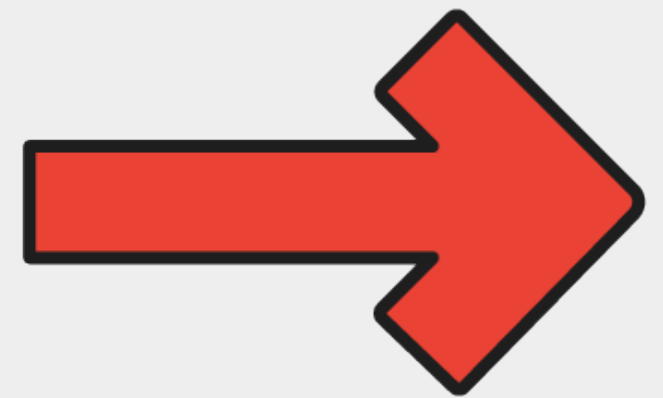


An Efficient All-round LLM-based Recommender System

GDGoC INU AI Part Paper Seminar

AI core 지원근



Large Language Models meet Collaborative Filtering: An Efficient All-round LLM-based Recommender System

Sein Kim*

rlatpdlsgns@kaist.ac.kr

Korea Advanced Institute of Science
and Technology
Daejeon, Republic of Korea

Hongseok Kang*

ghdtjr0311@kaist.ac.kr

Korea Advanced Institute of Science
and Technology
Daejeon, Republic of Korea

Seungyoon Choi

csyoon08@kaist.ac.kr

Korea Advanced Institute of Science
and Technology
Daejeon, Republic of Korea

Donghyun Kim

amandus.kim@navercorp.com

NAVER Corporation
Seongnam, Republic of Korea

Minchul Yang

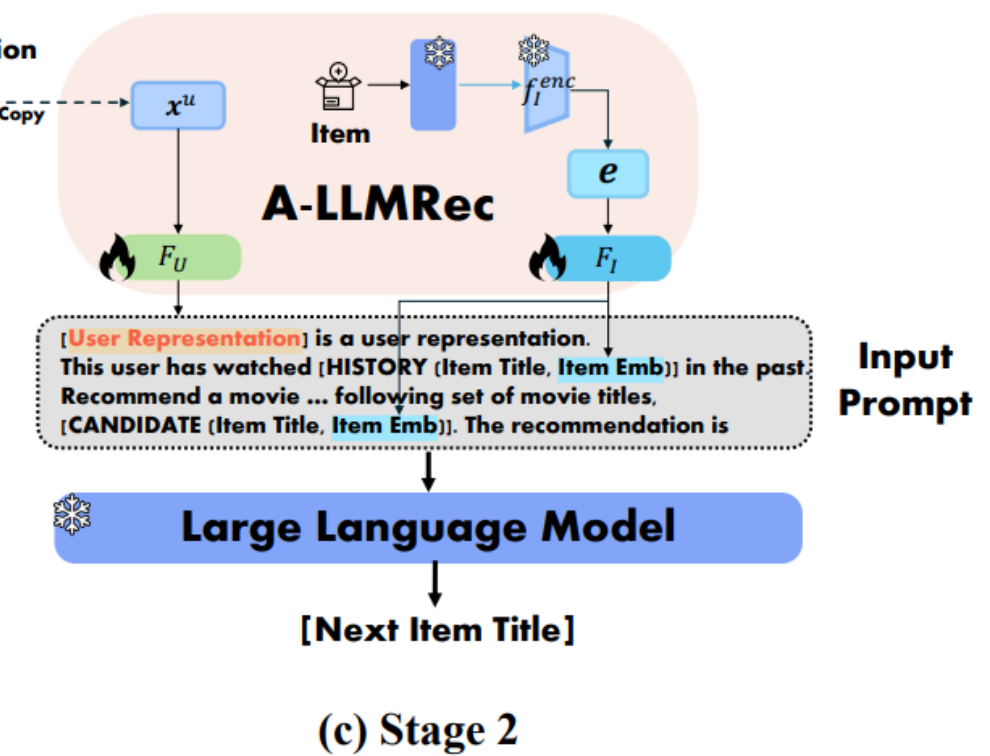
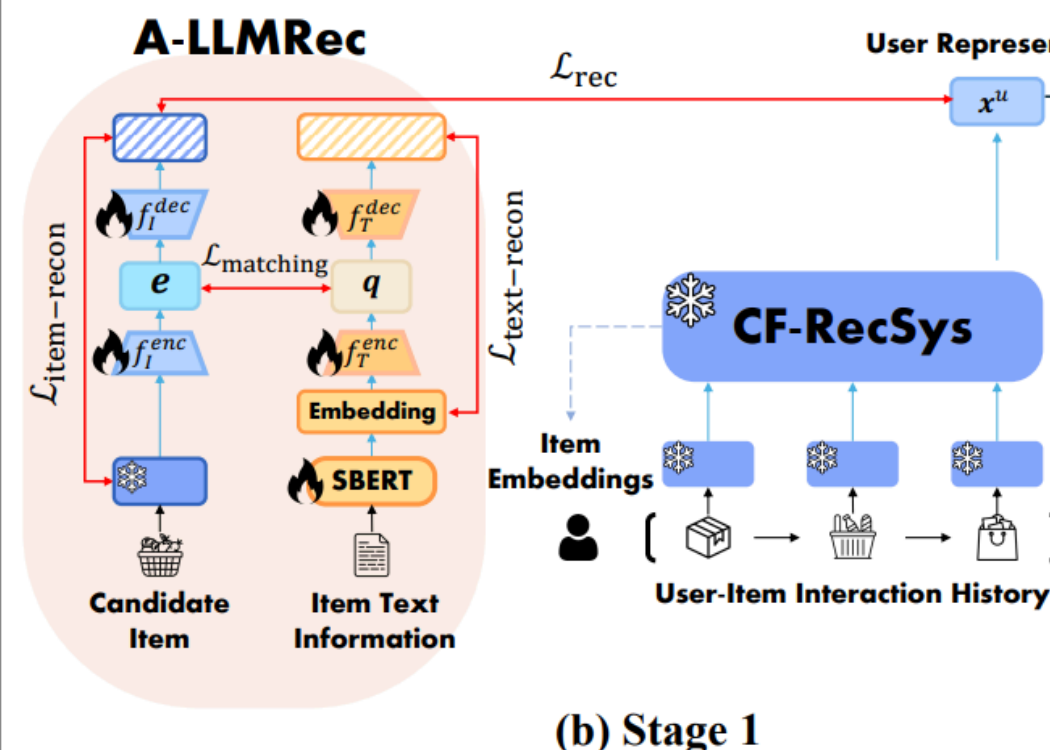
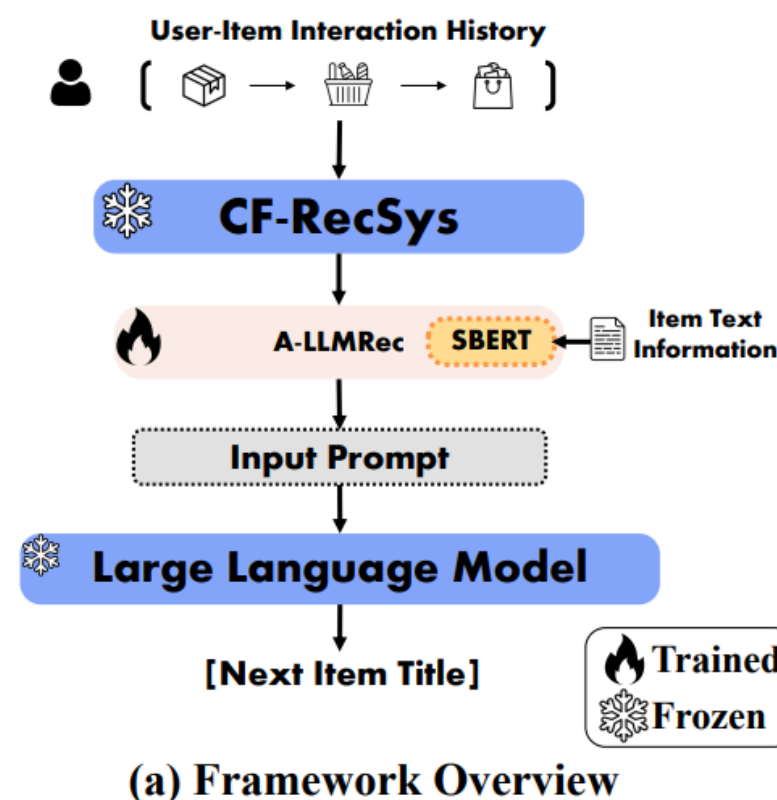
minchul.yang@navercorp.com

