Model Pre-training for Recommendation Systems

余阳

2023.10.23

Personal Information

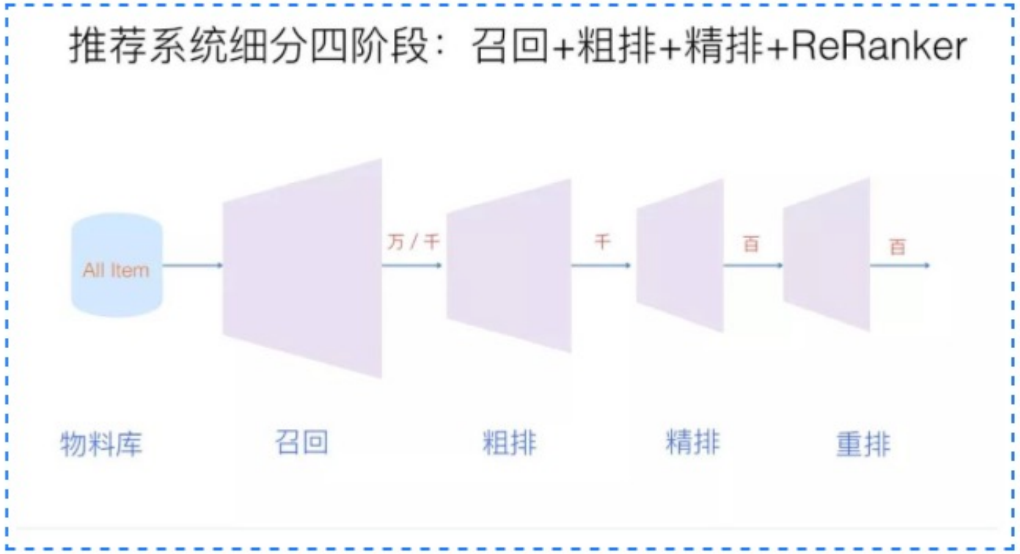
- 姓名：余阳
- 年级：研三
- 导师：刘淇 教授
- 研究方向：推荐系统，预训练模型
- 工位：B702
- 邮箱：yflyl613@mail.ustc.edu.cn

Outline

- Background
- Model Pre-training for Item Modeling
- Model Pre-training for User Modeling
- Future Directions

Background

■ Introduction to recommendation systems

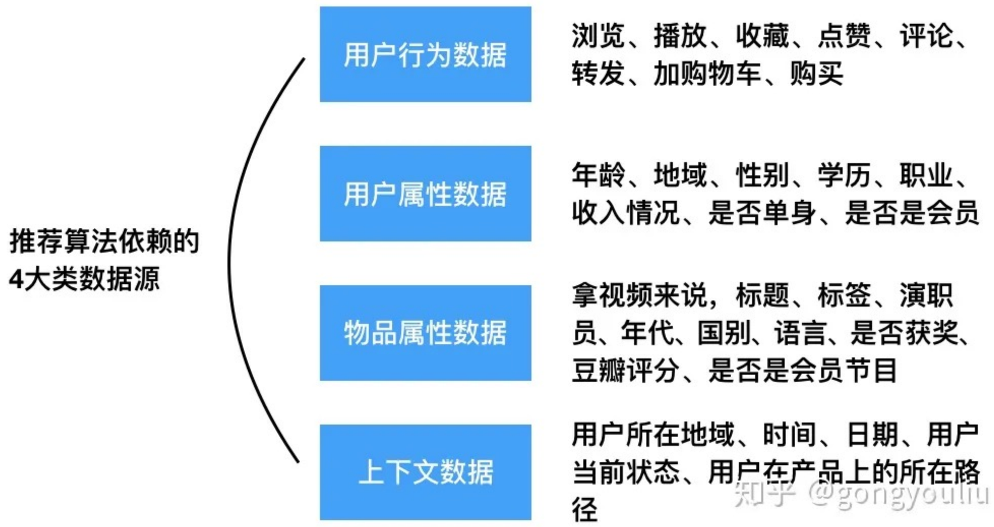


A general recommendation system pipeline.

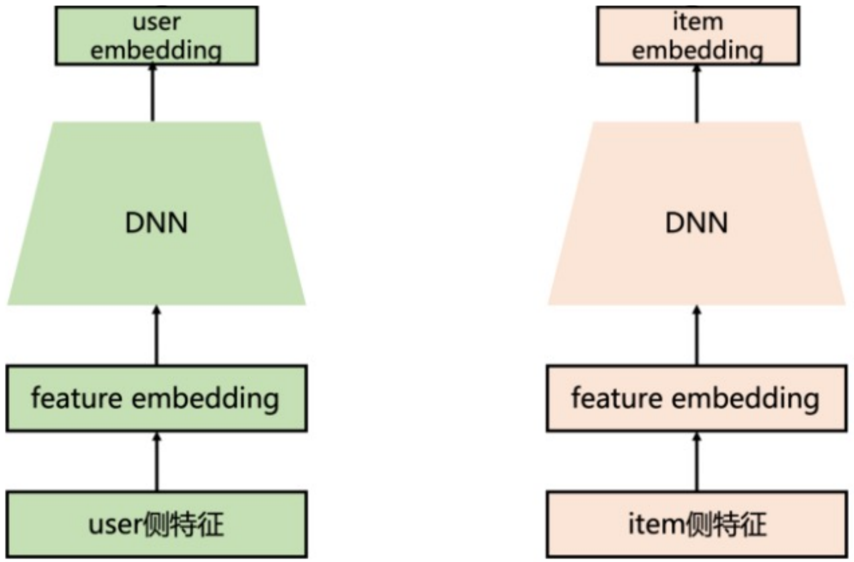
A real-world recommendation system framework.

Background

- **Introduction to recommender systems**
  - Two key elements: **item & user**
  - **Item modeling & user modeling**: Capture the characteristics of the item/user for specific tasks.
  - Previous method
    - Expertise-based feature engineering.



- Deep learning-based methods



ID, demographic information, historical behaviors, etc.

ID, text descriptions, images, etc.

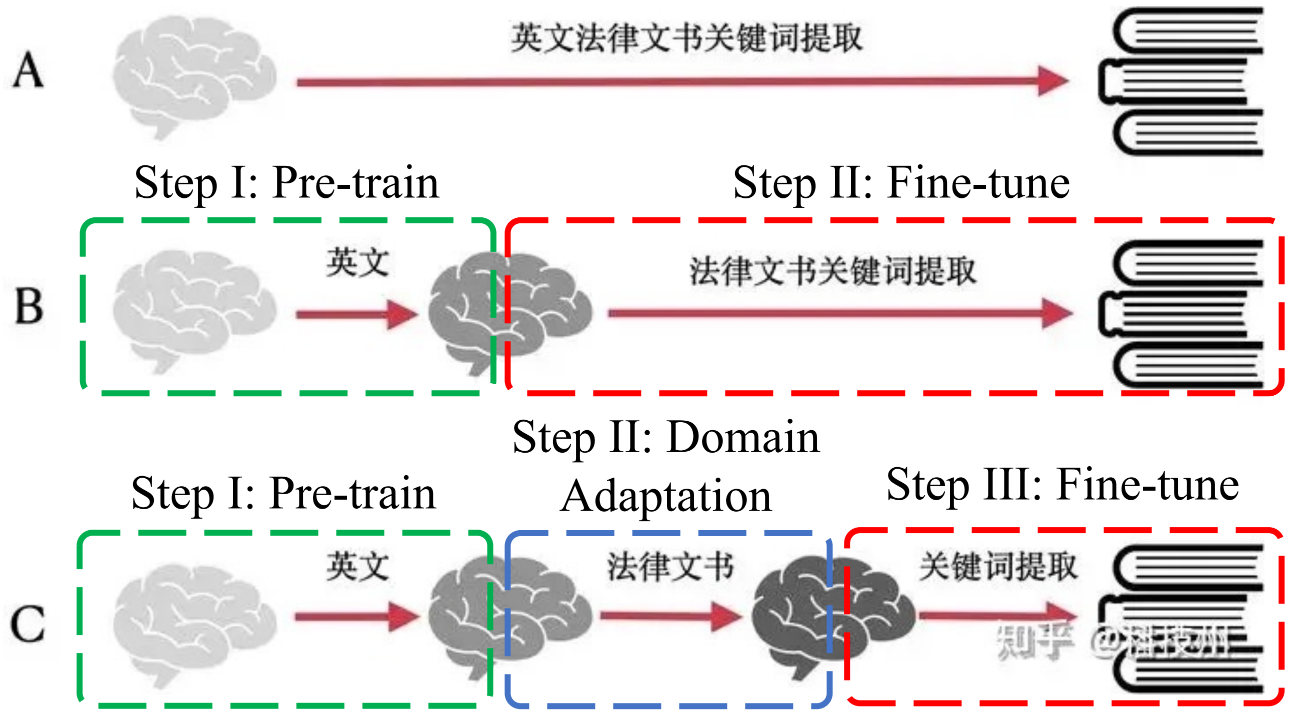
Background

- **Why do we need model pre-training? The data sparsity problem.**
  - Most existing methods heavily rely on task-specific labeled data (**supervised learning**).
  - Task-specific labeled data tend to be scarce in recommendation systems.
    - Some user-related labels are **privacy-sensitive** (e.g., age, income level).
    - Some behaviors are **naturally sparse** (e.g., traveling, purchasing).
    - The recommendation system **updates rapidly** (e.g., hundreds/thousands of new items/users are added to the system every day).
  - Limited labeled data leads to poor item/user modeling performance for supervised learning methods.

Background

■ The pre-training paradigm

- Labeled data are limited and expensive while **unlabeled data** are abundant and cheap.
- **Pre-train**: Train the model to capture general knowledge on massive unlabeled data in an **unsupervised/self-supervised** manner.
- **Fine-tune**: Transfer the general knowledge and learn task-specific knowledge in a **supervised** manner.



Outline

- Background
- Model Pre-training for Item Modeling
- Model Pre-training for User Modeling
- Future Directions

Model Pre-training for Item Modeling

- **How to capture transferable features for items?**
  - Domain-specific data
    - Only valid in a specific domain.
    - E.g., item ID, category ID, etc.
  - General data
    - Can be generalized across domains.
    - E.g., text descriptions, images, etc.

ID

texts
images
entities

Model Pre-training for Item Modeling

- **How to capture transferable features for items?**
  - Incorporate **texts / entities** into item modeling.

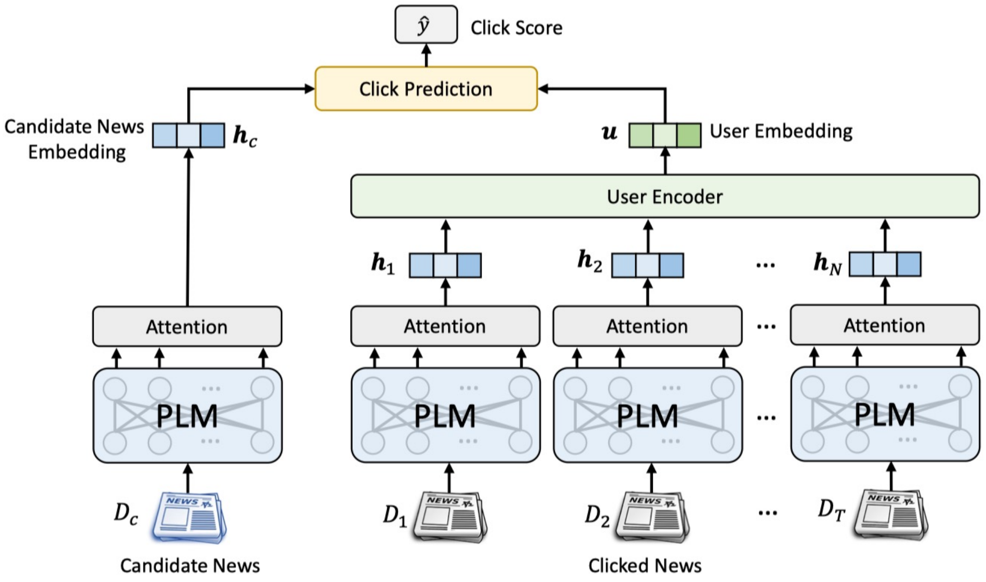


Figure 2: The framework of PLM empowered news recommendation.

markets with different languages

Methods	EN-US	DE-DE	FR-FR	IT-IT	JA-JP	ES-ES	KO-KR
EBNR-Single	62.08	59.94	61.66	60.27	61.57	58.30	63.53
EBNR-MUSE	62.26	60.19	61.75	60.44	61.74	57.53	63.78
EBNR-Unicoder	63.35	61.44	62.34	61.18	62.76	58.70	64.80
EBNR-InfoXLM	64.29	62.03	62.97	61.98	63.34	59.33	65.58
NAML-Single	62.05	59.89	61.56	60.21	61.54	58.21	63.5
NAML-MUSE	62.17	60.17	61.71	60.4	61.69	57.46	63.73
NAML-Unicoder	63.3	61.37	62.32	61.16	62.74	58.61	64.77
NAML-InfoXLM	64.27	61.98	62.94	61.91	63.29	59.33	65.49
NPA-Single	62.09	59.90	61.56	60.24	61.57	58.24	63.56
NPA-MUSE	62.23	60.21	61.78	60.44	61.75	57.47	63.71
NPA-Unicoder	63.32	61.41	62.35	61.20	62.77	58.64	64.80
NPA-InfoXLM	64.29	62.00	62.93	61.94	63.31	59.37	65.50
LSTUR-Single	62.09	59.95	61.58	60.22	61.58	58.22	63.57
LSTUR-MUSE	62.21	60.21	61.79	60.44	61.73	57.49	63.75
LSTUR-Unicoder	63.34	61.40	62.36	61.20	62.77	58.65	64.80
LSTUR-InfoXLM	64.31	62.03	62.96	61.95	63.32	59.38	65.54
NRMS-Single	62.11	59.94	61.62	60.28	61.57	58.30	63.64
NRMS-MUSE	62.33	60.29	61.86	60.54	61.90	57.62	63.93
NRMS-Unicoder	63.41	61.50	62.46	61.22	62.81	58.84	64.79
NRMS-InfoXLM	**64.34**	**62.05**	**63.04**	**61.98**	**63.40**	**59.44**	**65.58**

Wu et al. Empowering News Recommendation with Pre-trained Language Models. SIGIR 2021.

