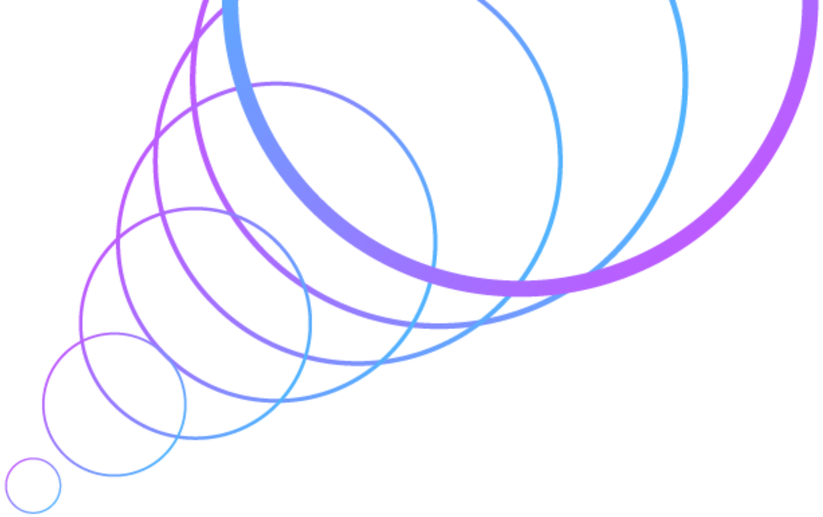


Bring Your Own Data: Business AI 고민?
HyperCLOVA에게 무엇이든 물어보살

김경민/곽하녹/신규용 NAVER CLOVA / ML X

CONTENTS

1.Business AI 제품 양산의 가능성

1.1 AI 모델 개발 파이프라인

고급레스토랑

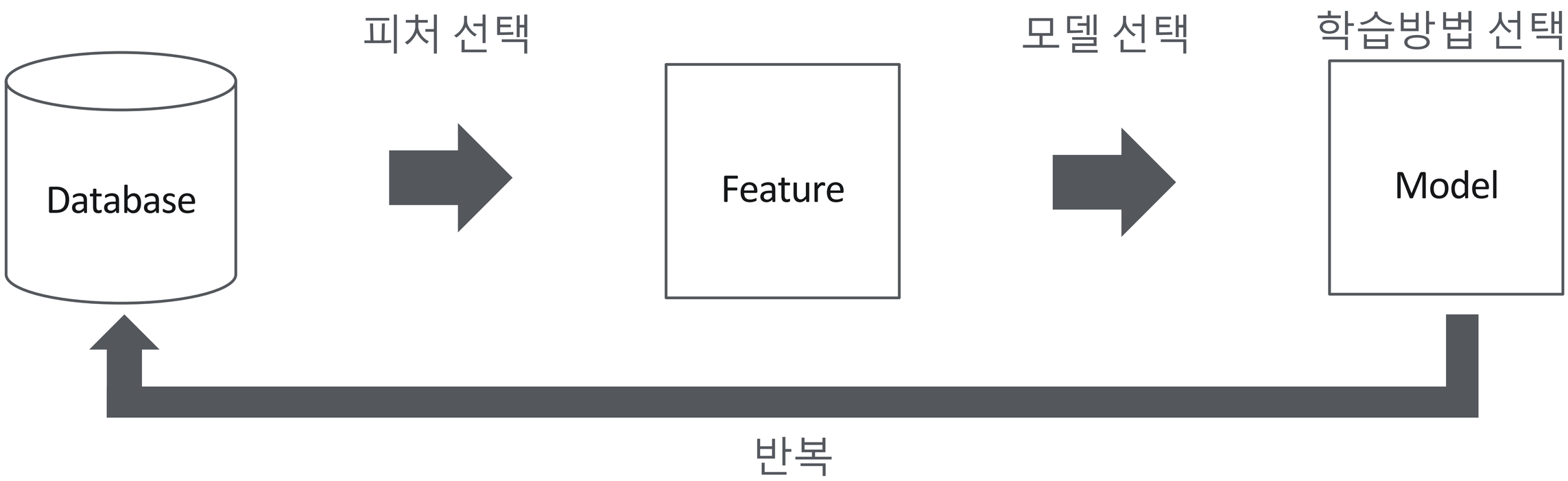
VS.

패스트푸드 식당

[출처: 위키피디아, CC 라이센스] The Ritz-Carlton Hong Kong Tin Lung Heen View from 103, https://commons.wikimedia.org/wiki/File:The_Ritz-Carlton_Hong_Kong_Tin_Lung_Heen_View_from_103.jpg
[출처: 위키피디아, CC 라이센스] Fast food restaurant in the port of Malinska, https://commons.wikimedia.org/wiki/File:Fast_Food_Restaurant_in_Malinska_im_Hafen.jpg

