

[논문 리뷰] What Changes Can Large-scale Language Models Bring? Intensive Study on HyperCLOVA

[통계](#) [수정](#) [삭제](#)

diddu · 10분 전

0

[NLP](#) [논문](#)

논문 리뷰

▼ [목록 보기](#)

1/1



💡 What Changes Can Large-scale Language Models Bring? Intensive Study on HyperCLOVA_ 논문 리뷰

📌 Introduction

연산, Model size, data size가 커질 수록 LM의 성능이 좋아진다는 연구의 논문이 발표된 이후, LM size를 키우는 연구들이 많이 나오게 되었다. 그 중 하나가 GPT-3이다. GPT-3는 In-Context Learning으로 Fine-Tuning없이 다양한 task를 수행할 수 있다.

하지만 이러한 능력이

- 1) 다른 언어에서도 똑같이 적용이 될지,
- 2) Hyper-scale LM으로 무엇을 할 수 있을지

위 두 가지 의문에 대한 논문이 바로 ***What Changes Can Large-scale Language Models Bring? Intensive Study on HyperCLOVA*** 이다.

GPT-3에는 아래의 issue들이 있다.

1. the language composition of the training corpus is **heavily skewed towards English** with 92.7%
2. while it is pragmatic and useful to know the capabilities of various sized models, **we only have access of models of 13B and 175B but none in between.**
3. **advanced prompt-based learning methods have not yet been experimented** for an in-context large-scale LM learner.

이 논문은 위 issue들을 다루고 있다.

Model Structure

GPT-3의 구조에서 크게 변화된 부분은 없다

- Transformer decoder 82B parameter size

Tokenization

- **Morph-aware byte-level BPE**

byte-level encoding + BPE + 형태소 분석기 → 문장을 공백과 형태소 기준으로 미리 나눈 후에, BPE를 사용

Tokenization은 어떻게 텍스트를 학습 가능한 토큰으로 나누는지의 문제이다.

HyperCLOVA는 뜻을 갖는 가장 작은 단위인 형태소(morpheme)를 기반으로 토큰을 나누어 좋은 성능을 낼 수 있게 하였다.

Data Description

GPT-3에서 Korean data가 차지하는 비율은 불과 약 0.02%로 매우 적다.

→ Large Korean corpus를 construct 하는 것이 중요하다.

→ 562B 한국어 데이터

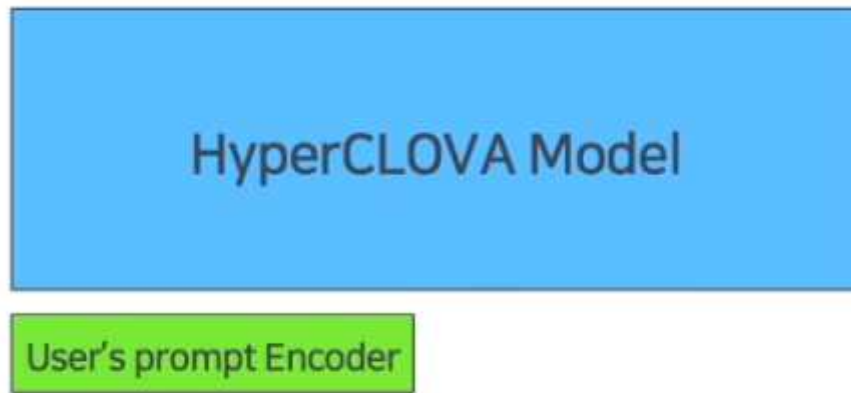
Name	Description	Tokens
Blog	Blog corpus	273.6B
Cafe	Online community corpus	83.3B
News	News corpus	73.8B
Comments	Crawled comments	41.1B
KiN	Korean QnA website	27.3B
Modu	Collection of five datasets	6.0B
WikiEn, WikiJp	Foreign wikipedia	5.2B
Others	Other corpus	51.5B
Total		561.8B

Table 1: Descriptions of corpus for HyperCLOVA

블로그, 카페, 뉴스 등에서 총 562B의 한국어 데이터로 훈련되었다.

📌 Model and Learning

- **Model** : based on megatron - LM
- **Trained on** : NVIDIA superpod
- **Optimizer** : AdamW with cosine learning rate scheduling and weight decay
- **Mini-batch size** : 1024
- **Model size** : set near exponential interpolation from 13B to 175B OpenAI GPT-3
GPT-3의 경우 중규모의 모델 실험이 X, 중규모 모델에선 어떤 성능이 나오는지 보고자 함.
- **Prompt Optimization** : Prompt-Based Tuning + P-tuning



(a) P-tuning

*P-tuning (전체 파라미터 입력단에 추가 파라미터를 두어 성능을 높이는 것)

"p-tuning enables HyperCLOVA to perform comparatives with no parameter update of the main model."