Model Pre-training for Item Modeling

■ How to capture transferable features for items?

- Incorporate **texts / entities** into item modeling.

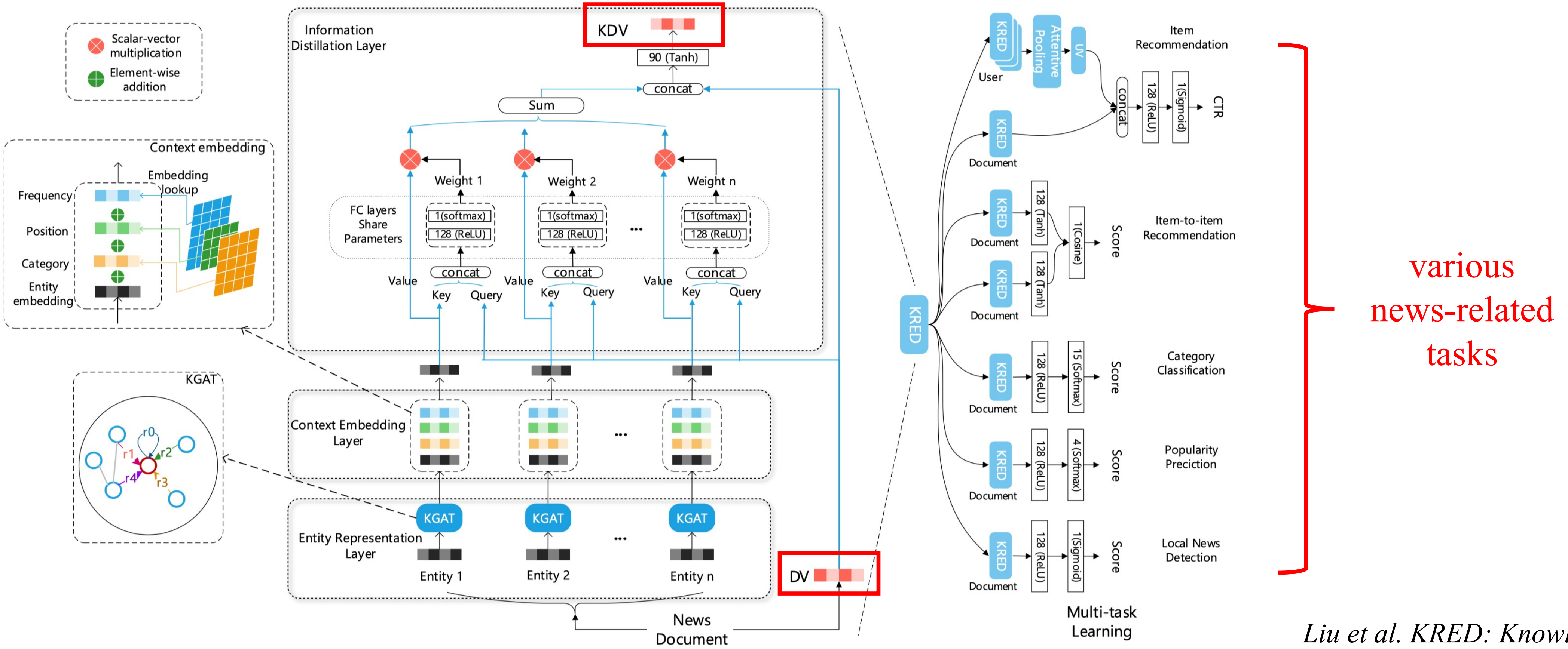


Figure 2: An overview of the proposed *KRED* model. DV indicates the (original) document vector. KDV indicates the knowledge-enhanced document vector produced by KRED. UV indicates the user vector.

Liu et al. KRED: Knowledge-Aware Document Representation for News Recommendations. RecSys 2020.

Model Pre-training for Item Modeling

- **How to capture transferable features for items?**
  - Incorporate **images / multi-modal features** into item modeling.

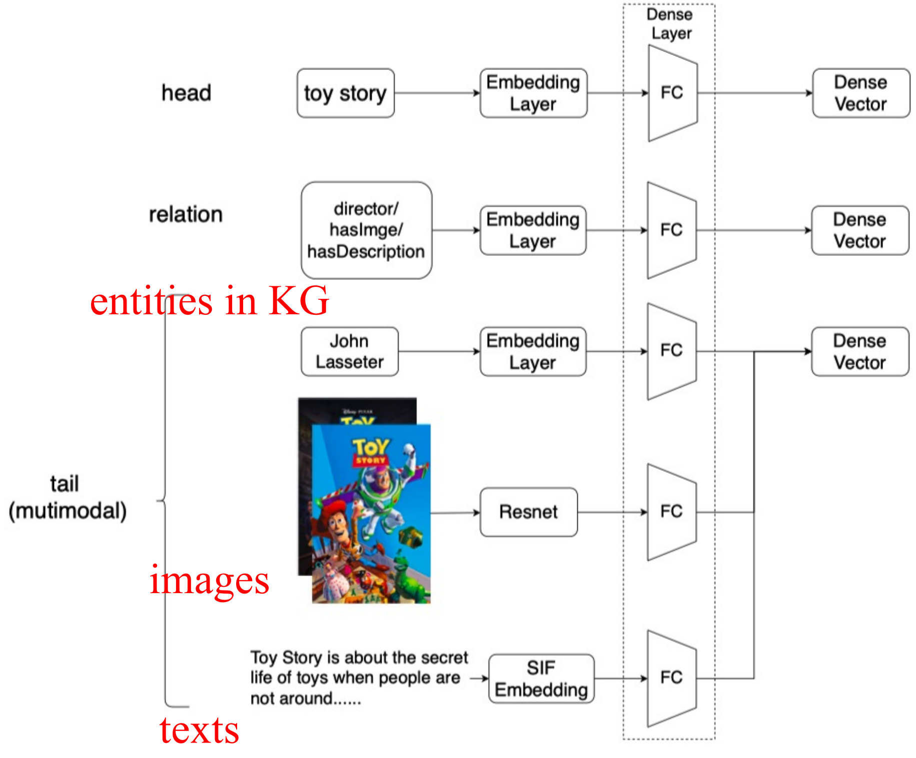


Figure 4: Multi-modal knowledge graph encoder.

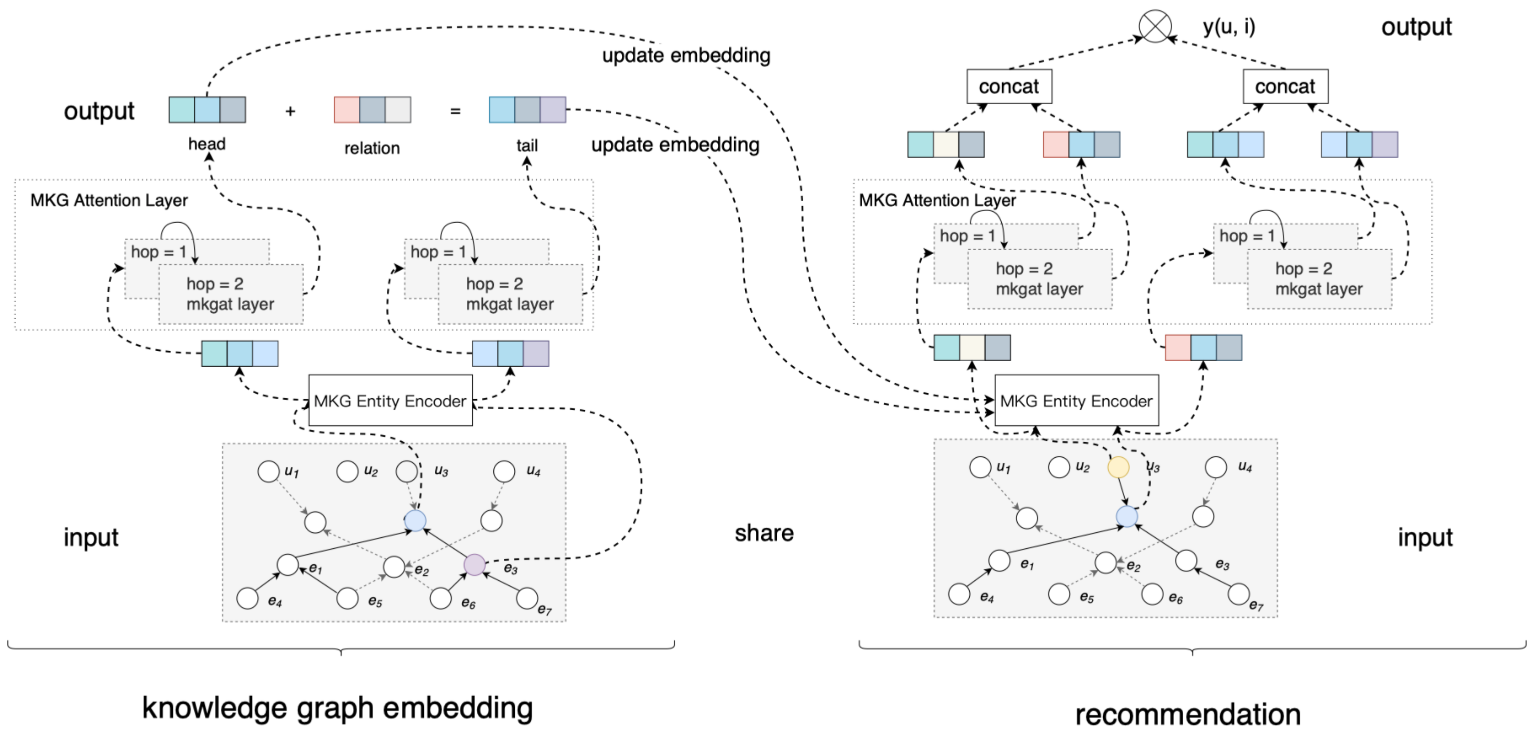


Figure 5: Framework overview of the MKGAT model.

Sun et al. Multi-modal Knowledge Graphs for Recommender Systems. CIKM 2020.

Model Pre-training for Item Modeling

- **How to capture transferable features for items?**
  - Incorporate **item-related reviews** into item modeling.

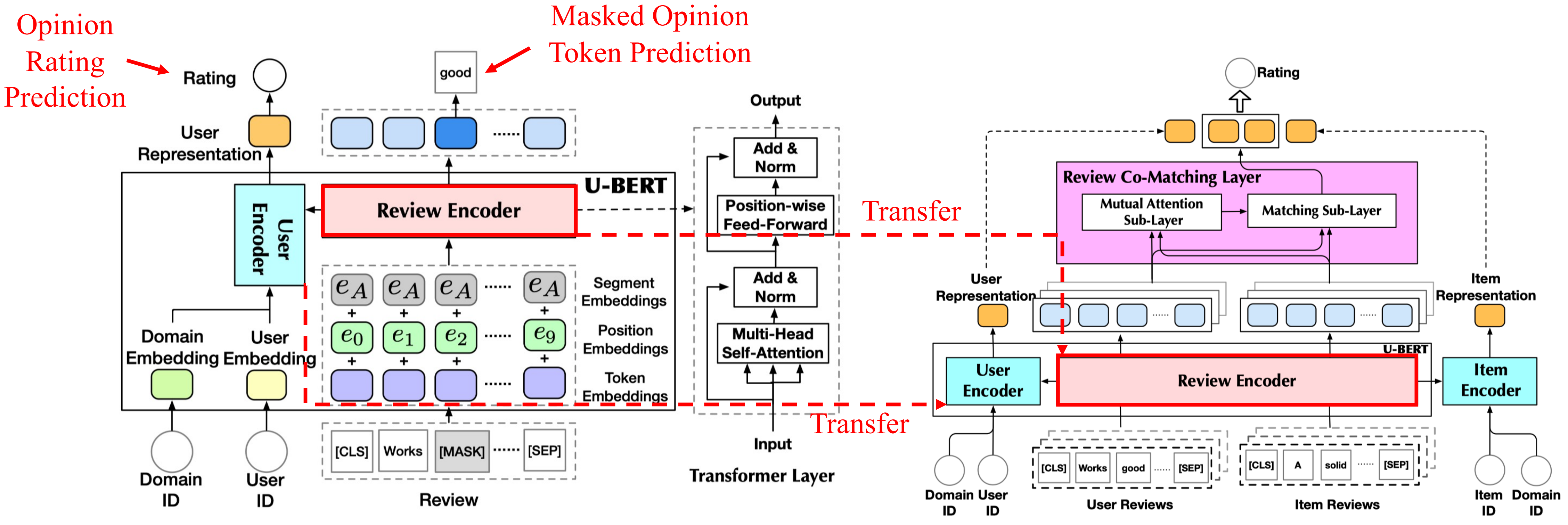


Figure 2: The pre-training stage.