1.1 AI 모델 개발 파이프라인

- AI 모델을 직접 만드는 과정
- 쉽지 않은 AI 요리

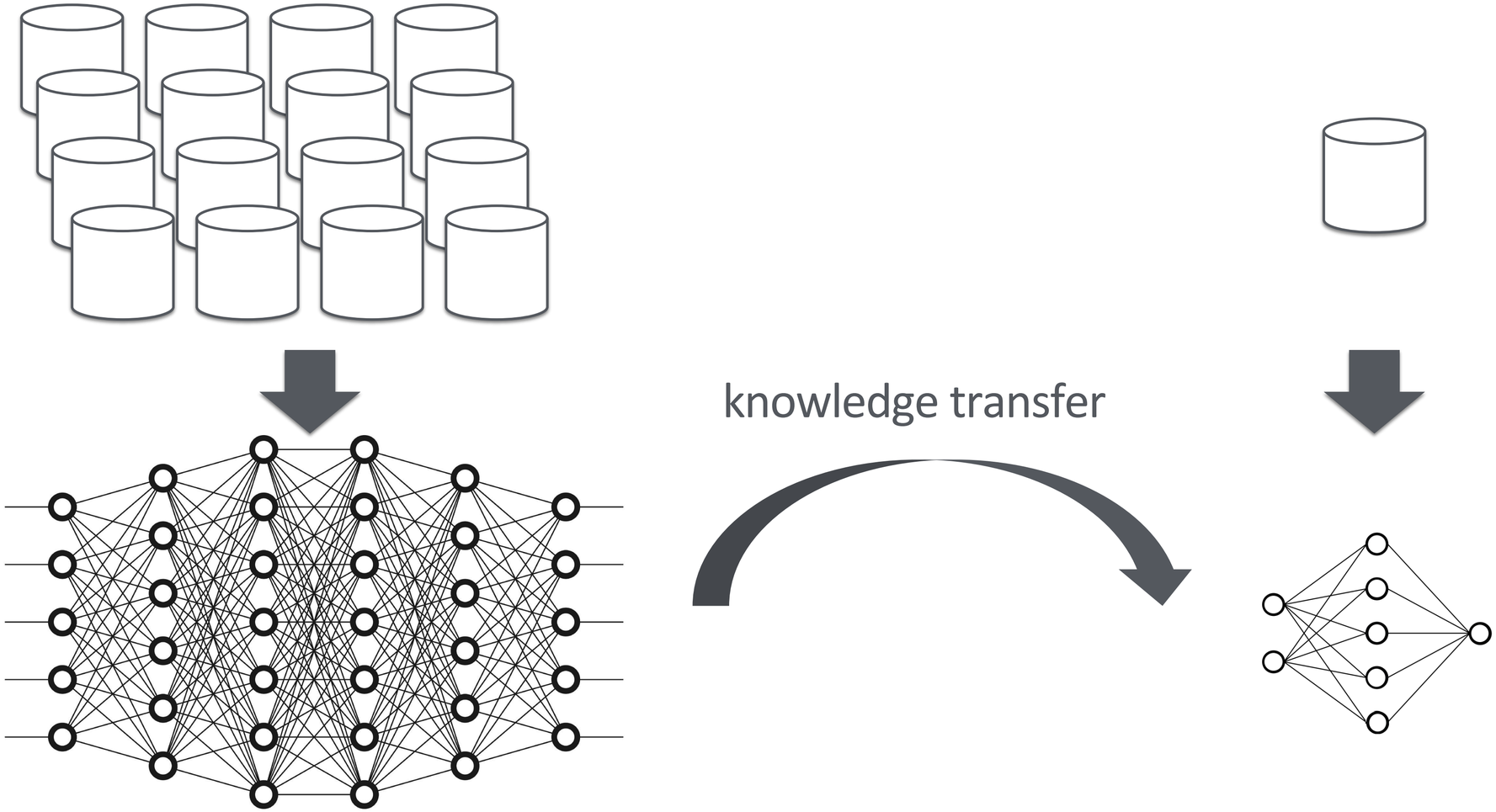


1.2 최신 AI 연구 트렌드 TL;DR

#선크게후고민

BIG

1.2 최신 AI 연구 트렌드: Pretrained & Finetuning



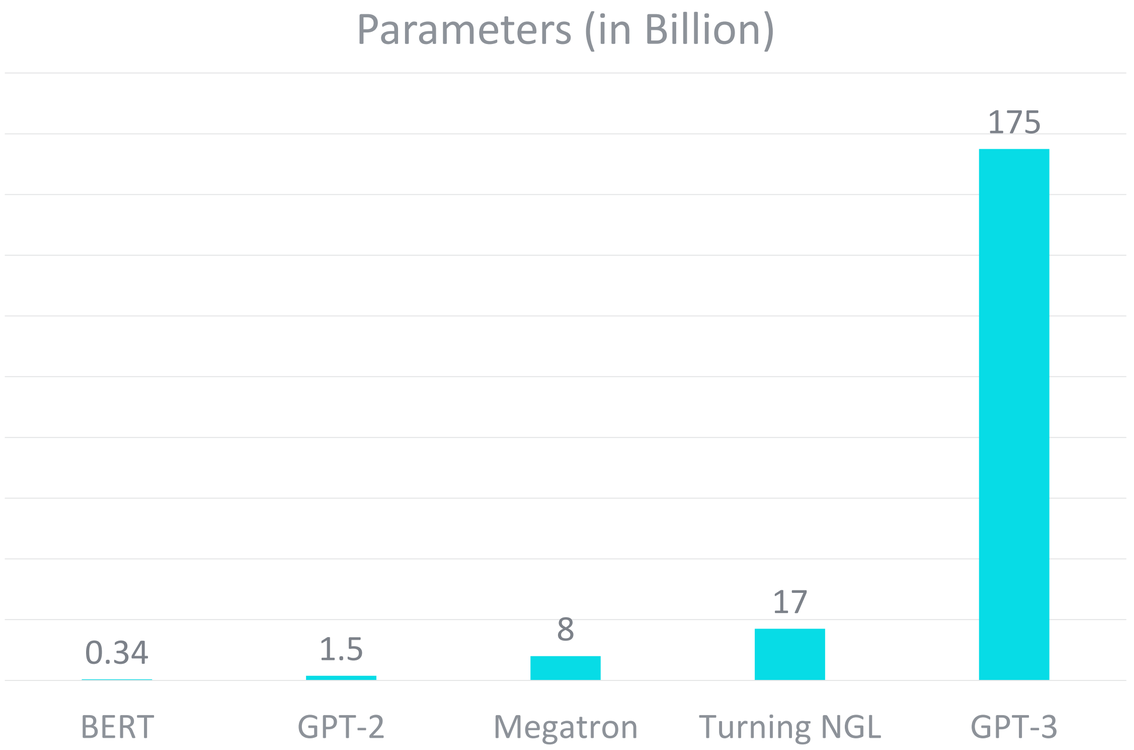
1.2 최신 AI 연구 트렌드: 육수 끓이기

#육수는_필수_but_모두가_끓일필요X

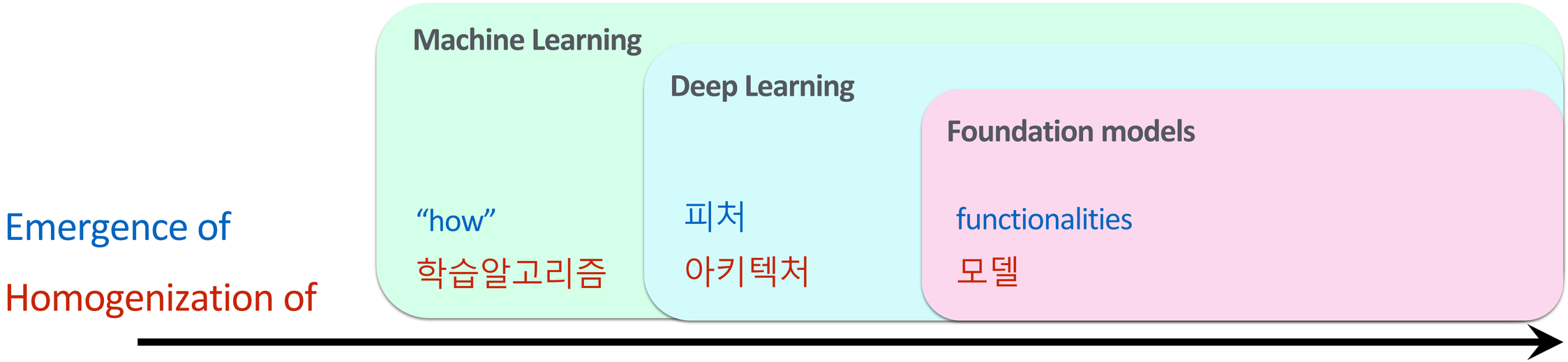
1.3 GPT-3가 보여준 놀라운 점

DEVIEW 2021

- 아주아주 큰 언어모델
- Transformer를 엄청 많이 쌓고 (175B!!) 그냥 다음 단어(Token) 맞추기
- 데이터는 엄청 많이 활용 (1T tokens)
- GPU 10,000 (v100)
- 기존 AI 가 하지 못했던 것들의 가능성을 보여줌
  - 소설도 쓰고
  - 시도 쓰고
  - 대본도 만들고
  - 레시피도 만들고
  - 레딧 댓글 일화



1.4 Foundation Model



#AI의 발전 방향

Emergence: 의도했던 능력 + 추가 능력

Homogenization: 비슷한 능력을 갖고 있는 그룹

https://arxiv.org/pdf/2108.07258.pdf

1.4 Foundation Model

DEVIEW 2021

	Emergence	Homogenization
룰베이스	없음	서로 다른 룰로 작성된 개별 프로그램
머신러닝		
딥러닝		
파운데이션 모델		

1.4 Foundation Model

	Emergence	Homogenization
룰베이스	없음	서로 다른 룰로 작성된 개별 프로그램
머신러닝	How (ex> 입출력 사이 관계)	학습 알고리즘 (ex> 결정트리, 신경망)
딥러닝		
파운데이션 모델		

1.4 Foundation Model

DEVIEW 2021

	Emergence	Homogenization
룰베이스	없음	서로 다른 룰로 작성된 개별 프로그램
머신러닝	How	학습 알고리즘
딥러닝	피처 (ex> CNN ImageNet 피처)	아키텍처 (ex> CNN, LSTM, Transformer)
파운데이션 모델		

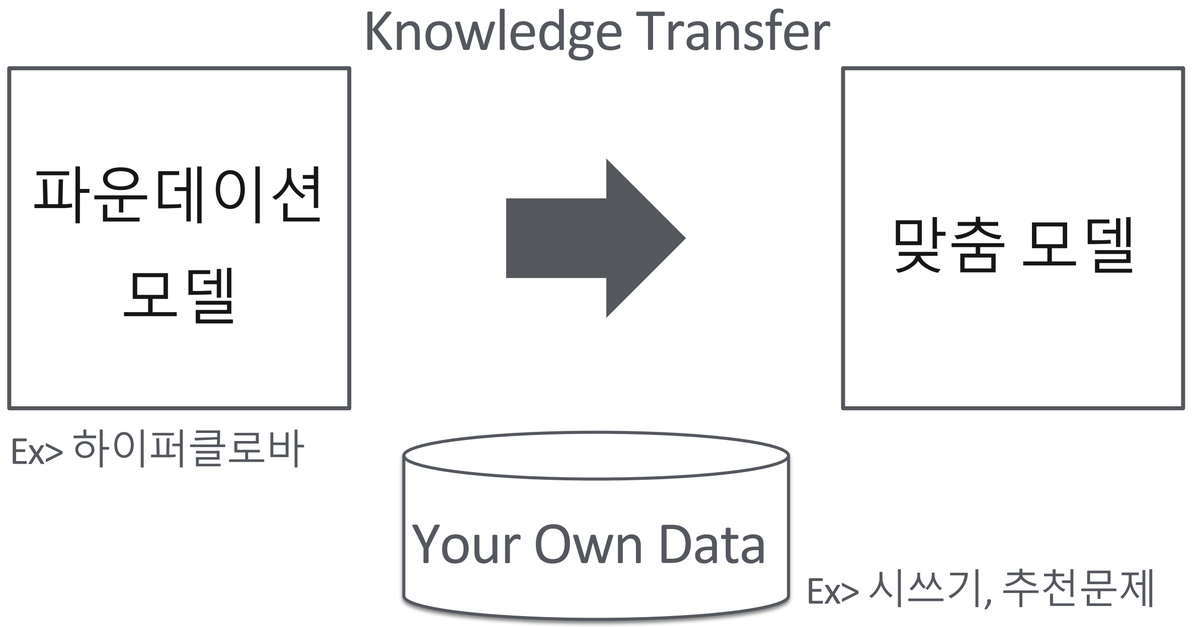
1.4 Foundation Model

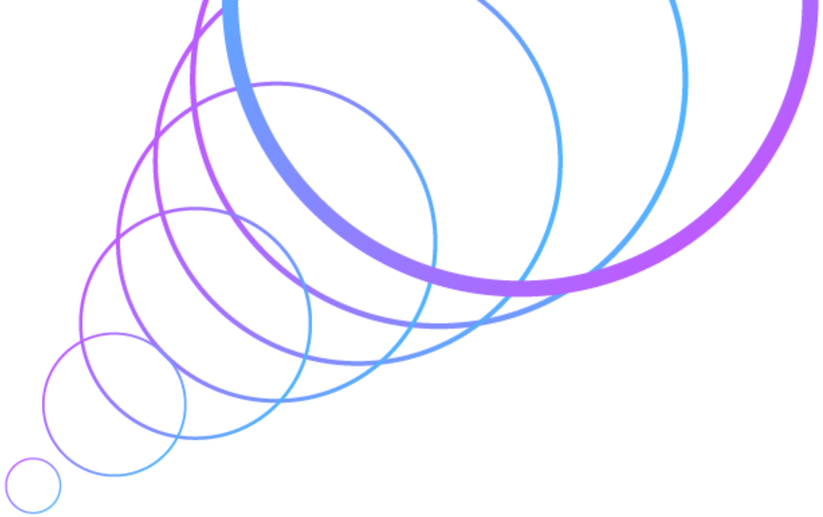
DEVIEW 2021

	Emergence	Homogenization
룰베이스	없음	서로 다른 룰로 작성된 개별 프로그램
머신러닝	How	학습 알고리즘
딥러닝	피처	아키텍처
파운데이션 모델	Functionalities (ex> 시를 쓰는 GPT-3)	모델 (ex> GPT-3, T5)

