

네이버 AI플랫폼 CLOVA 그리고 초대규모 AI HyperCLOVA

하정우 (NAVER AI Lab, NAVER CLOVA), 박흥석 · 이바도 · 황민제 (NAVER CLOVA)

목 차	1. 서 론
	2. AI 선행연구 성과
	3. CLOVA AI 플랫폼 및 서비스
	4. HyperCLOVA
	5. 결 론

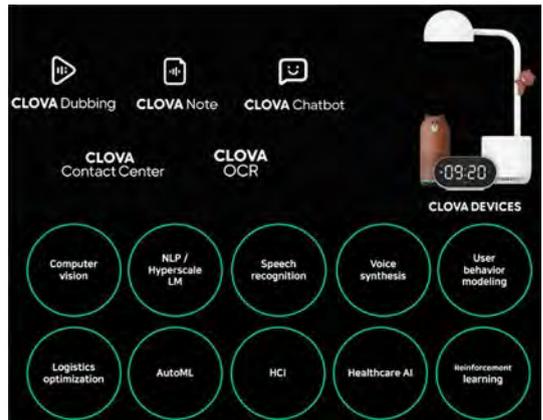
1. 서 론

2010년대 초반 음성과 이미지 인식에서 보여준 딥러닝의 놀라운 성과와 2016년 알파고 쇼크 이후로 인공지능(Artificial Intelligence: AI)은 이제 기술을 넘어 사용자들의 일상 생활의 많은 곳에 녹아들어 사람들이 일하고 살아가는 방식을 완전히 변화시키고 있다. 인공지능 스피커, 이커머스와 동영상 서비스에서의 추천시스템, 번역기앱, 자동차 주행보조기술과 같은 사용자 서비스 뿐 아니라 제조, 금융등 기존 산업에서의 디지털 트랜스포메이션(digital transformation: DT)에서도 AI는 핵심적인 기술로서 활용되고 있다.

이에 따라 구글, 페이스북, 아마존, 마이크로소프트, 엔비디아 등의 글로벌 기술 기업은 물론 바이두, 알리바바와 같은 중국의 기술 기업들 또한 AI연구에 대규모 투자를 진행하고 이 결과물들을 활용한 다양한 제품과 서비스를 제공하고 있다.

국내에서도 네이버를 비롯하여 삼성전자, 카카오, SKT, LG AI연구원, KT 등 많은 기업들이 연구개발한 AI기술을 다양한 서비스에 성공적으로 적용 사례들이 보고되고 있다.

특히 네이버의 AI플랫폼인 CLOVA는 2017년 이후 사용자들의 일상생활을 편리하게 하고 중소기업(SME)와 크리에이터들의 성장을 돕는 강

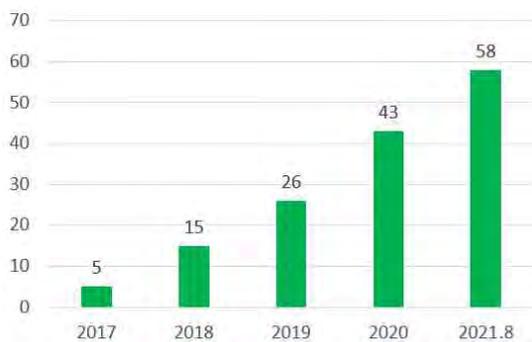


(그림 1) 네이버 CLOVA의 대표 AI기술과 서비스

력한 AI 도구 제공을 위해 그림 1과 같이 음성인식, 합성, 컴퓨터비전, 자연어처리 등 다양한 분야의 AI 기술 및 서비스 개발을 진행하고 있다. 본 논문에서는 네이버 CLOVA의 AI연구 결과와 이에 기반한 플랫폼 서비스 사례를 공유한다. 또한 최근 AI기술의 가장 중요한 화두인 글로벌 초대규모 AI 경쟁과 관련하여 올해 5월에 공개한 초대규모 한국어 언어모델인 HyperCLOVA를 소개하고 향후 방향성에 대해 논의한다.

2. AI 연구 성과

네이버는 지난 6년 이상 매출의 약 25% 이상을 연구개발에 꾸준히 투자해왔으며 그 연구개발의 중심에는 AI가 있다. 특히 네이버 CLOVA는 글로벌 AI기술 경쟁력 강화를 위해 조직 설립 초기인 2017년 초부터 실제적인 AI문제 발굴과 중장기 선행 연구에 기반한 기술 내재화를 위해 그림과 같은 “데이터 → 선행연구 → 엔진개발 → 서비스 및 제품화 → 사용자 → 데이터”로 이어지는 선순환 구조를 구축해왔다. 이러한 노력의 결과로 그림 2와 같이 CVPR, ICLR, NeurIPS, ICML, ACL, EMNLP, ICASSP, AAAI 등 다양한 AI 분



(그림 2) 네이버 CLOVA의 Top AI 학회 정규논문 발표 수

야의 세계 최고 학회에서 연간 수십 편 이상의 정규 논문을 발표하는 등 국내 기업들 중에서 독보적인 성과를 보여주고 있다. 이 중 NAVER AI Lab은 CLOVA의 중장기 선행연구를 담당하는 조직으로 대표적인 연구성과는 아래와 같다.

