언어 모델 기반의 범용 유저 임베딩과 이를 활용한 추천시스템 및 광고 타겟팅

곽하녹/신규용
NAVER Search

NAVER DEVIEW 2023

CONTENTS

1. 소개

- Business Foundation Models
- Feature-based Transfer Learning
- 추천 시스템과 광고 타겟팅 소개

2. 비즈니스 AI 모델에서 언어 모델링의 중추적 역할

- 언어 모델을 활용한 기존 유저 모델링 연구
- 분류 학습에 도움이 되는 언어 모델링
- Task-specific 및 Task-agnostic 유저 임베딩의 성능 향상
- 결과 분석

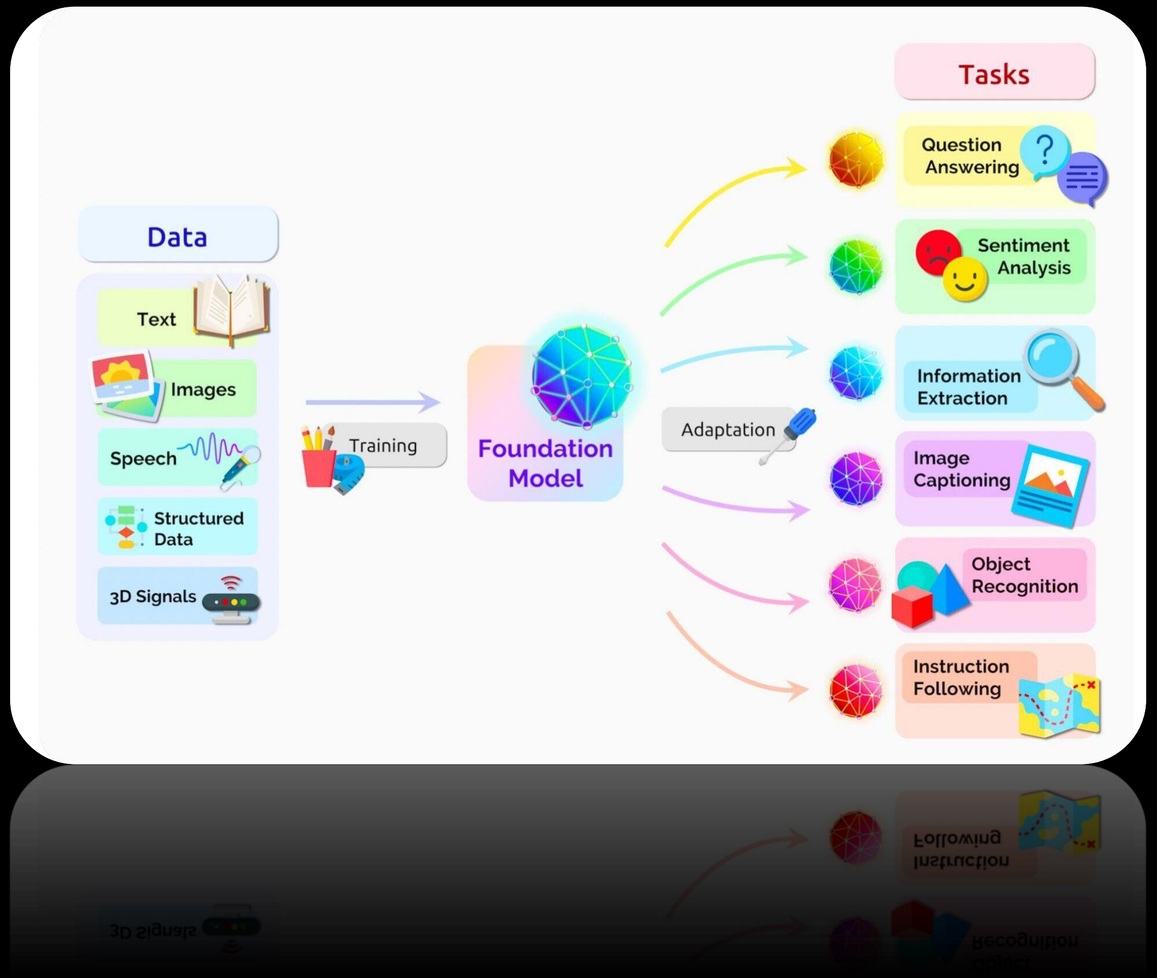
3. 서비스 적용 사례 및 성과

- 추천 시스템
- 광고 타겟팅

1. 소개

1.1 Business Foundation Models

NAVER DEVIEW 2023



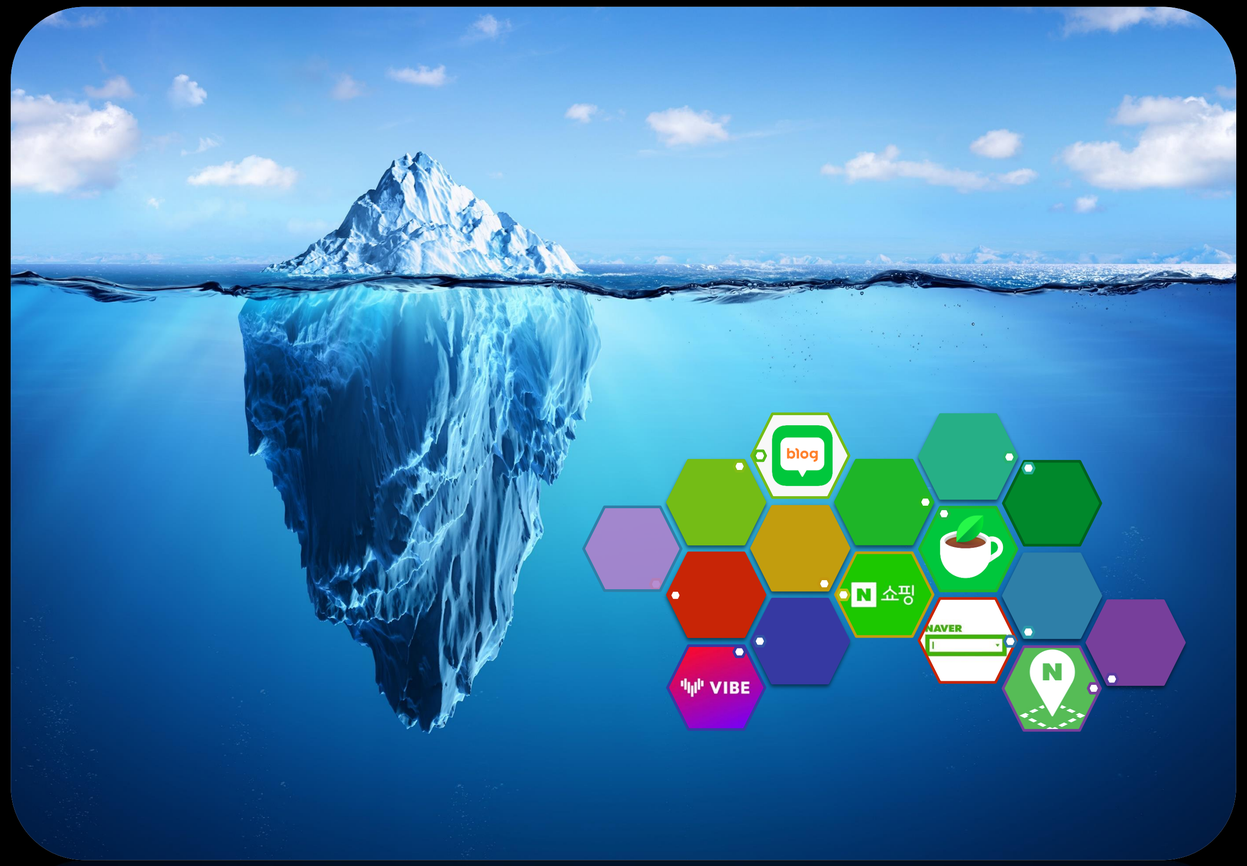
- 웹 상에서 쉽게 구할 수 있는 대규모 데이터를 수집
- 별도의 레이블링을 하지 않고 다음 단어 맞추기처럼 일반적인 목적 함수를 학습 (Self-Supervised Learning)
- 이렇게 학습된 모델을 다른 어플리케이션에 활용
  - Prompt Tuning (GPT-3)
  - Fine-Tuning
  - Feature-based Transfer Learning
- 많은 벤치마크 테스크에서 이러한 방식이 최고의 성능을 달성

https://arxiv.org/pdf/2108.07258

1.1 Business Foundation Models

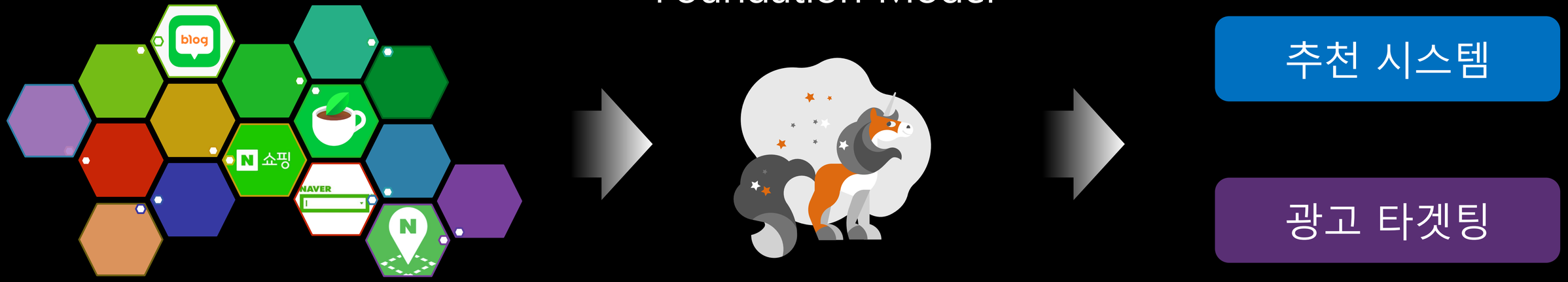
NAVER
DEVIEW
2023

- 네이버와 같은 종합 인터넷 플랫폼에는 정말 많고 다양한 데이터가 존재
- 각 서비스에서는 해당 서비스 데이터만 활용하는 경우가 많음
- 시스템과 법적인 이슈로 다른 서비스 데이터를 가져오기 힘들고 연관성을 사전에 파악하기 어려움



