A Contrastive Framework with User, Item and Review Alignment for Recommendation

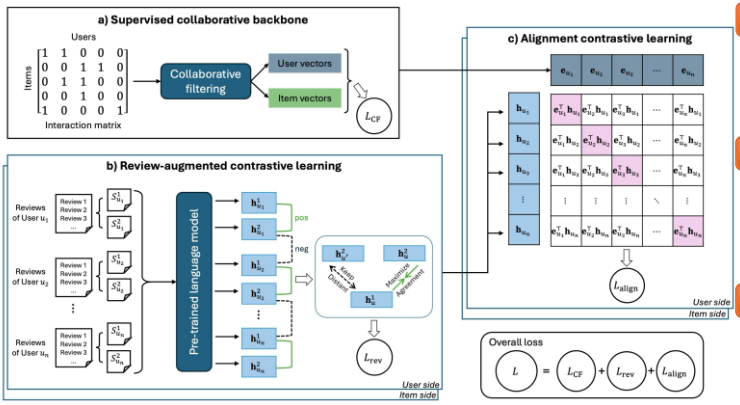
Hoang V. Dong, Yuan Fang, Hady W. Lauw
Singapore Management University

Contributions

In this works, our contributions can be summarized as follows:

- We identify the limitations inherent in treating review data merely as features and observe that reviews provide distinctive contrastive signals for both user and item sides.
- We propose a **Re**view-centric **C**ontrastive **A**lignment **F**ramework for **R**ecommendation (**ReCAFR**) framework that employs review data for augmentation to mitigate the sparsity problem and aligns the tripartite representations to improve robustness.

Framework: ReCAFR



CF backbone

$$y_{u,i} = \begin{cases} 1, & u \text{ has an observed interaction with } i \\ 0, & \text{otherwise.} \end{cases} \Rightarrow L_{\text{CF}} = \sum_{(u,i,j)\in O} - \ln \sigma(\hat{y}_{u,i} - \hat{y}_{u,j}),$$

Review-augmented contrastive learning

$$\mathbf{h}_u^k = \text{MEAN}(\{f(s) \mid s \in S_u^k\}), \quad \mathbf{h}_i^k = \text{MEAN}(\{f(s) \mid s \in S_i^k\}).$$

$$L_{\text{rev}}^{\text{user}} = \sum_{u\in U} - \log \frac{\exp(\text{sim}(\mathbf{h}_u^1, \mathbf{h}_u^2)/\tau)}{\sum_{u'\in U} \exp(\text{sim}(\mathbf{h}_u^1, \mathbf{h}_{u'}^2)/\tau))},$$

$$L_{\text{rev}}^{\text{item}} = \sum_{i\in I} - \log \frac{\exp(\text{sim}(\mathbf{h}_i^1, \mathbf{h}_i^2)/\tau)}{\sum_{i'\in I} \exp(\text{sim}(\mathbf{h}_i^1, \mathbf{h}_{i'}^2)/\tau))},$$

Alignment contrastive learning

$$L_{\text{align}}^{\text{user}} = \sum_{u\in U} - \log \frac{\exp(\text{sim}(\mathbf{e}_u, \mathbf{h}_u)/\tau)}{\sum_{u'\in U} \exp(\text{sim}(\mathbf{e}_u, \mathbf{h}_{u'}))/\tau))},$$

$$L_{\text{align}}^{\text{item}} = \sum_{i\in I} - \log \frac{\exp(\text{sim}(\mathbf{e}_i, \mathbf{h}_i)/\tau)}{\sum_{i'\in I} \exp(\text{sim}(\mathbf{e}_i, \mathbf{h}_{i'}))/\tau))},$$