- 1) StarGAN [6]: CycleGAN [30]과 같은 기존 이미지변환 모델은 변환하는 도메인이 증가하면 조합의 수만큼의 generator가 필요하다는 단점이 있었으나 하나의 generator와 discriminator만 사용하는 구조로 주어진 이미지를 다양한 도메인으로 변환을 가능하게 한 최초의 GAN 모델로서 CVPR 2018에서 구두로 발표하고 1800회가 넘는 인용수를 기록1).
- 2) Cutmix [28]: Cutmix는 기존의 Mixup [29] 대비 ImageNet 분류 정확도 개선 뿐 아니라 다양한 태스크의 전이학습과 적대적 공격에 대한 안정성에서 더욱 개선된 성능을 보인 데이터 증강 기법으로 ICCV 2019에서 구두 발표된 이후 인용수가 현재 700회에 이를 정도로 많은 이미지 인식연구에서 활용되고 있음.
- 3) AdamP [9]: 기존 널리 활용되는 SGD나 Adam이 batch normalization과 momentum을 함께 사용할 경우 비효율성이 있었음. 이를 위해 에러 감소에 영향을 주지 않는 gradient 성분을 제거하는 형태로 개선하여 다양한 컴퓨터 비전, 자연어 처리, 오디오 인식 등의 태스크에서 개선된 효과를 입증한 연구로서 ICLR 2021에서 발표. 옵티마이저(optimizer)는 인공지능망 연구에서 가장 근본적인 연구분야로 국내에서는 연구 사례가 극히 드문데 네이버 CLOVA의 중장기선행연구의 대표적 사례로 인식됨.

1) 모든 인용수는 Google scholar (<https://scholar.google.com/>) 기준

이 외에 많은 연구들이 발표되었으며 많은 연구들의 소스코드와 훈련모델, 데이터들이 CLOVA (<https://github.com/clovaai>)와 NAVER AI Lab (<https://github.com/naver-ai>) github를 통해 공개되고 있다. 또한 이 논문 연구 중 40% 이상이 클로바 노트, 클로바 더빙, OCR, 클로바케어콜 등 실제 네이버의 서비스에 직간접적으로 적용되어 사용자들의 생활에 가치를 제공하고 있다.

3. CLOVA AI 플랫폼 및 서비스

3.1 OCR

광학문자인식(Optical Character Recognition: OCR)은 이미지나 비디오등 시각 데이터로부터 텍스트를 추출 및 인식하는 기술로서, 로봇 프로세스 자동화(RPA: Robot Process Automation)의 기술 요소로써 기존의 문서 기반의 사무에서 DT에 핵심적인 역할을 담당하고 있다(그림 3).

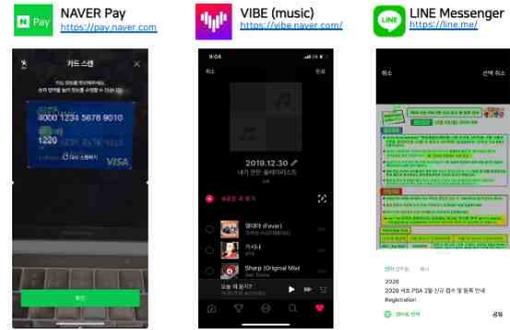
네이버 CLOVA OCR팀은 RPA에 필요한 다양한 코어 기술을 개발하고 있으며 크게는 이미지 문서에 대한 시각-언어 분석, 학습 또는 문서 이미지 합성을 위한 데이터 생성/합성, 정형화된 문서에 대하여 정보를 추출하는 양식 인식 기술을 보유하고 있다.



(그림 3) OCR과 RPA의 관계

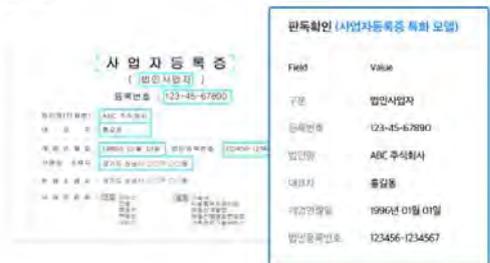
3.1.1 OCR과 Parsing

OCR은 텍스트 영영검출(detection)과 텍스트 인식(recognition)으로 이어지는 파이프라인이며 영역 검출의 경우 CRAFT [2] 모델을 사용하고 있다. 이 모델의 특징은 글자 단위의 검출이 가능하다는 점인데, 일반적으로 OCR 학습데이터에는 글자 단위의 Annotation 대신에 단어 단위의 Annotation이 일반적이다. 이러한 상황을 해결하기 위하여 CRAFT에서는 학습 중간 모델을 이용하여 단어 단위의 데이터를 글자 단위의 Annotation (Pseudo Label)로 변환하여 학습을 진행한다. 텍스트 인식을 위해 인식기의 각 단계별 요소들의 조합 성능을 파악하여 가장 좋은 성능을 내는 모델을 활용하고 있다[1].