1.1 Business Foundation Models

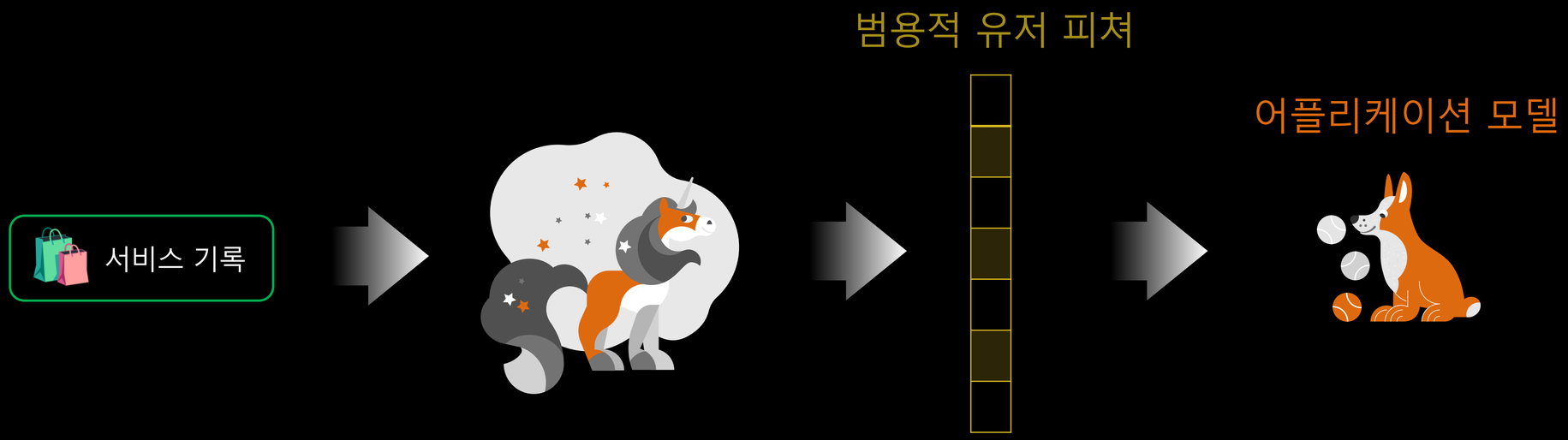
NAVER
DEVIEW
2023



- 네이버 내의 데이터들을 모아서 하나의 범용적인 비즈니스 모델을 학습
- 범용적인 모델을 네이버 내의 추천 시스템 또는 광고 타겟팅 등의 어플리케이션에 활용

1.2 Feature-based Transfer Learning

NAVER DEVIEW 2023



- Fine-tuning 방식을 추천 시스템에 적용하기 어려운 이유
  - 추천 시스템은 매일 많은 인퍼런스를 수행해야 하는데 Foundation Model은 인퍼런스 비용이 크다.
  - 네이버 내에 추천 시스템이 많다.
  - 어플리케이션 데이터의 형식이 호환이 안 되는 경우가 많다.
  - 서비스 부서마다 다양한 방식으로 추천 시스템을 커스터마이징 하는 경우가 많다.
- 유저의 피쳐를 제공하고 자유롭게 활용
  - 어플리케이션 모델에서 추가적인 유저 피쳐로서 활용 가능
  - 특히 어플리케이션 서비스를 신규 방문하는 유저에게 효과적
  - 거대한 서비스 기록 데이터를 적은 차원의 벡터로 함축

1.3 추천 시스템과 광고 타겟팅 소개

NAVER DEVIEW 2023

추천 시스템

에게 어떤 아이템들을 보여줄까?

결과:

- 유저들이 일상적으로 보게 되는 지면에서 자주 활용
- 접근이 많아서 데이터가 쉽게 쌓인다.
- 옛날이나 지금이나 유저와 아이템의 유사도를 학습

광고 타겟팅

"아바타2 할인 티켓" 광고를 누구에게 보낼까?

결과:

- 예) 문자 광고, 카카오톡 광고, 이메일 광고, …
- 구독 취소, 수신 차단 등을 회피하면서 높은 매출이 목표
- 광고 내용이 다채롭고 대규모로 보내는 광고도 있어서 마케터가 직접 타겟 유저들을 선정.

2. 비즈니스 AI 모델에서 언어 모델링의 중추적 역할

2.1 언어 모델을 활용한 기존 유저 모델링 연구

NAVER DEVIEW 2023

U-BERT: Pre-training User Representations for Improved Recommendation, **Tencent, AAAI 2021**

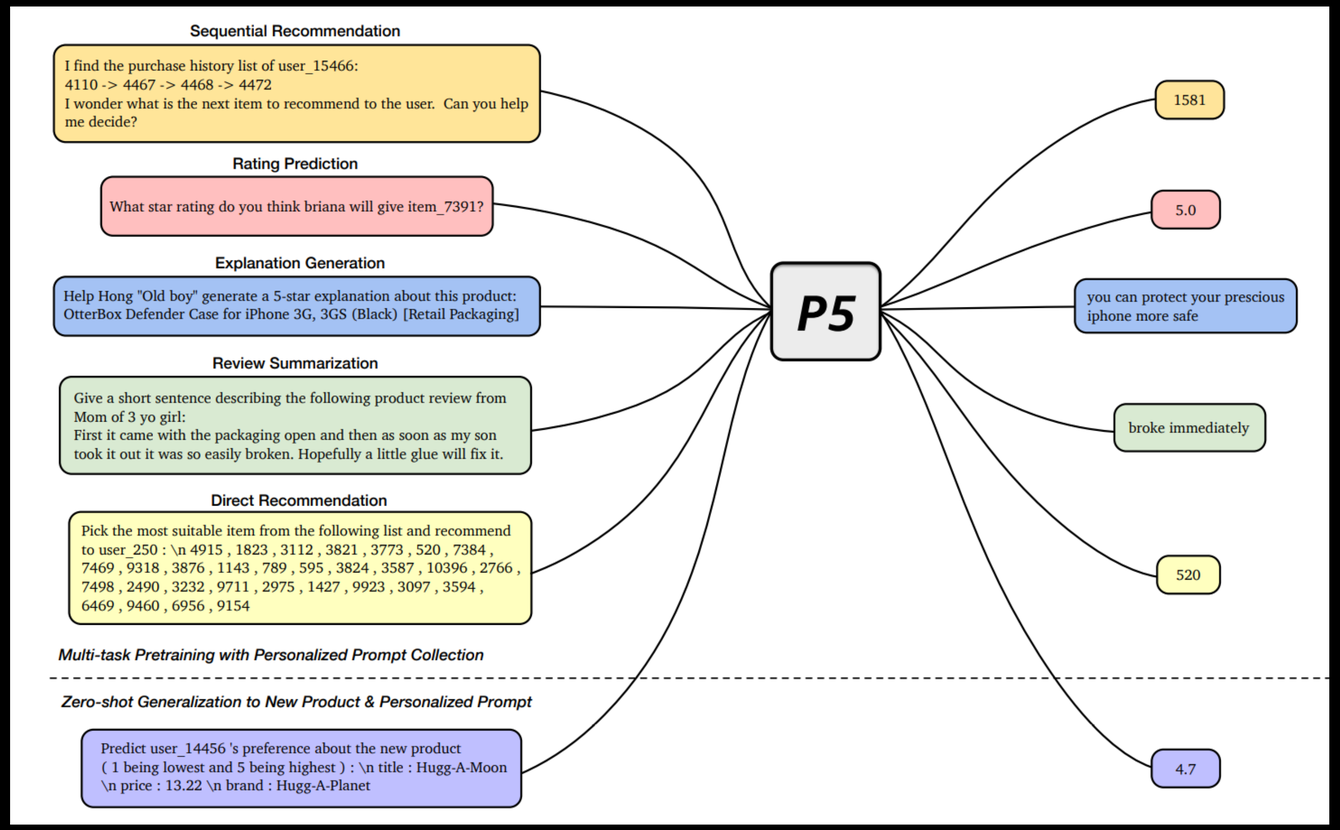
- 상품 리뷰가 유저의 특성을 반영하는 현상을 관찰
- Pretext Task
  - 리뷰에서 사용자의 선호도를 나타내는 단어 맞추기
  - 리뷰 점수 예측
- Downstream Tasks
  - Review Encoder에서 사용자가 단 리뷰와 아이템에 달린 리뷰를 이용해 recommendation tasks finetuning
- Datasets
  - Amazon product review dataset + Yelp challenge dataset

https://ojs.aaai.org/index.php/AAAI/article/view/16557

2.1 언어 모델을 활용한 기존 유저 모델링 연구

NAVER DEVIEW 2023

A Unified Pretrain, Personalized Prompt & Predict Paradigm (P5), **Rutgers Univ., RecSys 2022.**



