 발휘할 수 있다는 장점을 가지고 있다. 하지만, 한편으로 파라미터의 수가 너무 크기 때문에, inference하는데 있어서는 느린 시간과 많은 비용이 든다. 따라서 초대규모 언어 모델을 실용적으로 사용하기 위한 좋은 방법으로 데이터 증강 방법이 오랫동안 주목받아 왔다. 데이터 증강을 통해 얻어진 데이터는 작은 언어 모델을 학습하는 데에 사용되며, 이는 실제 빠른 서비스 inference 속도에 기여한다. 초대규모 언어 모델 기반 데이터 증강에 대해서는 학계의 주목할만한 연구가 초기부터 현재까지 꾸준히 나오고 있다 [20]. 본 장에서는 네이버의 데이터 증강 연구에 대해 소개한다 [21,22,24].

4.1 GPT3Mix

GPT3Mix는 초대규모 언어 모델이 등장한지 상대적으로 얼마 되지 않은 시기에 등장한 데이터 증강 방법 중 하나이다 [21]. GPT3Mix에서는 적은 데이터가 있을 때, 적은 데이터를 GPT-3의 prompt에 넣고, 이를 바탕으로 유사한 데이터를 생성하게끔 하는 알고리즘이다. 구체적인 알고리즘의 도식이 그림 4에 있다. GPT3Mix는 서로 다른 예제를 바탕으로 새로운 데이터와 그 label을 생성하고, 초대규모 언어 모델의

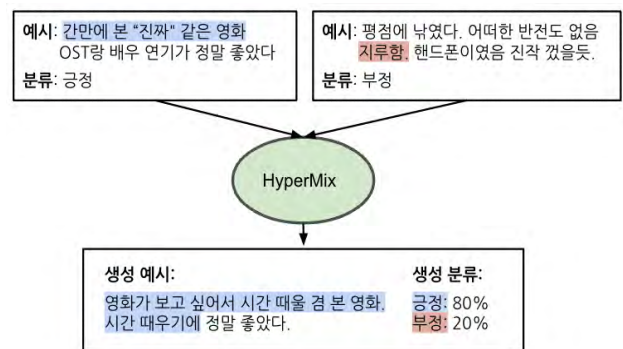


그림 4 GPT3Mix의 도식

1) <https://huggingface.co/bigscience/bloom>

label 생성 확률을 작은 언어 모델에 distillation한다. 즉, 작은 언어 모델은 생성된 데이터 뿐 아니라 초대규모 언어 모델의 label 확률을 바탕으로 학습을 수행하게 된다.

논문에서, GPT3Mix는 이전의 NLP에서의 데이터 증강 방법이었던 back-translation (데이터를 예컨대 파카고와 같은 번역기 모델로 다른 언어로 이동 시켰다가 다시 원 언어로 돌리는 방법) 보다 성능이 훨씬 좋다고 보고 하였다 [21]. 특히, 내부적인 연구 결과에 따르면, GPT3Mix는 특정한 몇몇 실제 서비스 상황에서의 데이터에 대하여서는 사람이 생성하는 것과 유사한 수준의 성능을 보일 수 있는 것으로 밝혀졌다.

4.2 대화 시스템을 위한 데이터 증가

[22] 논문에서는 HyperCLOVA의 데이터 증강 기능을 활용해, 대화 데이터를 생성하는 방법에 대해 다루고 있다. 먼저 HyperCLOVA의 prompt에 대화 예시를 넣고, 이를 바탕으로 대화를 생성한다. 그 뒤에 생성된 대화가 적절한지 여부를 crowdworker를 통하여 변별한다. 적절하지 않은 대화는 crowdworker가 직접 수정한다. 이렇게 대화시스템을 HyperCLOVA와 함께 human-in-the-loop 방식으로 만들게 되면, 짧은 시간

내에 대화 데이터를 생성해 낼 수 있다. 이러한 방법론을 바탕으로 네이버에서는 케어콜이라는 대화 시스템 서비스를 만들었다. 이 시스템은 어르신 분들에게 안부 전화를 드리고 자유롭게 말벗이 되어드리기 위한 챗봇이다. 해운대구를 시작으로, 현재 (2022년 11월) 시점에서 전국 수십여개의 자치단체에 서비스를 제공하고 있다. 그림 5는 이러한 케어콜 시스템의 제작 도식을 표현하고 있다.