1.5 앞으로 바뀔 AI 모델 개발 파이프라인

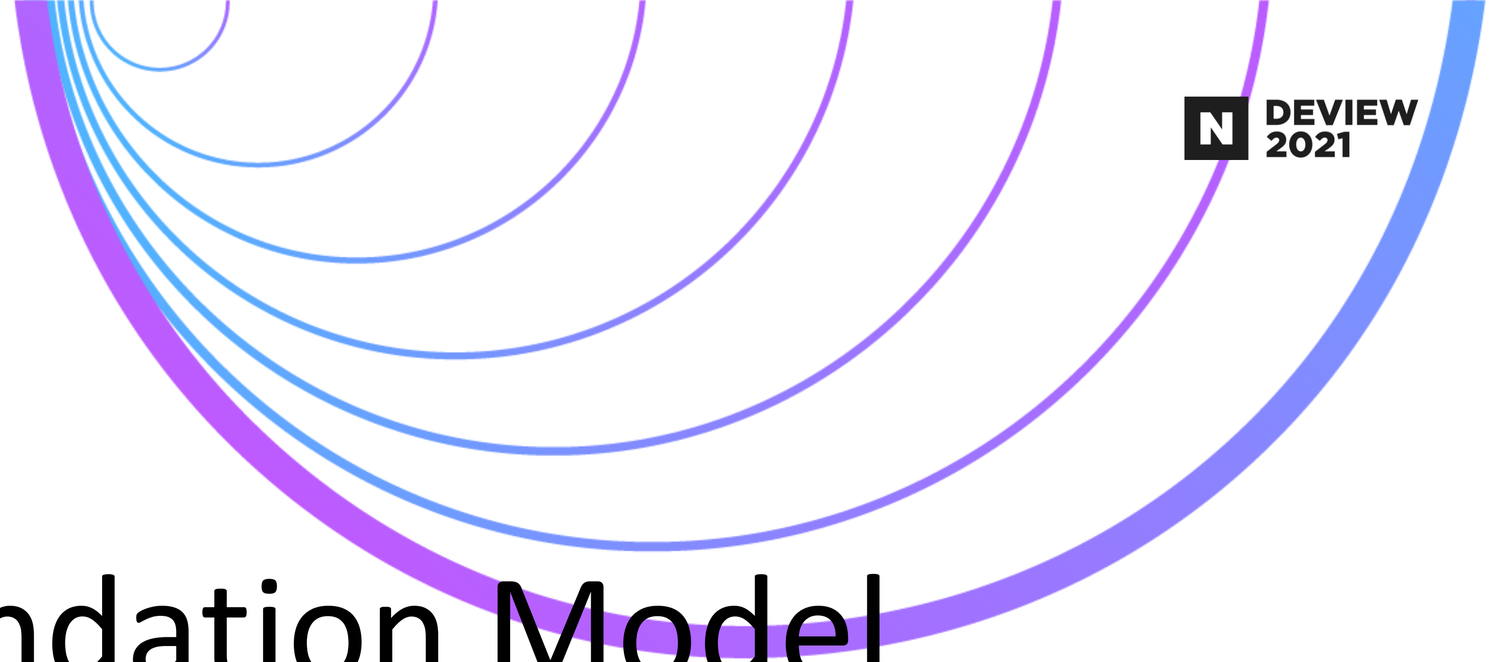
- AI를 몰라도 개발 가능
- Bring Your Own Data: knowledge transfer를 통한 AI 모델 간단 개발
- 프롬프트 수정, API 사용으로 가능
- 육수를 통해 맛있는 요리를 빠르게





Bring Your Own Data: Business AI 고민?
HyperCLOVA에게 무엇이든 물어보살

김경민/곽하녹/신규용 NAVER CLOVA / ML X



DEVIEW 2021

2. Foundation Model for Business

2.1 Related Works

최근 BERT, GPT 등의 Pre-text task를 응용해 business model을 만들려는 시도 활발 (Zhang et al., 2020; Xie et al., 2020; Gu et al., 2021).

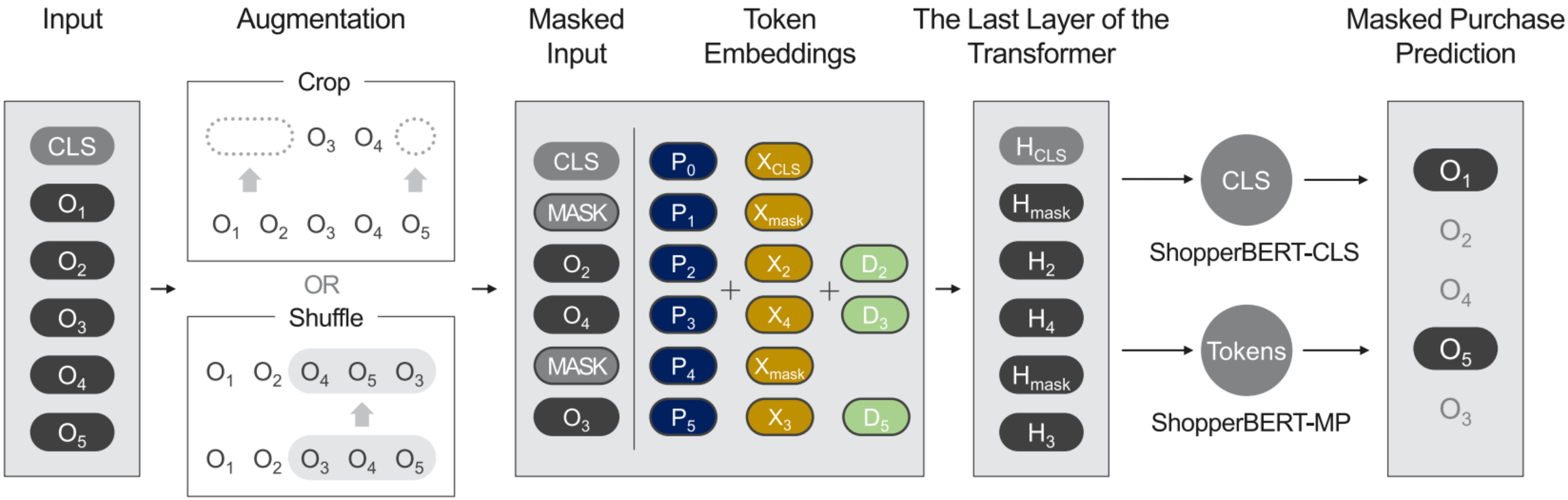
- Foundation Model로 확장되기엔 부족한 Scale
  - Model Size
  - Dataset Size
  - Diversity of Downstream Tasks
- Unimodal Dataset
- 다른 시스템으로의 전이 불가

2.2 ShopperBERT (initial apporach)

쇼핑 구매 기록을 Masked Language Model (MLM) 방식으로 학습

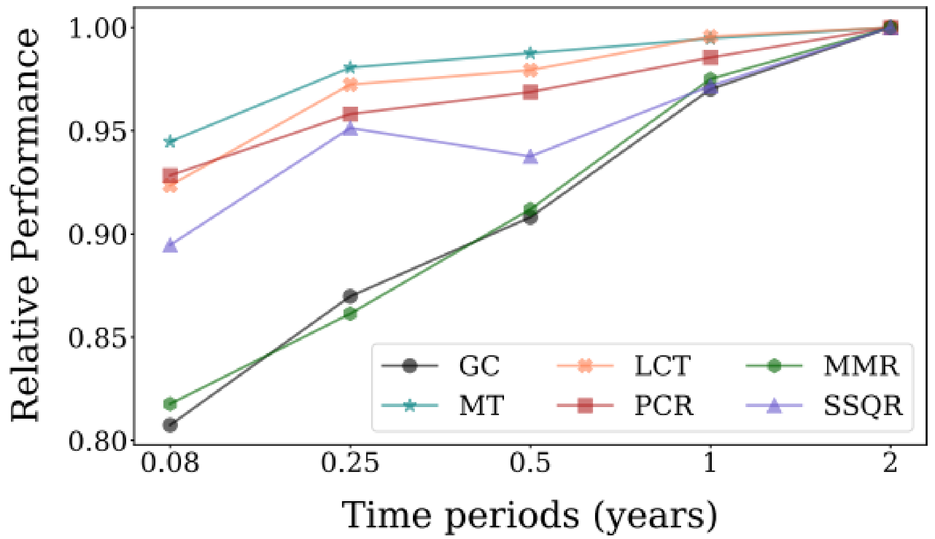
- 1,300만 유저와 4,800만개의 상품을 대상으로 2년치 구매 기록 (8억) 을 수집
  - SentenceBERT를 이용해 상품명을 embedding vector로 넣어줌
  - 상품의 계층별 카테고리를 supervision으로 사용 (총 7,945)
- 6개의 Downstream tasks에서 pre-trained된 user embedding을 feature-based MLP 방식으로 학습 시킨 후 성능을 평가, From scratch 추천 모델의 성능을 뛰어넘음