사업자 등록증

사업자 등록증 문서를 자동 분석하여 등록번호, 법인명, 대표사 등 사업자 등록증의 기재된 정보를 추출합니다.



(그림 4) OCR 및 Parsing Application 이 이용되는 다양한 사례

Parsing 기술의 경우 BERT [7] 기반의 언어모델을 활용하고 있으며, 문서로부터 인식된 다양한 텍스트들의 의미를 추출하는 텍스트 분류 문제를 풀어내는 기술이다. 실제 사용자가 필요로 하는 서비스 적용에 있어서 필수적인 기술이며 그림 4와 같이 영수증 인식, 명함 인식, 사업자등록증 인식, 신분증 인식 등 다방면에 활용되고 있다.

3.1.2 데이터 생성

OCR 학습을 위해서는 대량의 데이터가 필수적인데, 이에 해당하는 모든 데이터를 실제 이미지들로 구축하는 것은 시간적 물리적 비용적 한계가 존재한다. 이 때문에, 상당수의 데이터는 사전학습(pretraining)을 위하여 생성된 데이터를 활용하게 되는데, 이를 위하여 개발된 DM-Font [5], MX-Font [20] 등의 요소 기술들은 적은 양의 데이터로도 11,172 글자의 한글을 모두 지원하는 폰트 생성 서비스, 네이버의 대문을 책임지는 영수증 이미지 Captcha 등에 활용되고 있다.

3.1.3 양식, 테이블, 개체명 인식

양식 인식(form recognition)을 위해서는 문서에서 템플릿 매칭을 하기 위한 Key-value 추출, 템플릿 매칭, Image Unwarping 등 다양한 기술들이 사용된다. 일반적인 비정형 문서의 경우 구조 정보를 담고있는 테이블에 대한 인식이 RPA에서 중요한 역할을 담당하고 있는데, 네이버에서는 요소들을 검출하여 전체 구조를 잡아나가는 방식, 점과 선 등의 시각 정보들 또는 텍스트의 상대적인 위치 정보를 이용하여 테이블을 재구성하는 방식 등으로 테이블을 추출하는 기술도 보유하고 있다. 마지막으로 개체명(entity) 인식의 경우 문서 상에 존재할 수 있는 다양한 형태의 데이터

(Multibox, Checkbox 등)에 대하여 정보를 추출해낼 수 있는 기술이다. CLOVA Visual AI 팀에서는 앞서 설명된 기술들 뿐만 아니라, 문서의 종합적인 이해를 위한 다양한 최신 기술에 대한 개발을 진행하고 있다.

3.2 음성합성

TTS(text-to-speech) 시스템은 텍스트를 사람의 목소리로 변환하는 기술로서, 스크린 리더, 오디오북, 대화형 인공지능과 같은 다양한 음성 인터페이스의 핵심 기술이다. 네이버의 TTS 엔진인 nVoice는 UTS, HDTS, NES²⁾ 음성 합성 기술을 보유하고 있으며, 스마트스피커 CLOVA를 비롯하여 네이버 뉴스, 네이버 지도, 파파고, 오디오클럽 등 다양한 서비스에 명료하고 자연스런 합성음을 제공하고 있다.

이 중 HDTS 시스템은 딥러닝 기반의 AI 음성 합성 기술로, 적은 양의 녹음으로도 다양한 스타일의 고품질 음성을 생성하는 것을 목표로 하고 있다. HDTS 시스템은 그림 5와 같이 크게 (1) Text analyzer, (2) Acoustic model, (3) Neural vocoder의 세가지 모델로 구성된다.



(그림 5) HDTS 시스템 개요

3.2.1 텍스트 분석기

텍스트 분석기(text analyzer)는 자연어 처리 기술을 이용해 입력 텍스트를 분석하여, 음소

2) UTS: unit-selection TTS
 HDTS: high-quality DNN TTS
 NES: natural end-to-end speech synthesis

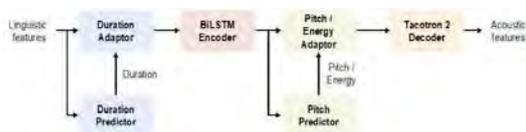
(phoneme) 정보, 성조(tone) 정보, 운율 정보, 끊어 읽기(break) 정보와 같은 언어적 특징을 나타내는 특징 벡터(linguistic feature) 를 추정하는 모델이다.

CLOVA의 음성 합성 팀은 text analyzer의 중요 시스템인 G2P (grapheme-to-phoneme)을 위해 신경망 기반의 NN-KoG2P 시스템 [13] 을 개발하였고, 이를 HDTS 시스템에 적용하여 음소 추정 정확도를 기존 룰 기반 G2P 시스템 대비 15% 가량 향상시켰다. 이 밖에도 딥러닝 기반 끊어 읽기 추정 시스템을 개발하여 합성기의 끊어 읽기 성능을 향상시켰다. 또한 최근에는 BERT [7] 나 GPT-3 [4] 와 같은 언어모델을 이용해 기쁨, 슬픔과 같은 문장의 감정 세기를 추정하고, 이렇게 추정된 각 감정의 정도를 조합해 감정 조절을 가능하게 하는 TTS 시스템을 구현하였다.