케어콜 시스템은 계속하여 기술을 발전시키고 있다. 최근에는 어르신 분들이 과거에 했던 말을 기억하거나 혹은 업데이트 하고 이를 바탕으로 어르신 분들에게 응대하는 시스템 역시 개발되었다 [23]. 이러한 기술의 대화의 몰입도를 높이고, 대화 시스템이 어르신 분들을 실제로 이해하고 있다는 느낌을 줄 수 있는 것을 목표로 한다.

4.3 CODA

앞선 4.1장과 4.2장에서는, 초대규모 언어 모델에 대한 추가 fine-tuning 없이 데이터를 생성하는 방법들을 소개하고 있다. 하지만 초대규모 언어 모델은 원래 목적이 특화된 데이터 분포의 데이터 증강이 아니기 때문에, 실제 우리가 풀고자 하는 데이터를 특화해서 생성하는 데에는 한계가 있다. 네이버에서는 이러한 문제를 개선하여, 데이터 일부를 받아서 HyperCLOVA를 학습하고 학습된 모델을 통해 더 양질의 데이터를 생성하는 CODA 프레임워크를 가지고 있다 [24]. 이 발표에서는 CODA를 활용하는 예시로 바탕으로 특정 페르소나를 가지는 자유 발화 대화 시스템 (open-domain dialogue) 을 어떻게 만들 수 있는 지에 대해 또한 소개하고 있다.

5. 초대규모 언어 모델에 대한 분석 연구

초대규모 언어 모델의 모델 크기는 커지고 다양한 창발적인 언어 능력들이 보고 되고 있지만, 상대적으로 초대규모 언어 모델의 동작 원리에 대해서는 적은 관심을 받고 있다. 이는 일부분은 가장 성능이 좋은 대기업의 초대규모 언어 모델들의 checkpoint가 공개되어 있지 않아 회사 내에서만 분석이 가능한 데 기인하고, 다른 일부분은 언어 능력들이 발현되는 원리보다는 scale-up을 바탕으로 어떤 창발적인 언어 능력들이 더 발현될 수 있는 지에 대한 탐색에 집중하는 최근 연구 분야의 트렌드에 기인한다. 하지만, 초대규모 언어 모델의 분석 연구 없이는 초대규모 언어 모델에 대한 학계의 체계적인 발전이 어려운 것도

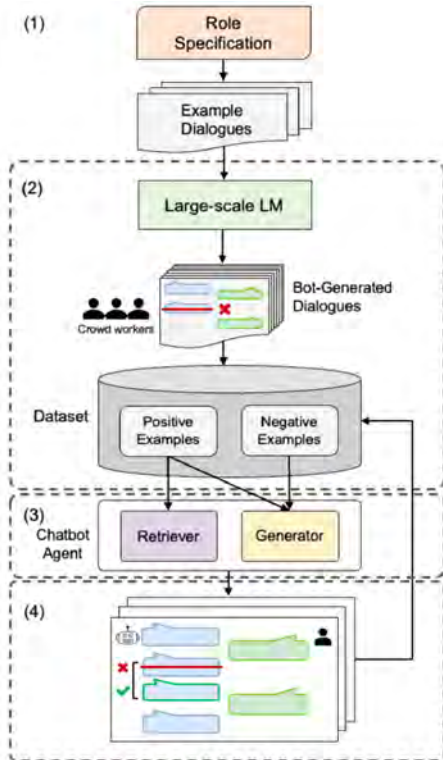


그림 5 케어콜 대화 시스템 제작의 도식. HyperCLOVA의 데이터 증강 기능이 대화 데이터 구축에 적극적으로 활용된다.

사실이다. 본 논문에서는 초대규모 언어 모델에 대한 네이버의 분석 연구들을 중점적으로 설명한다 [25, 5-1-2].