NAVER Corporation
Seongnam, Republic of Korea

Chanyoung Park[†]

cy.park@kaist.ac.kr

Korea Advanced Institute of Science
and Technology
Daejeon, Republic of Korea



INDEX

1. Introduction
2. Method
3. Experiments
4. Conclusion

1. Introduction

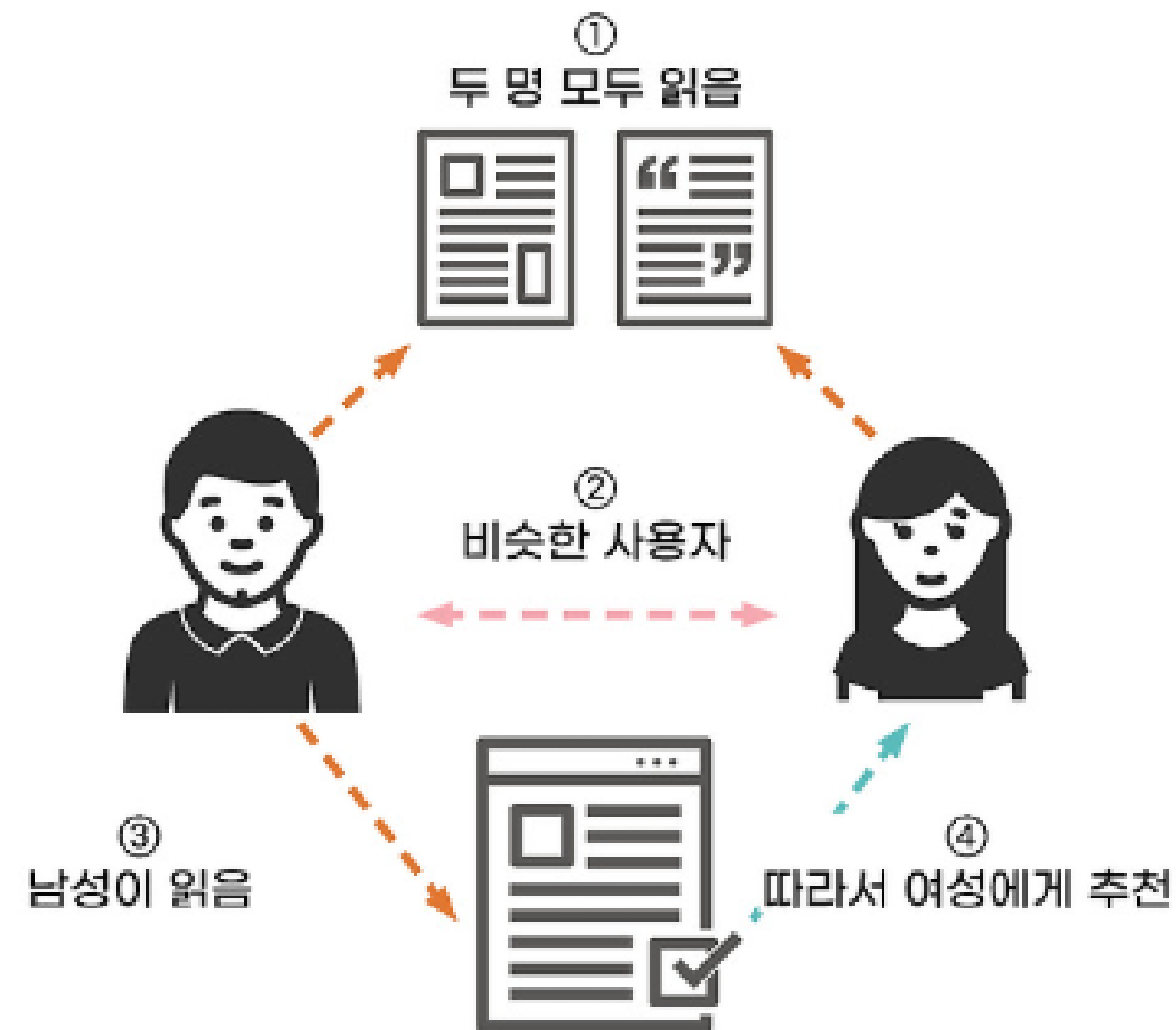
Introduction

추천 시스템 기본 개념

1. 협업 필터링 (Collaborative Filtering, CF): 사용자-아이템 상호작용 데이터를 기반으로 추천을 수행

협업 필터링

(Collaborative Filtering)