2.2 ShopperBERT (Concept Figure)



- 인풋을 augmentation 한 뒤, [MASK] 토큰을 맞추는 Pretext-task

2.2 ShopperBERT (analysis)



Tasks	Cold/Heavy	Metrics	T-Trans	U-MLP
PCR	Cold	HR@10	0.991	**1.000**
		NDCG@10	0.964	**1.000**
		MRR	0.953	**1.000**
	Heavy	HR@10	0.998	**1.000**
		NDCG@10	0.993	**1.000**
		MRR	0.991	**1.000**

- Pre-trained dataset의 수집 기간을 늘릴 수록 downstream tasks의 성능이 향상됨
- From scratch 추천 모델과 (T-Trans) pre-trained user embedding 모델의 (U-MLP) 성능을 비교 했을 때 Cold에서 더 큰 향상을 보임

2.2 ShopperBERT (findings)

- 쇼핑 데이터를 이용한 pretrained user embedding이 쇼핑과 연관된 다양한 task에서 task-specific 추천 모델을 뛰어넘음
  - Global user embedding의 가능성 발견
  - 학습된 user embedding을 이용하면 편리하고 빠르게 MLP 모델을 이용하여 복잡한 추천 모델의 성능을 뛰어넘을 수 있음 (2,458배 빠른 속도)
- 데이터 수집 기간에 따른 downstream tasks 성능 향상 발견

2.2 ShopperBERT (limitations)

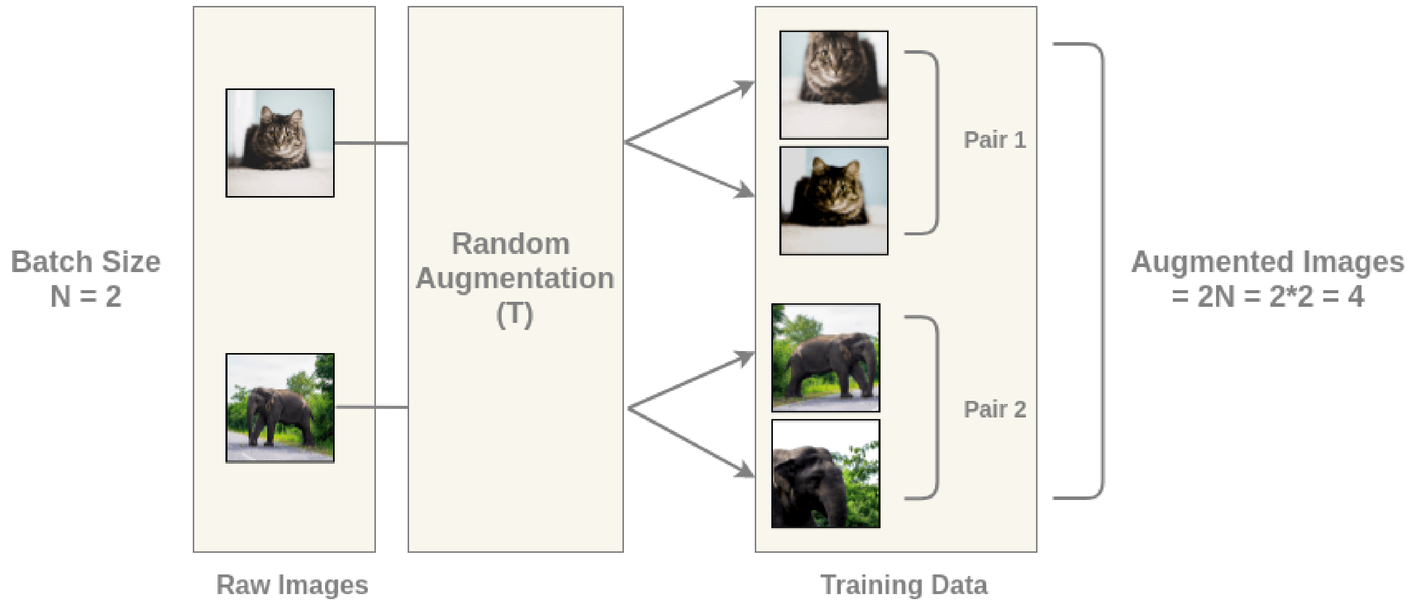
- 모델 크기를 늘려도 성능 향상이 두드러지지 않음
  - 상품의 카테고리를 맞추는 task가 상당히 쉽다
- 쇼핑과 연관된 도메인에 국한된 downstream task
- Unimodal dataset
- 다른 시스템으로의 전이 불가능

2.3 SimCLR (second apporach)

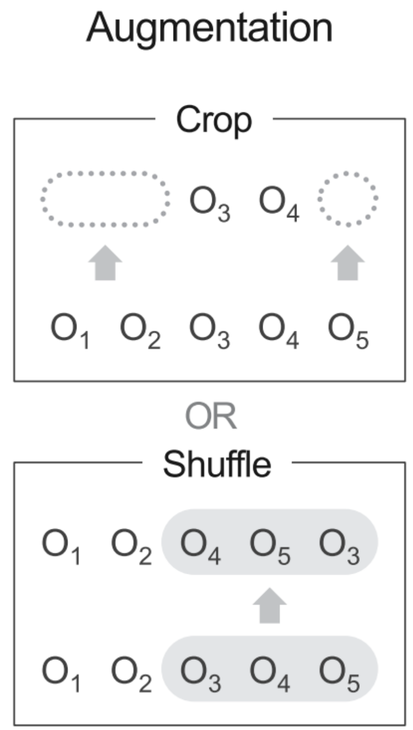
쇼핑 구매 기록을 일부분 바꿔치거나, 일정 기간으로 잘라 두 가지 augmentation data를 만들고 contrastive setup으로 학습

- Equivalent predictive objective가 아닌 contrastive objective를 사용함으로써 상품의 카테고리 등과 같은 부가 정보를 알 필요 없음
- ShopperBERT와 달리 상품명 텍스트 피쳐 자체를 학습함 (Wordpiece Tokdenizer)
- ShopperBERT에 비해 Downstream Tasks에서 7~12% 성능 상승

2.3 SimCLR (Concept Figure)



(출처: https://amitness.com/2020/03/illustrated-simclr)



- Augmenatation을 통해 인풋을 서로 다른 두 개의 데이터로 변형하고 이를 배치 내 모든 데이터에 대해 실행. 이 후 Contrastive Learning을 통해 유사도 maximize

2.3 SimCLR (findings)

- Contrastive objectiv를 사용함으로써 pre-text task가 더욱 어려워짐
  - 모델 크기를 키울 수록 성능 향상
- Pre-train하는데 있어 상품명만 필요함으로 다른 시스템으로의 전이 가능

2.3 SimCLR (limitations)

- 쇼핑과 연관된 도메인에 국한된 downstream task
- Unimodal dataset
- 정밀하게 가공된 augmentation 기법 필요