Figure 3: The rating prediction framework.

Qiu et al. U-BERT: Pre-training User Representations for Improved Recommendation. AAAI 2021.

Model Pre-training for Item Modeling

The domain adaptation problem for pre-trained models

- The PLM is usually pre-trained on general corpora.
  - BookCorpus
  - Wikipedia
  - CCNews, OpenWebText, CommonCrawl, etc.
- There is a **domain gap** between the pre-training corpora and the downstream task texts.
- Simply fine-tuning on limited labeled data may be insufficient to mitigate the **domain shift** problem.

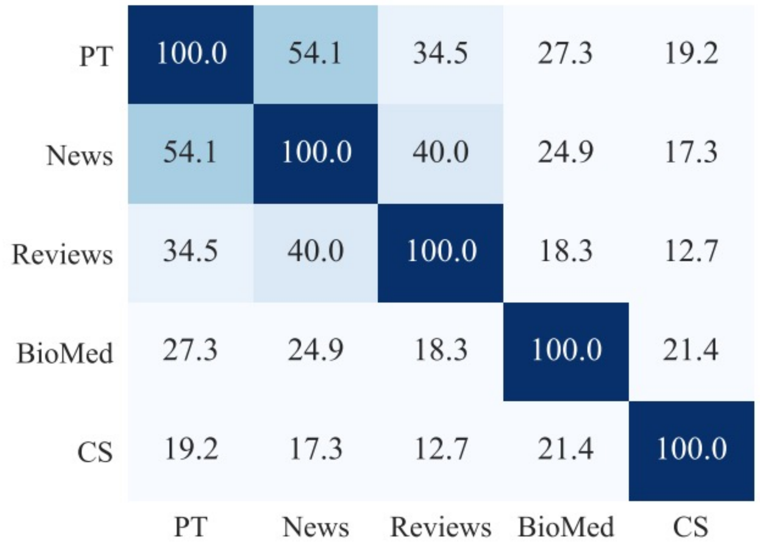


Figure 2: Vocabulary overlap (%) between domains. PT denotes a sample from sources similar to RoBERTA's pretraining corpus. Vocabularies for each domain are created by considering the top 10K most frequent words (excluding stopwords) in documents sampled from each domain.

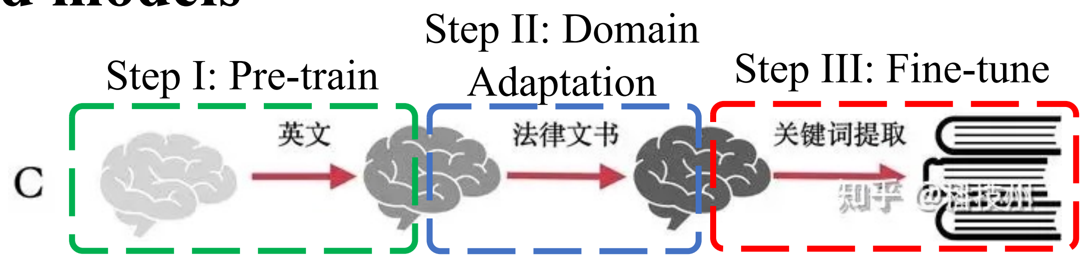
Domain	Pretraining Corpus	# Tokens	Size
BIOMED	2.68M full-text papers from S2ORC (Lo et al., 2020)	7.55B	47GB
CS	2.22M full-text papers from S2ORC (Lo et al., 2020)	8.10B	48GB
NEWS	11.90M articles from REALNEWS (Zellers et al., 2019)	6.66B	39GB
REVIEWS	24.75M AMAZON reviews (He and McAuley, 2016)	2.11B	11GB

Gururangan et al. Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. ACL 2020.

Model Pre-training for Item Modeling

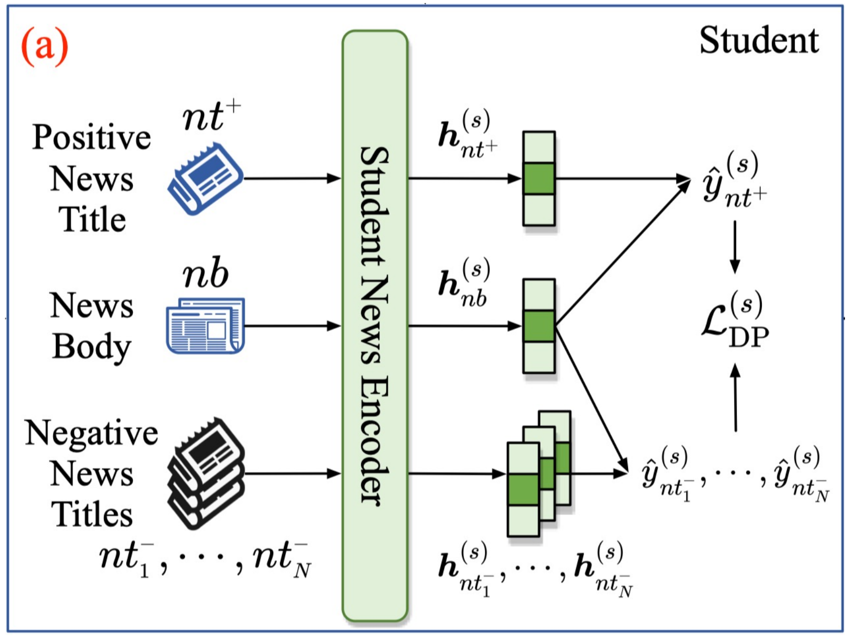
The domain adaptation problem for pre-trained models

- Domain-adaptive pre-training (DAPT)
- Task-adaptive pre-training (TAPT)
- Domain-specific post-training (DP)
  - Given a news body nb, train the PLM the identify the corresponding news title nt^+ from a set of candidates.



$$\mathcal{L}_{\mathrm{DP}} = -\log \frac{\exp(\hat{y}_{nt^+})}{\exp(\hat{y}_{nt^+}) + \sum_{i=1}^{N} \exp(\hat{y}_{nt_i^-})}$$

Model	MIND			Feeds			Model Size
	AUC	MRR	nDCG@10	AUC	MRR	nDCG@10	
PLM-NR$_{12}$ (FT)	69.72±0.15	34.74±0.10	43.71±0.07	67.93±0.13	34.42±0.07	45.09±0.07	109.89M
PLM-NR$_{12}$ (DAPT)	69.97±0.08	35.07±0.15	43.98±0.10	68.24±0.09	34.63±0.10	45.30±0.09	109.89M
PLM-NR$_{12}$ (TAPT)	69.82±0.14	34.90±0.11	43.83±0.07	68.11±0.11	34.49±0.12	45.11±0.08	109.89M
PLM-NR$_{12}$ (DP)	**71.02±0.07**	**36.05±0.09**	**45.03±0.12**	**69.37±0.10**	**35.74±0.11**	**46.45±0.11**	109.89M



Gururangan et al. Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. ACL 2020.
Yu et al. Tiny-NewsRec: Effective and Efficient PLM-based News Recommendation. EMNLP 2022.

Model Pre-training for Item Modeling

- **When item modeling meets LLM**
  - Utilize the powerful **text understanding ability** and **inherent world knowledge** of LLM to generate the item embedding.

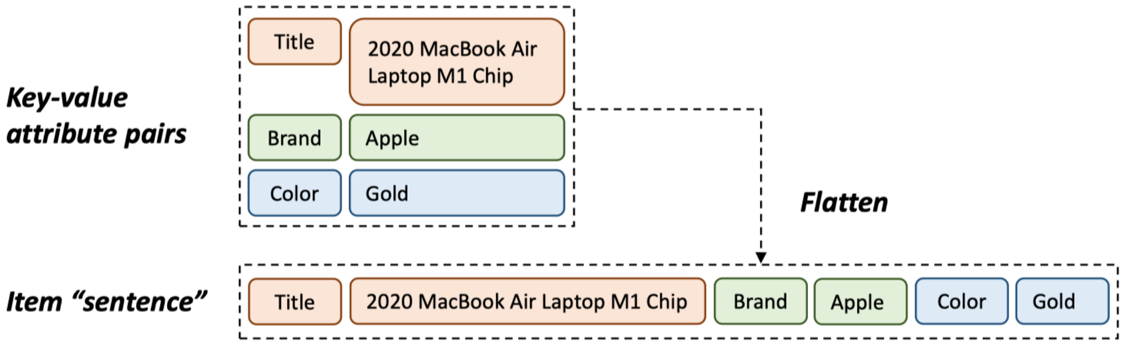
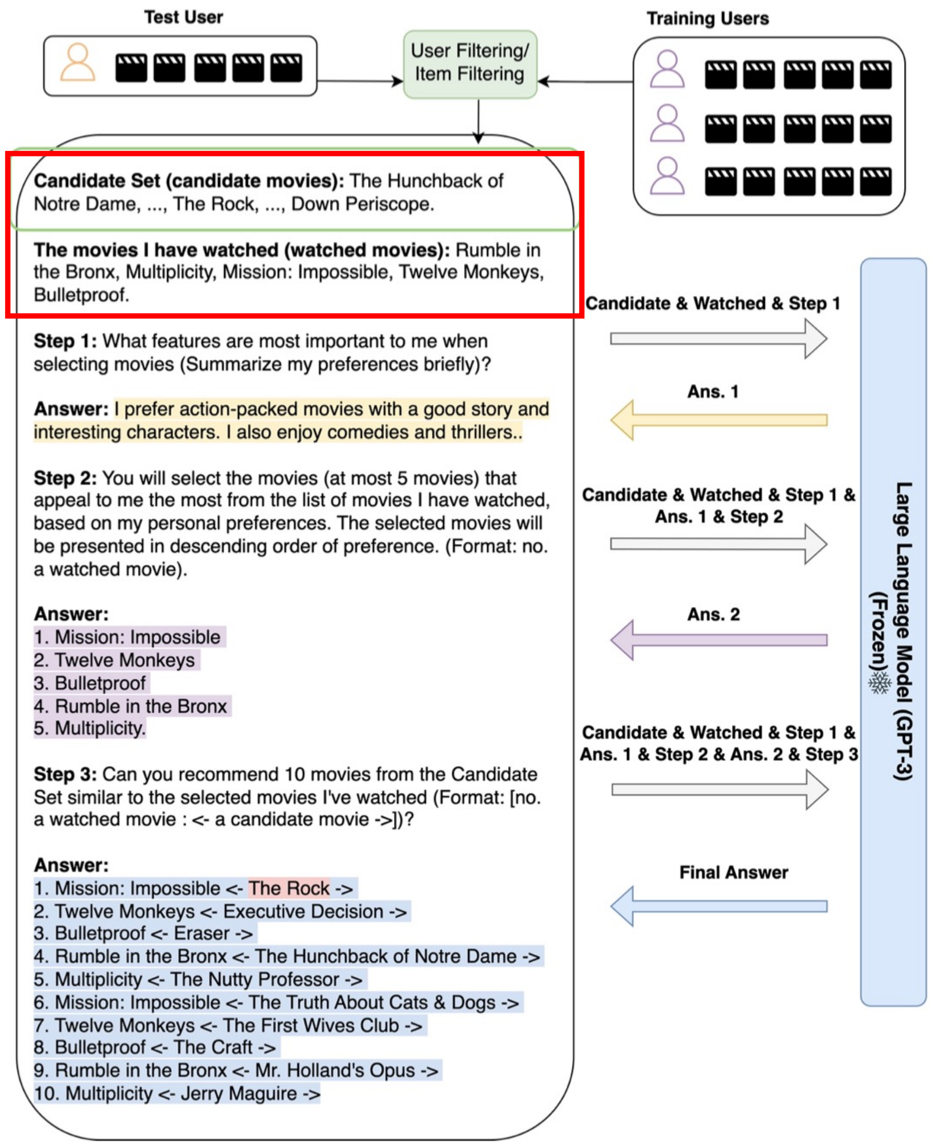


Figure 2: Model input construction. Flatten key-value attribute pairs into an item "sentence".