출처 : 마음을 읽는 개인화 마케팅의 시대

Introduction

추천 시스템 기본 개념

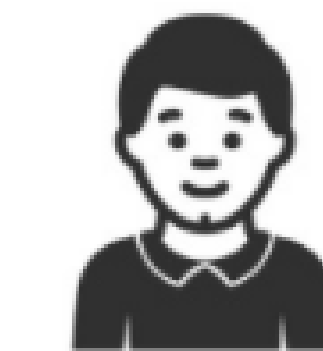
1. 협업 필터링 (Collaborative Filtering, CF): 사용자-아이템 상호작용 데이터를 기반으로 추천을 수행

협업 필터링

(Collaborative Filtering)



How recommendation?



New User!



New Item!

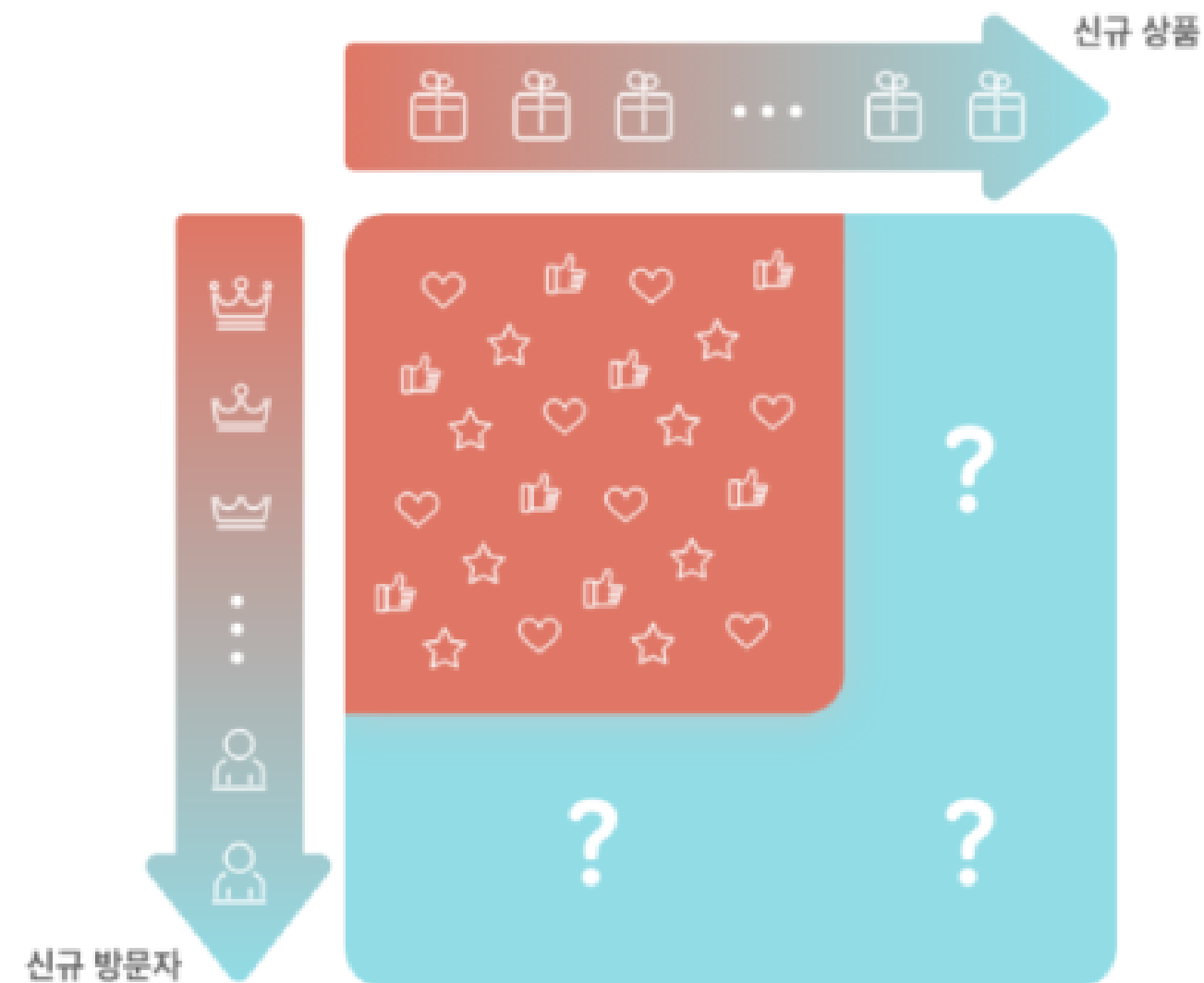
출처 : 마음을 읽는 개인화 마케팅의 시대

Introduction

추천 시스템 기본 개념

2. 콜드 스타트 문제 (Cold Start Problem): 신규 사용자/아이템의 데이터 부족으로 추천 품질 저하

콜드 스타트 문제 (Cold start Problem)



콜드스타트의 원인

출처 : [쇼핑몰 AI 상품 추천의 '콜드스타트' 문제 해결하기](#)