📌 Experimental Result

	NSMC (Acc)	KorQuAD (EA / F1)		AI Hub (BLEU) Ko→En En→Ko		YNAT (F1)	KLUE-STS (F1)
Baseline	89.66	74.04	86.66	40.34	40.41	82.64	75.93
137M	73.11	8.87	23.92	0.80	2.78	29.01	59.54
350M	77.55	27.66	46.86	1.44	8.89	33.18	59.45
760M	77.64	45.80	63.99	2.63	16.89	47.45	52.16
1.3B	83.90	55.28	72.98	3.83	20.03	58.67	60.89
6.9B	83.78	61.21	78.78	7.09	27.93	67.48	59.27
13B	87.86	66.04	82.12	7.91	27.82	67.85	60.00
39B	87.95	67.29	83.80	9.19	31.04	71.41	61.59
82B	88.16	69.27	84.85	10.37	31.83	72.66	65.14

다양한 task에 대해 Fine-tuned된 Baseline에 근접하는 성능을 보이고 있다.

다만, Ko→En task의 성능이 저조하게 나왔는데 이는 학습시킨 데이터에 English corpus가 부족했기 때문일 것 같다.

Discussion on Industrial Impacts

1. HyperCLOVA Studio

: the place for building and communicating the shared artifact generated by HyperCLOVA

- provide a GUI interface
- support API end point

2. Case Studies on HyperCLOVA Studio

- Character bot
- Zero-shot transfer data augmentation
- Title of promoting event

###

태그: 유아동복, 주니어쇼핑몰, 아동원피스,
아동맨투맨, 아동바지, 아동레깅스, 아동모자,
아동가방, 아동양말, 아동신발

(Tags: children's clothes, junior shopping mall, children's
dresses, children's sweatshirts, children's pants,
children's leggings, children's hats, children's bags,
children's socks, children's shoes)

날짜: 2월 26일

(Date: February 26)

제목: 주목받는 신학기 코디제안

(Title: New semester style suggestion for attracting
attention)

###

태그: 스마트워치, 스마트밴드, 웨어러블디바이스
(Tags: Smart watch, smart band, wearable device)

날짜: 12월 7일

(Date: December 7)

제목: 시계도 스마트하게 사용해봐

(Title: Try to use your watch smartly)

###

태그: 커피머신, 에스프레소머신, 커피메이커
(Tags: coffee machine, espresso machine, coffee maker)

날짜: 4월 13일

(Date: April 13)

제목: 커피 한 잔의 여유를 위한 커피머신

(Title: Coffee machine for relaxing with a cup of coffee)

Table 14: Prompt for title of online special sales event.

3. No/Low Code AI Paradigm

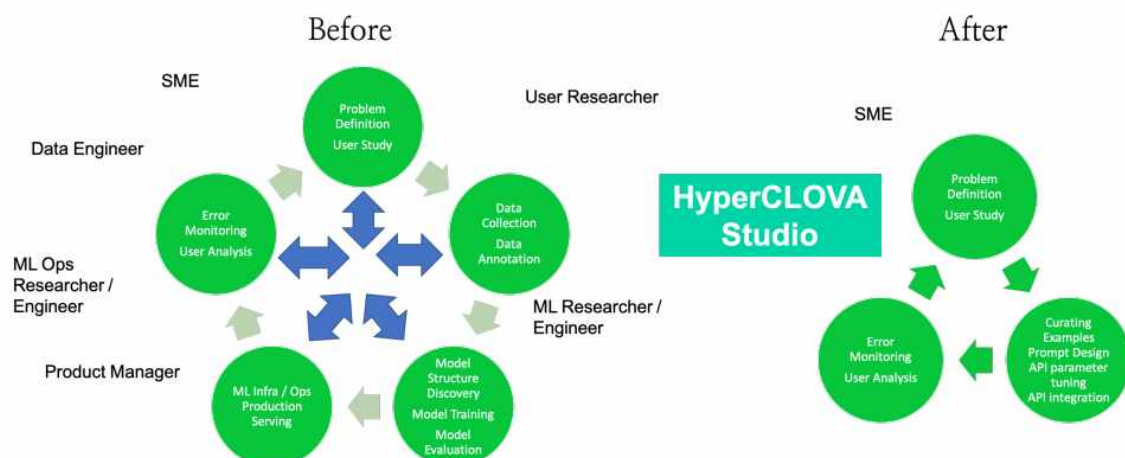
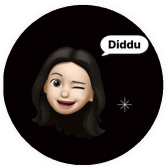


Figure 4: No Code AI paradigm in HyperCLOVA Studio.

📌 Conclusion

- Introduced HyperCLOVA, a **large-scale Korean in-context learning based LM** with nearly 100B parameters, **by constructing a large Korean-centric corpus** of 360B tokens
- Discovered the effect of **language-specific tokenization on large-scale in-context LMs for training corpus of non-English languages**
- Explored **zero-shot and few-shot capabilities of mid-size HyperCLOVA** with 39B and 92B parameters and find that **prompt-based tuning can enhance the performance** outperforming state-of-the-art on downstream tasks
- Argued **the possibility of realizing No Code AI** by designing and applying HyperCLOVA Studio to three in-house applications



Diddu

0개의 댓글

댓글을 작성하세요

댓글 작성