5.1 GPT-3 말뭉치에 대한 분석

기존 초대규모 언어 모델 연구에서는 학습 데이터 관련 in-context learning의 발현 지점에 대한 연구가 제한적으로 이루어졌다. 예컨대, 웹에서 모을 수 있는 데이터들을 최대한 넣어서 학습을 수행했지만, 데이터의 품질에 대한 filtering 정도의 논의가 있었을 뿐 [1,5], 어떠한 데이터들이 들어갔을 때 in-context learning이 발현되는가에 대해서는 발현 지점의 추측이 이루어지지 않고 있었다. 하지만 학습 데이터는 언어 모델 성능에 큰 영향을 미치며, 학습 데이터에 대한 연구는 경제적인 데이터 수집 관점에서 중요하다.

[25]에서는 HyperCLOVA 말뭉치 (Pretraining Corpus) 내의 일부분을 학습했을 때, 어떠한 성능 변화가 일어나는 지에 대해 다루었다. 말뭉치는 7가지의 종류로 구분하였는데, 네이버 블로그, 네이버 카페, 네이버 뉴스, 네이버 댓글, 네이버 지식인, 모두의 말뭉치, 위키피디아이다.

그림 6은 HyperCLOVA 세부 말뭉치의 few-shot 성능 결과를 보고하고 있다. 블로그 말뭉치를 학습했을 때는 전체 말뭉치를 학습했을 때 근접한 성능이 나오는 반면, 카페 말뭉치나 뉴스 말뭉치만 학습해서는 few-shot 성능이 잘 나오지 않고, 심지어 댓글이나 모두 말뭉치 대비 성능이 나오지 않는다. 이를 통하여 단순히 말뭉치의 크기만으로 in-context learning의 성능 향상이 일어나는 것은 아니라는 것을 보여준다.

이 연구는 그 외에 기존의 in-context learning 성능과 데이터의 관계에 대한 여러 통념들에 대해 반



그림 6 HyperCLOVA 말뭉치의 세부 말뭉치들만 학습했을 때의 in-context few-shot learning 성능 결과

예시 A	예시 B
예시: 간만에 본 "진짜" 영화같은 영화 OST랑 배우 연기가 정말 좋았다 분류: 긍정	예시: 간만에 본 "진짜" 영화같은 영화 OST랑 배우 연기가 정말 좋았다 분류: 부정
예시: 평점에 낮았다. 어떠한 반전도 없음 지루함. 핸드폰으로 봤다면 진작 꺼버렸겠지 분류: 부정	예시: 평점에 낮았다. 어떠한 반전도 없음 지루함. 핸드폰으로 봤다면 진작 꺼버렸겠지 분류: 긍정
예시: 영화가 보고 싶어서 시간 때울 겸 본 영화. 시간 때우기에 정말 좋았다 분류: 긍정	예시: 영화가 보고 싶어서 시간 때울 겸 본 영화. 시간 때우기에 정말 좋았다 분류: 부정

그림 7 In-context learning에서 분류 label token의 의미가 정반대인 경우의 prompt 예시. 논문 [27]에서는 “긍정”이 나올 것이라고 주장하고, [28]에서는 “부정”이 나올 것이라고 주장한다.

박하고 있다. 자세한 사항들은 논문에서 확인할 수 있다.

5.2 In-context Learning의 prompt에 대한 분석

초대규모 언어 모델의 in-context learning 성능에 있어, prompt는 중요한 역할을 차지한다. 그림 2에서 볼 수 있는 prompt에서 Task의 설명을 어떻게 적는가, prompt의 구조를 어떻게 바꾸는가가 성능을 크게 좌우한다. 따라서, 어떠한 자연어 prompt가 좋은 prompt인지에 대하여 많은 언어 모델 연구들에서 관심을 가지고 탐색해 나가고 있다 [26]. 본 장에서는 prompt의 task 설명과 label의 의미가 in-context learning에 미치는 영향에 대한 서로 다른 입장을 가진 두 연구 [27, 28]에 대해 소개한다.