3.2.2 Acoustic model

Acoustic model은 그림 6과 같이 앞 단계에서 추정된 linguistic feature 로 부터 음색 (voice color), 목소리의 높낮이(pitch), 유/무성 정보 (voicing information), 음성 에너지(energy) 등을 표현하는 음향 특징 벡터(acoustic feature) 를 추정하는 시스템이다.

HDTS 시스템의 acoustic model 은 Tacotron2 [23] 의 구조에 기반하고 있다. Tacotron과 같은 어텐션 attention) 메커니즘 기반의 acoustic model 은 간헐적으로 발생하는 어텐션 에러로 인해 시스템이 불안정해지는 문제를 겪고 있다 [16].



(그림 6) HDTS 시스템의 acoustic model 구조

이러한 문제를 해결하기 위해 HDTS 시스템에서는 기존의 어텐션 메커니즘을 발화 길이(duration) predictor 로 대체하였다. 각 음소의 발화 길이를 딥러닝 모델로 추정할 수 있으므로, 추정된 발화 길이로 linguistic feature 과 acoustic feature 사이의 시간 정보를 직접 동기화 할 수 있게 되었다. 따라서 어텐션 에러의 위험을 미연에 방지하면서도 Tacotron 보다 높은 품질의 음성을 안정적으로 생성할 수 있게 되었다.

최근에는 HDTS 시스템에 FastSpeech2 시스템 [22] 에서 제안된 것과 같이 energy, pitch, 그리고 발화 길이 정보를 직접 조절 가능한 variance adaptor 기술을 적용하였고, 이를 통해 사용자가 직접 음성의 운율적 특징을 조절할 수 있는 음성 합성기를 구현하였다.

3.2.3 Neural vocoding

Neural vocoding 단계에서는 앞서 추정된 acoustic feature 를 음성의 파형 신호로 변환시켜 주는 과정이 수행된다.

HDTS 시스템에서는 음성 합성 태스크를 실시간 (real-time) 생성과 비-실시간 생성으로 구분한 후, 태스크에 따라 다른 종류의 뉴럴 보코더 (neural vocoder) 를 사용한다. 먼저 비-실시간 음성 합성에서는 속도가 느리더라도 높은 품질의 합성음을 생성 가능한 자기 회기(auto-regressive) WaveNet 보코더 [17]를 이용한다. 이 때 보코더의 품질을 더욱 향상시키기 위해 Clova 음성 합성 팀에서는 전통적 음성 신호처리 기법인 소스-필터 (source-filter) 이론을 WaveNet 에 적용시킨 LP-WaveNet 보코더 [10]를 개발하였고, 결과적으로 기존 WaveNet 보코더 대비 합성음 품질을 34% 향상시킬 수 있었다.

최근에는 AiCall 과 같은 AI for Contact Center

[8] 서비스에서 실시간 음성 합성 서비스의 수요가 증가하는 추세이고, 이에 따라 실시간 음성 합성이 가능한 뉴럴 보코더를 개발하였는데, non-causal WaveNet 보코더와 GAN(generative adversarial networks) 프레임워크를 결합시킨 PWG (Parallel WaveGAN) 보코더 [26]가 그 예시이다. PWG는 LP-WaveNet 대비 18,000 배 빠른 합성속도를 보여주며 실시간 합성이 가능하단 장점이 있지만, 그만큼 낮은 품질의 음성 파형을 생성한다는 문제가 있다. 이러한 문제를 해결하기 위해 Clova 음성 합성 팀에서는 (1) 사람 목소리의 주파수 영역 특징을 PWG 의 loss 에 반영한 훈련법 [24], (2) 유/무성음을 독립적으로 판별하는 voicing-aware discriminator [27], (3) HN (harmonic-plus-noise) 파형 모델을 적용한 HN-PWG [11] 등 다양한 시스템을 개발해 왔으며, 결과적으로 실시간 합성 속도를 유지하며 WaveNet 보다 더 좋은 품질을 제공하는 합성기를 개발하는 것에 성공하였다 [11].