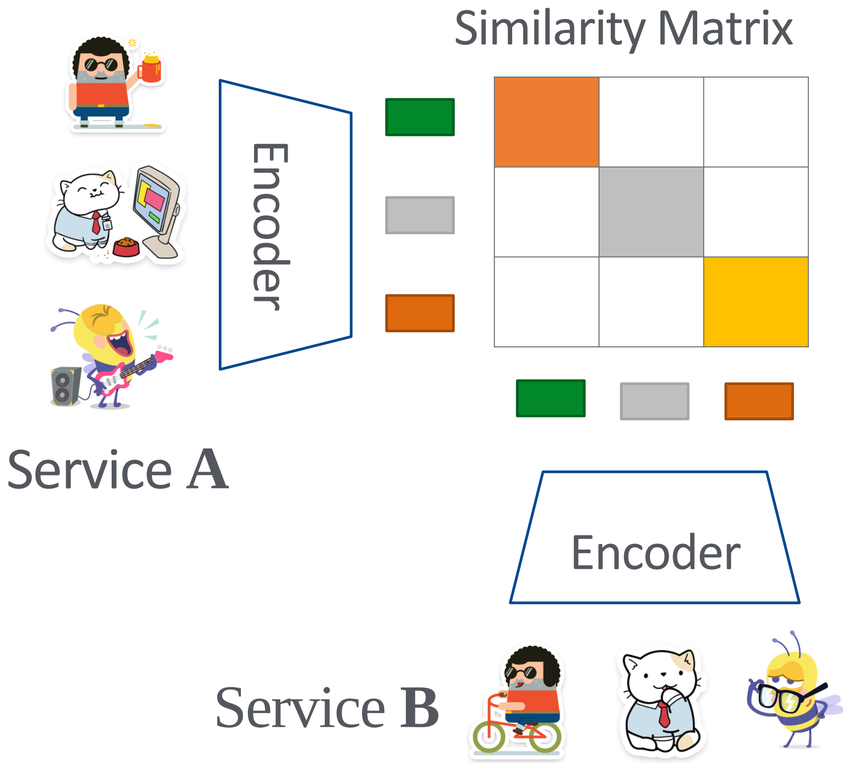
DEVIEW 2021

Next-Generation Business Foundation Model

2.4 Contrastive Learning User Embedding (CLUE)

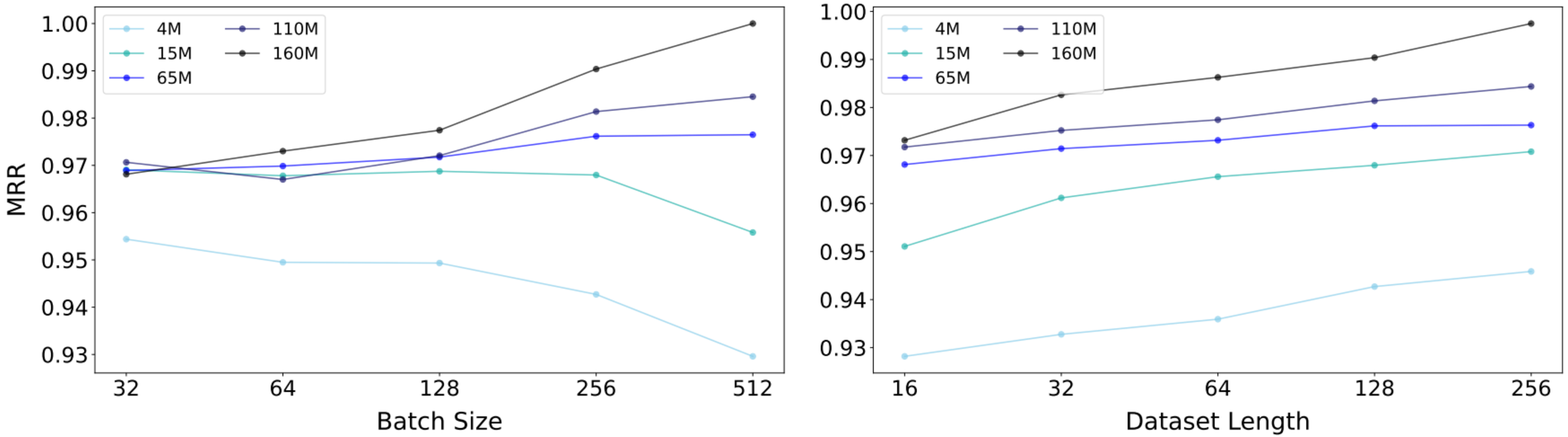
- 여러 서비스의 로그를 모아 contrastive setup으로 학습
- 725 million 파라미터 모델과, 50 billion 서비스 로그로 기존에 없던 규모의 business model 구축
- 텍스트화 된 인풋을 사용함으로써 domain과 system을 넘나드는 transfer learning 가능 (Byte-level BPE 사용)
- 여러 Downstream Tasks에서 SimCLR 모델에 비해 10~20% 높은 성능 기록

2.4 CLUE (Concept Figure)



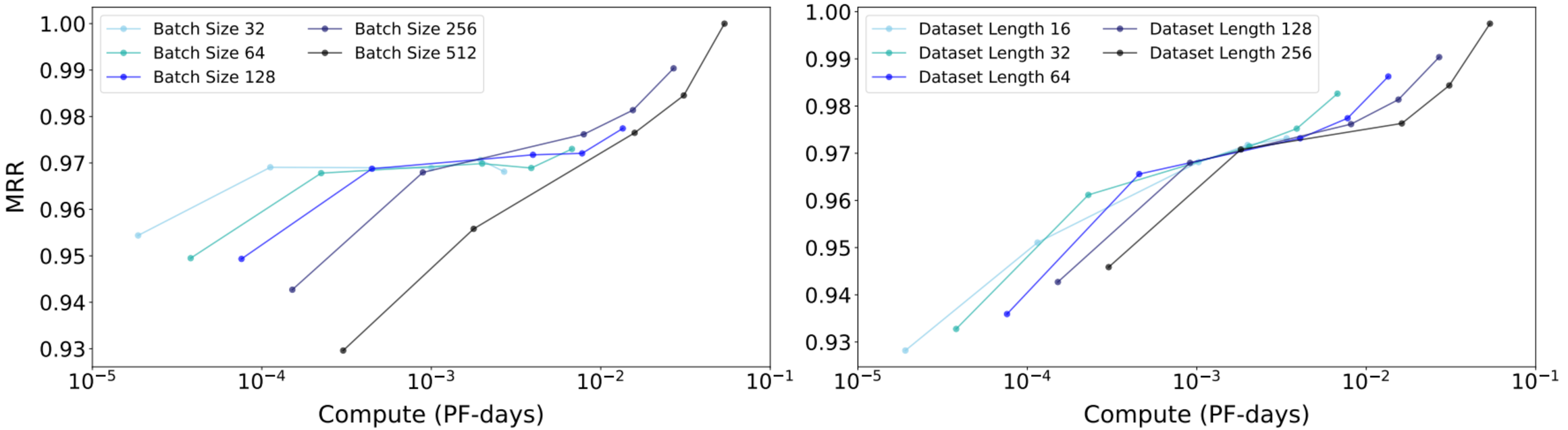
- 다양한 training data를 통해 풍부한 downstream task로 trasnfer learning
- 텍스트로 변형된 서비스 로그를 encoding 한 뒤, 유사도 maximize

2.4 CLUE (analysis)



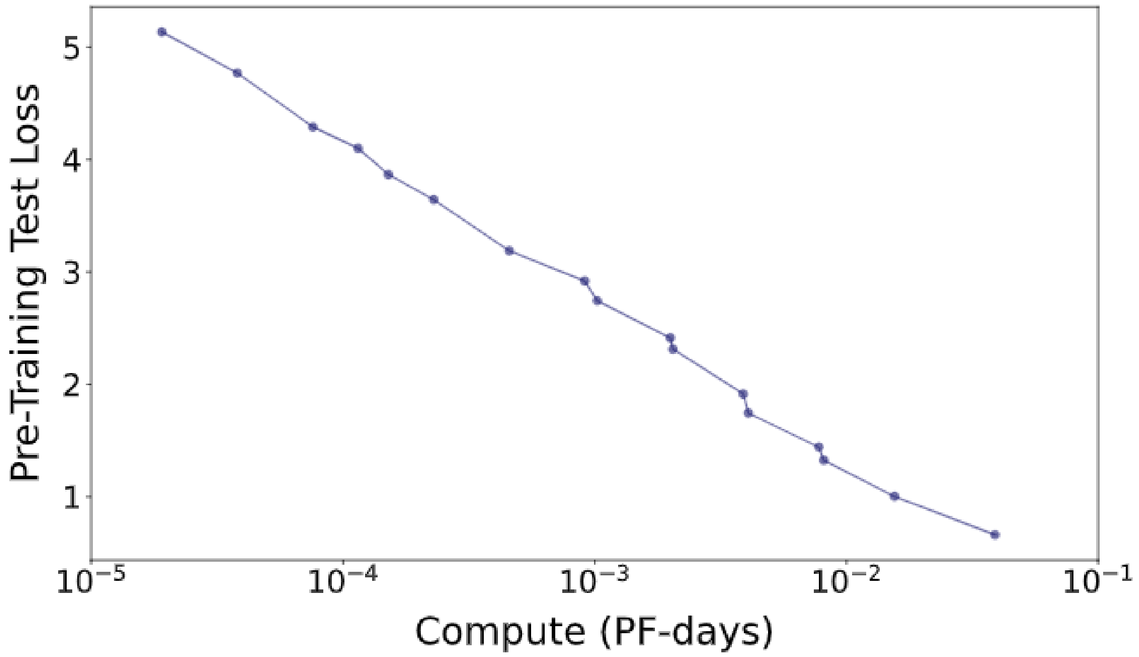
- 배치 사이즈와 모델 사이즈 모두를 적절하게 같이 증가시켜야 성능이 향상
- 데이터 길이를 증가시키면 항상 성능이 향상됨