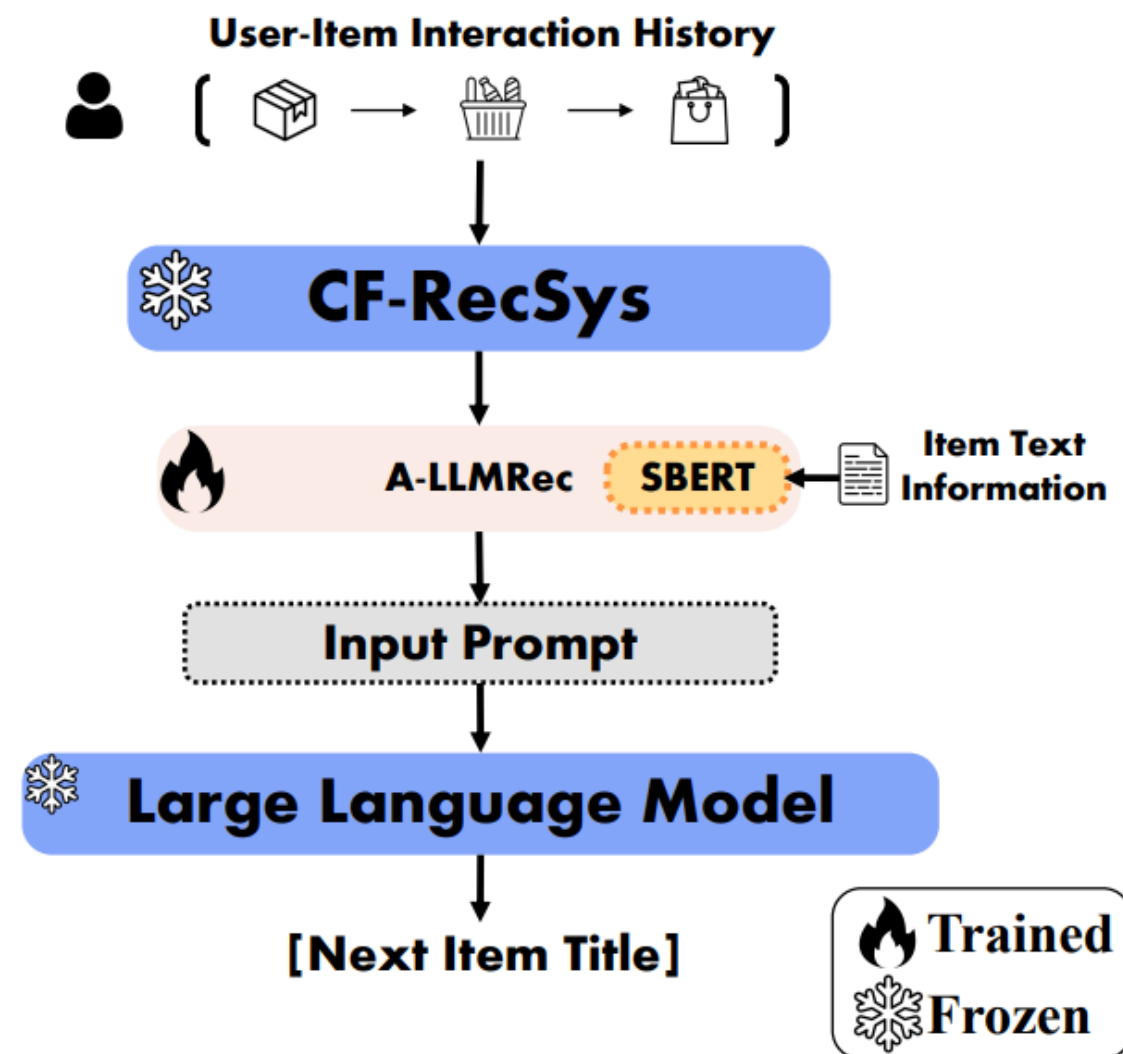
Introduction

추천 시스템 기본 개념

3. 모달리티 정보 활용 (Modality-aware RecSys)

텍스트, 이미지 등 아이템 속성 정보를 활용하여 콜드 스타트 문제 해결 시도

모달리티 정보 활용 (Cold start Problem)



(a) Framework Overview

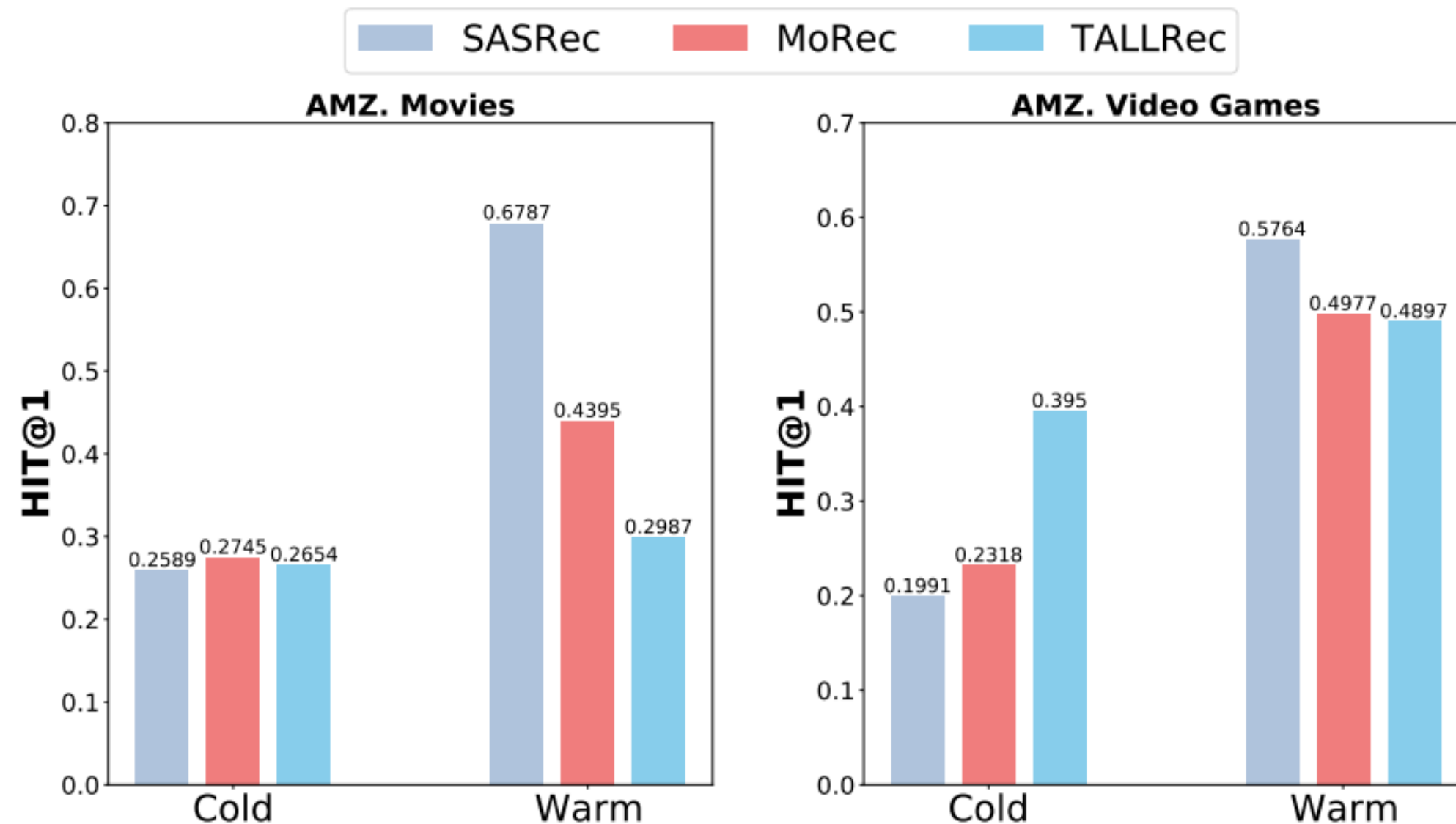
사용자/아이템 협업 정보
+
LLM의 방대한 지식
=
더 나은 추천
콜드 스타트 문제 해결