Experiments

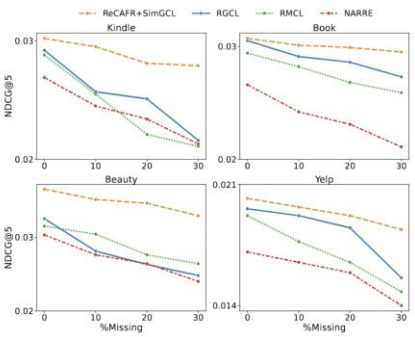
Comparison to baselines w/o review data

Methods	Kindle			Book			Beauty			Yelp		
	Recall@5	Prec@5	NDCG@5	Recall@5	Prec@5	NDCG@5	Recall@5	Prec@5	NDCG@5	Recall@5	Prec@5	NDCG@5
BPR	.0203±.0003	.0223±.0008	.0242±.0005	.0142±.0009	.0227±.0006	.0236±.0001	.0264±.0007	.0272±.0009	.0310±.0003	.0109±.0006	.0157±.0007	.0166±.0002
ReCAFR+BPR	**.0226**±.0001	**.0241**±.0009	**.0245**±.0001	**.0143**±.0002	**.0228**±.0001	**.0237**±.0008	**.0277**±.0003	**.0287**±.0002	**.0313**±.0008	**.0115**±.0008	**.0162**±.0004	**.0169**±.0004
LightGCN	.0232±.0001	.0253±.0007	.0281±.0001	.0152±.0005	.0244±.0001	.0254±.0006	.0254±.0008	.0261±.0005	.0298±.0000	.0112±.0006	.0159±.0004	.0168±.0002
ReCAFR+LightGCN	**.0243**±.0005	**.0265**±.0009	**.0293**±.0001	**.0187**±.0005	**.0299**±.0002	**.0311**±.0001	**.0266**±.0006	**.0276**±.0004	**.0300**±.0006	**.0125**±.0006	**.0168**±.0002	**.0175±.0005**
SGL	.0237±.0006	**.0259**±.0007	.0286±.0007	.0169±.0001	.0270±.0008	.0281±.0005	.0269±.0002	**.0276**±.0009	.0316±.0008	.0101±.0004	.0143±.0001	.0151±.0004
ReCAFR+SGL	**.0242**±.0009	.0257±.0005	**.0291**±.0005	**.0171**±.0006	**.0274**±.0007	**.0285**±.0007	**.0276**±.0007	.0273±.0004	**.0319**±.0009	**.0106**±.0001	**.0148**±.0008	**.0158**±.0001
DirectAU	.0255±.0009	.0271±.0003	.0309±.0007	.0167±.0009	.0269±.0007	.0281±.0009	.0298±.0007	.0284±.0007	.0338±.0006	**.0124**±.0009	.0161±.0008	.0169±.0005
ReCAFR+DirectAU	**.0262**±.0000	**.0273**±.0006	**.0317**±.0009	**.0172**±.0006	**.0271**±.0006	**.0285**±.0007	**.0301**±.0008	**.0286**±.0008	**.0341**±.0007	.0121±.0006	**.0179**±.0001	**.0172**±.0008
SimGCL	.0253±.0002	.0277±.0004	**.0306**±.0003	.0180±.0007	.0289±.0006	.0301±.0007	.0288±.0000	.0295±.0002	.0339±.0007	.0145±.0001	.0181±.0002	.0188±.0002
ReCAFR+SimGCL	**.0269**±.0002	**.0285**±.0001	.0302±.0003	**.0184**±.0003	**.0295**±.0006	**.0307**±.0001	**.0296**±.0007	**.0304**±.0005	**.0365**±.0001	**.0155**±.0007	**.0191**±.0009	**.0202**±.0004

Comparison to review-based baselines

Method	Kindle			Book			Beauty			Yelp		
Using all available reviews												
	Recall@5	Prec@5	NDCG@5	Recall@5	Prec@5	NDCG@5	Recall@5	Prec@5	NDCG@5	Recall@5	Prec@5	NDCG@5
NARRE	.0238±.0009	.0235±.0005	.0269±.0007	.0160±.0008	.0257±.0003	.0266±.0001	.0266±.0006	.0276±.0002	.0303±.0009	.0129±.0002	.0182±.0009	.0171±.0006
RGCL	.0242±.0006	.0264±.0009	.0292±.0001	.0183±.0003	.0293±.0003	.0305±.0003	.0269±.0006	.0279±.0002	.0325±.0004	.0142±.0002	.0189±.0006	.0196±.0005
RMCL	.0236±.0008	.0259±.0006	.0288±.0001	.0172±.0003	.0286±.0001	.0294±.0009	.0273±.0006	.0265±.0006	.0315±.0004	.0144±.0007	.0186±.0003	.0192±.0006
ReCAFR+SimGCL	**.0269**±.0004	**.0285**±.0008	**.0302±.0003**	**.0184**±.0005	**.0295**±.0005	**.0307±.0003**	**.0296**±.0003	**.0304**±.0001	**.0365**±.0008	**.0155**±.0009	**.0191**±.0008	**.0202**±.0007
Removing 30% of the reviews												
	Recall@5	Prec@5	NDCG@5	Recall@5	Prec@5	NDCG@5	Recall@5	Prec@5	NDCG@5	Recall@5	Prec@5	NDCG@5
NARRE	.0206±.0007	.0229±.0009	.0213±.0008	.0141±.0005	.0214±.0008	.0211±.0006	.0239±.0007	.0261±.0007	.0240±.0006	.0102±.0006	.0145±.0005	.0140±.0007
RGCL	.0215±.0009	.0238±.0004	.0216±.0009	.0152±.0002	.0258±.0009	.0273±.0006	.0247±.0005	.0274±.0008	.0248±.0008	.0109±.0001	.0152±.0005	.0156±.0001
RMCL	.0226±.0007	.0248±.0002	.0211±.0005	.0163±.0005	.0247±.0003	.0259±.0001	.0251±.0001	.0279±.0006	.0264±.0008	.0116±.0005	.0163±.0004	.0148±.0005
ReCAFR+SimGCL	**.0241**±.0007	**.0254**±.0005	**.0279**±.0007	**.0171**±.0001	**.0261**±.0005	**.0295**±.0004	**.0274**±.0008	**.0289**±.0009	**.0329**±.0006	**.0121**±.0005	**.0175**±.0005	**.0184**±.0001