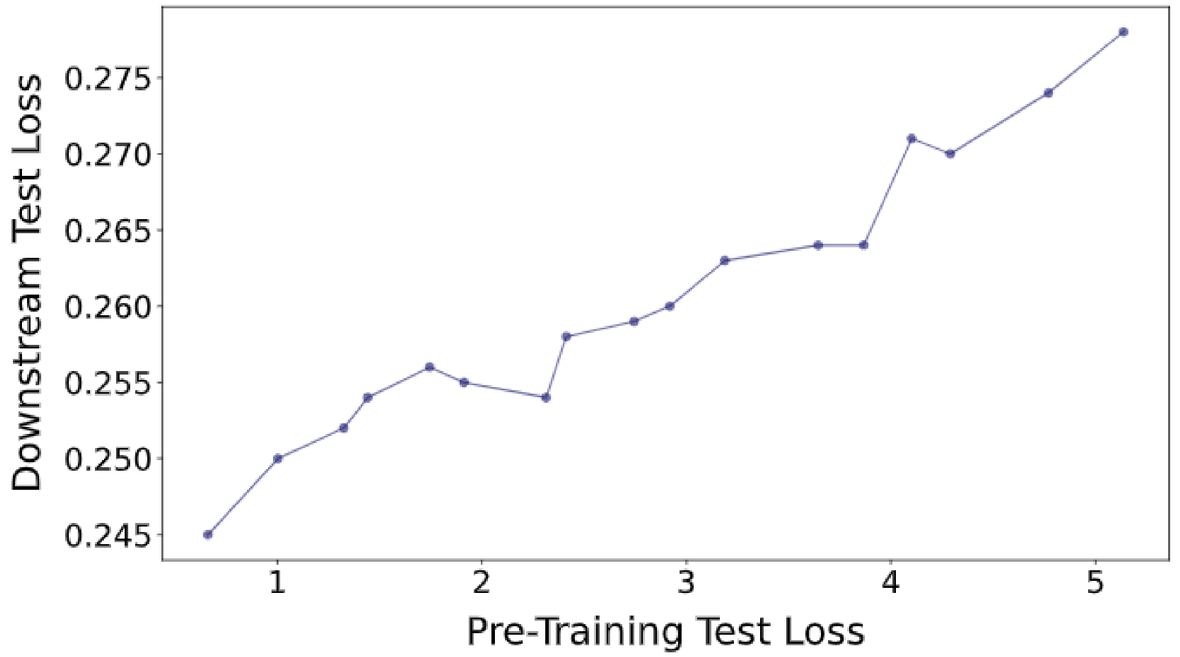
2.4 CLUE (analysis)



- 배치 사이즈가 작다면 Compute Resource를 투자해 모델 크기를 키워도 성능이 향상되지 않음. 배치 사이즈가 크다면 모델 크기를 증가시켜 성능 향상이 가능
- Compute Resource는 데이터셋 길이보다 모델 크기에 먼저 투자하는 게 좋음

2.4 CLUE (analysis)





- Compute Resources가 많이 투입될 수록 Test Loss가 줄어듬 (linear)
- Pre-text task의 test loss가 작다면 downstream의 test loss도 작은 경향이 나타남

2.4 CLUE (analysis)

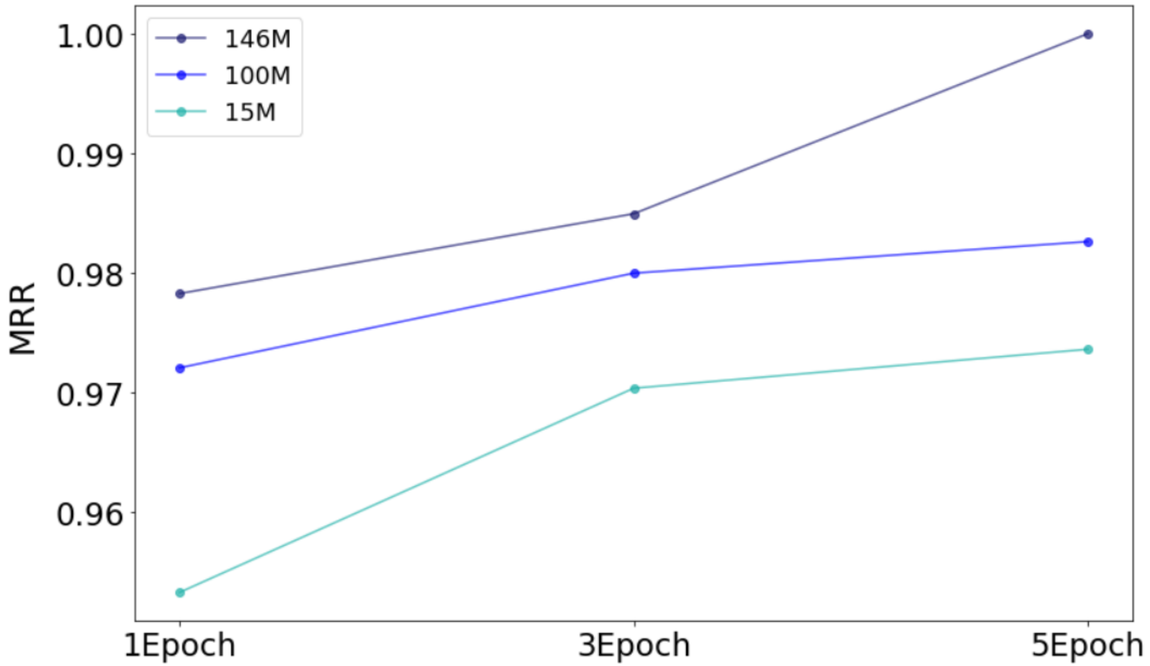


Table 6. Performance comparisons on output dimension

Model (CLUE 100M)	MRR	HR@1
300D	1.000	1.000
3000D	0.998	0.996

- 더 많은 Epoch을 학습할 수록, 모델 사이즈가 클 수록 Downstream 성능이 향상됨
- User Embedding의 Output Dimension이 Downstream 성능에 영향을 미치지 않음

2.4 CLUE (findings)

- 모델 크기를 키울수록 확실한 성능 향상
- 모든 유형의 서비스 로그를 텍스트화 해 사용함으로 Domain과 System간의 Trasnfer 용이

2.4 CLUE (limitations)

- Equivalent predictive objectives를 사용하는 모델들과 달리

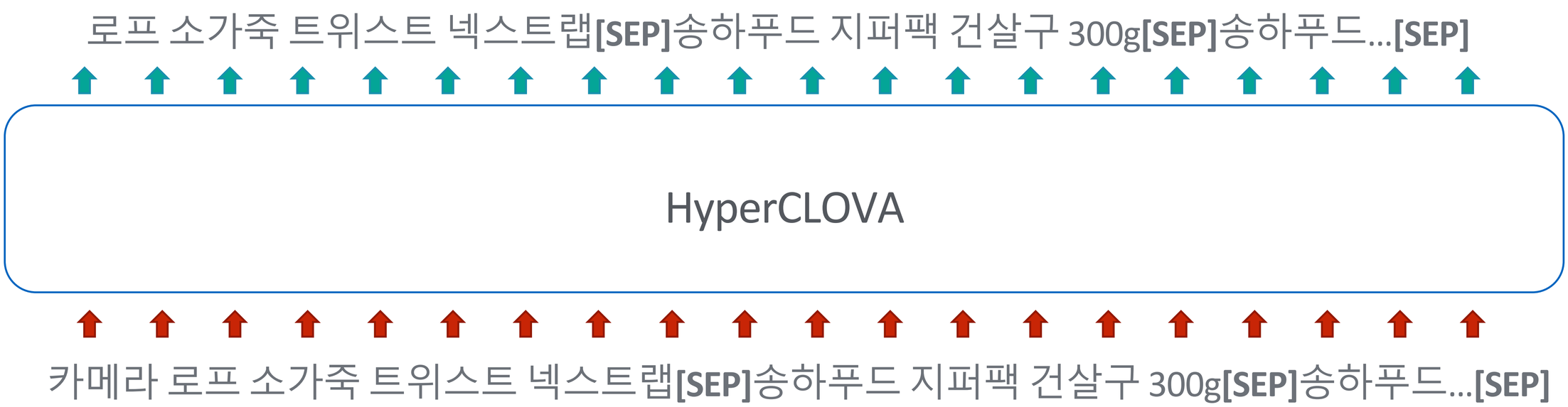
  Contrastive setup으로 인해 Complicated Scaling Law

  - Batch Size vs Model Size

2.5 HyperCLOVA-Biz

- 언어모델인 HyperCLOVA-LM을 Finetuning한 뒤 feature를 뽑아 user embedding으로 사용
- CLUE 대비 2~3% 성능 향상, CLUE 피쳐와 함께 사용됐을 때 모든 Downstream Tasks에서 SOTA대비 5~10% 성능 향상

2.5 HyperCLOVA-Biz (Concept Figure)



- HyperCLOVA-LM을 NAVER 쇼핑 데이터로 Finetuning
- 실제 사용될 때는 해당 Task의 Task-specific 데이터로 한 번 더 Finetuning
- 마지막 [SEP] 토큰을 [CLS] 토큰처럼 사용