그림 7은 감정 분류 문제를 다루는 in-context learning prompt의 두가지의 예시를 표현하고 있다. 예시 A에서는 예시에 대응되는 분류의 semantic이 올바르게 달려 있지만, 예시 B에서는 반대로 달려 있다. 예시 B에서는 HyperCLOVA가 부정이라는 label이라는 예측을 할까? 긍정이라는 예측을 할까? 이 질문은 초대규모 언어 모델이 예시에 대한 적절한 semantic을 더 중요하게 생각할 지, label 예시에 대한 syntax를 더 중요하게 생각할 지에 대한 질문이다. [27]에서는 각 예시에 달린 label 자체는 덜 중요하게 생각하기에 결국 “긍정” label이 나올 것이라는 기존 연구자들의 가정에 반하는 주장을 하여 주목을 받았다. 한편, [28]에서는 각 예시에 달린 label도 상당 경우 중요하게 동작한다고 주장하고 있다. 이에 따르면, 예시 B에서는 “부정” label이 나올 가능성이 높다. 각, [27]와 [28]에서는 어떤 조건에서 어떠한 결과가 나오는 지에 대해 논의하고 있다.

실제 HyperCLOVA에서 실험해본 결과, 예시 B에 대해서는 “부정”이라는 label이 달려, [28]의 주장을 좀 더 뒷받침했다. 다만, 앞에 이 예시가 어떠한 예시 인지를 설명하는 경우에는 label이 “긍정”이 나오기도

“부정” 이 나오기도 하였다. 이를 통해 task에 대한 설명 역시 prompt와 in-context learning의 동작에 있어 중요한 역할을 한다는 점을 알 수 있다.

6. 토의 및 결론

본 논문에서는 네이버의 초대규모 언어 모델에 대한 작년과 올해의 연구들을 [8,13,14,21,22,23,24,25,28] 중심으로 초대규모 언어 모델의 최근 연구 동향에 대하여 살펴보았다. 2020년 5월 GPT-3가 나온지로부터 2년 반이 지났지만, 초대규모 언어 모델에 대한 연구 성과들은 꾸준히 이어지고 있다. 초대규모 언어 모델은 여전히 발전 중이다. 편향성 해소, 혐오 표현 생성 등 AI윤리 문제 해결이나 비즈니스 적용을 위한 경량화 운영 효율화를 위한 연구 노력들도 많이 진행되고 있다. 초대규모 AI는 하나의 기술을 넘어 인터넷이 나 모바일 같은 새로운 생태계를 탄생시키는 기술이고 실제로 국내에서도 300여개의 스타트업들이 하이퍼클로바 API를 그들의 서비스를 위해 활용하고 있다. 초대규모 AI의 시대의 한국이 First mover로서 자리잡을 수 있도록 더욱 많은 협력이 필요한 시점이다.