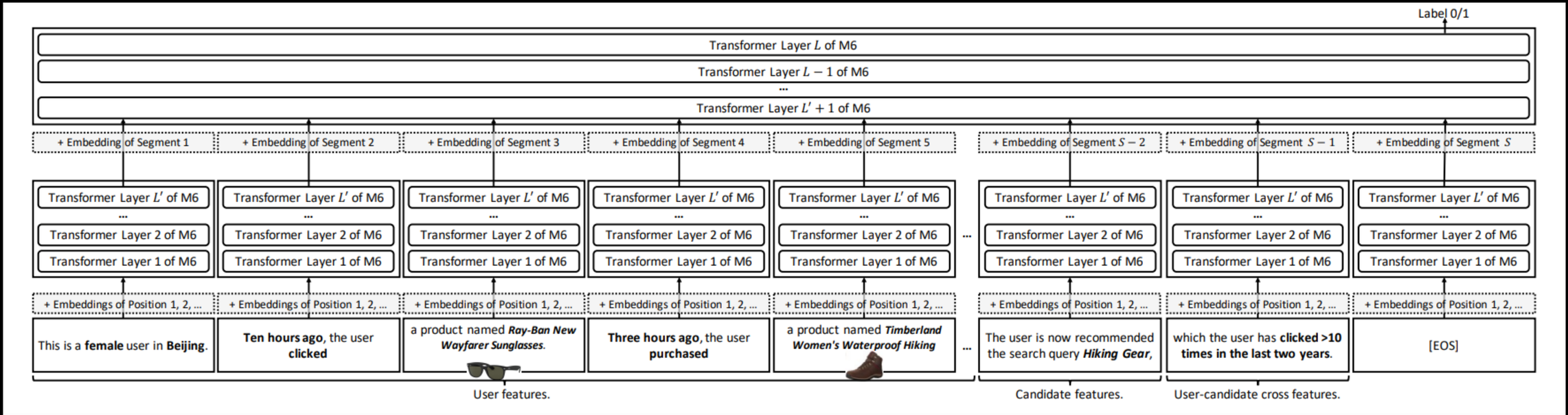
- NLP의 T5와 같은 unified user modeling framework 제안
- Pretext Task
  - 적절한 prompt engineering을 통해 정제한 다양한 user modeling tasks
- Downstream Tasks
  - User modeling tasks
  - Ability to perform zero-shot learning
- Datasets
  - Amazon product review dataset + Yelp challenge dataset

https://dl.acm.org/doi/abs/10.1145/3523227.3546767

2.1 언어 모델을 활용한 기존 유저 모델링 연구

NAVER DEVIEW 2023

M6-Rec: Generative Pretrained Language Models are Open-Ended Recommender Systems, **Alibaba, Arxiv 2022.**



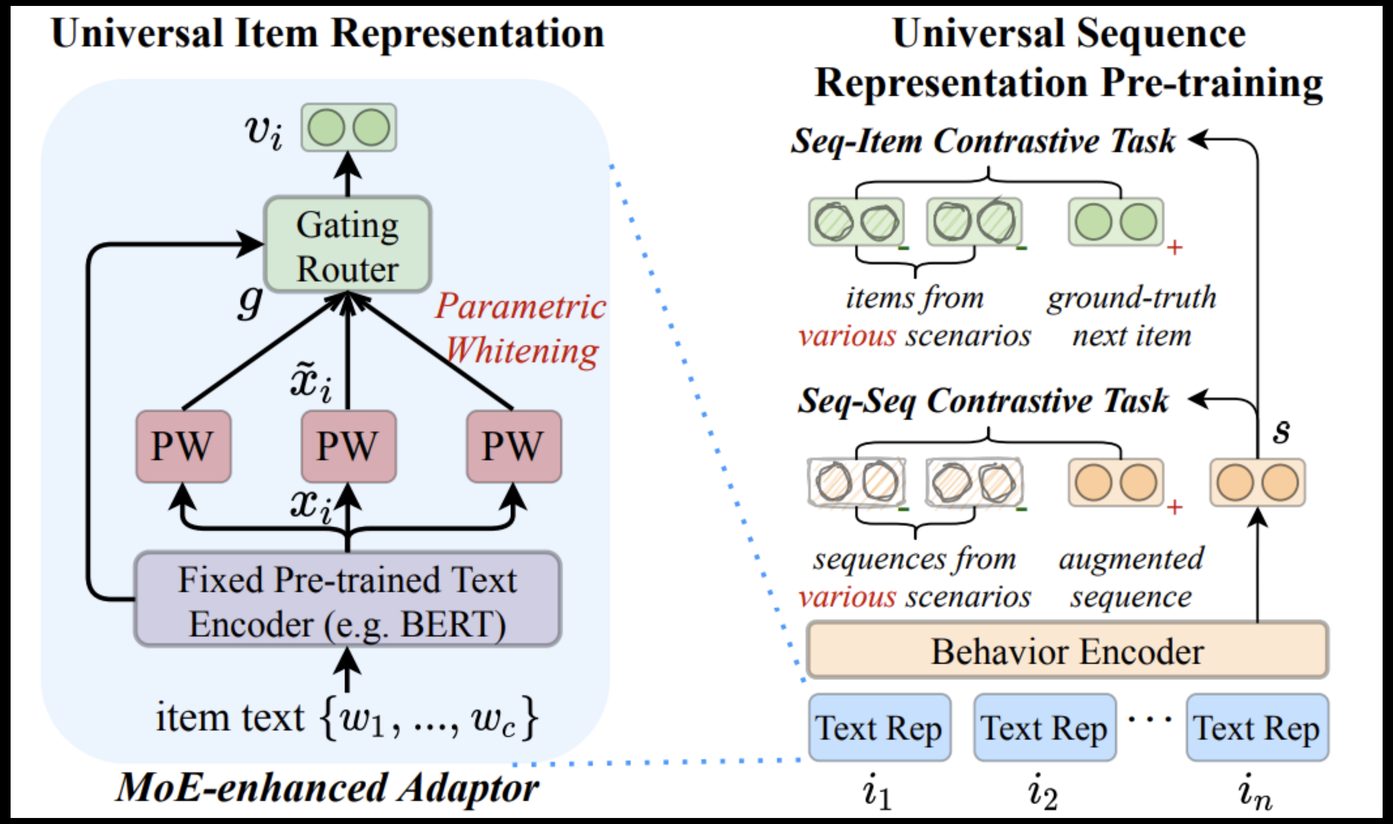
- 유연한 확장이 가능한 언어 모델 기반의 unified framework
- Pretext Task
  - 자연어 형태로 표현할 수 있는 거의 모든 user modeling tasks
- Downstream Tasks
  - Scoring, generation, explanation 등 다양함
- Datasets
  - Billion 단위의 in-house dataset

https://arxiv.org/abs/2205.08084

2.1 언어 모델을 활용한 기존 유저 모델링 연구

NAVER DEVIEW 2023

Towards Universal Sequence Representation Learning for Recommender Systems, **Alibaba, KDD 2022.**



- 도메인 및 회사 간 transfer learning을 위한 sequential recommendation framework 제안
- Pretext Task
  - 사용자의 seq-to-seq 및 seq-item contrastive learning
- Downstream Tasks
  - 다양한 도메인 및 시스템의 recommendation tasks
- Datasets
  - Amazon product review dataset + UK online retail data

https://dl.acm.org/doi/10.1145/3534678.3539381

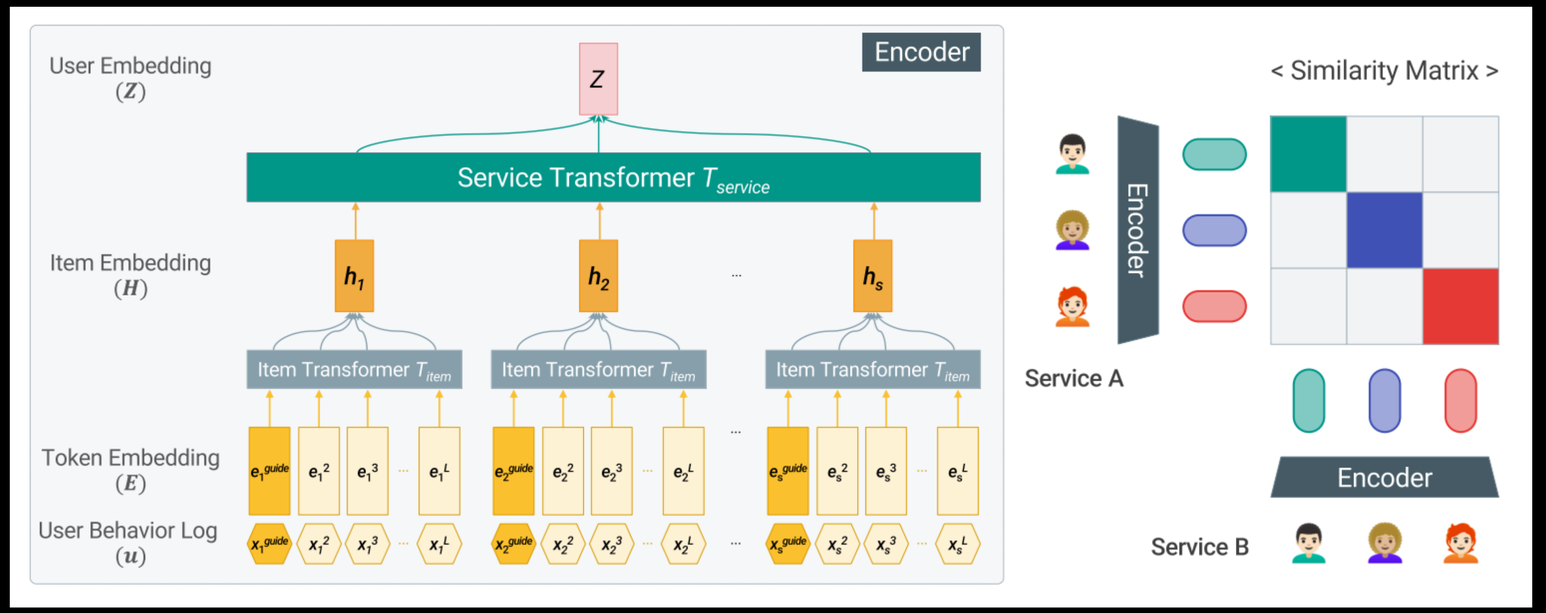
2.1 언어 모델을 활용한 기존 유저 모델링 연구

NAVER DEVIEW 2023

Scaling Law for Recommendation Models: Towards General-purpose User Representations, **NAVER, AAAI 2023.**

Search query
Mapo Han River Metro Xi → Pororo Camping Car → Geomdan Kumho New Town → Squid Game → Secret Forest Broadcast Time → Bookshelf → Home Surveillance Camera → Mapo Ehak Galbi