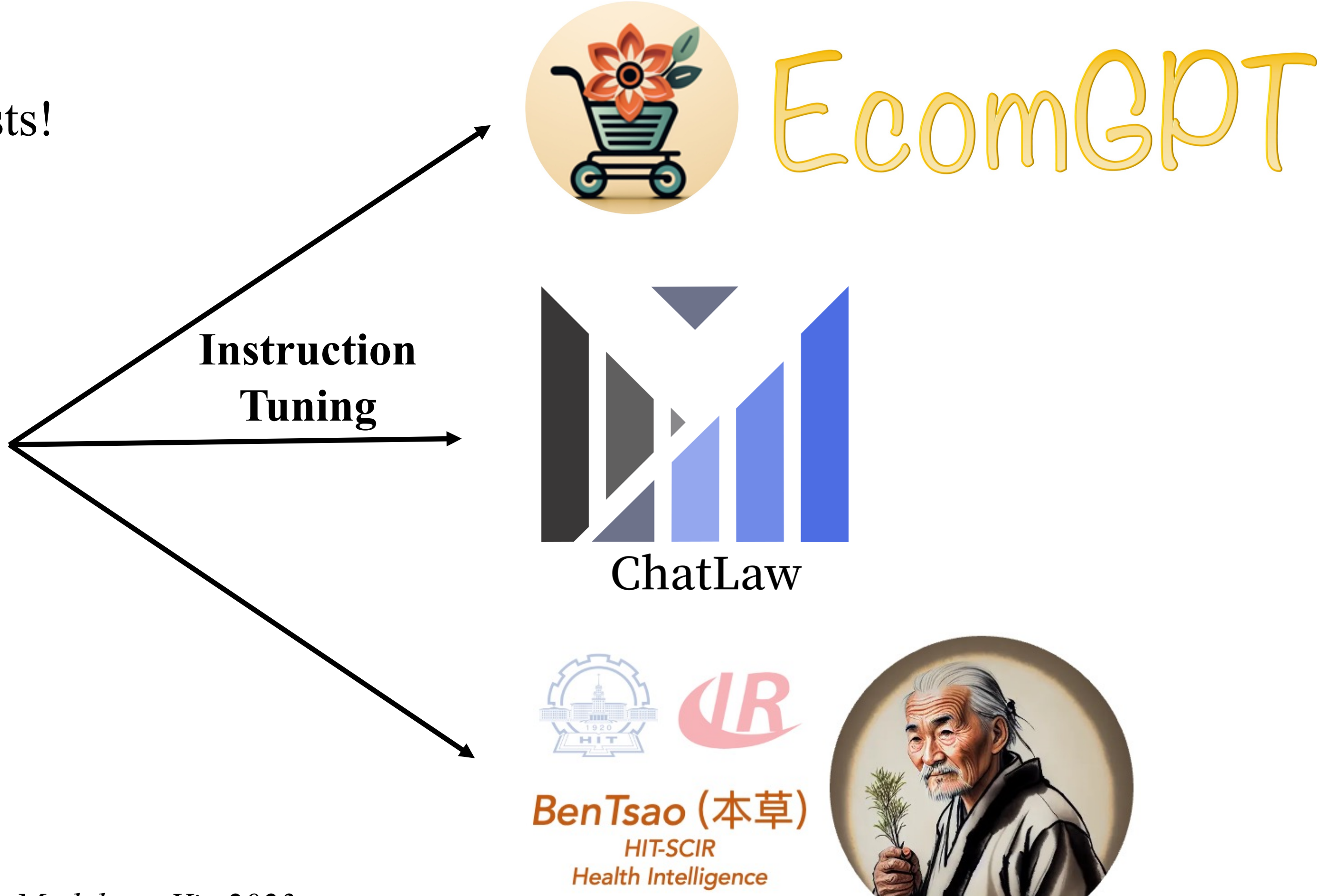


Li et al. Text Is All You Need: Learning Language Representations for Sequential Recommendation. KDD 2023.
Wang et al. Zero-Shot Next-Item Recommendation using Large Pretrained Language Models. arXiv 2023.

Model Pre-training for Item Modeling

- **When item modeling meets LLM**
  - The domain adaptation problem still exists!

Dataset	Sampling prop.	Epochs	Disk size
CommonCrawl	67.0%	1.10	3.3 TB
C4	15.0%	1.06	783 GB
Github	4.5%	0.64	328 GB
Wikipedia	4.5%	2.45	83 GB
Books	4.5%	2.23	85 GB
ArXiv	2.5%	1.06	92 GB
StackExchange	2.0%	1.03	78 GB



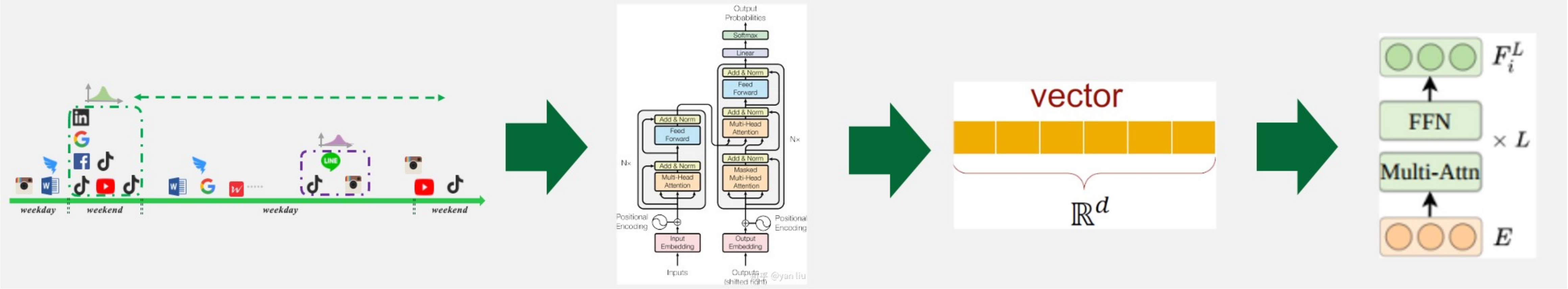
Touvron et al. LLaMA: Open and Efficient Foundation Language Models. arXiv 2023.
Cui et al. ChatLaw: Open-Source Legal Large Language Model with Integrated External Knowledge Bases. arXiv 2023.
Li et al. EcomGPT: Instruction-tuning Large Language Models with Chain-of-Task Tasks for E-commerce. arXiv 2023.
Wang et al. HuaTuo: Tuning LLaMA Model with Chinese Medical Knowledge. arXiv 2023.

Outline

- Background
- Model Pre-training for Item Modeling
- Model Pre-training for User Modeling
- Future Directions

Model Pre-training for User Modeling

- **How to capture transferable features for users?**
  - ID-based data ❌
  - Users do not have general data such as text descriptions or images.
  - The user model is first pre-trained on massive **unlabeled user behavior data** (e.g., browsing history) in a source domain and then transferred to various downstream user-oriented tasks.
- **Challenges**
  - Humans are more complex than texts/images. User interests are **diverse and versatile**.
  - User behavior sequences are **discrete** and contain **heavy noise** as well.



User behavior sequence | Pre-trained user model | Pre-trained user representation | Downstream task model

Model Pre-training for User Modeling

- **Behavior prediction-based methods**
  - Apply NLP pre-training methods to user behavior sequences.
  - **Next behavior prediction** (GPT): predict the next behavior based on previous behaviors.

$$p(\mathbf{x}^u;\Theta) = \prod_{i=1}^{n} p(x_i^u \mid x_1^u, .., x_{i-1}^u;\Theta) \qquad L(\mathcal{S};\Theta) = -\sum_{(u,\mathbf{x}^u)\in\mathcal{S}} \log p(\mathbf{x}^u;\Theta)$$

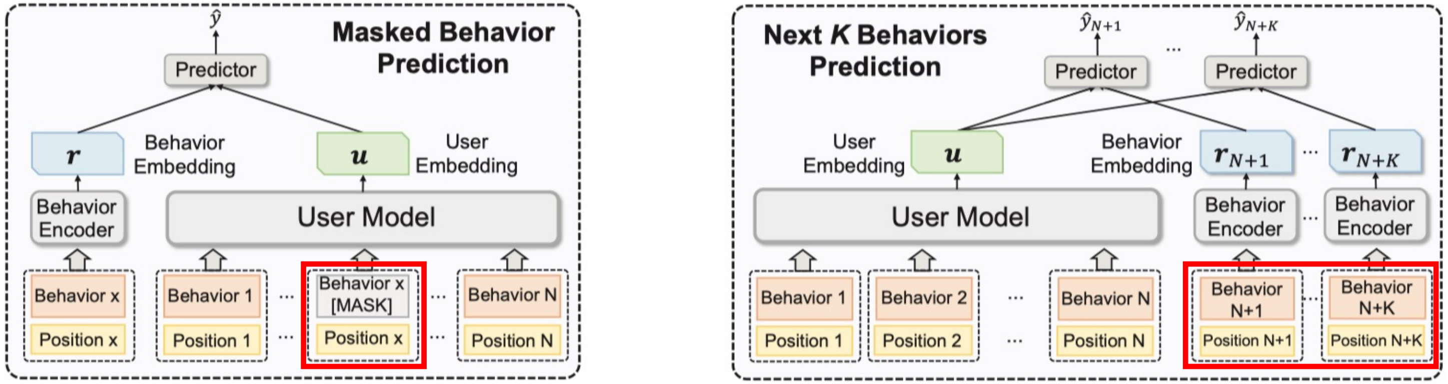
  - **Masked behavior prediction** (BERT): predict the masked behavior based on the remaining behaviors.

$$p(\mathbf{x}_\triangle^u;\Theta) = \prod_{i=1}^{m} p(x_{\triangle_i}^u \mid \tilde{\mathbf{x}}^u;\Theta) \qquad G(\mathcal{S};\Theta) = -\sum_{(u,\mathbf{x}^u)\in\mathcal{S}} \log p(\mathbf{x}_\triangle^u;\Theta)$$

Yuan et al. Parameter-Efficient Transfer from Sequential Behaviors for User Modeling and Recommendation. SIGIR 2020.

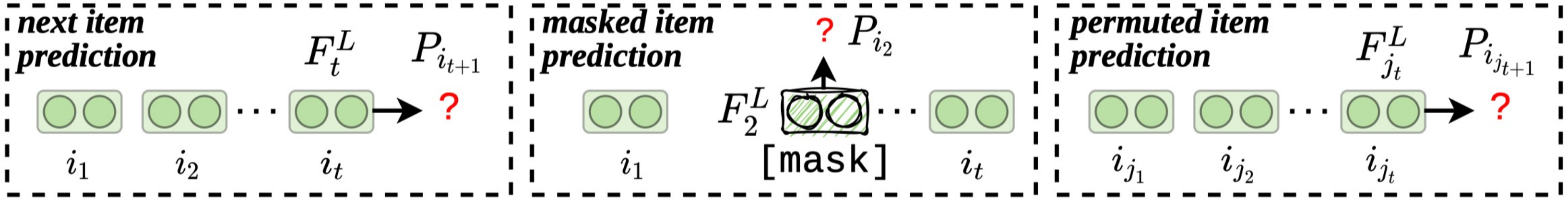
Model Pre-training for User Modeling

- **Behavior prediction-based methods**
  - **PTUM**: masked behavior prediction + next *K* behaviors prediction.



(a) Masked Behavior Prediction (MBP) task. (b) Next K Behaviors Prediction (NBP) task.

  - **IDA-SR**: next behavior prediction + masked behavior prediction + **permuted behavior prediction**



Wu et al. PTUM: Pre-training User Model from Unlabeled User Behaviors via Self-supervision. Findings of EMNLP 2020.
Mu et al. ID-Agnostic User Behavior Pre-training for Sequential Recommendation. CCIR 2022.

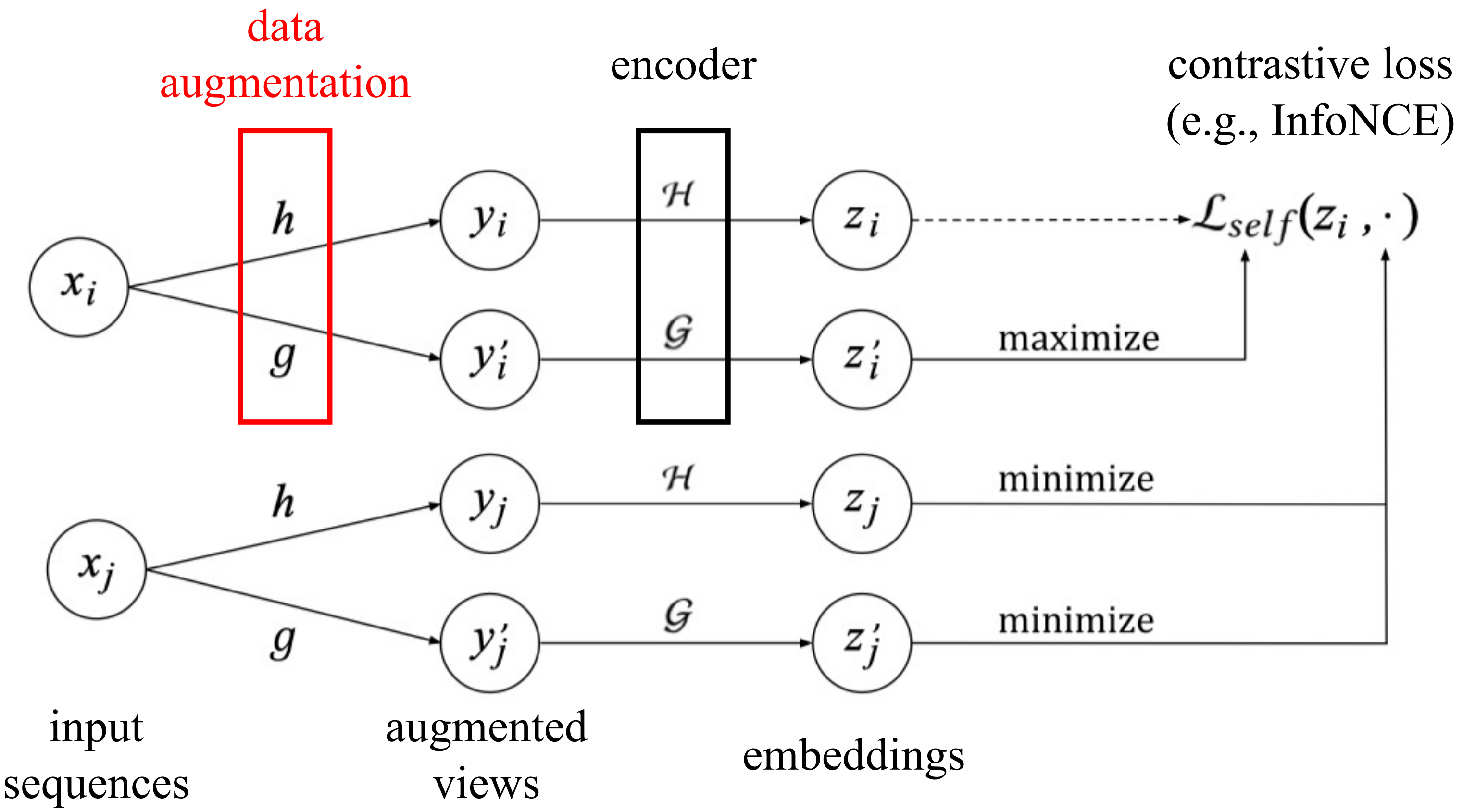
Model Pre-training for User Modeling

- **Behavior prediction-based methods**
  - Predict the **exact behavior** given a context.
  - There is a gap between the **behavior-level prediction task** and the **sequence-level user modeling task.**
  - Existing methods try to model the correlation between behaviors while lacking **sequence-level representation learning**.
  - The behavior set (>1M) can be much larger than the vocabulary set in PLMs (32K~250K), which makes the prediction task much more difficult.