7. 참고 문헌

- [1] T. Brown et al., “Language models are few-shot learners,” NeurIPS, 2020.
- [2] L. Ouyang et al., “Training Language Models to Follow Instructions with Human Feedback,” arXiv, 2022.
- [3] A. Chowdhery et al., “PaLM: Scaling Language Modeling with Pathways,” arXiv, 2022.
- [4] H. W. Chung et al., “Scaling Instruction-Finetuned Language Models,” arXiv, 2022.
- [5] J. W. Rae et al., “Scaling Language Models: Methods, Analysis & Insights from Training Gopher,” arXiv, 2022.
- [6] J. Hoffmann et al., “Training Compute-Optimal Large Language Models,” arXiv, 2022.
- [7] S. Zhang et al., “OPT: Open Pre-trained Transformer Language Models,” arXiv, 2022.
- [8] B. Kim et al., “What Changes can Large-scale Language Models Bring? Intensive Study on Hyperclova: Billions-scale Korean Generative Pretrained Transformers,” EMNLP, 2021.
- [9] A. Vaswani et al., “Attention is All You Need,” NIPS, 2017.
- [10] A. Radford et al., “Language Models are Unsupervised Multitask Learners,” 2019.
- [11] M. Chen et al., “Evaluating Large Models Trained on Code,” arXiv, 2021.
- [12] Wang et al., “Deepnet: Scaling transformers to 1,000 layers,” arXiv, 2022.
- [13] 신성진, 김보섭, “Multilingual HyperCLOVA”, DEVIEW, 2021.
- [14] 김형석, 이상우, “HyperCLOVA – Korean GPT3: 하이퍼스케일, 그리고 그 후,” DEVIEW, 2021.
- [15] S. Black et al., “GPT-NeoX-20B: An Open-Source Autoregressive Language Model,” ACL Workshop on Challenges & Perspectives in Creating Large Language Models, 2022.
- [16] I. Kim et al., “KoGPT: KakaoBrain Korean(hangul) Generative Pre-trained Transformer”, 2021.
- [17] H. Ko et al., “Polyglot-Ko: Open-Source Korean Autoregressive Language Model,” 2022.
- [18] Liu et al., “GPT understands, too,” arXiv, 2021.
- [19] Hu et al., “Lora: Low-rank adaptation of large language models,” arXiv, 2021.
- [20] Z. Dai et al., “Promptagator: Few-shot Dense Retrieval From 8 Examples,” arXiv, 2022.
- [21] K. M. Yoo et al., “GPT3Mix: Leveraging Large-scale Language Models for Text Augmentation,” Findings of EMNLP, 2022.
- [22] S. Bae et al., “Building a Role Specified Open-Domain Dialogue System Leveraging Large-Scale Language Models,” NAACL, 2022.
- [23] S. Bae et al., “Keep Me Updated! Memory Management in Long-term Conversations,” Findings of EMNLP, 2022.
- [24] 가순원 외, “100억 대화 데이터 없이도 가능해요! - HyperCLOVA로 만드는 캐릭터 챗봇,” 2021, DEVIEW.
- [25] S. Shin et al., “On the Effect of Pretraining Corpora on In-context Learning by a Large-scale Language Model,” NAACL, 2022.
- [26] L. Reynolds and K. McDonell, “Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm,” arXiv, 2021.
- [27] S. Min et al., “Rethinking the Role of Demonstrations: What Makes In-Context Learning Work?,” EMNLP, 2022.
- [28] K. M. Yoo et al., “Ground-Truth Labels Matter: A Deeper Look into Input-Label Demonstrations,” EMNLP, 2022.



이 상 우

2012 서울대학교 컴퓨터공학부 졸업 (학사)
2018 서울대학교 컴퓨터공학부 졸업 (박사)
2018~2021 NAVER AI Lab Research Scientist
2020~현재 NAVER CLOVA Tech Lead
2021~현재 NAVER AI Lab 리더
2021~현재 KAIST AI 겸직교수

관심분야: 대화 시스템, 언어 모델, 자연어 처리

Email: sang.woo.lee@navercorp.com



이 기 창

2010 서울대학교 국어국문학과 졸업 (학사)
2018 고려대학교 산업경영공학과 졸업 (석사)
2018~현재 NAVER CLOVA AI/ML 개발자
2020~2021 NAVER CLOVA Tech Lead
2021~현재 NAVER CLOVA 리더
관심분야: 언어모델 임베딩, 대화시스템, 거대 모델 학습

Email: gichang.lee@navercorp.com



하 정 우

2004 서울대학교 컴퓨터공학부 졸업(학사)
2015 서울대학교 전기컴퓨터공학부 졸업(박사)
2015~2016 네이버랩스 책임연구원
2016~2017 네이버랩스 Tech Lead
2017~2020 네이버 CLOVA AI Research 리더
2020~현재 네이버 AI Lab 연구소장 (이사)

2021~현재 서울대-네이버 초대규모 AI연구센터 공동센터장

2021~현재 KAIST-네이버 초창의적 AI연구센터 공동센터장

2021~현재 AI미래포럼 공동의장

2022~현재 한국공학한림원 컴퓨팅분과 일반회원

2022~현재 대통령직속 디지털플랫폼정부위원회 AI/Data 분과위원장

관심분야: 인공지능, 기계학습, 자연어처리, 컴퓨터비전, 음성언어처리

Email: jungwoo.ha@navercorp.com