Purchased product description
Latex Powder Free Gloves → Conte quilted Vest zip up → Multi-use Super Silicone Broom → Children Kids Junior Basic T-Shirts → Comfortable Corduroy Leggings Pants for Weekly → High Quality Expert Olive Oil → Illy Coffee Capsules → Children's Bedroom Slippers Chandelier



- User representation learning을 활용해 추천 도메인에서 scaling law 탐구
- Pretext Task
  - 동일한 사용자의 서로 다른 서비스 로그 contrastive learning

- Downstream Tasks
  - User Profiling, User Targeting, Recommendation Tasks
- Datasets
  - Billion 단위의 in-house dataset, Amazon review datasets

https://arxiv.org/abs/2111.11294

2.2 분류 학습에 도움이 되는 언어 모델링

NAVER DEVIEW 2023

$$X = \{x_1, x_2, \dots, x_N\}$$

$$\mathcal{L}(\theta, X) = -\sum_{i=1}^{N} \log P(x_i | x_{i-k}, \dots, x_{i-1}; \theta)$$

- Language model pretraining은 downstream NLP classification tasks에 어떠한 영향을 주는 것인가?
  1. Pretraining corpora와 downstream corpora에 대한 비교분석
  2. 수학적 해석

2.2 분류 학습에 도움이 되는 언어 모델링

NAVER DEVIEW 2023

- [1] 웹 텍스트 기반 언어 모델이 아닌 task-specific (target task) 혹은 task-agnostic (other downstream) corpus를 활용한 언어 모델의 우수함에 대해 논의 [Gururangan et al., 2020; Lee et al., 2020; krishna2022downstream]

- [2] Language model pretraining 자체가 downstream task를 푸는데 도움이 되는 pretraining objectives 라는 것을 수학적으로 보이고 있음 [Saunshi et al., 2021; Wei et al., 2021]. 직관적인 아이디어는, classification tasks를 sentence completion tasks로 reformulate

- 적절한 형태의 language corpus에 language model pretraining을 하는 것은 downstream task에 긍정적인 영향을 준다!

2.2 분류 학습에 도움이 되는 언어 모델링

NAVER DEVIEW 2023

Task-specific data

네이버 예약
텍사스 데 브라질 → 더 플레이스 (서울 스퀘어 지점) → 스시 라쿠 → [EOS]

Amazon Scientific
3D 프린터용 유리 → 의료보조 도구 → [EOS]

Task-agnostic data

카페 방문 기록
더 콥스 - Liverpool Football Club → 아영이네행복주택 → 중고나라 → [EOS]

Amazon Movie & TV
트랜스포머 4 → 반 헬싱 → 닥터 하우스 시즌 1 → [EOS]

<인공적으로 만든 데이터셋 예시>

- 사용자 로그를 자연어 형태로 엮어서 tokenize 해준 뒤 모델 인풋으로 사용

- Task-specific data는 target downstream task의 user history logs data를, task-agnostic data는 target downstream이 아닌 다른 user history log data를 지칭

2.2 분류 학습에 도움이 되는 언어 모델링

NAVER DEVIEW 2023

Method	네이버 예약		Amazon Scientific	
	Recall@10	NDCG@10	Recall@10	NDCG@10
$LM_{webtext}$	0.83	0.82	0.80	0.68
$LM_{agnostic}$	0.83	0.81	0.78	0.65
$LM_{specific}$	1.0	1.0	1.0	1.0

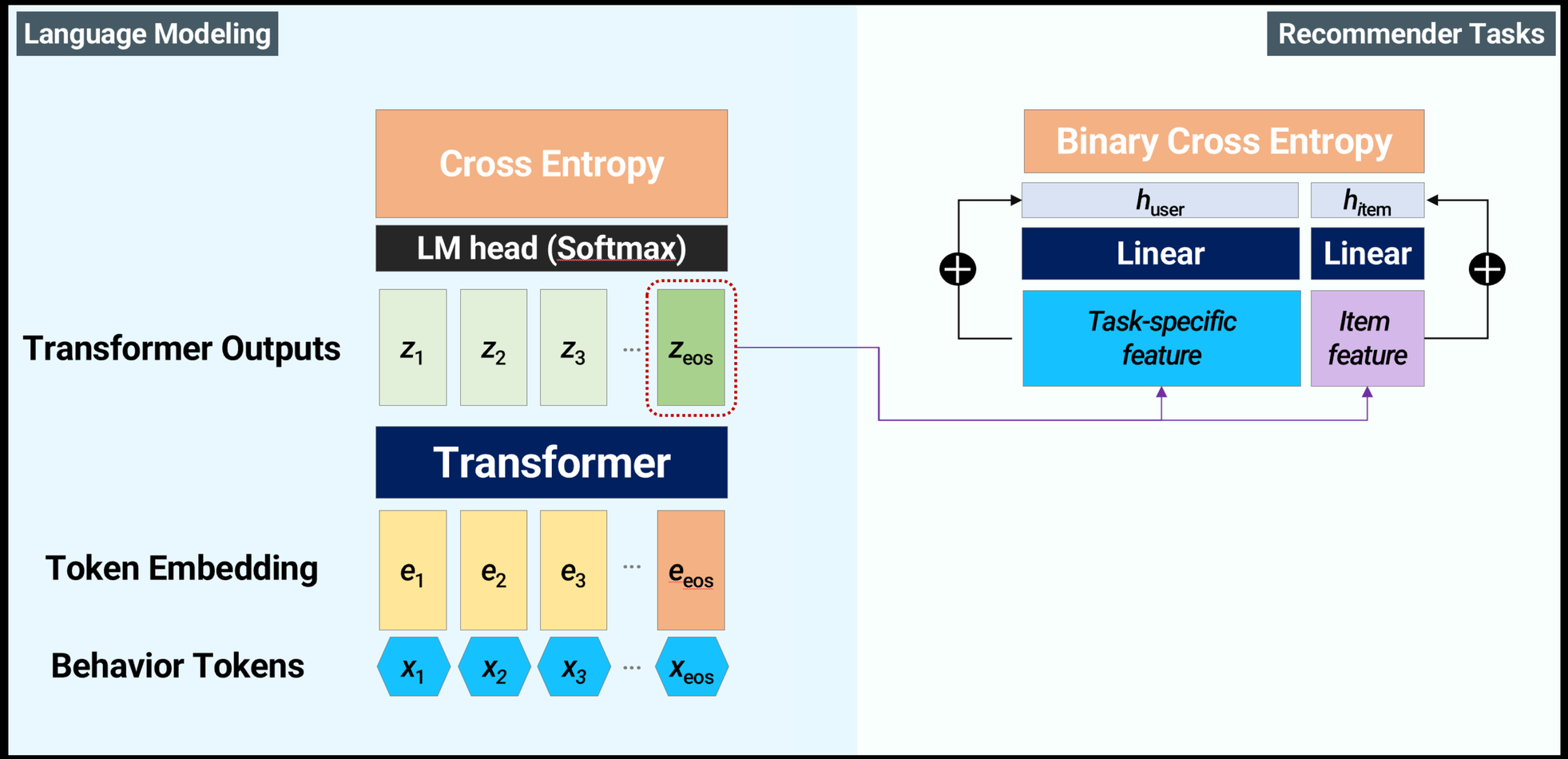
- 서로 다른 corpus에 대해 학습된 language model embedding을 활용해 downstream recommendation task에 linear probe한 결과

- Task-specific이 가장 우수했으며, task-agnostic의 데이터 양이 webtext에 비해 매우 적음에도 이 두 데이터셋의 성능이 비슷한 걸 확인