앞서 설명한 HDTS 음성 합성 시스템은 네이버의 뉴스 본문 듣기 서비스, CCAI 서비스, 그리고 네비게이션 서비스 등의 다양한 음성 활용 서비스에 적용되고 있다. 끝으로, Clova 음성 합성 팀은 expressive TTS, style transfer, 혹은 voice

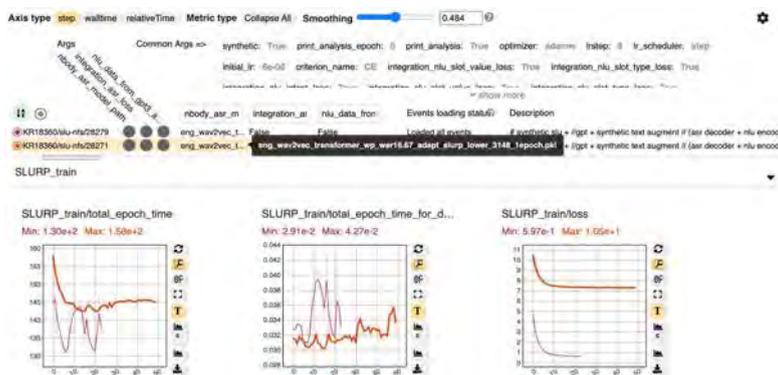
conversion 등과 같은 다양한 최신 음성 합성 기술의 연구 개발 진행하고 있고, 이를 AI스피커, 네이버 뉴스 읽기, CLOVA Dubbing (<https://clovadu.bbing.naver.com/>) 등 음성 서비스에 적용해 새로운 가치를 창출하고 있다.

3.3 NSML

NSML은 머신러닝 연구자들의 모델 연구 및 개발에 최적화된 클라우드 기반의 머신러닝 플랫폼이다 [12, 25]. 일반적인 머신러닝 모델 개발 과정부터 하이퍼파라미터 최적화를 위한 대규모 반복 및 분산 실험 환경을 지원하며, 연구자들과 활발한 협업을 위한 리더보드, URL 기반의 편리한 연구 과정 및 결과 공유 환경 등을 지원하고 있다.

NSML은 일반적인 머신러닝 플랫폼의 기능들을 제공할 뿐만 아니라, 고도화된 여러 시각화 시스템 (Visual Analytics) 을 제공하여 효율적인 모델 분석 환경을 제공한다.

Scalar 시각화 시스템은 머신러닝 모델이 학습 과정에서 기록한 여러 성능 지표들을 시간에 따라 변화하는 과정들을 시각화하여 모델의 하이퍼파라미터에 따른 학습 양상을 분석할 수 있는 시각화 분석 시스템이다 (그림 7). 또한 Scalar 시각화



(그림 7) NSML 상에서 각종 실험결과 시각화 예시

시스템은 다수의 서로 다른 머신러닝 모델들의 하이퍼파라미터들을 직관적으로 비교할 수 있도록 시각화하여 연구자들이 하이퍼파라미터의 효과를 분석하고, 실험 전략을 효율적으로 제어할 수 있는 환경을 제공한다. [15].

HyperTendril 시각화 시스템은 NSML 에서 제공하고 있는 AutoML의 결과를 분석할 수 있는 시스템이다 [15,18,19]. HyperTendril은 AutoML이 탐색한 search space의 결과를 단순히 시각화할 뿐만 아니라, 최적화 하려는 모델의 성능에 영향력이 높은 하이퍼파라미터와 하이퍼파라미터의 효과적인 범위의 값들을 시각적으로 가이드를 제공해주는 특징을 가지고 있다. 또한, 서로 다른 특징을 가진 여러 탐색 algorithm 들이 어떠한 방식으로 search space를 탐색하는지, 높은 성능의 지속적 기록 여부와 같은 시각화 모듈을 제공하여 사용자의 모델에 적합한 AutoML configuration을 설정할 수 있도록 통찰을 제공해주는 특징을 가지고 있다.

끝으로, NSML은 그 사용성과 효율성을 검증하기 위해 사용자들의 clickstream, 1 on 1 in-depth interview, focused-group interview 등 여러 HCI 방법론을 통해 지속적으로 플랫폼의 사용성을 평가하고 있으며, 연구자들의 숨은 요구를 발굴하여 효율적인 모델 개발에 필요한 환경들을 개발하고 있다.

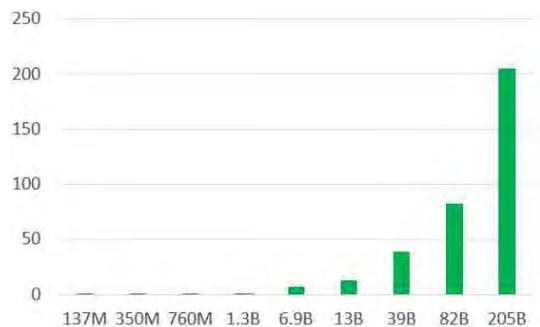
3.4 그 외 중요 CLOVA AI 서비스

소개된 3가지 기술외에도 CLOVA 는 다양한 AI 서비스를 클라우드 혹은 앱형태로 제공하고 있다. 먼저 end-to-end 음성인식기술 (NEST) 을 적용하여 회의가 진행되는 동안의 음성대화를 참가자별로 구분하여 자동으로 전사하는 앱인 CLOVA Note가 있다. 그리고 콜센터에서 사람대

신 정해진 응답을 하거나 예약을 대신 받아주는 AI인 CLOVA AICC 기술은 신한은행, 골프존 등 다양한 파트너기업에 적용되어 있다. 특히 AICC 기술은 비즈니스를 넘어 코로나19 능동감시자들의 건강상태를 확인하는 기술로 적용되었다. 이 CLOVA 케어콜 서비스는 2020년 3월 성남시에 처음 적용된 이후 많은 지방자치단체에 적용되어 있으며 이를 더욱 확장하여 독거노인들의 안부전화를 대신 걸어주는 AI인 CLOVA 안부콜은 부산 해운대구와 현재 시범서비스를 진행중이다. 이러한 CLOVA 케어콜과 안부콜은 ESG관점에서 AI for Social Good의 대표적인 사례로 인식되고 있다. 또한 네이버와 전략적 파트너를 맺은 CJ의 물류 수요예측을 위한 AI인 CLOVA 포캐스트도 수요예측 최적화를 통해 비용절감에 기여하고 있다.