Introduction

LLM을 활용한 추천 시스템 연구 흐름

- 전통적인 협업 필터링 모델 (SASRec 등): 데이터가 많을 때(Warm Scenario) 강력한 성능 발휘
- 모달리티 기반 모델 (MoRec 등): 콜드 스타트에서 성능이 좋지만 협업 지식 부족으로 Warm Scenario에서는 성능 저하
- LLM 기반 추천 시스템 (TALLRec 등): LLM을 활용해 추천하지만, 협업 지식 활용 부족 문제 존재

Comparisons between collaborative filtering model

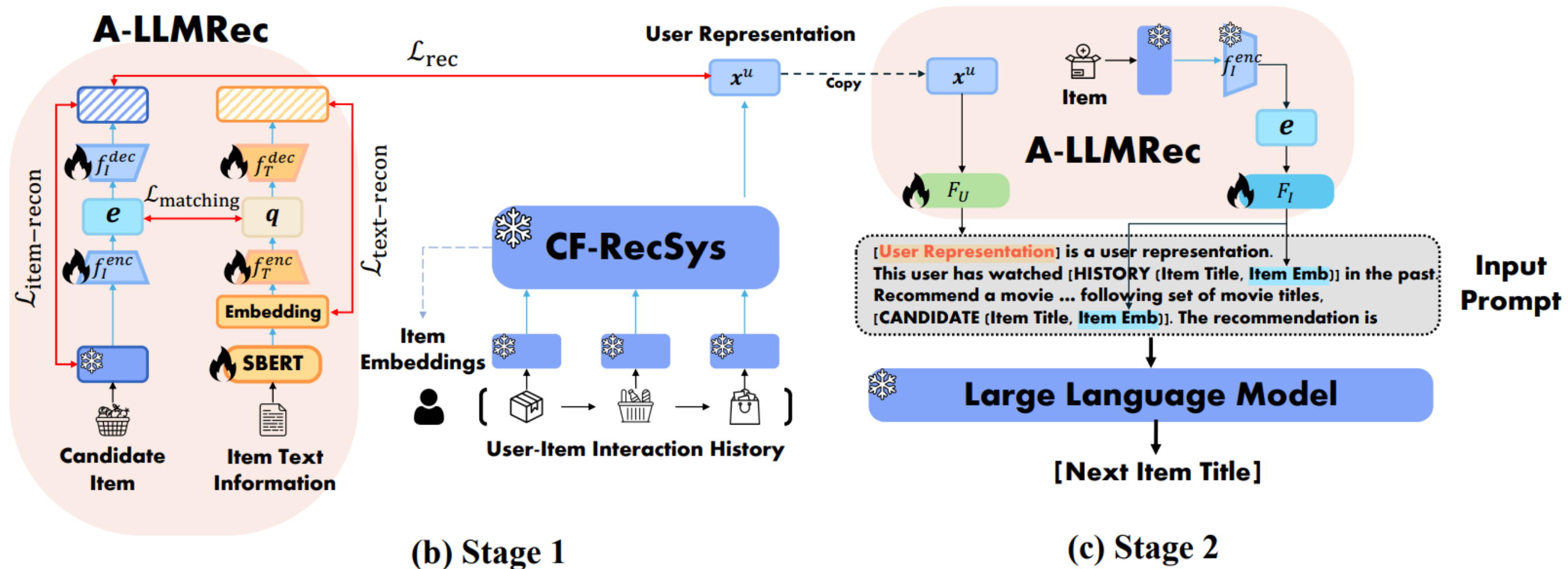


2. Method

Method

주요 아이디어

1. 협업 필터링 모델의 사용자/아이템 임베딩을 LLM에 전달
2. LLM의 강력한 언어 이해력과 CF 모델의 협업 지식을 결합
3. LLM이나 CF 모델을 직접 미세조정(fine-tuning)하지 않고 효율적으로 연계