Impact of missing reviews



Ablation Study

Table 4: Ablation study on ReCAFR, reporting NDCG@5.

Variants	Kindle	Book	Beauty	Yelp
ReCAFR+SimGCL	**.0302**	**.0307**	**.0365**	**.0202**
w/o text emb. init.	.0294	.0286	.0351	.0199
w/o user CL	.0281	.0281	.0346	.0186
w/o item CL	.0276	.0284	.0331	.0182

LLM enhancement

You will serve as an assistant to help me determine which types of books a specific user is likely to enjoy.
I will provide you with information about books that the user has purchased, as well as his or her review of that book.
Here are the instructions:
1. Each purchased book will be described in JSON format, with the following attributes:
{ "title": "the title of the book", (if there is no title, I will set this value to "None")
"description": "a description of what types of users will like this book",
"review": "the user's review on the book" (if there is no review, I will set this value to "None")}
2. The information I will give you:
PURCHASED ITEMS: a JSON string describing the item that the user has purchased.

Requirements:
1. Please provide your decision in JSON format, following this structure:
{ "summarization": "A summarization of what types of books this user is likely to enjoy" (if you are unable to summarize it, please set this value to "None")
"reasoning": "briefly explain your reasoning for the summarization}
2. Please ensure that the "summarization" is no longer than 100 words.
3. The "reasoning" has no word limits.
4. Do not provided any other text outside the JSON string.

Instruction

{"title": "The Mythic Guide to Characters: Writing Characters Who Enchant and Inspire",
"description": "The Mythic Guide to Characters is, in short, a remarkable book. It's intelligent and insightful but at the same time easily approachable and user friendly..."
"review": "The Mythic Guide to Characters is a well-organized effort to assist writers, especially those early in their careers, in creating strong and varied characters..."}

Input Prompt

{ "summarization": "This user is likely to enjoy books focused on writing and character development, particularly those aimed at helping aspiring writers improve their craft.",
"reasoning": "The user praised The Mythic Guide to Characters' for its organizational structure and practical advice for early-career writers, indicating a strong interest in writing as a form of expression. The appreciation for examples from popular works suggests a desire for relatable references and a focus on character creation. This aligns with an enjoyment for instructional and insightful literature that aids in the writing process."}

Generated from LLM

Table 5: Demonstration of ReCAFR with LLM-enhanced reviews on the Beauty dataset.

Methods	Recall@5	Prec@5	NDCG@5
BPR	.0264	.0272	.0310
ReCAFR+BPR	**.0277**	.0287	.0313
ReCAFR+BPR (LLM)	.0269	**.0289**	**.0315**
LightGCN	.0254	.0261	.0298
ReCAFR+LightGCN	.0266	.0276	.0300
ReCAFR+LightGCN (LLM)	**.0269**	**.0277**	**.0311**
DirectAU	.0298	.0284	.0338
ReCAFR+DirectAU	.0301	.0286	.0341
ReCAFR+DirectAU (LLM)	**.0306**	**.0295**	**.0349**
SimGCL	.0288	.0295	.0338
ReCAFR+SimGCL	.0296	.0304	.0365
ReCAFR+SimGCL (LLM)	**.0298**	**.0313**	**.0376**

ACKNOWLEDGEMENT

This research / project is supported by the Ministry of Education, Singapore, under its Academic Research Fund Tier 2 (MOET2EP20122-0041). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the Ministry of Education, Singapore.

For complete references please refer to:

Paper

FANG Lab@SMU: https://www.yfang.site/group
Email: ds.hoangdv@gmail.com