4. HyperCLOVA

HyperCLOVA [14]는 네이버 클로바가 자체적으로 개발한 한국어 중심의 초대규모 언어모델이다. 2020년 OpenAI가 공개한 이후 GPT-3[4]는 전세계 많은 AI 연구자들에게 엄청난 충격을 선사했다. 특히 1750억개 매개변수를 가진 GPT-3가 보여주는 매우 다양한 자연어이해 및 생성 태스크에서 추가적인 fine-tuning 학습없이 프롬프트만



(그림 8) 다양한 크기의 HyperCLOVA 모델들

으로 in context few-shot learning을 통해 기존 모델 수준의 성능을 보여줌으로써 기존의 학습 패러다임을 변화시키고 있다는 평가를 받고 있다. 그러나 기존 GPT-3는 학습데이터의 대부분이 영어로 구성되어 있고 한국어 비중은 0.016%에 불과하여 한국어로 된 데이터나 문제에 활용하는 것은 사실상 불가능하다.

GPT-3의 뛰어난 성능을 고려하면 국내의 사용자들에게 혁신적 기술과 AI도구 제공을 위해 수백억개 이상의 매개변수를 포함하는 한국어 중심의 언어모델을 구축하는 것이 필요했다. 이에 네이버는 기술 플랫폼 기업으로서 AI주권 수호 관점에서 2020년 9월부터 대규모 인프라 투자와 한국어 중심의 데이터 구축 그리고 연구를 통해 초대규모 한국어 생성 언어모델인 HyperCLOVA를 개발하고 지난 5월에 그 결과를 공개했다³⁾.

HyperCLOVA 학습을 위해 네이버 자체적으로 구축한 데이터는 5600억개의 token을 포함하고 있으며 한국어의 비중은 97%를 차지하여 전세계에서 가장 뛰어난 한국어 이해 및 생성 성능을 자랑한다. 그림 8과 같이 최대 2050억개의 매개변수를 가진 모델을 포함하여 다양한 크기의 모델들이 개발되었고 이 가운데에 390억개와 820억개 매개변수 모델은 이미 검색, 쇼핑 등의 서비스에 적용되어 있다. 또한 기계학습이나 인공지능 지식이나 경험이 거의 없는 개발자나 기획자들도 손쉽게 HyperCLOVA를 활용하여 AI서비스 시제품을 만들 수 있는 HyperCLOVA-Studio를 개발하여 사내 베타테스트를 진행하고 있으며 No Code AI의 가능성을 확인하고 있다. 표 1은 네이버 쇼핑의 기획전을 제목 생성을 위해 해당 상품의 태그와 날짜가 입력으로 주었을 때 HyperCLOVA (39B)이 생성해낸 제목들을 보여준다. 현재 기획자들이 판단

한 서비스 노출 적합도 채택율이 99%에 이른다.

또한 개발된 HyperCLOVA의 정확한 성능 평가 및 한국어 자연어처리 연구 생태계 저변확대를 위해 카이스트, 뉴욕대, 업스테이지, 워이드, 한국경제신문, 셀렉트스타, 카카오 등과 함께 최초의 한국어이해 평가 데이터인 KLUE를 제작하고 공개했다 [21].

초대규모 언어모델은 그 놀라운 성능에도 불구하고 편향성, Hallucination 등과 같은 한계점을 내포하고 있으며 이미지나 비디오와 같은 멀티모달로의 확장 등 더 많은 개선 가능성을 갖고 있다 [3]. 이를 위해 네이버는 서울대, KAIST와 각각 공동연구센터를 설립하고 협력연구를 수행 중이다. 또한 AI의 성능이 개선되면서 신뢰가능성이 더욱 중요해짐에 따라 서울대 SAPI와 AI 윤리준칙을⁴⁾ 만들어 지난 2월에 공개했고 University of Tuebingen과 함께 신뢰가능한 AI 공동연구를 수행하고 있다.

