You Are What You Bought: Generating Customer Personas for E-commerce Applications

Yimin Shi, Yang Fei, Shiqi Zhang, Haixun Wang, Xiaokui Xiao

SIGIR 2025 Padova ITALY | NUS National University of Singapore | EvenUp | PyroWis

15 July | ANTENORE | 10:30-12:30

E-commerce with LLM

LLM capabilities

- personalization, real-world knowledge, customer intention understanding

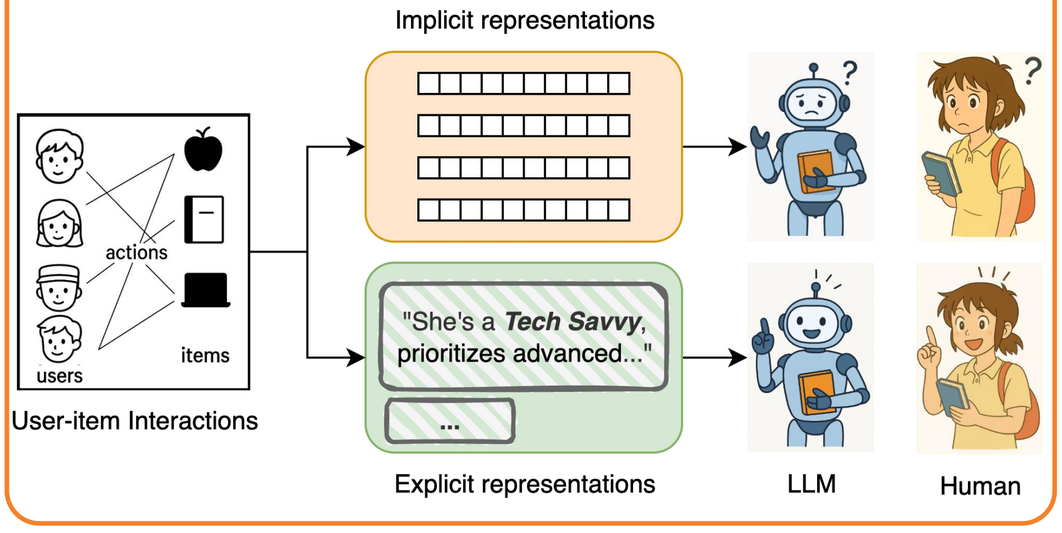
Scenario: online bookstore

- provides explainable book recommendation based on reading history, personal interests, and contextual preferences, etc.

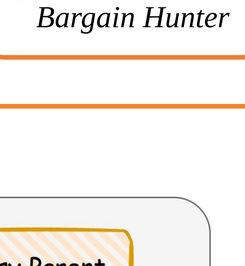
Customer Representation Gap

From implicit to explicit

- traditional: map user behavior data to vectors
- need: explicit representations comprehensible to both LLM and human

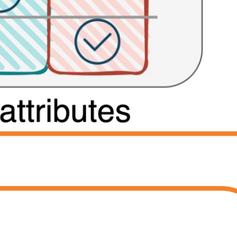


Persona



Busy Parents

Health Enthusiast



Bargain Hunter

Representation by Personas

Persona properties: Informativeness, Readability, Robustness

Persona generation pipeline:

- serialize user-item interactions into text with a template
- LLM labels the user based on a pre-defined persona set

Potential downstream applications:

- Product recommendation, Customer Segmentation, Customer-centric Search Navigation, etc.