Model Pre-training for User Modeling

Contrastive learning-based methods

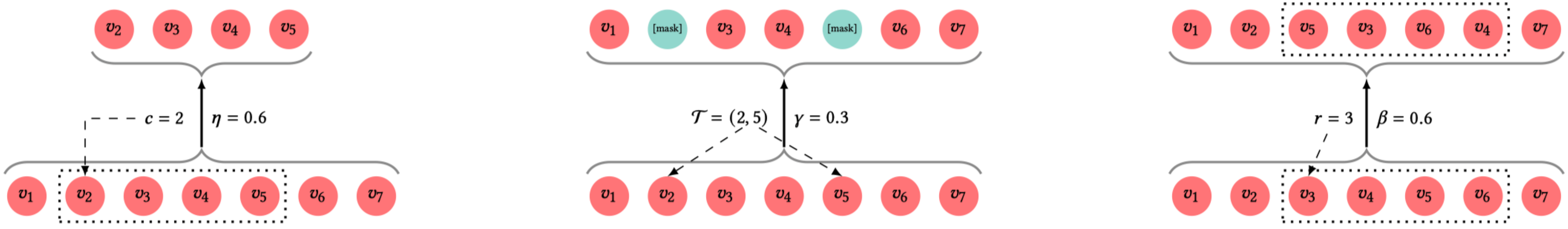
- Apply **contrastive learning** to user behavior sequences.



Model Pre-training for User Modeling

Contrastive learning-based methods

- Apply **contrastive learning** to user behavior sequences.
- **CL4SRec**: mask + crop + reorder



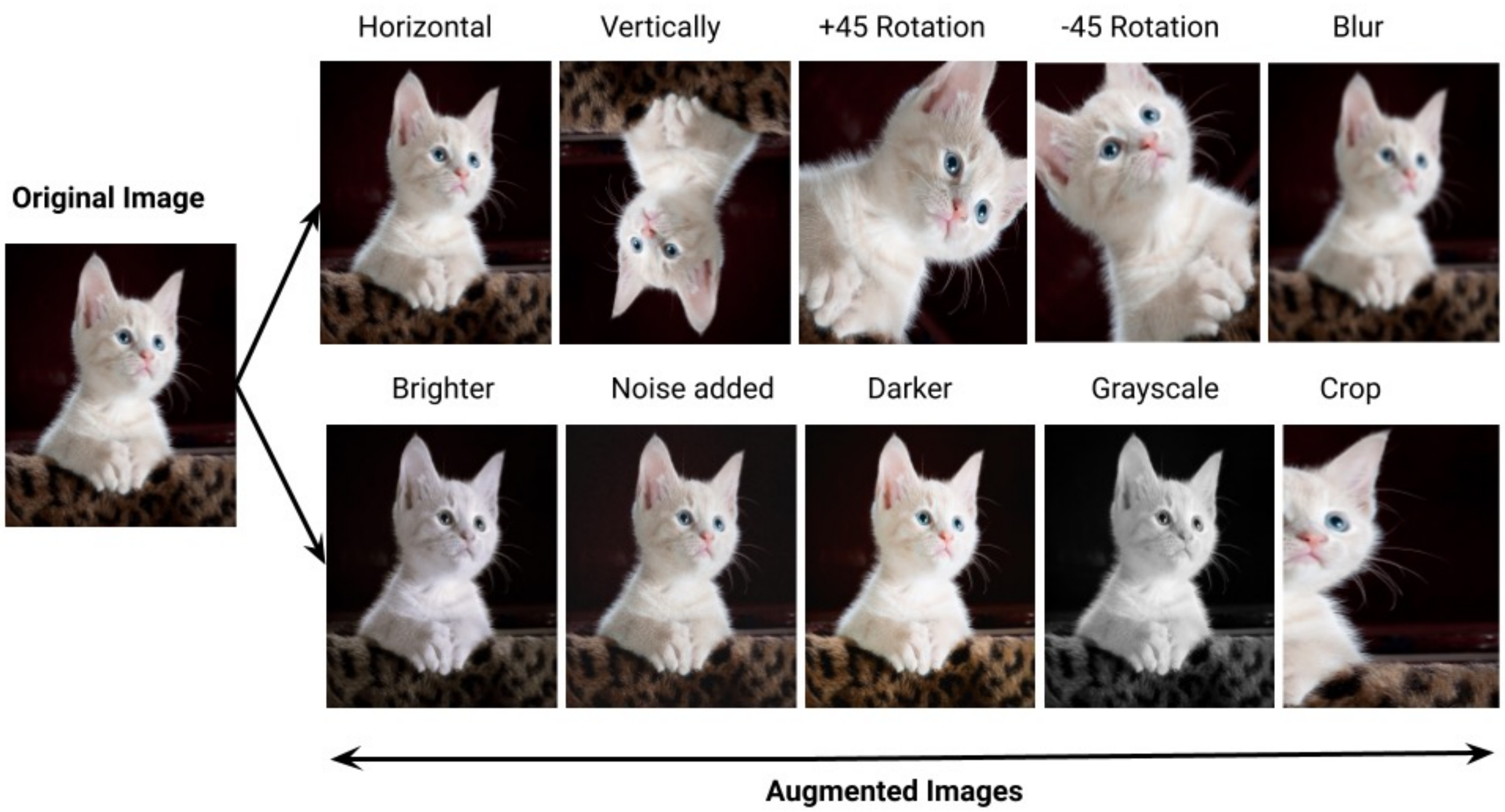
(a) item crop (b) item mask (c) item reorder

- May not hold the basic assumption of contrastive learning: the augmented views should be **semantically consistent**.

Xie et al. Contrastive Learning for Sequential Recommendation. ICDE 2022.

Model Pre-training for User Modeling

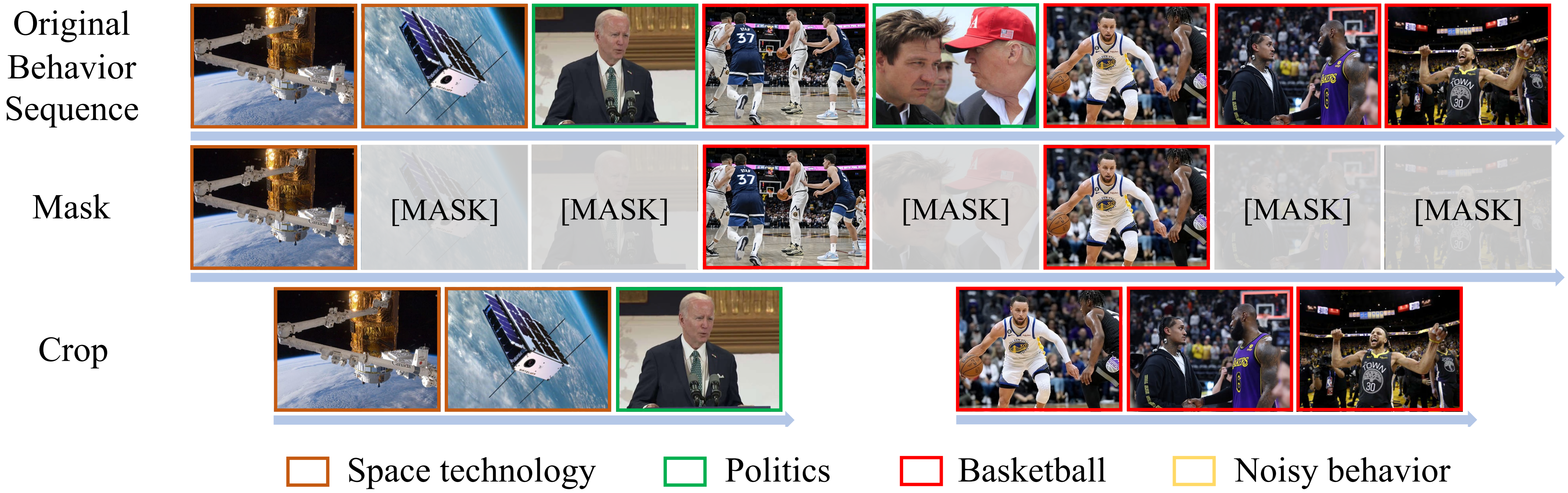
- **Contrastive learning-based methods**
  - The **semantic inconsistency problem** for contrastive learning-based pre-training methods.



Model Pre-training for User Modeling

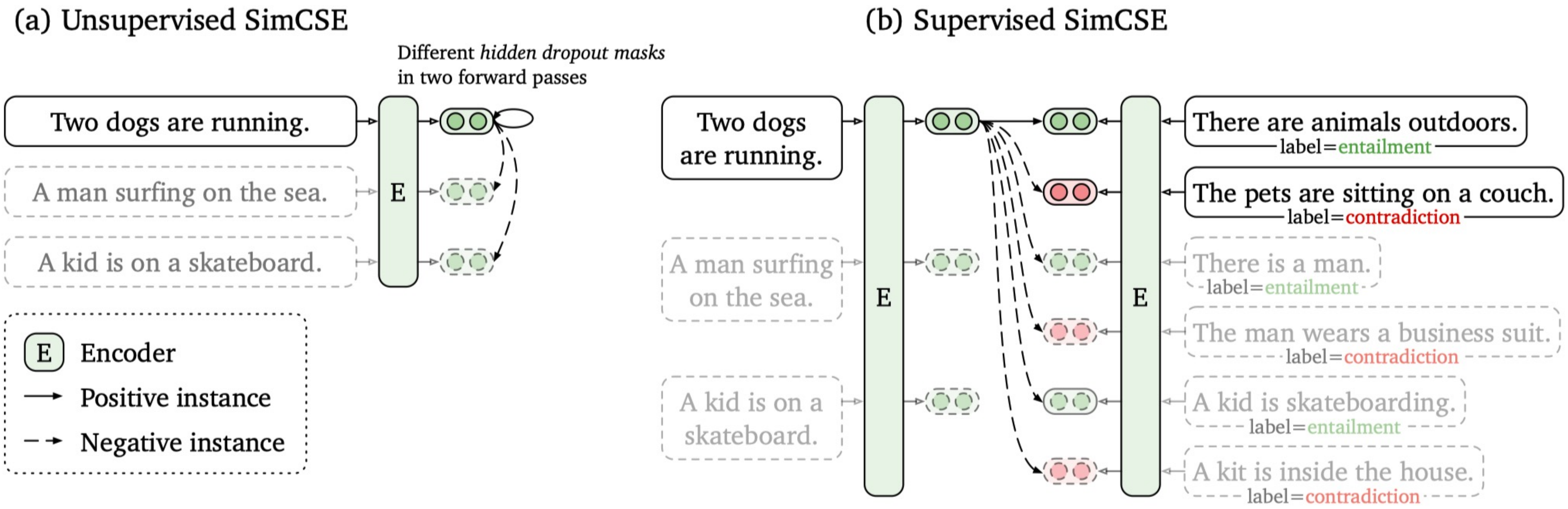
■ Contrastive learning-based methods

- The **semantic inconsistency problem** for contrastive learning-based pre-training methods.



Model Pre-training for User Modeling

- **Contrastive learning-based methods**
  - The **semantic inconsistency problem** for contrastive learning-based pre-training methods.