〈표 1〉 HyperCLOVA로 생성한 네이버 쇼핑의 기획전 제목

태그: 땀땀이웃, 크리스마스, 따뜻한강아지이웃, 강아지코스튬

날짜: 12월 23일

생성제목: 겨울시즌 땀땀이를 위해

#####

태그: 명품구두, 여자들의로망, 여름구두

날짜: 7월 7일

생성제목: 보기만해도 행복한 명품 슈즈

#####

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

날짜: 2월 26일

생성제목: 주목받는 신학기 코디제안

3) <https://naver-ai-now.kr/>

4) <https://www.navercorp.com/value/aiCodeEthics>

5. 결 론

많은 전문가들이 AI는 PC, 인터넷, 모바일 기술이 가져온 변화 이상으로 사람들의 삶을 혁신적으로 바꿀 것으로 예상하고 있다. 이로 인해 전세계의 국가와 기업들간의 AI 기술 경쟁이 점점 더 심화되고 있는 상황이다. 네이버 CLOVA는 국내 대표 AI플랫폼기업으로서 꾸준한 연구개발 투자를 통해 음성인식, 음성합성, 컴퓨터비전, 자연어 처리, 사용자이해 및 AI인프라 기술과 서비스를 플랫폼 사용자들과 구성원들에게 제공함으로써 지속가능한 동반성장을 추구하고 있다. 그리고 나날이 격화되고 있는 글로벌 경쟁 속에서 AI주권 수호를 위해 국내외 대학, 연구소, 기업들과 함께 초대규모 AI 생태계를 구축하고 북미 중국과 어깨를 나란히 하는 제3의 AI 세력을 만들기 위해 노력을 경주하고 있다.

참 고 문 헌

- [1] Jeonghun Baek, Geewook Kim, Junyeop Lee, Sungrae Park, Dongyoon Han, Sangdoo Yun, Seong Joon Oh, Hwalsuk Lee. What Is Wrong With Scene Text Recognition Model Comparisons? Dataset and Model Analysis. ICCV 2019.
- [2] Youngmin Baek, Bado Lee, Dongyoon Han, Sangdoo Yun, Hwalsuk Lee. Character Region Awareness for Text Detection. CVPR 2019.
- [3] Rishi Bommasani et al. On the Opportunities and Risks of Foundation Models. arXiv:2108.07258, 2021.
- [4] Tom Brown et al., "Language Models are Few-shot Learners," NeurIPS 2020.
- [5] Junbum Cha, Sanghyuk Chun, Gayoung Lee, Bado Lee, Seonghyeon Kim, Hwalsuk Lee. Few-shot Compositional Font Generation with Dual Memory. ECCV 2020.
- [6] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, Jaegul Choo. "StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation." CVPR 2018.
- [7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." NAACL 2019.
- [8] Jung-Woo Ha et al. ClovaCall: Korean Goal-Oriented Dialog Speech Corpus for Automatic Speech Recognition of Contact Centers. Interspeech 2020.
- [9] Byeongho Heo, Sanghyuk Chun, Seong Joon Oh, Dongyoon Han, Sangdoo Yun, Gyuwan Kim, Youngjung Uh, Jung-Woo Ha. "AdamP: Slowing Down the Slowdown for Momentum Optimizers on Scale-invariant Weights." ICLR 2021.
- [10] Min-Jae Hwang, Frank Soong, Eunwoo Song, Xi Wang, Hyeonjoo Kang, Hong-Goo Kang. "LP-WaveNet: Linear Prediction-based WaveNet Speech Synthesis." APSIPA 2020.
- [11] Min-Jae Hwang, Ryuichi Yamamoto, Eunwoo Song, Jae-Min Kim. "High-fidelity Parallel WaveGAN with Multi-band Harmonic-plus-Noise Model." Interspeech, 2021.
- [12] Hanjoo Kim, Minkyu Kim, Dongjoo Seo, Jinwoong Kim, Heungseok Park, Soeun Park, Hyunwoo Jo, KyungHyun Kim, Youngil Yang, Youngkwan Kim, Nako Sung, Jung-Woo Ha. "NSML: Meet the MLaaS platform with a real-world case study." arXiv:1810.09957, 2018.