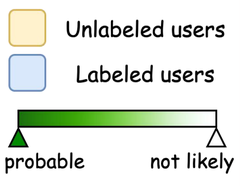
Scalability issue:

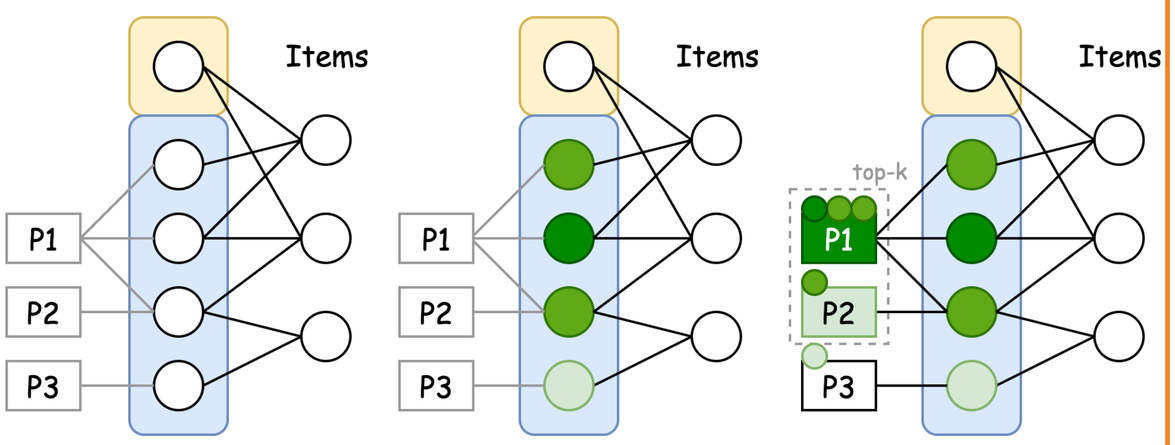
- e-commerce platforms easily have millions of users
- user representations need to be dynamically updated
- LLM inferences are expensive, we want to reduce the cost

GPLR

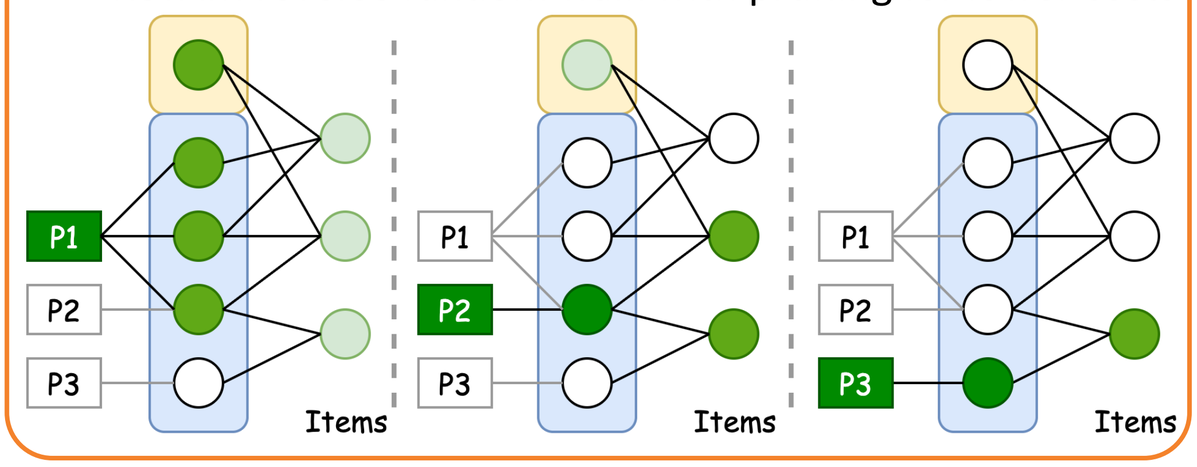
Method overview

- DU-Sampling: selectively label 5% customers
- Random walk: for each unlabeled, find their labeled neighbors, aggregate their personas