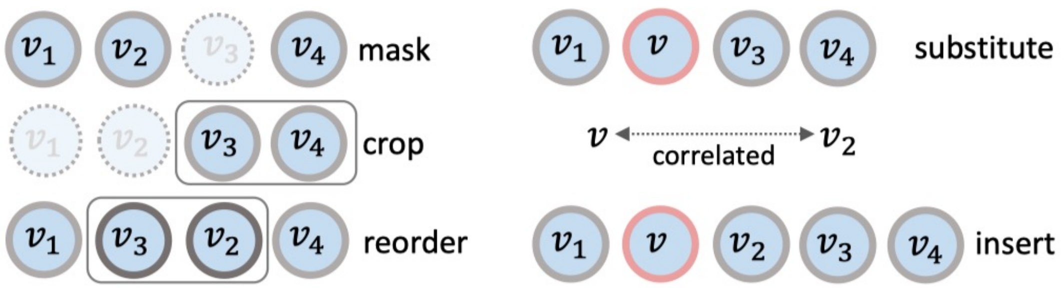


Gao et al. SimCSE: Simple Contrastive Learning of Sentence Embeddings. EMNLP 2021.

Model Pre-training for User Modeling

Contrastive learning-based methods

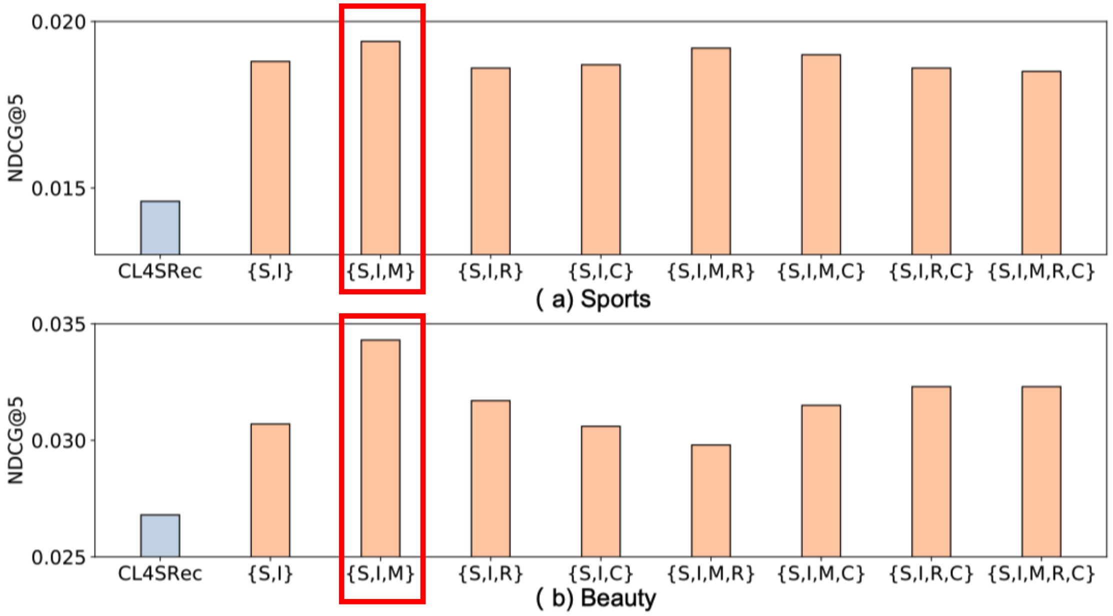
- The **semantic inconsistency problem** for contrastive learning-based pre-training methods.
- **CoSeRec**: Take **item correlation** into consideration



(a) Random (b) Informative

memory-based correlation $$\mathrm{Cor}_o(i,j) = \frac{1}{\sqrt{|\mathcal{N}(i)||\mathcal{N}(j)|}} \sum_{u \in \mathcal{N}(i) \cap \mathcal{N}(j)} \frac{1}{\log(1+|\mathcal{N}(u)|)}$$

model-based correlation $$\mathrm{Cor}_e(i,j) = \mathbf{e}_i \cdot \mathbf{e}_j$$



Liu et al. Contrastive Self-supervised Sequential Recommendation with Robust Augmentation. arXiv 2021.

Model Pre-training for User Modeling

■ Contrastive learning-based methods

- The **semantic inconsistency problem** for contrastive learning-based pre-training methods.
- **CCL**: Generate high-quality positive samples with **mask-and-fill**

$$s' = g_{\text{mask}}(s_u) = [i'_1, i'_2, \ldots, i'_n],$$
$$i'_t = \begin{cases} i_t \in s_u, & t \notin \mathcal{M}; \\ [\text{mask}], & t \in \mathcal{M}. \end{cases}$$

$$z = g_{\text{fill}}(s') = [i'_1, i'_2, \ldots, i'_n],$$
$$i'_t = \begin{cases} i_t \in s, & t \notin \mathcal{M}; \\ i \sim \boxed{\mathrm{P}}(\cdot | s_{u,\neg k}, \mathcal{A}_u), & t \in \mathcal{M}. \end{cases}$$

an MLM model

Models	HR@5	NDCG@5	HR@10	NDCG@10
$\text{CCL}_{\neg\text{Aug}}$	0.3796	0.2855	0.4898	0.3136
$\text{CCL}_{\neg\text{Attr}}$	0.4060	0.3035	0.5152	0.3389
$\text{CCL}_{\neg\text{Curriculum}}$	0.3937	0.2915	0.5014	0.3277
$\text{CCL}_{\neg\text{Annealing}}$	0.4044	0.3035	0.5126	0.3401
$\text{CCL}_{\text{Reorder}}$	0.3998	0.3002	0.5085	0.3354
CCL	**0.4156**	**0.3125**	**0.5234**	**0.3466**

Models	File Search	File Download	Video Player	News Search
Origin	0.0594	0.0636	0.0640	0.0678
Origin+CCL	**0.0650**	**0.0650**	**0.0676**	**0.0696**
Improvement	+2.6%	+2.2%	+4.7%	+2.7%

Bian et al. Contrastive Curriculum Learning for Sequential User Behavior Modeling via Data Augmentation. CIKM 2021.

Model Pre-training for User Modeling

■ Contrastive learning-based methods

- □ The **semantic inconsistency problem** for contrastive learning-based pre-training methods.

Yu et al. AdaptSSR: Pre-training User Model with Augmentation-Adaptive Self-Supervised Ranking. NIPS 2023.

Model Pre-training for User Modeling

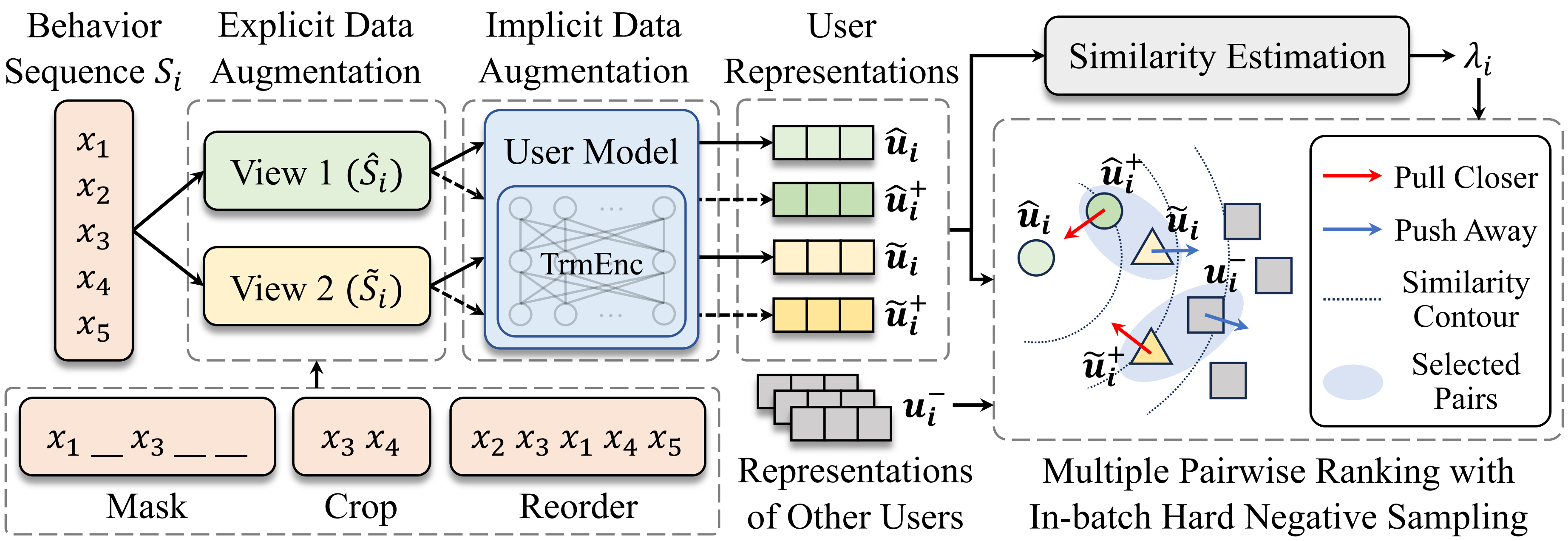
- **Contrastive learning-based methods**
  - The **semantic inconsistency problem** for contrastive learning-based pre-training methods.
  - **AdaptSSR**: Replace the contrastive learning task with a **self-supervised ranking task**.
    - User model $\mathcal{M}$, user behavior sequence $S = \{x_1, x_2, \ldots, x_n\}$
    - Input S into $\mathcal{M}$ twice with different independently sampled dropout masks → $\boldsymbol{u}, \boldsymbol{u}^+$ (implicit data augmentation)
    - Input the augmented behavior sequence $\hat{S}$ into $\mathcal{M}$ → $\hat{\boldsymbol{u}}$ (explicit data augmentation)
    - Input the behavior sequence of another user into $\mathcal{M}$ → $\boldsymbol{u}^-$
    - Pre-training objective: $\mathrm{sim}(\boldsymbol{u}, \boldsymbol{u}^+) \geq \mathrm{sim}(\boldsymbol{u}, \hat{\boldsymbol{u}}) \geq \mathrm{sim}(\boldsymbol{u}, \boldsymbol{u}^-)$

Yu et al. AdaptSSR: Pre-training User Model with Augmentation-Adaptive Self-Supervised Ranking. NIPS 2023.

Model Pre-training for User Modeling

Contrastive learning-based methods

- The **semantic inconsistency problem** for contrastive learning-based pre-training methods.
- **AdaptSSR**: Replace the contrastive learning task with a **self-supervised ranking task**.



Yu et al. AdaptSSR: Pre-training User Model with Augmentation-Adaptive Self-Supervised Ranking. NIPS 2023.

Model Pre-training for User Modeling

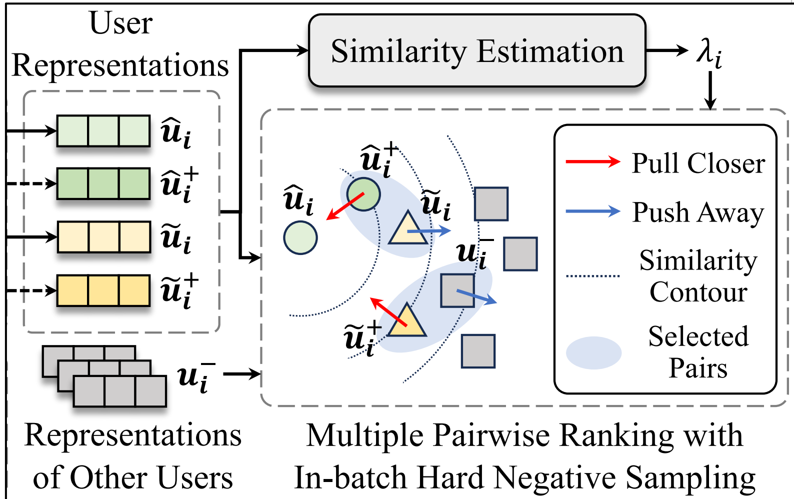
■ Contrastive learning-based methods

- The **semantic inconsistency problem** for contrastive learning-based pre-training methods.
- **AdaptSSR**: Replace the contrastive learning task with a **self-supervised ranking task**.
  - Pre-training objective: $\mathrm{sim}(\boldsymbol{u},\boldsymbol{u}^{+}) \geq \mathrm{sim}(\boldsymbol{u},\hat{\boldsymbol{u}}) \geq \mathrm{sim}(\boldsymbol{u},\boldsymbol{u}^{-})$
  - Multiple pairwise ranking loss (MPR) with in-batch hard negative sampling

$$\hat{\mathcal{L}}_i = -\log\sigma\left[\lambda\boxed{\left(\mathrm{sim}(\hat{\boldsymbol{u}}_i,\hat{\boldsymbol{u}}_i^{+}) - \max_{\boldsymbol{v}\in\{\tilde{\boldsymbol{u}}_i,\tilde{\boldsymbol{u}}_i^{+}\}}\mathrm{sim}(\hat{\boldsymbol{u}}_i,\boldsymbol{v})\right)} + (1-\lambda)\boxed{\left(\min_{\boldsymbol{v}\in\{\tilde{\boldsymbol{u}}_i,\tilde{\boldsymbol{u}}_i^{+}\}}\mathrm{sim}(\hat{\boldsymbol{u}}_i,\boldsymbol{v}) - \max_{\boldsymbol{w}\in\mathbf{U}_i^{-}}\mathrm{sim}(\hat{\boldsymbol{u}}_i,\boldsymbol{w})\right)}\right]$$

  - Augmentation-adaptive fusion
    - ✓ The effects of data augmentation vary significantly across different behavior sequences.