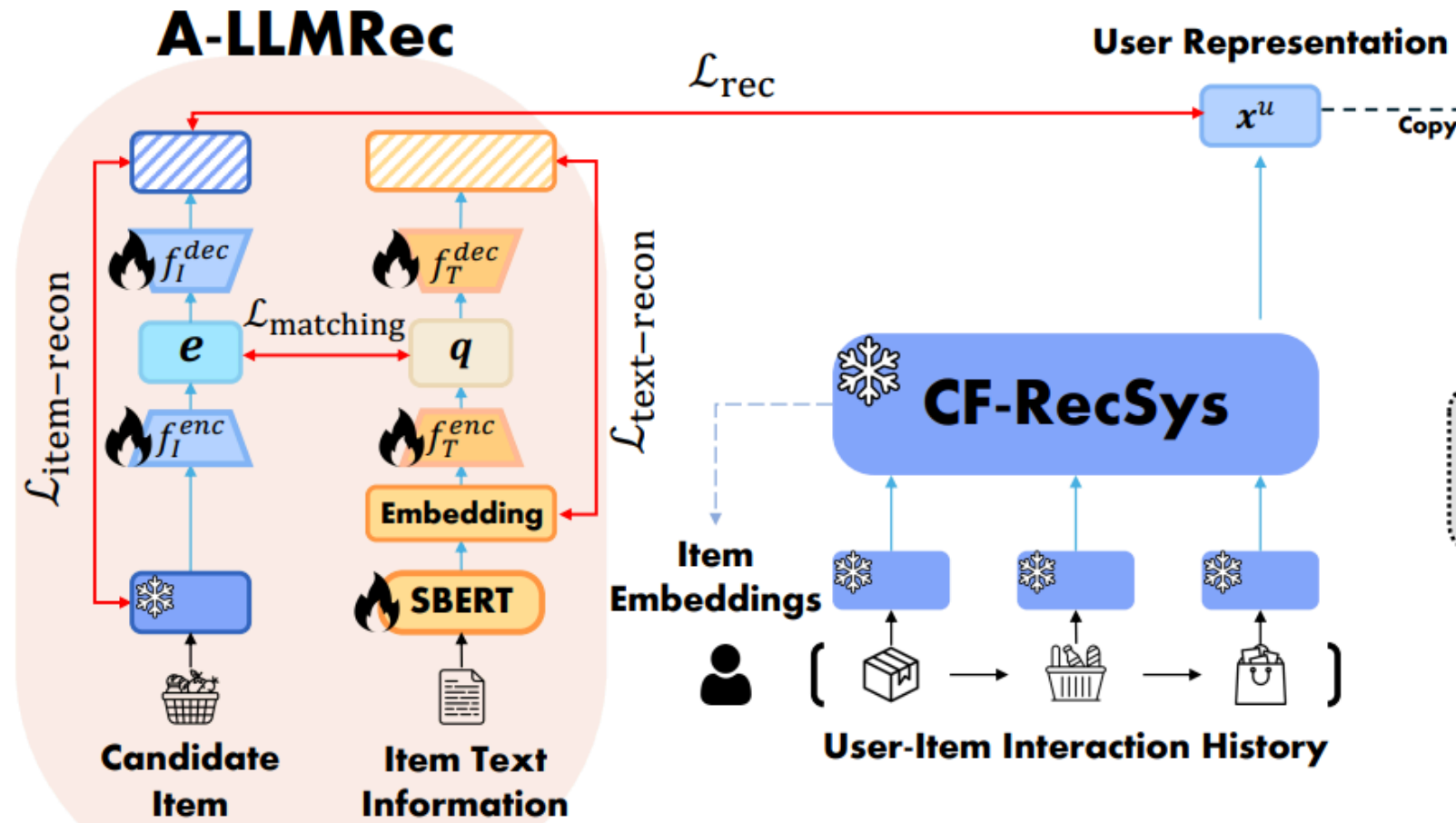


Method

A-LLMRec의 두 가지 학습 단계

(1) Collaborative-Text Embedding Alignment (Stage-1)

- 협업 필터링(CF) 모델의 아이템 임베딩을 SBERT 텍스트 임베딩과 정렬
- 모달리티 간 정렬을 위한 Matching Loss, Reconstruction Loss 적용
- 이 과정에서 CF-RecSys의 협업 지식이 LLM에 전달될 준비 완료



(b) Stage 1

$$\begin{aligned}\mathcal{L}_{matching} &= \mathbb{E}_{S^u \in \mathcal{S}} \left[\mathbb{E}_{i \in S^u} [MSE(\mathbf{e}_i, \mathbf{q}_i)] \right] \\ &= \mathbb{E}_{S^u \in \mathcal{S}} \left[\mathbb{E}_{i \in S^u} [MSE(f_I^{enc}(\mathbf{E}_i), f_T^{enc}(\mathbf{Q}_i))] \right]\end{aligned}$$

$$\mathcal{L}_{item-recon} = \mathbb{E}_{S^u \in \mathcal{S}} \left[\mathbb{E}_{i \in S^u} [MSE(\mathbf{E}_i, f_I^{dec}(f_I^{enc}(\mathbf{E}_i)))] \right]$$

$$\mathcal{L}_{text-recon} = \mathbb{E}_{S^u \in \mathcal{S}} \left[\mathbb{E}_{i \in S^u} [MSE(\mathbf{Q}_i, f_T^{dec}(f_T^{enc}(\mathbf{Q}_i)))] \right]$$

$$\begin{aligned}\mathcal{L}_{rec} &= - \sum_{S^u \in \mathcal{S}} [\log(\sigma(s(\mathbf{x}_{|S^u|-1}^u, f_I^{dec}(f_I^{enc}(\mathbf{E}_{i_{|S^u|}^u)))))) \\ &\quad + \log(1 - \sigma(s(\mathbf{x}_{|S^u|-1}^u, f_I^{dec}(f_I^{enc}(\mathbf{E}_{i_{|S^u|}^u, -})))))]\end{aligned}$$