- RevAff: reverse random walk with pruning for lower costs



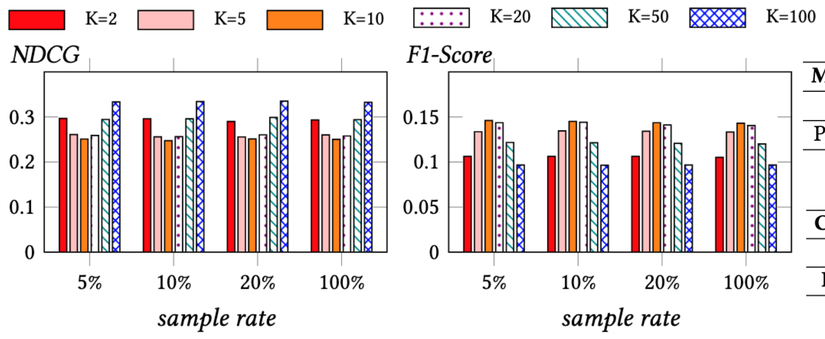
Experiments

Application 1: product recommendation

- solution: construct user-persona-item tripartite graph
- apply GNN-recommendation models on the tripartite
- competitors: MF, GCMC, LCFN, LightGCN, LGCN, AFDGCF

	OnelineRetail						Instacart					
	MF	Light	LGCN	AFD	LGCN3	A-LGCN3	MF	Light	LGCN	AFD	LGCN3	A-LGCN3
N@2	0.2391	0.2801	0.2686	0.2898	0.2933	**0.2940**	0.1166	0.1477	0.1405	0.1535	0.1570	**0.1634**
N@5	0.2143	0.2497	0.2356	0.2578	**0.2602**	0.2549	0.1006	0.1273	0.1208	0.1308	0.1319	**0.1357**
N@10	0.2104	0.2443	0.2274	0.2489	**0.2503**	0.2497	0.0916	0.1180	0.1106	0.1205	0.1197	**0.1230**
N@20	0.2221	0.2575	0.2383	0.2595	0.2577	**0.2617**	0.0934	0.1198	0.1117	0.1222	0.1206	**0.1242**
N@50	0.2612	0.2942	0.2772	0.2978	0.2940	**0.2996**	0.1150	0.1448	0.1351	0.1476	0.1451	**0.1491**
N@100	0.3011	0.3337	0.3167	0.3346	0.3325	**0.3392**	0.1389	0.1733	0.1618	0.1761	0.1732	**0.1766**
F@2	0.0859	0.1080	0.0955	0.1066	0.1052	**0.1092**	0.0326	0.0415	0.0375	0.0427	0.0411	**0.0433**
F@5	0.1123	0.1333	0.1207	0.1348	0.1332	**0.1349**	0.0478	0.0609	0.0564	**0.0629**	0.0611	0.0628
F@10	0.1225	0.1416	0.1352	**0.1436**	0.1431	0.1432	0.0560	0.0716	0.0672	0.0730	0.0722	**0.0738**
F@20	0.1239	0.1398	0.1341	0.1393	0.1406	**0.1415**	0.0602	0.0752	0.0710	0.0768	0.0758	**0.0780**
F@50	0.1081	0.1181	0.1159	0.1186	0.1201	**0.1210**	0.0565	0.0684	0.0644	0.0692	0.0681	**0.0701**
F@100	0.0886	0.0955	0.0934	0.0945	0.0967	**0.0976**	0.0470	0.0562	0.0528	0.0568	0.0557	**0.0570**

- even 5% sample achieves performance comparable to full



Method	# of Consistent Customers
RFM	4 (1.3%)
Persona	54 (18%)

(1)

Clusters#	5	15	25	35
RFM	0.366	0.404	0.431	0.445
Persona	**0.451**	**0.671**	**0.771**	**0.788**

(2)

Application 2: customer segmentation

- solution: encode each customer using one-hot, reduce the dimension via PCA, then cluster with KMeans
- dimensions: (1) robustness, (2) customer clustering quality

Acknowledgement, technical report, code repository, etc.

This work is supported by the Ministry of Education, Singapore, under its MOE AcRF TIER 3 Grant (MOE-MOET32022-0001). For the technical report and code, please scan the QR code and visit Yimin's homepage.

Yimin Shi