- [13] Hwa-Yeon Kim, Jong-Hwan Kim, Jae-Min Kim. "NN-KoG2P: A Novel Grapheme-to-Phoneme Model for Korean Language." ICASSP, 2021.
- [14] Boseop Kim et al. What Changes Can Large-scale Language Models Bring? Intensive Study on Billions-scale Korean Generative Pretrained Transformers. EMNLP 2021.
- [15] Jinwoong Kim, Minkyu Kim, Heungseok Park, Ernar Kusdavletov, Dongjun Lee, Adrian Kim, Ji-Hoon Kim, Jung-Woo Ha, Nako Sung. "CHOPT: Automated hyperparameter optimization framework for cloud-based machine learning platforms." arXiv:1810.03527. 2018.
- [16] Takuma Okamoto, Tomoki Toda, Yoshinori Shiga, Hisashi Kawai. "Tacotron-based Acoustic Model using Phoneme Alignment for Practical Neural Text-to-Speech Systems." ASRU 2019.
- [17] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu. "WaveNet: A Generative Model for Raw Audio." 9th ISCA Speech Synthesis Workshop, 2016.
- [18] Heungseok Park, Jinwoong Kim, Minkyu Kim, Ji-Hoon Kim, Jaegul Choo, Jung-Woo Ha, Nako Sung. "VisualHyperTuner: Visual Analytics for User-driven Hyperparameter Tuning of Deep Neural Networks." Demo@SysML Conference. 2019.
- [19] Heungseok Park, Yoonsoo Nam, Ji-Hoon Kim, Jaegul Choo. "HyperTendrill: Visual Analytics for User-Driven Hyperparameter Tuning of Deep Neural Networks." IEEE Trans. on Visualization and Computer Graphics 27(2). 2021.
- [20] Song Park, Sanghyuk Chun, Junbum Cha, Bado Lee, Hyunjung Shim. Multiple Heads are Better than One: Few-shot Font Generation with Multiple Localized Experts. AAAI 2021.
- [21] Sungjoon Park et al. KLUE: Korean Language Understanding Evaluation. arXiv:2105.09680. 2021.
- [22] Yi Ren, Chenxu Hu, Xu Tan, Tao Qin, Sheng Zhao, Zhou Zhao, Tie-Yan Liu. "FastSpeech 2: Fast and High-quality End-to-End Text to Speech." ICLR 2021.
- [23] Jonathan Shen, Ruoming Pang, Ron J. Weiss, Mike Schuster, Navdeep Jaitly, Zongheng Yang, Zhifeng Chen, Yu Zhang, Yuxuan Wang, RJ Skerry-Ryan, Rif A. Saurous, Yannis Agiomyrgiannakis, Yonghui Wu. "Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions." ICASSP 2018.
- [24] Eunwoo Song, Ryuichi Yamamoto, Min-Jae Hwang, Jin-Seob Kim, Ohsung Kwon, Jae-Min Kim. "Improved Parallel WaveGAN vocoder with Perceptually Weighted Spectrogram Loss." SLT 2021.
- [25] Nako Sung, Minkyu Kim, Hyunwoo Jo, Youngil Yang, Jingwoong Kim, Leonard Lausen, Youngkwan Kim, Gayoung Lee, Donghyun Kwak, Jung-Woo Ha, Sunghun Kim. "NSML: A machine learning platform that enables you to focus on your models." MLSYS WS@NIPS 2017.
- [26] Ryuichi Yamamoto, Eunwoo Song, Jae-Min Kim. "Parallel WaveGAN: A Fast Waveform Generation Model based on Generative Adversarial Networks with Multi-resolution Spectrogram." ICASSP, 2020.
- [27] Ryuichi Yamamoto, Eunwoo Song, Min-Jae Hwang, Jae-Min Kim. "Parallel

Waveform Synthesis Based on Generative Adversarial Networks with Voicing-aware Conditional Discriminators.” ICASSP 2021.

- [28] Sangdoon Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, Youngjoon Yoo. “CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features.” ICCV 2019.
- [29] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, David Lopez-Paz. mixup: Beyond Empirical Risk Minimization. ICLR 2018.
- [30] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros. “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks.” ICCV 2017.



박 흥 석

이메일 : heungseok.park@navercorp.com

- 2016년 아주대학교 미디어학과 (학사)
- 2018년 KAIST 문화기술대학원 (석사)
- 2018년~현재 NAVER CLOVA software engineer
- 관심 분야: Visual analytics, Machine learning, Explainable AI



이 바 도

이메일 : bado.lee@navercorp.com

- 2003년 3월~2010년 2월 서울대학교 전기공학부 (학사)
- 2010년 3월~2012년 2월 서울대학교 컴퓨터공학부 Biointelligence Lab (석사)
- 2012년 8월~2017년12월 삼성전자 무선사업부 Multimedia 개발그룹
- 2018년 1월~현재 Naver Clova OCR 리더
- 관심분야: computer vision, OCR, document understanding

저 자 약 력



하 정 우

이메일 : jungwoo.ha@navercorp.com

- 2004년 서울대학교 컴퓨터공학부 (학사)
- 2015년 서울대학교 전기컴퓨터공학부 (박사)
- 2015년~2016년 네이버랩스 Tech Lead
- 2017년~2020년 네이버 CLOVA AI Research 리더
- 2020년~현재 네이버 AI Lab 연구소장
- 2021년~현재 서울대-네이버 초대규모 시연구센터 공동 센터장
- 2021년~현재 카이스트-네이버 초창의적 시연구센터 공동 센터장
- 2021년~현재 시미래포럼 공동의장
- 관심분야: AI, machine learning, computer vision, NLP, audio signal modeling, recommender system



황 민 제

이메일 : min-jae.hwang@navercorp.com

- 2011년 3월~2015년 8월 연세대학교 전기전자 공학부 (학사)
- 2015년 8월~ 2020년 2월 연세대학교 전기전자 공학부 대학원 (박사)
- 2017년 12월~2017년 12월 Naver Clova, Voice 팀, 인턴
- 2018년 1월~2018년 11월 Microsoft Research Asia, Speech 팀, 인턴
- 2019년 4월~현재 Naver Clova, Voice&Avatar 팀, Research scientist
- 관심 분야: Deep generative model-based speech synthesis, Neural waveform generation