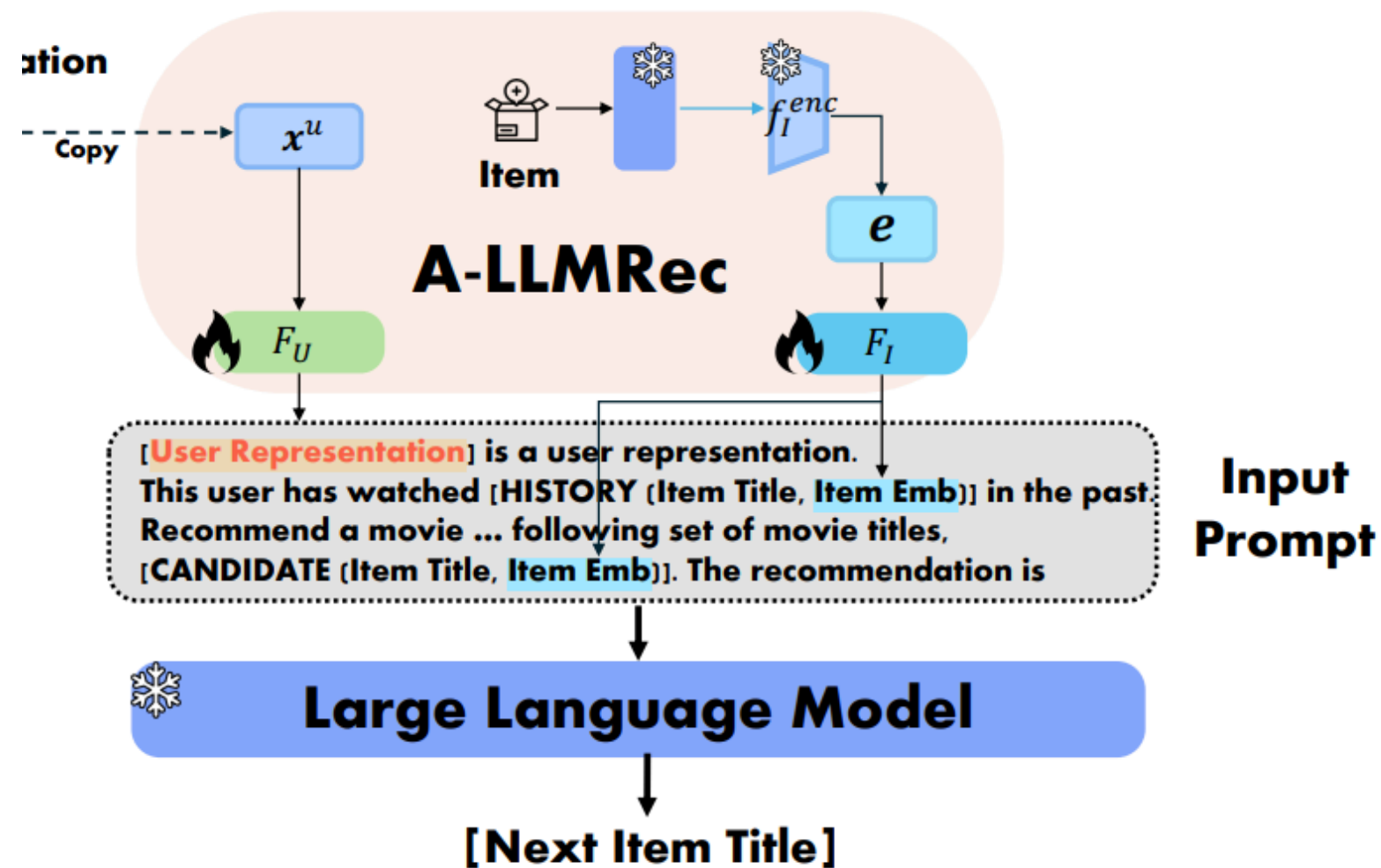
$$\mathcal{L}_{stage-1} = \mathcal{L}_{matching} + \alpha \mathcal{L}_{item-recon} + \beta \mathcal{L}_{text-recon} + \mathcal{L}_{rec}$$

Method

A-LLMRec의 두 가지 학습 단계

(2) Alignment with LLM & Recommendation (Stage-2)

- Stage-1에서 정렬된 임베딩을 LLM의 토큰 공간으로 매핑
- LLM의 프롬프트에 사용자/아이템 임베딩을 삽입하여 협업 지식을 전달
- 추가 학습 없이 LLM이 추천을 수행할 수 있도록 최적화



(c) Stage 2

LLM Input: [User Representation] is a user representation.
This user has watched [HISTORY (Item Titles, Item Emb)]
in the past. Recommend a movie for this user to watch
next from the following set of movie titles,
[CANDIDATE (Item Titles, Item Emb)]. The
recommendation is

LLM Output: [Next Item Title]

$$\max_{\theta} \sum_{S^u \in \mathcal{S}} \sum_{k=1}^{|y^u|} \log(P_{\theta, \Theta}(y_k^u | p^u, y_{<k}^u))$$

3. Experiments

Experiments

사용 벤치마크 데이터셋

A-LLMRec의 성능을 평가하기 위해 Amazon 데이터셋을 사용하여 실험 진행.
이 데이터셋은 다양한 도메인의 사용자-아이템 상호작용 데이터를 포함하고 있어 추천 시스템의 일반화 성능을 평가하기 적절한 환경을 제공.

Datasets	#Users	#Items	#Interactions.	Avg. Len
Movies and TV	297,498	59,944	3,409,147	11.46
Video Games	64,073	33,614	598,509	8.88
Beauty	9,930	6,141	63,953	6.44
Toys	30,831	61,081	282,213	9.15

각 데이터셋에는 아이템의 제목(title)과 설명(description) 등의 텍스트 정보가 포함되어 있으며, A-LLMRec은 텍스트와 협업 피식을 결합하여 추천 성능을 높이는 방법을 실험적으로 검증함.

Experiments

비교 대상 모델 (Baseline Models)

A-LLMRec의 성능을 기존의 추천 시스템과 비교하기 위해 세 가지 유형의 모델과 성능을 비교함.

Collaborative Filtering

NCF
NextItNet
GRU4Rec
SASRec

Modality aware

MoRec
CTRL
RECFORMER

LLM based

LLM-Only
TALLRec
MLP-LLM

Experiments

성능 평가

사용 평가 지표 = Hit Ratio@K

: K개의 predict 중 next item이 포함되어있을 확률

-> 이 논문에선 K=1, 즉 1개 예측했는데 그게 정답일 확률

Overall model performance (Hit@1) over various datasets. The best performance is denoted in bold.

	Collaborative filtering				Modality-aware			LLM-based			
	NCF	NextItNet	GRU4Rec	SASRec	MoRec	CTRL	RECFORMER	LLM-Only	TALLRec	MLP-LLM	A-LLMRec
Movies and TV	0.4273	0.5855	0.5215	0.6154	0.4130	0.3467	0.4865	0.0121	0.2345	0.5838	0.6237
Video Games	0.3159	0.4305	0.4026	0.5402	0.4894	0.2354	0.4925	0.0168	0.4403	0.4788	0.5282
Beauty	0.2957	0.4231	0.4131	0.5298	0.4997	0.3963	0.4878	0.0120	0.5542	0.5548	0.5809
Toys	0.1849	0.1415	0.1673	0.2359	0.1728	0.1344	0.2871	0.0141	0.0710	0.3225	0.3336

- A-LLMRec이 모든 데이터셋에서 가장 높은 성능을 기록
- SASRec이 Collaborative Filtering 모델 중 가장 높은 성능, 웹 시나리오에서 강점을 보임.
- MoRec과 같은 모달리티 기반 모델은 콜드 스타트 문제 해결에 효과적이나 전체 성능은 낮음
- LLM-Only 모델은 가장 낮은 성능, 단순한 LLM 기반 접근 방식만으로는 추천 성능이 낮음을 시사

Experiments

성능 평가

A-LLMRec 모델과 다른 추천 모델들의 cold/warm item 시나리오에서의 성능을 비교

Results (Hit@1) on cold/warm item scenario.

