

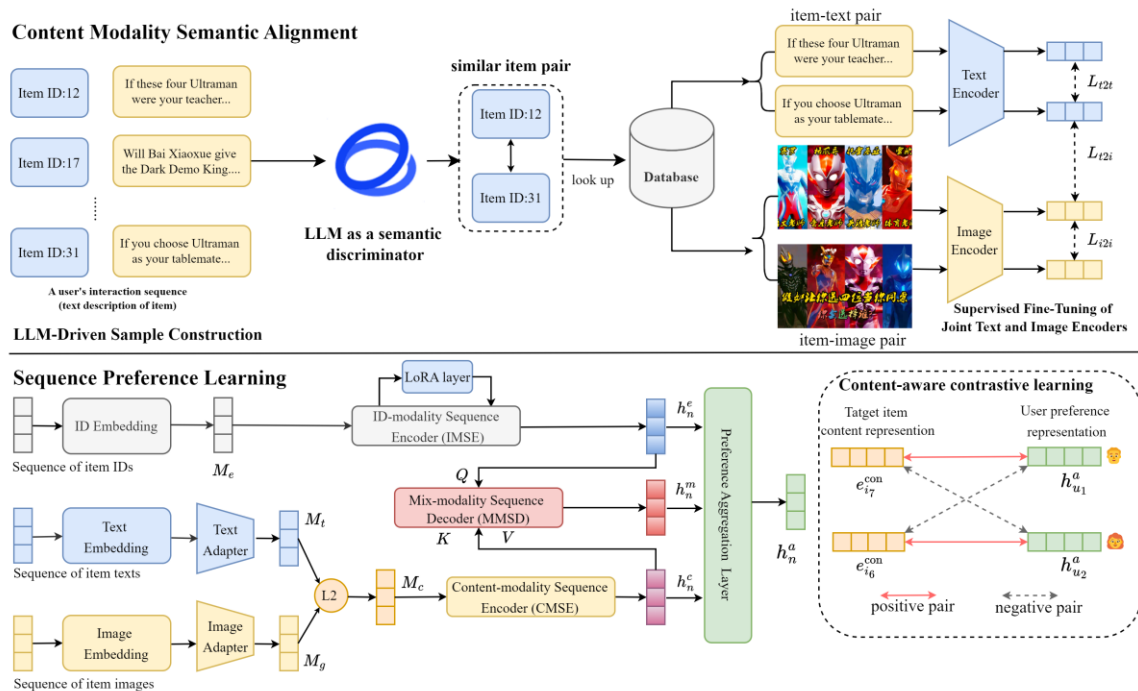
Self-Supervised Representation Learning with ID-Content Modality Alignment for Sequential Recommendation

Donglin ZHOU, Weike PAN, Zhong MING

Frontiers of Computer Science, DOI: [10.1007/s11704-025-50269-4](https://doi.org/10.1007/s11704-025-50269-4)

Problems & Ideas

- Problems of sequential recommendation:
 - We have three user interaction sequences, including the item-ID sequence, the item-text sequence, and the item-image sequence. Our goal is to exploit these sequences to predict the user's next preferred item.
- Ideas:
 - We design a content-modality semantic alignment module to reduce the semantic gap between different modalities and a Transformer-based architecture to learn user sequential preferences.



Main Contributions

- Contributions:
 - We propose a novel content-modality semantic alignment method, reducing the semantic gap between different content-modality representations.
 - We design a Transformer-based encoder-decoder architecture to model user behaviors, content preferences, and their inherent relationships.
 - We further introduce an effective two-step training strategy to combine content dependency and item collaborative dependency, facilitating efficient representation learning.

Input Type & Model →		ID			T			T+ID			T+V+ID			Improved
Dataset	Metric	GRU4Rec	SASRec	BERT4Rec	UniSRec	VQ-Rec	MISSRec	LLMESR	UniSRec	MISSRec	LLMESR	MISSRec	SICSR	
Cartoon	Hit@5	0.0678	0.0845	0.0354	0.0754	0.0470	0.0465	0.0713	0.0801	<u>0.0856</u>	0.0716	0.0730	0.0875	2.22%
	Hit@10	0.1074	0.1318	0.0628	0.1293	0.0909	0.0683	0.1120	<u>0.1327</u>	0.1296	0.1130	0.1166	0.1344	1.28%
	NDGC@5	0.0431	<u>0.0474</u>	0.0213	0.0432	0.0280	0.0241	0.0400	<u>0.0464</u>	0.0454	0.0415	0.0430	0.0531	12.03%
	NDCG@10	0.0558	0.0626	0.0300	0.0607	0.0421	0.0308	0.0531	<u>0.0634</u>	0.0589	0.0548	0.0564	0.0682	7.57%
Dance	Hit@5	0.1395	0.1489	0.0914	0.1387	0.1029	0.0914	0.1379	0.1420	0.1266	0.1392	<u>0.1535</u>	0.1578	2.80%
	Hit@10	0.2067	<u>0.2251</u>	0.1495	0.2155	0.1670	0.1409	0.2088	0.2150	0.1876	0.2102	0.2248	0.2295	1.95%
	NDGC@5	0.0928	0.0933	0.0583	0.0883	0.0660	0.0505	0.0877	0.0897	0.0696	0.0887	<u>0.0965</u>	0.1074	11.30%
	NDCG@10	0.1143	0.1179	0.0770	0.1131	0.0866	0.0657	0.1106	0.1133	0.0883	0.1115	<u>0.1184</u>	0.1304	10.14%
Food	Hit@5	0.0866	0.1356	0.0437	0.0556	0.0779	0.0739	0.1244	0.0999	0.1283	0.1266	0.1448	<u>0.1367</u>	-5.59%
	Hit@10	0.1376	0.1993	0.0765	0.1026	0.1460	0.1222	0.1849	0.1587	0.1855	0.1881	<u>0.2023</u>	0.2037	0.69%
	NDCG@5	0.0561	0.0758	0.0265	0.0332	0.0440	0.0388	0.0711	0.0582	0.0631	0.0719	<u>0.0837</u>	0.0849	1.43%
	NDCG@10	0.0724	0.0964	0.0370	0.0481	0.0658	0.0535	0.0906	0.0771	0.0807	0.0917	<u>0.1013</u>	0.1064	5.03%
Movie	Hit@5	0.0618	<u>0.0784</u>	0.0378	0.0761	0.0545	0.0415	0.0694	0.0765	0.0687	0.0734	0.0764	0.0810	3.32%
	Hit@10	0.0966	0.1247	0.0623	0.1229	0.0953	0.0631	0.1102	<u>0.1248</u>	0.1107	0.1105	0.1155	0.1260	0.96%
	NDCG@5	0.0404	0.0434	0.0226	0.0445	0.0332	0.0236	0.0383	<u>0.0461</u>	0.0369	0.0412	0.0446	0.0495	7.38%
	NDCG@10	0.0516	0.0583	0.0304	0.0595	0.0464	0.0301	0.0515	<u>0.0617</u>	0.0498	0.0532	0.0565	0.0640	3.73%