- 이 결과는 이전 연구의 결과와 맥이 닿으며, 연구에 큰 모티베이션이 됨

2.2 분류 학습에 도움이 되는 언어 모델링

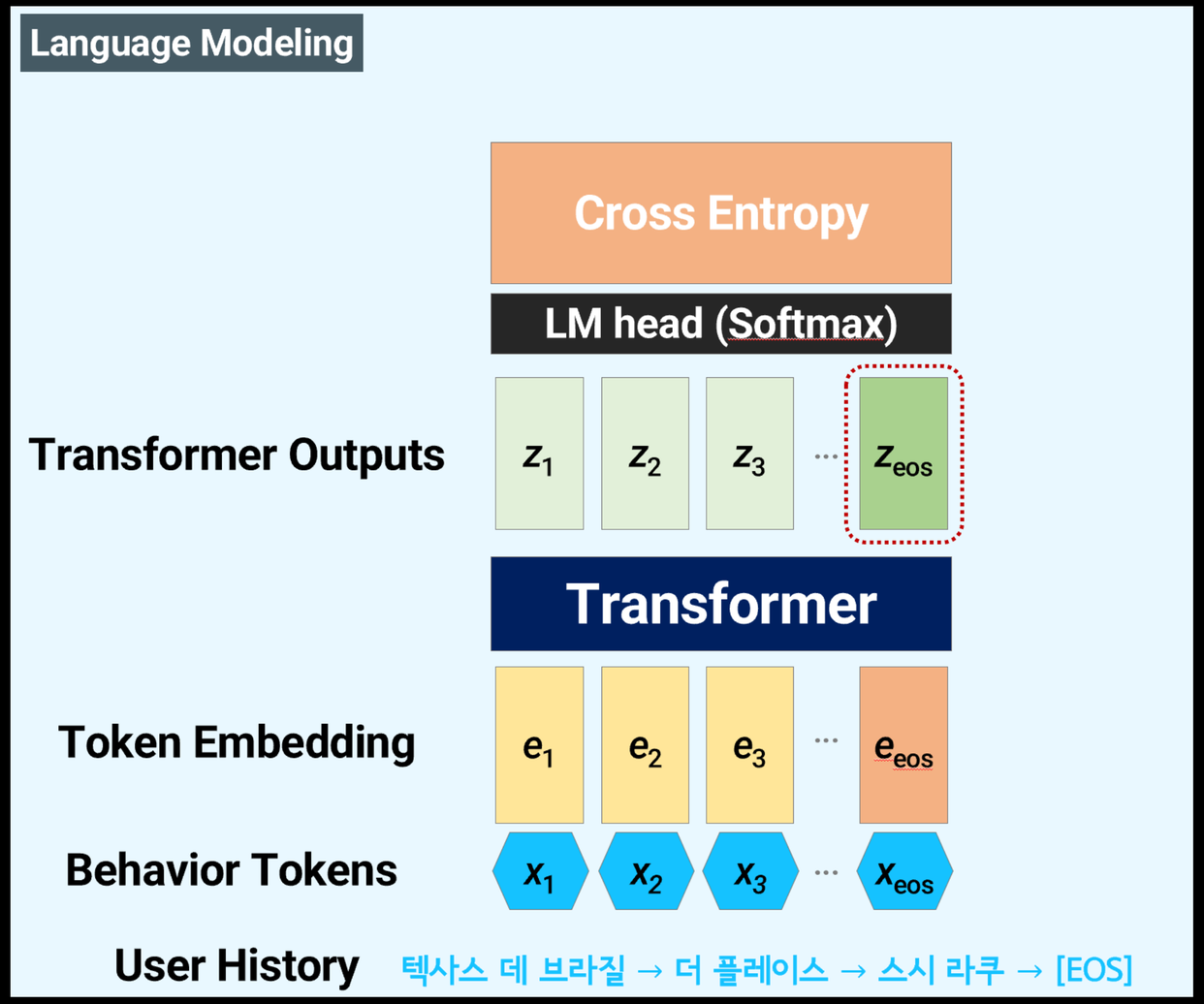
NAVER DEVIEW 2023



<모델 개요>

2.2 분류 학습에 도움이 되는 언어 모델링

NAVER DEVIEW 2023

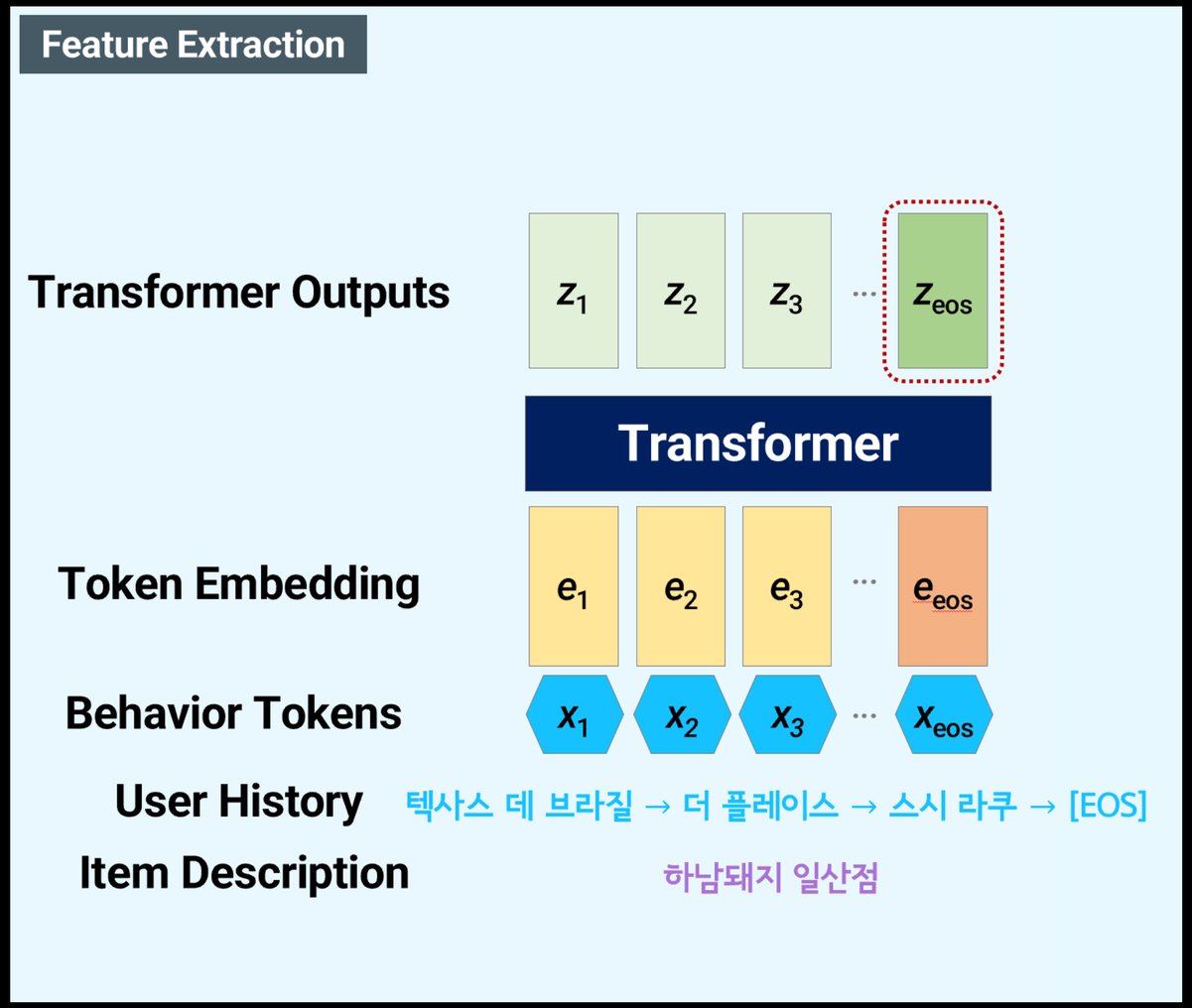


- 사용자 행동 기록을 이용해 Casual Language Modeling (i.e., next token prediction)을 진행

- 사용자 행동 기록의 경우 task-specific data를 사용하고, 경우에 따라서 task-agnostic data도 사용할 수 있음

2.2 분류 학습에 도움이 되는 언어 모델링

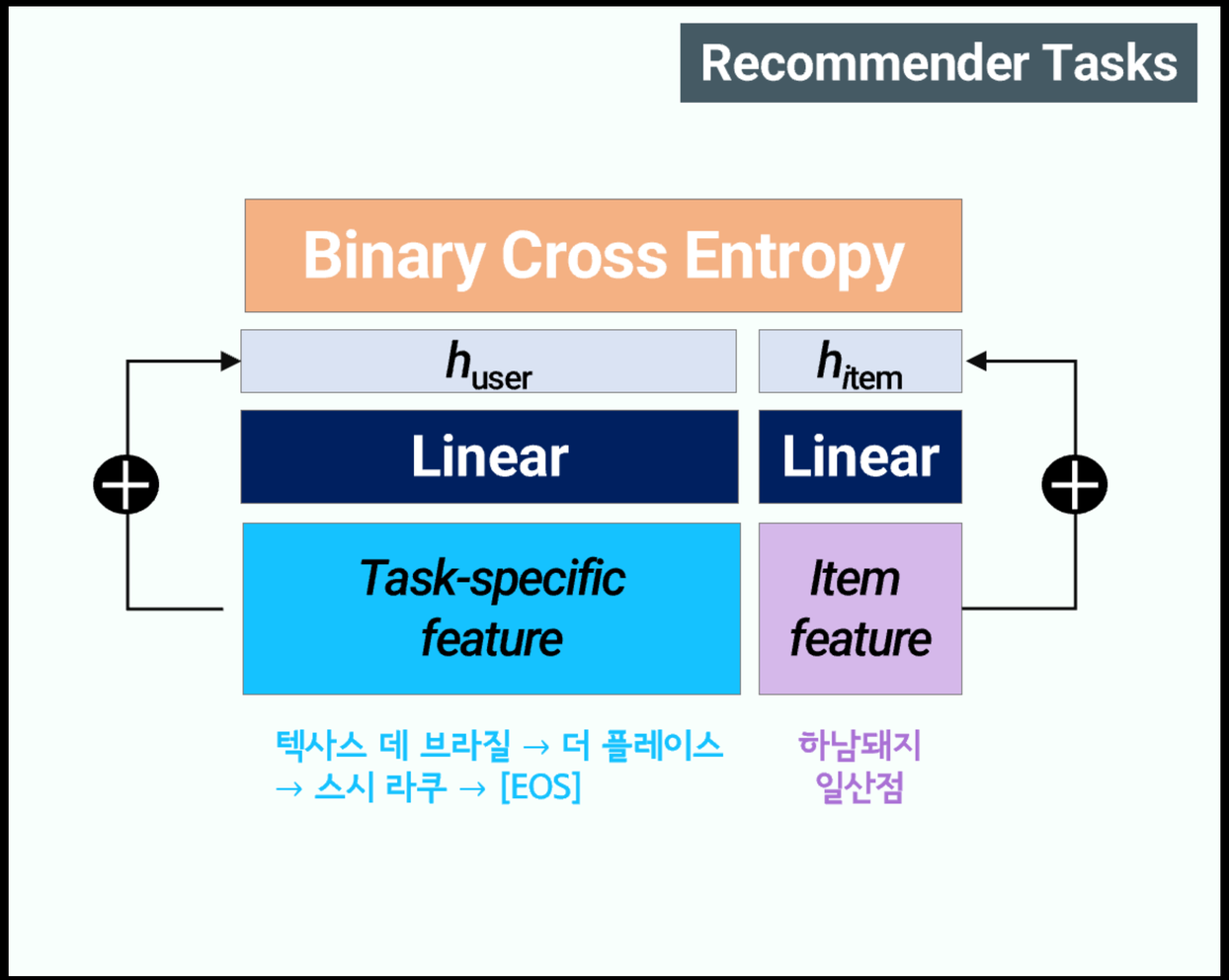
NAVER DEVIEW 2023



- 사용자 행동 기록 마지막에 EOS 토큰을 두고 이 토큰의 Transformer output feature를 user feature로 사용
- 마찬가지로, 아이템 기술 텍스트 끝에 EOS 토큰을 추가하고 이 EOS 토큰 feature를 item feature로 사용