2.5 HyperCLOVA-Biz (analysis)

DEVIEW 2021

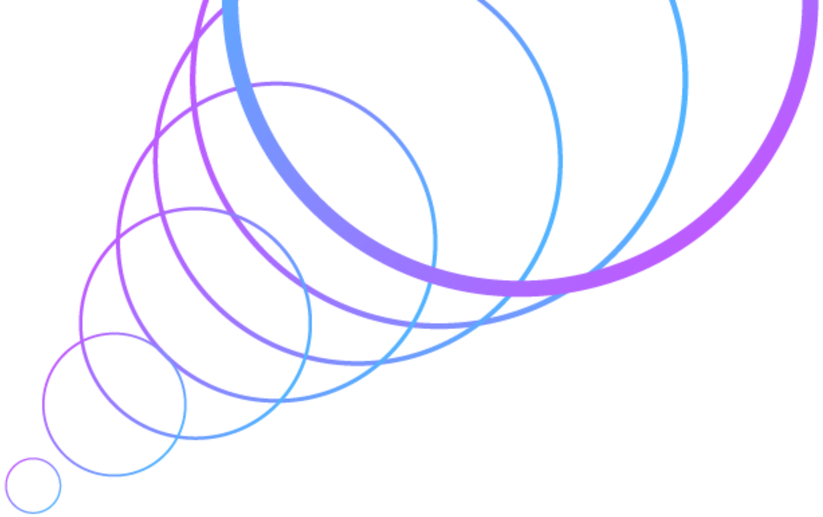
Input	Output	Ground truth
네이처엠 웰바이츠 견과 20g X 30팩 조은술 세종 우도 땅콩 전통주 6도 750ml x 12병 3500 A4 방안연습장 모눈노트 모눈종이 그리드 격자 (CM)위생페이퍼 (42cm x 42cm)400매/검진위생지/위생 베게카바/ 속눈썹 파마재료 / 롤리킹펌세트+글로비비 토닉&에센 스 세트 / 롤리킹펌 / 속눈썹펌(이쑤시게 면봉 100개입 셀라인 위생 블랙 마스크 50매 일회용 속눈썹 연장 고급 하이드로겔 아이패치 (10ea) 눈밑보 호패치 겔 유키반 초저자극 속눈썹 테이프 1ea / 종이반창고 / 속 눈썹 연장 부자재 커버랩 랩커터 뉴트리코어 임산부 엽산제 800 임신초기 임신전 레몬추 출엽산 ZOGGS 조그스 어린이&성인 수경 물안경 닥터썬데이D 비타민D 임산부 어린이 수이사쿠 베타 침대 베타쉼터 [네오샘플증정]API 베타픽스 [50ml] 베타약품/베타치료 제/베타	백셀 건전지 AA 2개 미니어처 인형 모음 / 구체관절인형 / 피규어 미니어처 인형 놀이 - DIY 미니어처 하우스 / 룸박스 / VIP 핑크 미니어처 모형 / 크리스마스 트리 인형 / 장식 / 미니어 쳐 / 데코소품 휴대용 접이식 장바구니 에코백 시장바구니 소형 대형 [프로쉬] 오늘만 특가! 무료배송 독일 친환경 식기세척 기 전용세제 2개SET [프로쉬]독일 친환경 맨손세제/주방세제 2+1 셀프시공 펫트 모노륨 바닥 장판 E26 LED 미니전구 5W 전구(노란)색 [ONLY] 마켓비 FIHA 서랍장 4.1단 [귀염장착] 800종 크록스 지비츠 / 3D입체 지비츠 (고퀄 리티) / 크록스 물빠지는 목욕바구니 수영 스파 헬스 목욕 사우나 가방 사우나 큰사이즈 New 아이팜 클래식 아기욕조 목욕의자 세트	겨울왕국 소피아공주 어린이 악세서리세트 귀걸이 반 지 보석상자 어항히터기 아마존히터 어항온도조절기 어항 스포이드 54CM 연장스포이드 남자 여자 공용 패션 무지양말 23col 1+1 가을 기본 긴팔티 8가지 색상 코튼 긴팔 스트랩 셔츠원피스 (2컬러) 9살 아홉 살 마음, 느낌, 함께, 내 사전 선택 유아동아기 면 컬러 돌돌이 골지양말 세트 II 고무뱅크 유아 아동 아기 머리고무줄 머리끈 칼라고무 줄 대용량 [프랑스자수] 오가닉 티아나 배냇저고리 만들기 세트 태교바느질 DIY(사계절 여름 선택)

2.5 HyperCLOVA-Biz (findings)

- HyperCLOVA-LM을 이용해 성공적으로 사용자 피쳐 추출
- 초기 단계이긴 하지만, 사용자의 다음 상품 구매 생성
- 추천 시스템에서 안정적으로 안착한 언어모델의 Transferability

2.5 HyperCLOVA-Biz (limitations)

- 복잡한 Finetuning 과정
- Causal Language Model에서의 embedding 추출 방법 연구 필요
- Computational Cost



Bring Your Own Data: Business AI 고민?
HyperCLOVA에게 무엇이든 물어보살

김경민/곽하녹/신규용 NAVER CLOVA / ML X

3.Business Transfer Learning

3.1 Business Transfer Learning Requirements

DEVIEW 2021

데이터 일반성

- 다양한 데이터 형식이 존재
- 예를 들어, 네이버 쇼핑에서 사용하는 상품 카테고리 체계는 다른 서비스와 다름
- **주요 정보를 비정형 텍스트로 치환**

유저 확장성

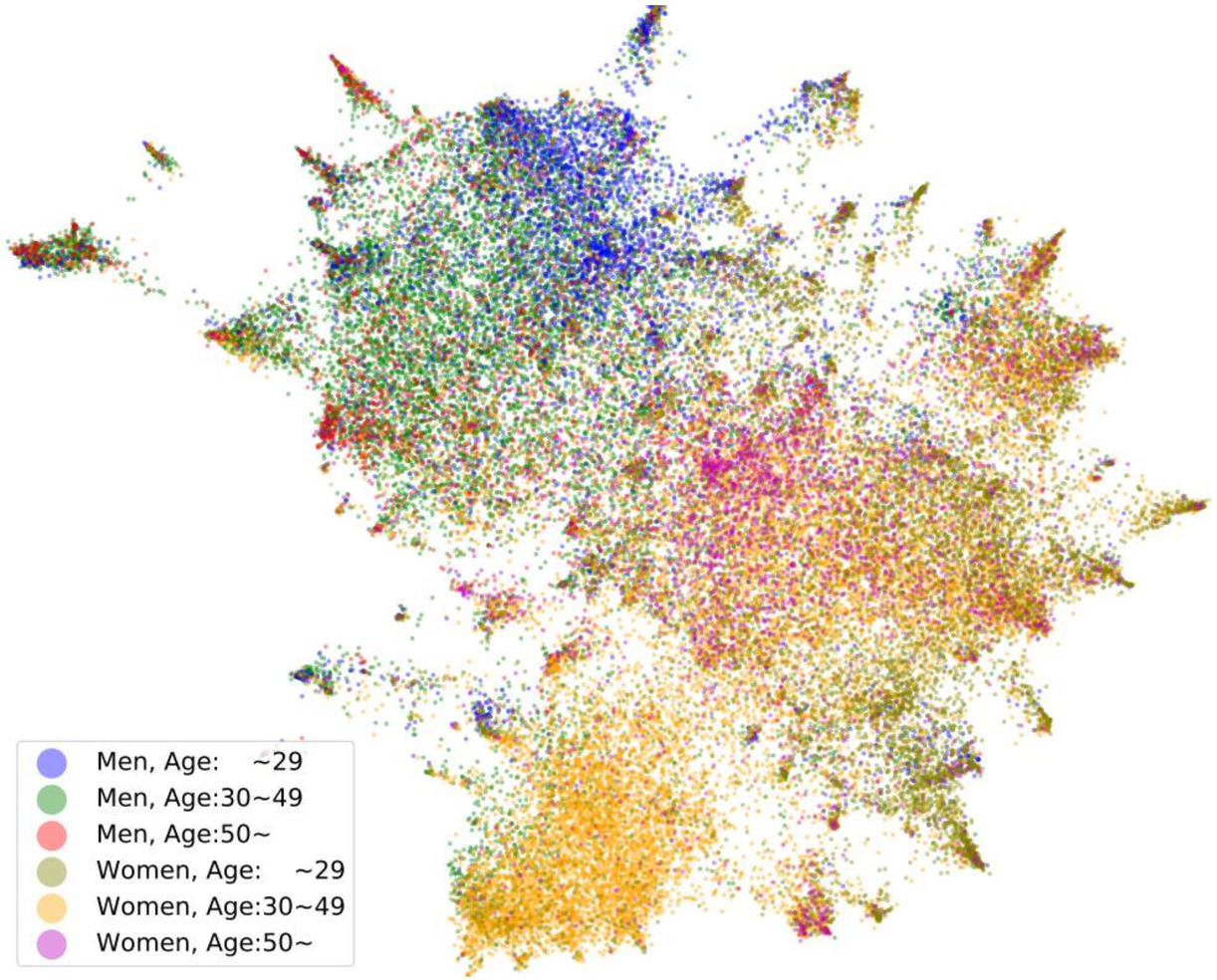
- 새로운 유저와 휴면 유저가 많음
- 보안 상의 이유로 유저를 식별하지 못하는 경우가 존재
- **유저 ID를 사용하지 않고 유저 자체의 정보를 활용**