$$\lambda_i = 1 - \frac{1}{4}\sum_{\hat{\boldsymbol{s}}\in\{\hat{\boldsymbol{u}}_i,\hat{\boldsymbol{u}}_i^{+}\}}\ \sum_{\tilde{\boldsymbol{s}}\in\{\tilde{\boldsymbol{u}}_i,\tilde{\boldsymbol{u}}_i^{+}\}}\max(\mathrm{sim}(\hat{\boldsymbol{s}},\tilde{\boldsymbol{s}}),0)$$



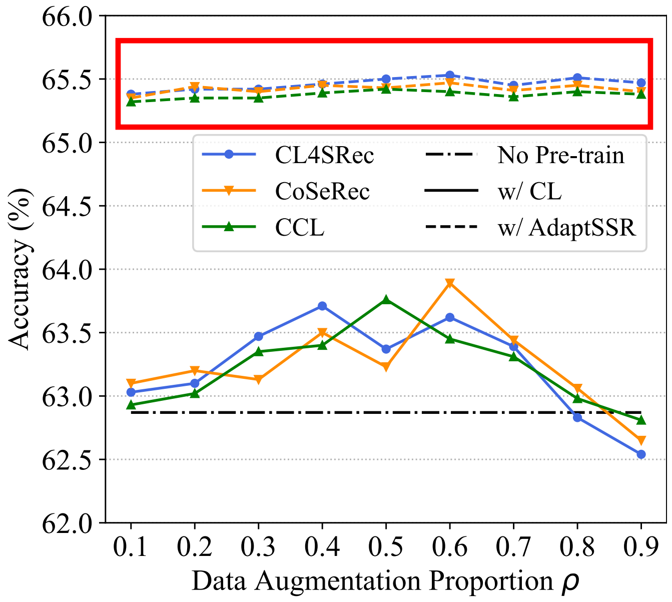
Yu et al. AdaptSSR: Pre-training User Model with Augmentation-Adaptive Self-Supervised Ranking. NIPS 2023.
Yu et al. Multiple Pairwise Ranking with Implicit Feedback. CIKM 2018.

Model Pre-training for User Modeling

Contrastive learning-based methods

- The **semantic inconsistency problem** for contrastive learning-based pre-training methods.
- **AdaptSSR**: Replace the contrastive learning task with a **self-supervised ranking task**.

	Pre-train Method	$\mathcal{T}_1$ age prediction Acc	Impr	$\mathcal{T}_2$ life status prediction Acc	Impr	$\mathcal{T}_3$ click recommendation NDCG@10	Impr	$\mathcal{T}_4$ thumb-up recommendation NDCG@10	Impr	$\mathcal{T}_5$ gender prediction AUC	Impr	$\mathcal{T}_6$ CVR prediction AUC	Impr
	None	62.87±0.05	-	52.24±0.16	-	1.99±0.03	-	2.87±0.07	-	78.63±0.06	-	75.14±0.14	-
behavior prediction-based methods	PeterRec	63.62±0.11	1.19	53.14±0.07	1.72	2.37±0.02	19.10	3.06±0.08	6.62	79.61±0.13	1.25	76.04±0.10	1.20
	PTUM	63.21±0.14	0.54	53.05±0.04	1.55	2.29±0.03	15.08	2.96±0.03	3.14	79.48±0.11	1.08	75.82±0.13	0.90
contrastive learning-based methods	CLUE	63.38±0.10	0.81	53.23±0.05	1.90	2.38±0.02	19.60	3.05±0.21	6.27	79.90±0.06	1.62	76.03±0.16	1.18
	CCL	63.76±0.11	1.42	53.37±0.09	2.16	2.43±0.02	22.11	3.32±0.13	15.68	80.22±0.07	2.02	77.35±0.10	2.94
	IDICL	63.88±0.04	1.61	53.45±0.05	2.32	2.46±0.02	23.62	3.42±0.04	19.16	80.34±0.05	2.17	77.92±0.08	3.70
different data augmentation methods	CL4SRec	63.71±0.14	1.34	53.43±0.05	2.28	2.41±0.03	21.11	3.29±0.06	14.63	80.14±0.08	1.92	77.02±0.05	2.50
	CoSeRec	63.89±0.03	1.62	53.53±0.09	2.47	2.44±0.02	22.61	3.33±0.05	16.03	80.48±0.06	2.35	77.71±0.09	3.42
	DuoRec	63.50±0.09	1.00	53.26±0.06	1.95	2.39±0.01	20.10	3.11±0.16	8.36	80.03±0.09	1.78	76.85±0.09	2.28
	AdaptSSR	**65.53**±0.04	**4.23**	**54.41**±0.02	**4.15**	**2.61**±0.03	**31.16**	**3.73**±0.03	**29.97**	**82.30**±0.03	**4.67**	**79.92**±0.05	**6.36**

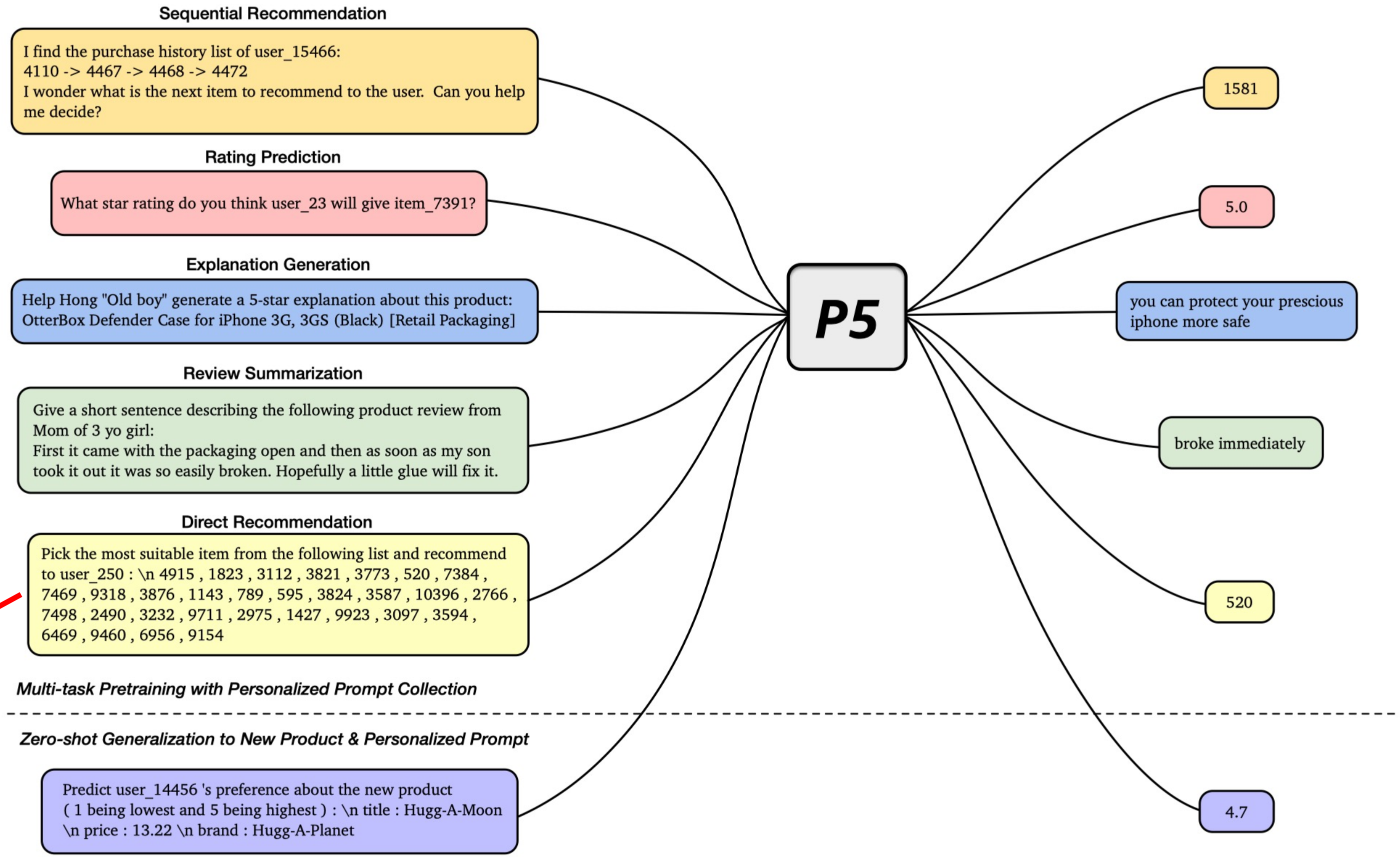


Performance with different data augmentation methods

Yu et al. AdaptSSR: Pre-training User Model with Augmentation-Adaptive Self-Supervised Ranking. NIPS 2023.

Model Pre-training for User Modeling

- **When user modeling meets LLM**
  - **P5**: Unify recommendation tasks into the text generation task via prompt learning.

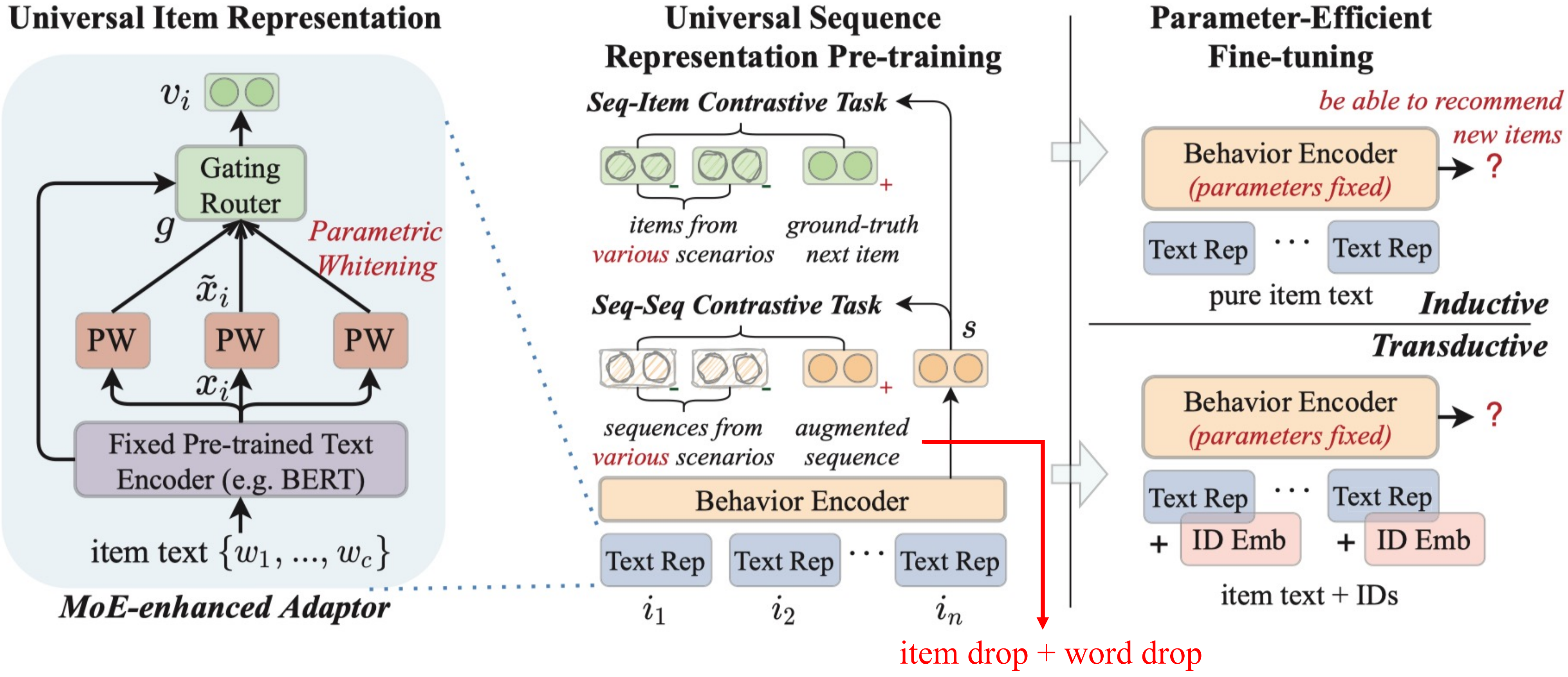


ID-based: limited within the same domain

Geng et al. Recommendation as Language Processing (RLP): A Unified Pretrain, Personalized Prompt & Predict Paradigm (P5). RecSys 2022.