2.2 분류 학습에 도움이 되는 언어 모델링

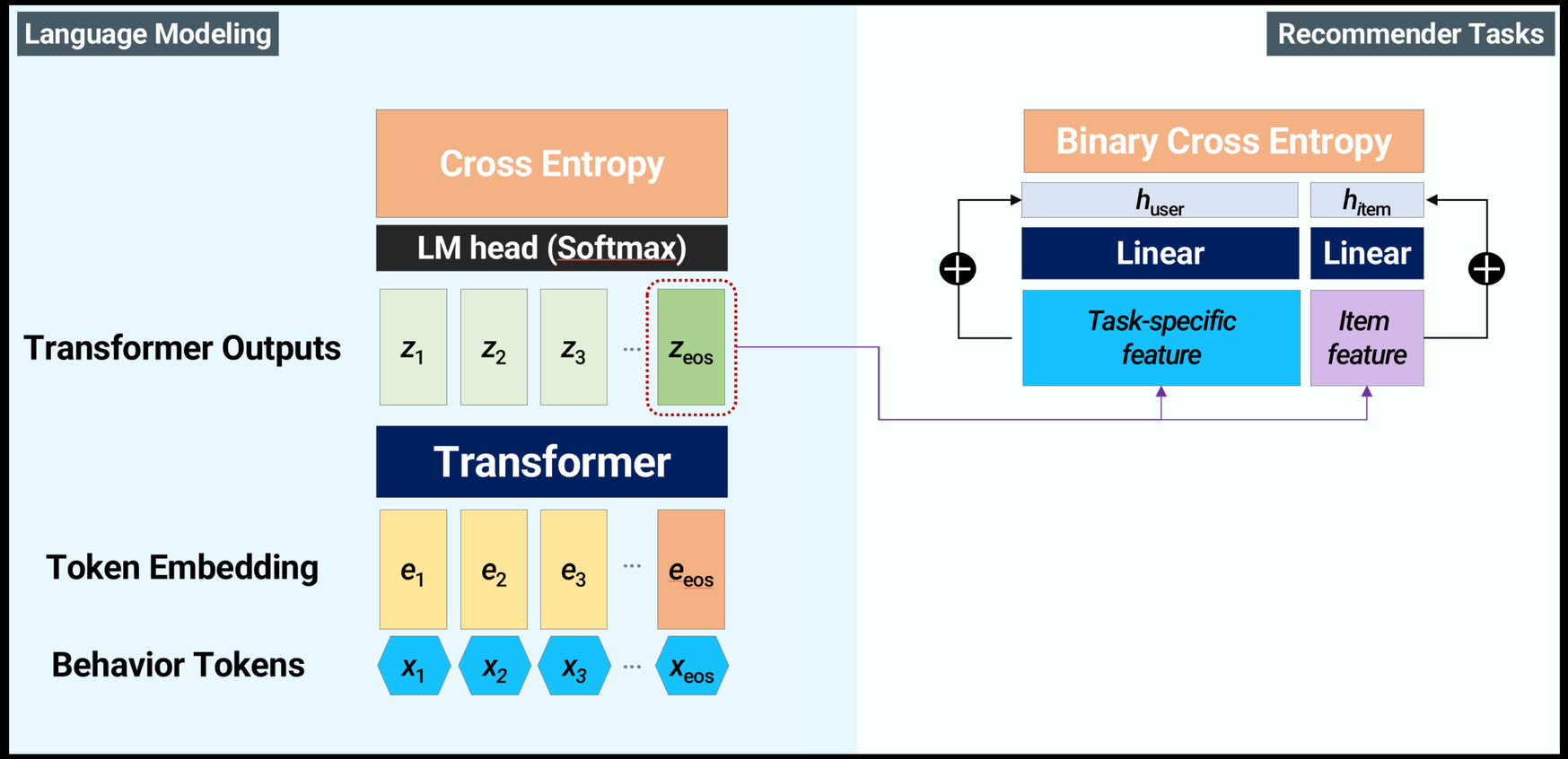
NAVER DEVIEW 2023



- 이렇게 추출한 user feature와 item feature로 downstream recommendation task를 학습
- 모델은 language modeling의 LM Head와 마찬가지로 간단한 linear layer를 사용

2.2 분류 학습에 도움이 되는 언어 모델링

NAVER DEVIEW 2023



- Catastrophic forgetting을 피하기 위해 language modeling과 recommendation task를 동시에 학습 [Chronopoulou et al., 2019; Karouzos 178 et al., 2021]

2.2 분류 학습에 도움이 되는 언어 모델링

NAVER DEVIEW 2023

Downstream Tasks	Task-specific data만 사용				추가적으로 task-agnostic data 사용			
	BST	LightGCN	$Ours_{-lm}$	$Ours$	UniSRec	CLUE	M6Rec	$Ours_{+agn.}$
네이버 예약	0.92	0.91	0.92	0.96	0.94	0.90	0.91	1.0
네이버 쇼핑	0.86	0.70	0.86	0.95	0.87	0.73	0.92	1.0
Amazon Scientific	0.49	0.42	0.74	0.98	0.93	0.70	0.73	1.0
Amazon Pantry	0.56	0.59	0.90	1.0	0.91	0.87	0.93	0.99

<Relative Recall@10>

- "$Ours$"는 task-specific data만 사용했음에도 task-agnostic data를 사용한 다른 모델들을 앞섬

- "$Ours_{-lm}$"는 "$Ours$" 에서 language modeling을 제외하고 학습한 모델로, 단순히 language modeling을 보조 objective로 학습하는 것이 추천 모델의 성능을 향상 시킴을 확인

2.3 Task-specific 및 Task-agnostic 유저 임베딩의 성능 향상

NAVER DEVIEW 2023

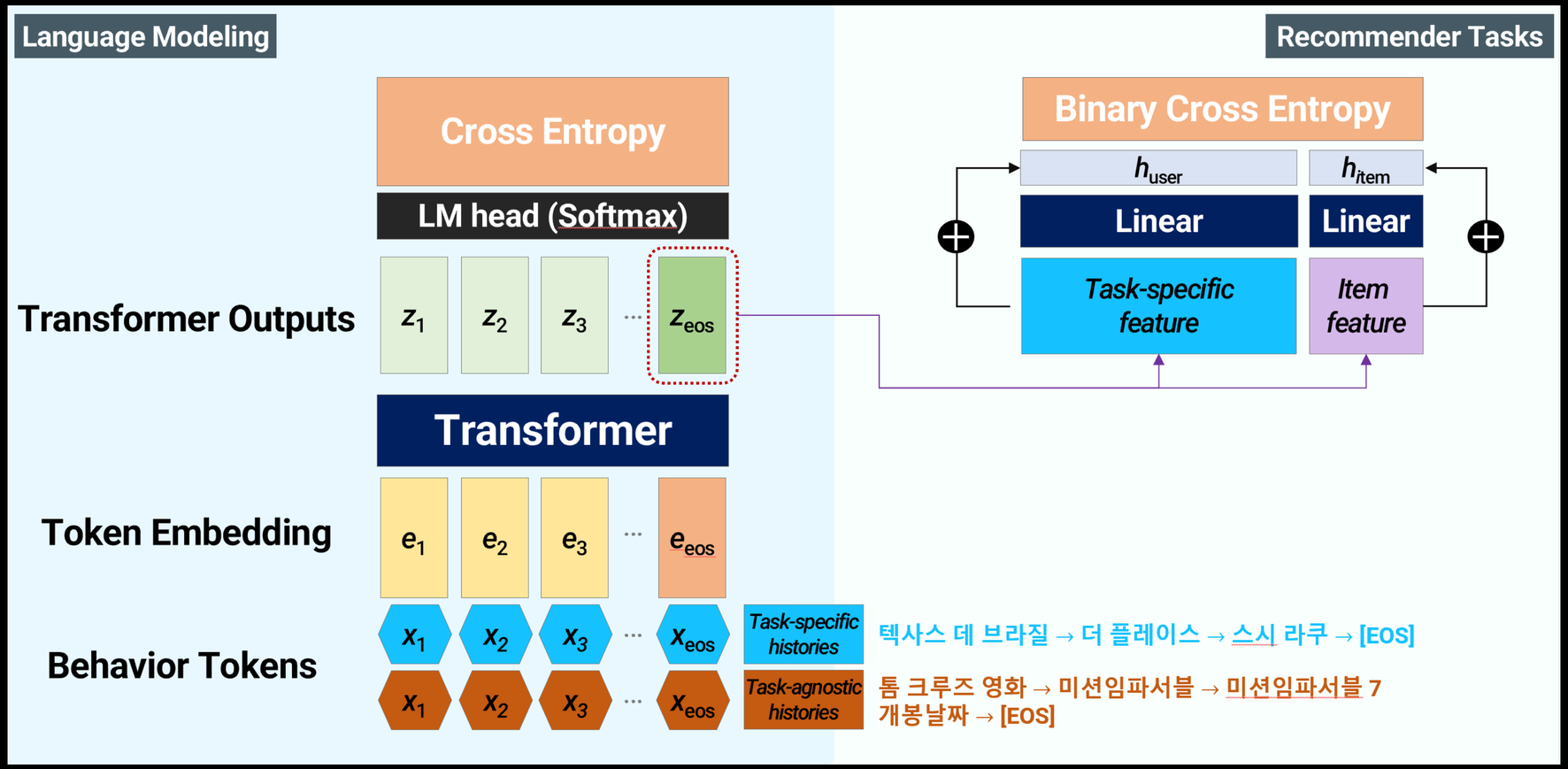
- **Task-agnostic data의 활용**: Downstream task의 데이터만을 사용하는 것은 unified frameworks로써 추천 모델의 잠재력을 제한

- **Transfer learning**: 실제 서빙 상황에서의 어려움. (1) 어떤 target downstream task에서 foundation model이 활용될지 미리 알 수 없으며, (2) user ID로 user log를 매칭하는 방법은 같은 회사의 서비스 로그에 대해서만 가능하고, (3) 큰 모델을 직접적으로 사용하는 데엔 컴퓨팅 리소스 제약이 따름