3.2 User Features

3.3 Downstream Data based User Features

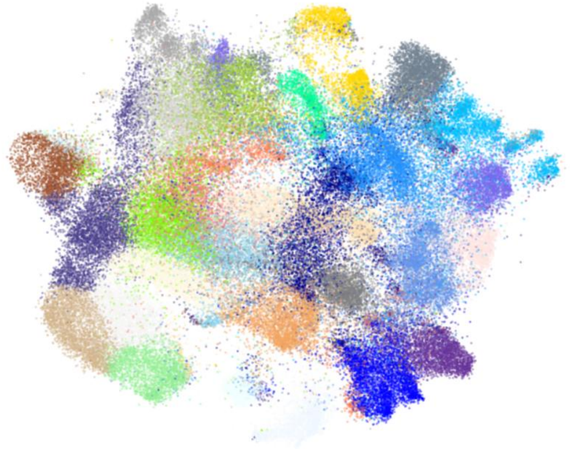
- 해당 다운스트림 내의 과거 기록을 사용하여 유저 피쳐 추출

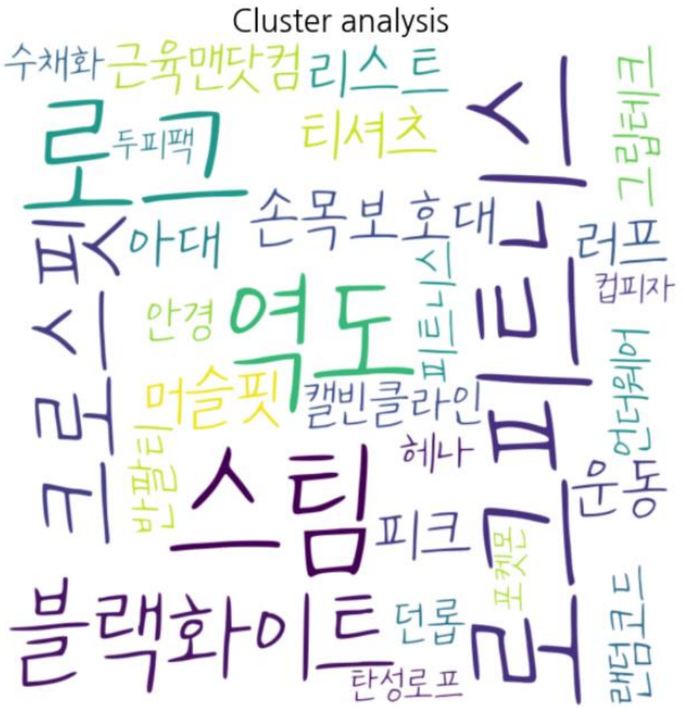
3.4 유저 임베딩 시각화



ShopperBERT 임베딩 시각화

CLUE 임베딩 시각화





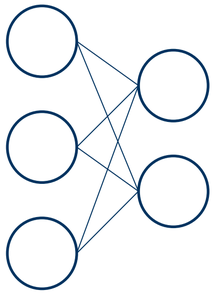
3.5 다운스트림 문제 종류

DEVIEW 2021

추천 문제

유저

아이템

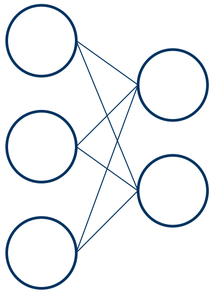


0.86

긍정 점수

분류 문제

유저



남자

3.6 Business Transfer Learning

다운스트림 데이터

User ID	Item ID	Description	Label
…	…		…
126	53	정말 맛있는 치킨	구매
…	…		…
…	…		…

0.86 (구매할 확률)

Inner Product

[0.4, -0.1, …, -0.8]

[-0.3, 0.7, …, 0.1]

User Encoder

Item Encoder

BUSINESS MODEL

Feature Table

User ID	Feature
…	…
126	[0.1, 0.7, …, -0.2]
…	…
…	…

“정말 맛있는 치킨”

3.7 Business Transfer Learning: 분류 문제

DEVIEW 2021

- Business Model로 Transfer Learning을 하여 많은 분류 문제를 해결
- 유의미한 수준의 정확도 달성

	Accuracy (Baseline 대비 성능)
성별 분류 문제	**+76.7%**
멤버쉽 가입할지 예측하는 문제	**+18.9%**
쇼핑 라이브를 사용하게 될지 예측하는 문제	**+23.2%**

label 비율 1:1

3.8 추천 문제 평가 지표: MRR

- Ranking Metrics
- 100 Random Negative Samples & 1 Ground Truth

$$MRR = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{rank_i}$$

3.9 다른 서비스의 기록이 도움이 되는가?

DEVIEW 2021

쇼핑 검색어 추천

	MRR (+%)
CLUE (다양한 서비스 데이터)	**+21.02**
SimCLR (A 서비스)	+8.45
SimCLR (B 서비스)	+5.18
SimCLR (C 서비스)	+3.67
인기순	0.00

뉴스 추천

	MRR (+%)
CLUE (다양한 서비스 데이터)	**+2.33**
SimCLR (A 서비스)	+1.60
SimCLR (C 서비스)	+0.55
인기순	0.00

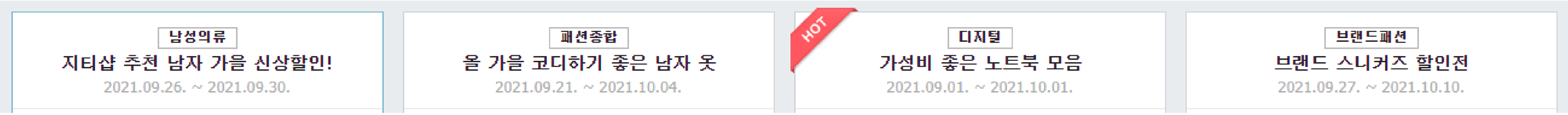
웹툰 추천

	MRR (+%)
CLUE (다양한 서비스 데이터)	**+20.08**
SimCLR (A 서비스)	+11.87
SimCLR (C 서비스)	+10.71
SimCLR (D 서비스)	+9.24
SimCLR (B 서비스)	+3.37
인기순	0.00

카페 추천

	MRR (+%)
CLUE (다양한 서비스 데이터)	**+8.91**
SimCLR (A 서비스)	+7.53
SimCLR (D 서비스)	+5.30
SimCLR (B 서비스)	+2.51
인기순	0.00

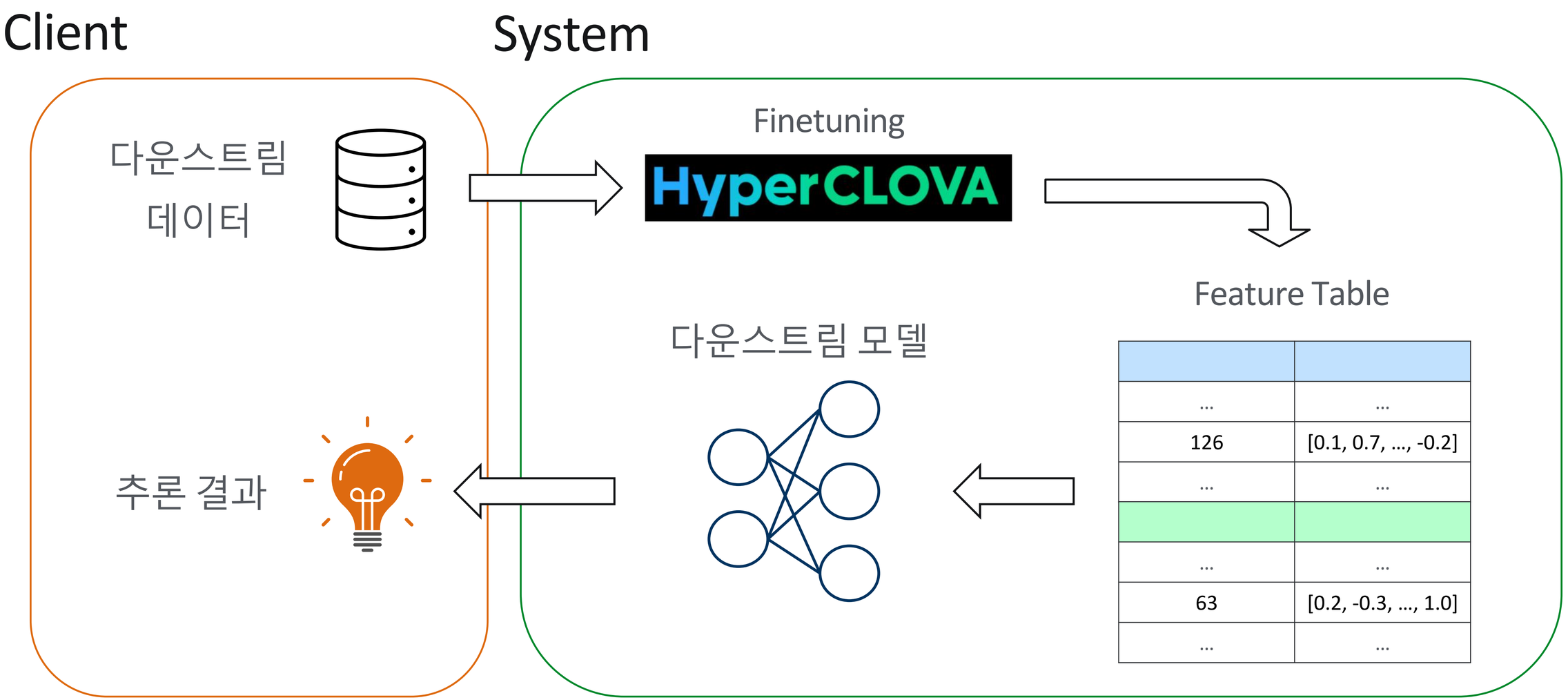
3.10 네이버 쇼핑기획전 추천 온라인 실험



기존 추천 모델 대비 온라인 성능 향상

	상대 성능 (CTR +%)
CLUE	**+13.37**
다른 모델1	+8.60
다른 모델2	+7.45
다른 모델3	+2.56
베이스라인	0.0

3.13 HyperCLOVA for Biz System (개발중)



We are hiring..

DEVIEW 2021

clova-jobs@navercorp.com

Thank You