Model Pre-training for User Modeling

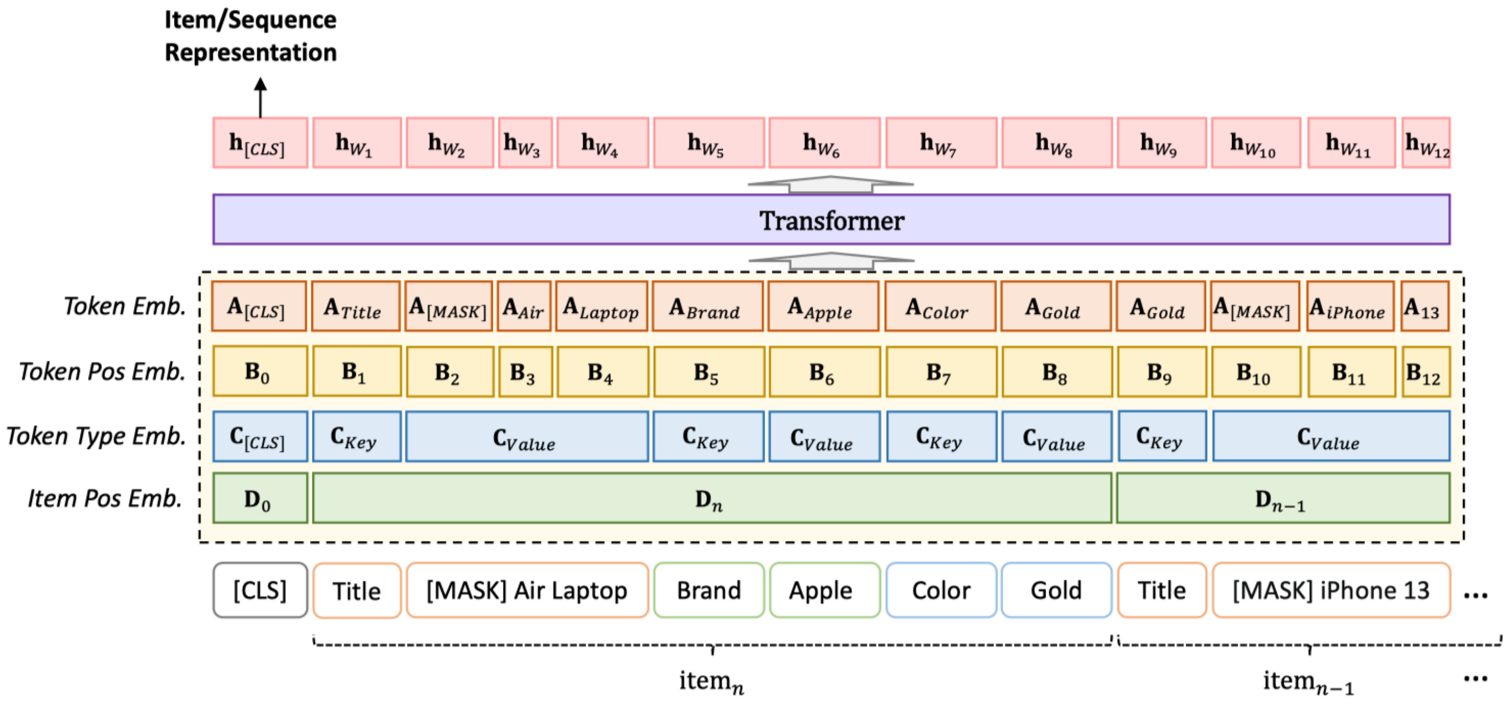
- **When user modeling meets LLM**
  - **UniSRec**: Learn universal item/user representations based on text descriptions.



Hou et al. Towards Universal Sequence Representation Learning for Recommender Systems. KDD 2022.

Model Pre-training for User Modeling

- **When user modeling meets LLM**
  - **Recformer**: Model user preferences and item features as language representations.



(a) Recformer Model Structure

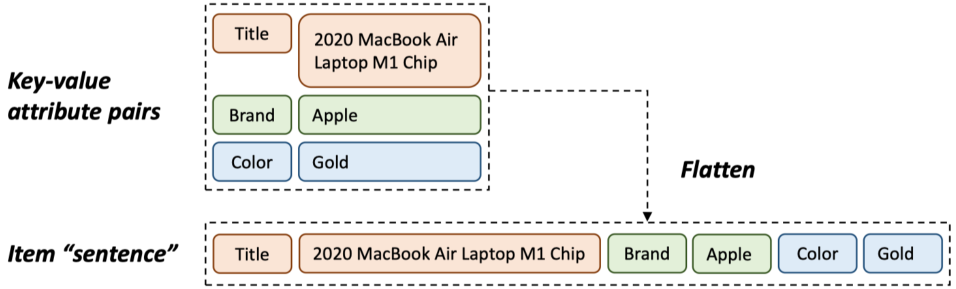
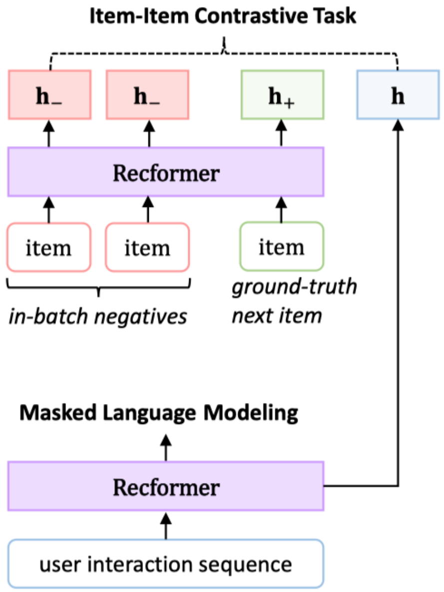


Figure 2: Model input construction. Flatten key-value attribute pairs into an item "sentence".



$$\mathcal{L}_{\mathrm{IIC}} = -\log \frac{e^{\mathrm{sim}(\mathbf{h}_s, \mathbf{h}_i^+)/\tau}}{\sum_{i \in \mathcal{B}} e^{\mathrm{sim}(\mathbf{h}_s, \mathbf{h}_i)/\tau}}$$

$$\mathbf{m} = \mathrm{LayerNorm}(\mathrm{GELU}(\mathbf{W}_h \mathbf{h}_w + \mathbf{b}_h))$$
$$p = \mathrm{Softmax}(\mathbf{W}_0 \mathbf{m} + \mathbf{b}_0)$$
$$\mathcal{L}_{\mathrm{MLM}} = -\sum_{i=0}^{|\mathcal{V}|} y_i \log(p_i)$$

(b) Pretraining

Li et al. Text Is All You Need: Learning Language Representations for Sequential Recommendation. KDD 2023.

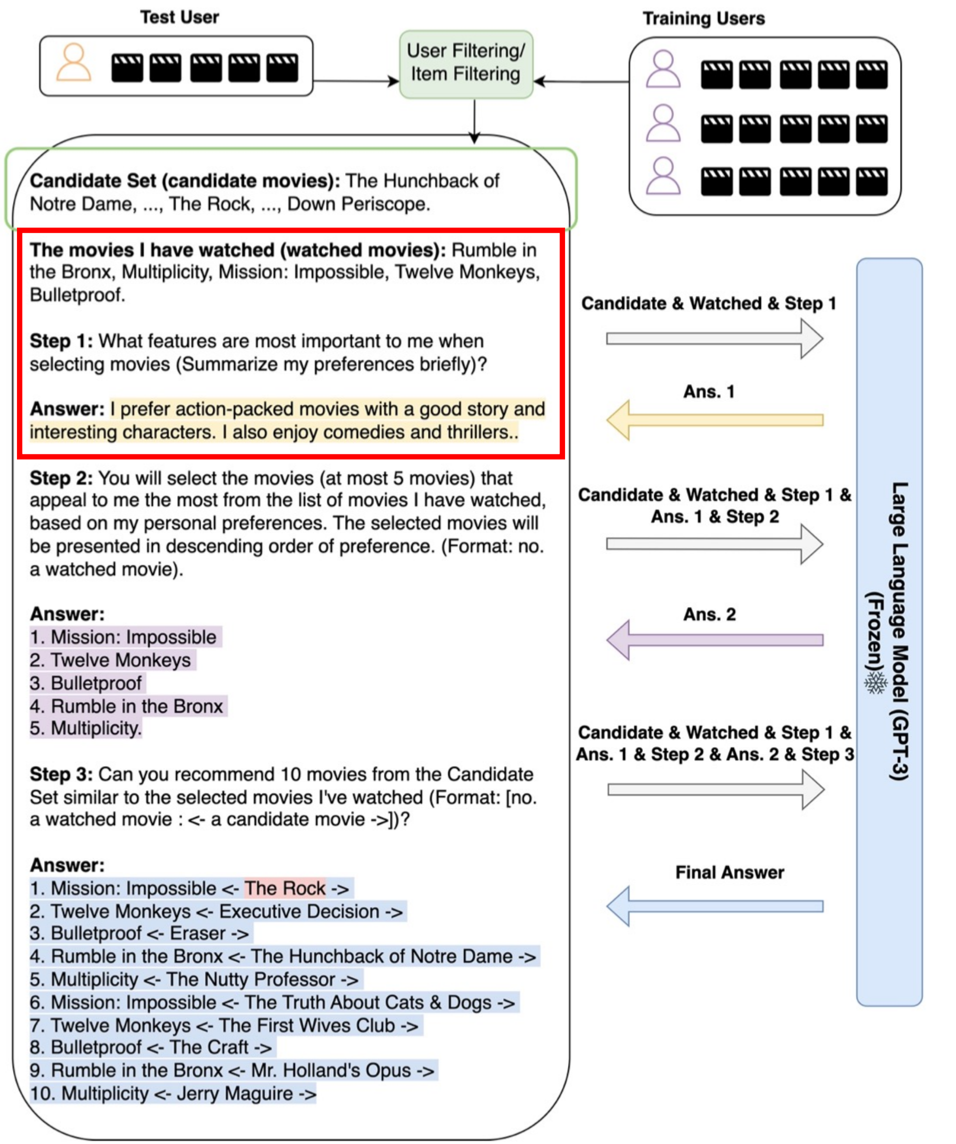
Model Pre-training for User Modeling

- **When user modeling meets LLM**
  - **NIR**: Let the LLM summarize the user's preferences based on user's historical behaviors.

The movies I have watched (watched movies): Rumble in the Bronx, Multiplicity, Mission: Impossible, Twelve Monkeys, Bulletproof.

Step 1: What features are most important to me when selecting movies (Summarize my preferences briefly)?

Answer: I prefer action-packed movies with a good story and interesting characters. I also enjoy comedies and thrillers..



Wang et al. Zero-Shot Next-Item Recommendation using Large Pretrained Language Models. arXiv 2023.

Outline

Future Directions

- Should we train large user models as LLMs?
  - Do we need new pre-training tasks/model structure?
  - Scaling law and emergent ability.
- How will the next-generation LLM-based recommendation system look like?
  - From the perspective of user-system interaction, the cost of dialogue is too high.
  - Learn to ask? Seems can be connected with CAT.
- Beyond recommendation systems…
  - The adaptation of LLMs in vertical domains.
  - User capability modeling pre-training.

Reference

- Wu et al. Empowering News Recommendation with Pre-trained Language Models. *SIGIR* 2021.
- Liu et al. KRED: Knowledge-Aware Document Representation for News Recommendations. *RecSys* 2020.
- Sun et al. Multi-modal Knowledge Graphs for Recommender Systems. *CIKM* 2020.
- Qiu et al. U-BERT: Pre-training User Representations for Improved Recommendation. *AAAI* 2021.
- Gururangan et al. Don’t Stop Pretraining: Adapt Language Models to Domains and Tasks. *ACL* 2020.
- Yu et al. Tiny-NewsRec: Effective and Efficient PLM-based News Recommendation. *EMNLP* 2022.
- Li et al. Text Is All You Need: Learning Language Representations for Sequential Recommendation. *KDD* 2023.
- Touvron et al. LLaMA: Open and Efficient Foundation Language Models. *arXiv* 2023.
- Cui et al. ChatLaw: Open-Source Legal Large Language Model with Integrated External Knowledge Bases. *arXiv* 2023.
- Li et al. EcomGPT: Instruction-tuning Large Language Models with Chain-of-Task Tasks for E-commerce. *arXiv* 2023.
- Wang et al. HuaTuo: Tuning LLaMA Model with Chinese Medical Knowledge. *arXiv* 2023.
- Wang et al. Zero-Shot Next-Item Recommendation using Large Pretrained Language Models. *arXiv* 2023.
- Yuan et al. Parameter-Efficient Transfer from Sequential Behaviors for User Modeling and Recommendation. *SIGIR* 2020.
- Wu et al. PTUM: Pre-training User Model from Unlabeled User Behaviors via Self-supervision. *Findings of EMNLP* 2020.

Reference

- Mu et al. ID-Agnostic User Behavior Pre-training for Sequential Recommendation. *CCIR* 2022.
- Xie et al. Contrastive Learning for Sequential Recommendation. *ICDE* 2022.
- Gao et al. SimCSE: Simple Contrastive Learning of Sentence Embeddings. *EMNLP* 2021.
- Liu et al. Contrastive Self-supervised Sequential Recommendation with Robust Augmentation. *arXiv* 2021.
- Bian et al. Contrastive Curriculum Learning for Sequential User Behavior Modeling via Data Augmentation. *CIKM* 2021.
- Yu et al. AdaptSSR: Pre-training User Model with Augmentation-Adaptive Self-Supervised Ranking. *NIPS* 2023.
- Yu et al. Multiple Pairwise Ranking with Implicit Feedback. *CIKM* 2018.
- Geng et al. Recommendation as Language Processing (RLP): A Unified Pretrain, Personalized Prompt & Predict Paradigm (P5). *RecSys* 2022.
- Hou et al. Towards Universal Sequence Representation Learning for Recommender Systems. *KDD* 2022.
- Li et al. Text Is All You Need: Learning Language Representations for Sequential Recommendation. *KDD* 2023.
- https://zhuanlan.zhihu.com/p/370859857

Thanks!
Q&A