- **Task-agnostic user feature**: Cross-domain data를 추천에 활용하는 것은 많이 시도되는 방법. Language modeling 덕분에 task-agnostic data를 강력한 추가 user feature로 활용할 수 있음

2.3 Task-specific 및 Task-agnostic 유저 임베딩의 성능 향상

NAVER DEVIEW 2023



- Task-agnostic data를 활용하는 방법은 매우 간단하다. Language model objectives에 대해서만 task-agnostic data를 활용했고, 이는 여러 면에서 모델 성능에 도움을 주는 것으로 확인

2.3 Task-specific 및 Task-agnostic 유저 임베딩의 성능 향상

NAVER DEVIEW 2023

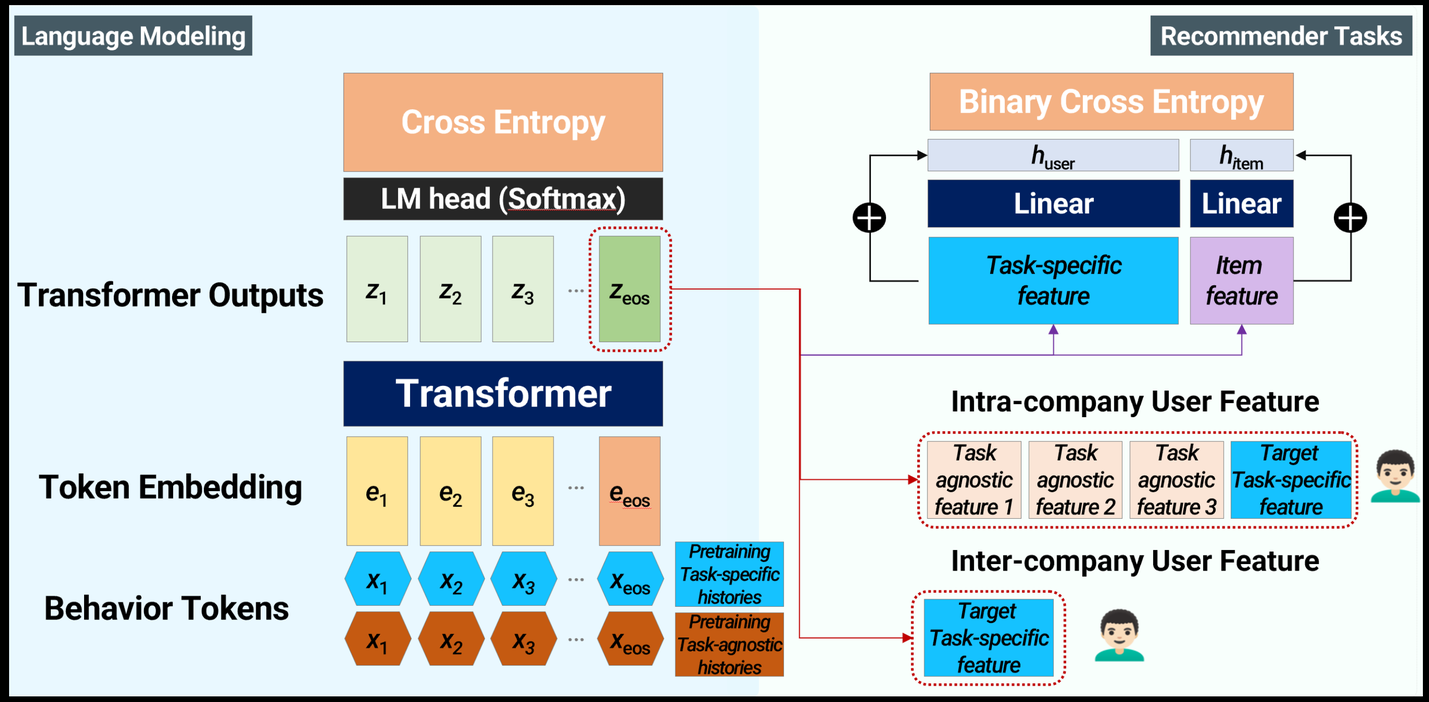
Downstream Tasks	Task-specific data만 사용				추가적으로 task-agnostic data 사용			
	BST	LightGCN	$Ours_{-lm}$	$Ours$	UniSRec	CLUE	M6Rec	$Ours_{+agn.}$
네이버 예약	0.92	0.91	0.92	0.96	0.94	0.90	0.91	1.0
네이버 쇼핑	0.86	0.70	0.86	0.95	0.87	0.73	0.92	1.0
Amazon Scientific	0.49	0.42	0.74	0.98	0.93	0.70	0.73	1.0
Amazon Pantry	0.56	0.59	0.90	1.0	0.91	0.87	0.93	0.99

<Relative Recall@10>

- "$Ours_{+agn.}$"는 language model objectives에 추가적으로 task-agnostic data를 활용해준 모델을 지칭
- 추가적인 task-agnostic data의 활용은 모델이 특정한 language corpus에 overfitting하는 것을 방지함으로써 일반화를 강화

2.3 Task-specific 및 Task-agnostic 유저 임베딩의 성능 향상

NAVER DEVIEW 2023



- Transfer learning의 경우, target downstream task가 아닌 다른 task로 pretraining 하고 Transformer를 freeze한 후 feature을 뽑아 feature-based transfer learning을 진행 (linear probe)

- 이 때 intra-company 회사라면, user ID로 서비스 로그를 매칭해 task-agnostic data를 추가적인 user feature로 활용

2.3 Task-specific 및 Task-agnostic 유저 임베딩의 성능 향상

NAVER DEVIEW 2023

모델	인풋	추론 스피드	파라미터
Transformer	History logs	1	125M
LightGCN	History logs	× 34	2M
$Ours^{TL}$	Pretrained user repr.	× 157	1.2M

- 다른 모델들의 경우 pretrain-then-finetune 방식을 따르고 있어, transfer learning은 여전히 매우 많은 리소스 필요

- 반면, 우리 모델의 경우 pretrained user representation을 추출한 후 linear probe 하기 때문에 빠르고 가볍게 downstream task에 transfer learning 가능

2.3 Task-specific 및 Task-agnostic 유저 임베딩의 성능 향상

NAVER DEVIEW 2023

Downstream Tasks	Task-specific user feature 사용			Task-agnostic user feature 사용			
	$Ours_{-lm}^{TL}$	$Ours^{TL}$	$Ours_{+agn.}^{TL}$	UniSRec	M6Rec	$Ours^{TL}$	$Ours_{+agn.}^{TL}$
네이버 예약	0.75	0.96	1.0	0.90	0.96	0.85	1.0
네이버 쇼핑	0.45	0.93	1.0	0.91	0.94	0.95	1.0

<Relative Recall@10>

- "$Ours_{-lm}^{TL}$" 은 transfer learning 능력이 매우 떨어짐
- Task-agnostic data를 language modeling에 활용하면, task-specific user feature에 대한 transfer learning 성능도 상승
- Task-agnostic data의 학습여부에 따라, task-agnostic data를 추가적인 user feature로 활용했을 때 성능 차이가 심한 경우 15%이상 남

2.4 분석

NAVER DEVIEW 2023

모델	네이버 예약
$Ours_{+agn.}$ $(0\%: 100\%)$	0.95
$Ours_{+agn.}$ $(30\%: 70\%)$	0.97
$Ours_{+agn.}$ $(50\%: 50\%)$	0.99
$Ours_{+agn.}$ $(70\%: 30\%)$	1.0
$Ours$	0.98

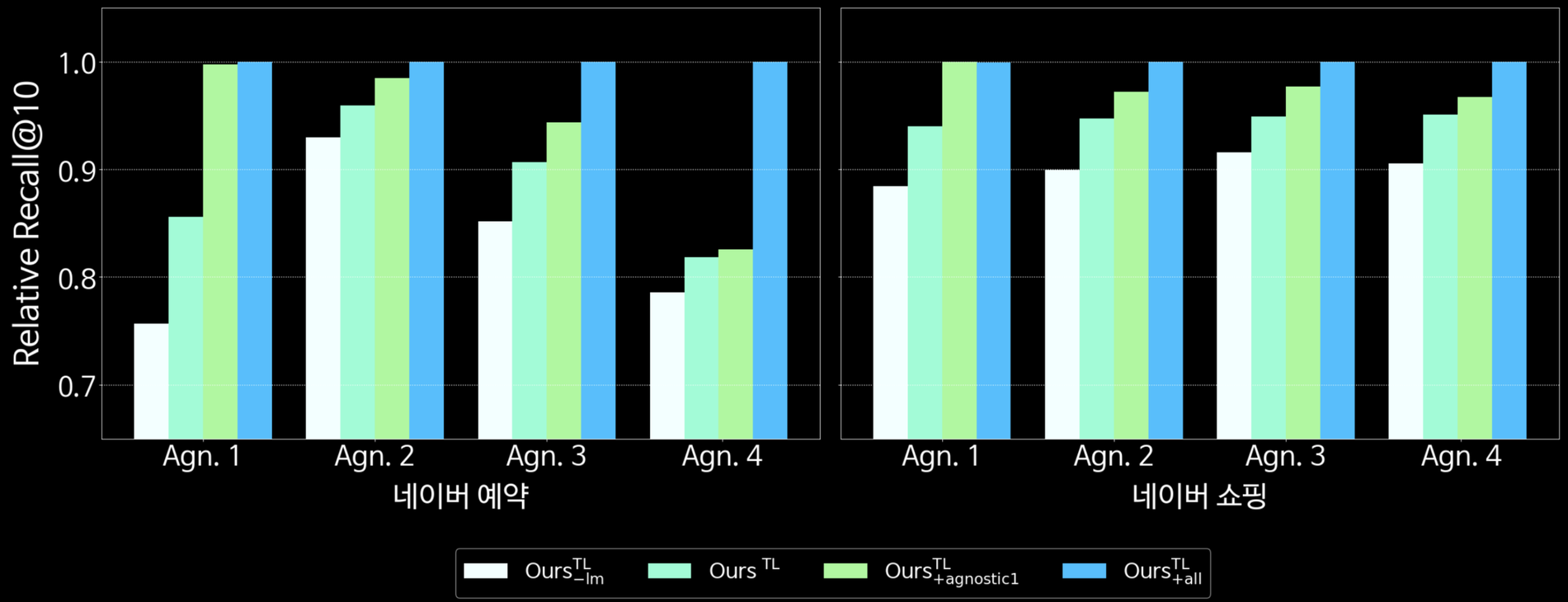
<Relative Recall@10>

- 데이터 양을 제한하고, task-specific data와 task-agnostic data 비율이 모델 성능에 어떠한 영향을 끼치는지에 대한 실험