	Movies and TV		Video Games		Beauty	
	Cold	Warm	Cold	Warm	Cold	Warm
SASRec	0.2589	0.6787	0.1991	0.5764	0.1190	0.6312
MoRec	0.2745	0.4395	0.2318	0.4977	0.2145	0.5425
CTRL	0.1517	0.3840	0.2074	0.2513	0.1855	0.4711
RECFORMER	0.3796	0.5449	0.3039	0.5377	0.3387	0.5133
TALLRec	0.2654	0.2987	0.3950	0.4897	0.5462	0.6124
A-LLMRec	0.5714	0.6880	0.4263	0.5970	0.5605	0.6414
A-LLMRec (SBERT)	0.5772	0.6802	0.4359	0.5792	0.5591	0.6405

A-LLMRec이 Cold & Warm Scenario에서 모두 강력한 성능을

A-LLMRec (SBERT)는 Cold Scenario에서 더 나은 성능, 콜드 스타트 문제 해결에 효과적.

콜드 시나리오에서는 SASRec의 성능 크게 저하

Experiments

성능 평가

A-LLMRec 모델과 다른 추천 모델들의 cold user 시나리오에서의 성능을 비교

Results (Hit@1) on cold user scenario.

	Movies and TV	Video Games	Beauty
SASRec	0.2589	0.4048	0.4459
MoRec	0.3918	0.3572	0.4815
CTRL	0.2273	0.1737	0.3902
RECFORMER	0.4481	0.3989	0.4644
TALLRec	0.2143	0.3895	0.5202
MLP-LLM	0.4909	0.3960	0.5276
A-LLMRec	0.5272	0.4160	0.5337

A-LLMRec이 Cold User Scenario에서도 가장 뛰어난 성능.

협업 필터링 기반 모델은 Cold User 환경에서 취약.

A-LLMRec은 LLM과 협업 필터링 정보를 효과적으로 결합하여 성능을 극대화

Experiments

성능 평가

A-LLMRec 모델과 다른 추천 모델들의 few-shot, cross domain 시나리오에서의 성능을 비교

Results (Hit@1) on the few-shot training scenario

	K	SASRec	MoRec	TALLRec	A-LLMRec	A-LLMRec (SBERT)
Movies and TV	256	0.2111	0.2208	0.1846	0.2880	0.2963
	128	0.1537	0.1677	0.1654	0.2518	0.2722
Video Games	256	0.1396	0.1420	0.2321	0.2495	0.2607
	128	0.1089	0.1157	0.1154	0.1608	0.1839
Beauty	256	0.2243	0.2937	0.3127	0.3467	0.3605
	128	0.1813	0.2554	0.2762	0.3099	0.3486

Results (Hit@1) on a cross-domain scenario

	SASRec	MoRec	RECFORMER	TALLRec	A-LLMRec	A-LLMRec (SBERT)
Movies and TV → Video Games	0.0506	0.0624	0.0847	0.0785	0.0901	0.1203

Experiments

성능 평가

A-LLMRec 모델의 Stage-1에서 수행된 Ablation Study(소거 실험) 결과

Ablation studies on Stage-1 of A-LLMRec (Hit@1).

Ablation	Movies and TV	Beauty	Toys
A-LLMRec	0.6237	0.5809	0.3336
w/o $\mathcal{L}_{\text{matching}}$	0.5838	0.5548	0.3225
w/o $\mathcal{L}_{\text{item-recon}} \& \mathcal{L}_{\text{text-recon}}$	0.5482	0.5327	0.3204
w/o \mathcal{L}_{rec}	0.6130	0.5523	0.1541
Freeze SBERT	0.6173	0.5565	0.1720

**A-LLMRec의 모든 구성 요소가
모델 성능 향상에 기여.**

**특히 \mathcal{L}_{rec} 와
 $\mathcal{L}_{\text{item-recon}}$ & $\mathcal{L}_{\text{text-recon}}$
이 중요한 역할을 함.**

Experiments

성능 평가

A-LLMRec 모델의 Stage-2에서 수행된 Ablation Study(소거 실험) 결과

Ablation study on Stage-2 of A-LLMRec (Hit@1).

Row	Ablation	Movies and TV	Video Games	Beauty	Toys
(1)	A-LLMRec	0.6237	0.5282	0.5809	0.3336
(2)	A-LLMRec w/o user representation	0.5925	0.5121	0.5547	0.3217
(3)	A-LLMRec w/o joint embedding	0.1224	0.4773	0.5213	0.2831
(4)	A-LLMRec with random joint embedding	0.1200	0.4729	0.5427	0.0776

**A-LLMRec의 사용자 표현과 조인트 임베딩이 LLM이 추천을 수행하는 데 중요한 역할.
LLM이 단순한 언어 모델이 아니라, 협업 필터링 정보와 결합될 때 강력한 추천 성능 발휘.**

4. Conclusion

Conclusion

연구 기여

- ✓ 기존 CF 및 LLM 모델의 단점을 극복한 새로운 접근법 제안
- ✓ 협업 지식을 효과적으로 LLM에 전달하는 정렬 네트워크 개발
- ✓ 콜드 & 웜 시나리오에서 모두 강력한 성능을 발휘하는 모델 구현
- ✓ LLM 미세조정 없이 효율적으로 사용할 수 있는 실용적인 모델 제안

향후 연구 방향

- 프롬프트 최적화 연구: Chain-of-Thought 방식 도입 검토
- 다양한 추천 도메인으로 확장: 음악, 뉴스 등에서의 성능 평가
- 멀티모달 정보 활용 확대: 텍스트뿐만 아니라 이미지, 영상 정보 활용 가능성 탐색