- 데이터 양이 제한된 조건에서 task-specific data의 양이 증가할 수록 성능이 상승 했지만, 적절한 비율 (7:3)을 넘어서면 성능이 하락

2.4 분석

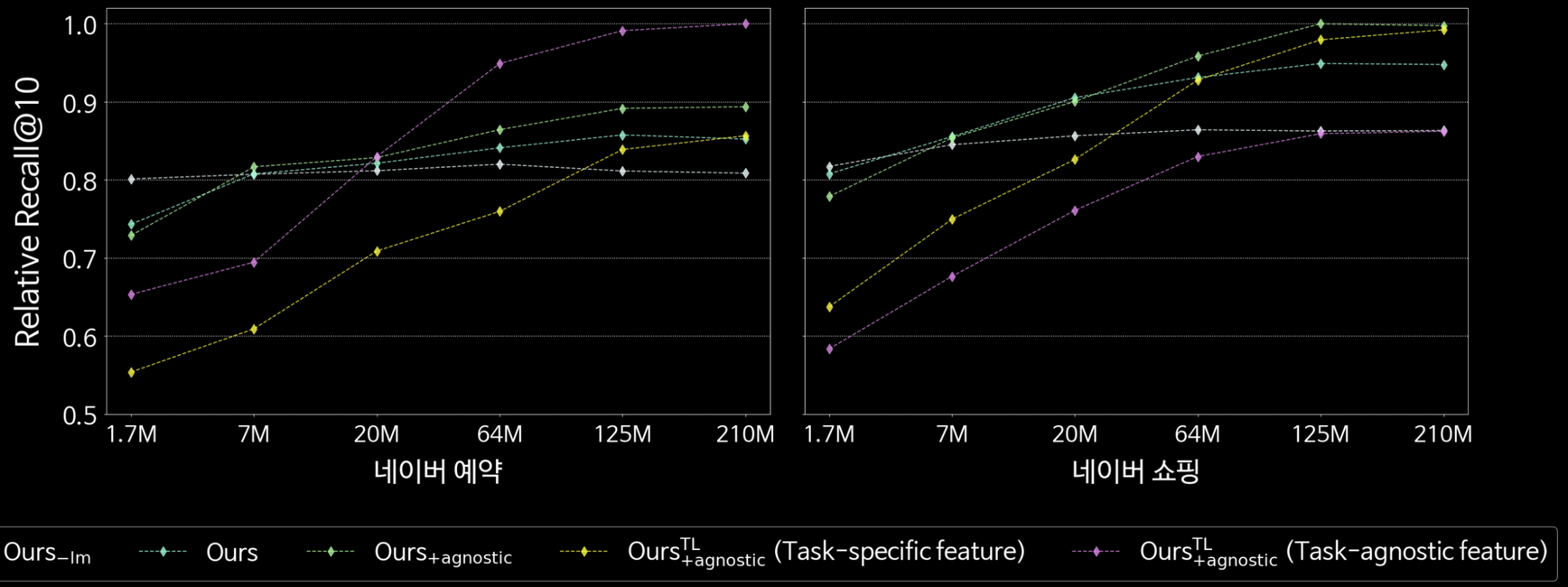
NAVER
DEVIEW
2023



- Pretraining task-agnostic corpus은 task-agnostic user feature에 어떠한 영향을 끼치는지에 대한 실험
- Task-agnostic을 pretraining시에 활용했냐 하지 않았냐에 따라 확연한 성능 차이를 보임

2.4 분석

NAVER DEVIEW 2023



- [Ardalani et al., 2022; Shin et al., 2023] 에서 추천시스템에서의 scaling law에 대해 논의하고 있음. 우리 연구도 비슷한 경향성을 보이지만, 일정 모델 사이즈 이상에서 더 이상 성능이 증가하지 않음. 이는 [Ardalani et al., 2022] 에서 언급한 power law plus a *constant* relationship (irreducible error) 을 따르는 것으로 보임

- "$Ours_{-lm.}$"은 모델 사이즈가 증가한다 해도 성능이 변하지 않음. Model complexity가 영향을 주지 못하는 것으로 보임

3. 서비스 적용 사례 및 성과

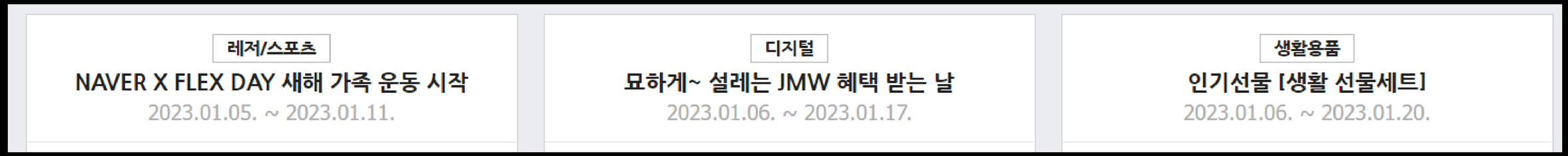
3.1 추천 시스템

NAVER
DEVIEW
2023

유저 인코더와 아이템 인코더의 두 결과 임베딩을 결합하여 유사도를 계산.

3.1 추천 시스템: 쇼핑 기획전 개인화 추천

NAVER DEVIEW 2023



	CTR		GMV	
	Cold User	Total	Cold User	Total
기존 ML 전문가가 만든 모델 (Baseline)				
범용적 유저 피쳐 활용 모델 (2021년 ver)	+ 52%	+ 14%	+ 8%	+ 2%
범용적 유저 피쳐 활용 모델 (2022년 ver)	+ 76%	+ 24%	+ 12%	+ 4%

범용적 유저 피쳐를 활용하여 전반적인 성능이 향상되었고 특히 Cold User 성능이 많이 올랐다.

3.1 추천 시스템

NAVER DEVIEW 2023

범용적 유저 피쳐 기반 추천 시스템을 다양한 서비스에 적용

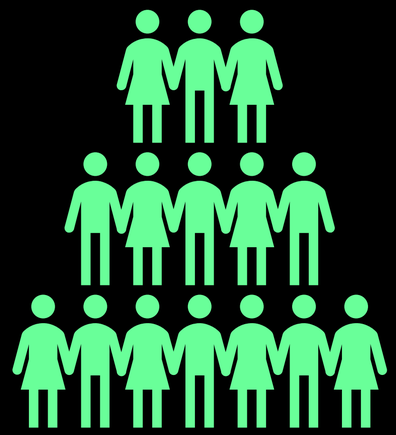
	범용적 유저 피쳐를 활용한 모델	메시지, 알림 타겟팅 (추천 시스템과 유사한 방법 사용)	
혜택 피드	+ 346%	톡톡 타겟 메시지 마케팅	+ 450%
커머스솔루션 고객 맞춤 (구독 수)	+ 15%	네이버 알림	+ 5%
커머스솔루션 비슷한 상품	+ 14%	원뿔딜 알림	+ 40%
패션 타운	+ 110%		

단순 Baseline 대비 성능 향상 (CTR)

3.2 광고 타겟팅

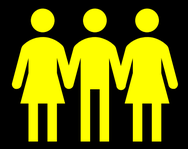
NAVER
DEVIEW
2023

1. 이번에 **BTS**가 NOW에 출연해서 광고를 하고 싶다! 누구에게 보내야 하나?
2. NOW에서 아이돌 관련 채널을 구독한 유저들에게 보내자! **(시드유저)**
3. 하지만.. BTS를 좋아하면서 NOW를 사용하지 않는 사람들이 많을텐데..

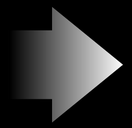


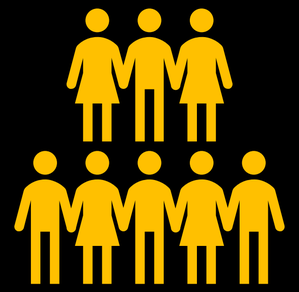
네이버 전체 유저

시드 유저그룹



시드 유저와 유사한
유저들 탐색





광고 타겟팅 유저그룹

This cover has been designed using images from Flaticon.com

3.2 광고 타겟팅

NAVER
DEVIEW
2023

- 딥러닝 피쳐는 코사인 거리와 같은 단순한 메트릭으로 의미를 알기 어렵다.
- 미리 만들어놓은 **범용적 유저 피쳐** 사용
- MLP와 같은 모델을 사용해서 시드 유저를 구분하는 분류기를 학습
- 분류기의 출력값(0 ~ 1)을 기준으로 타겟팅 유저들을 선정

This cover has been designed using images from Flaticon.com

3.2 광고 타겟팅

NAVER DEVIEW 2023

	비교군	지표	상대적 성과
A 광고 (쇼핑 광고)	나이 타겟팅*	매출 / 클릭	+ 585%
B 광고 (쇼핑 광고)	기존 유저 세그먼트 도구	매출 / 클릭	+ 215%
C 광고 (가입 광고)	랜덤	가입 / 클릭	+ 268%

범용적 유저 피쳐를 활용하여 랜덤 또는 기존 타겟팅 방식 대비 상당한 성능 차이를 보였다.

* 광고 날짜가 다름

Q & A

Thank You

Contact Point

곽하녹: hanock.kwak2@navercorp.com | https://careerly.co.kr/profiles/579426

신규용: ky.shin@navercorp.com