PEARL: A Review-driven Persona-Knowledge Grounded Conversational Recommendation Dataset

Minjin Kim*[1], Minju Kim*[1], Hana Kim[1], Beong-woo Kwak[1], SeongKu Kang[2], Youngjae Yu[1], Jinyoung Yeo[1], Dongha Lee[1]

[1]Yonsei University, [2]University of Illinois at Urbana-Champaign

ACL 2024 Bangkok, Thailand

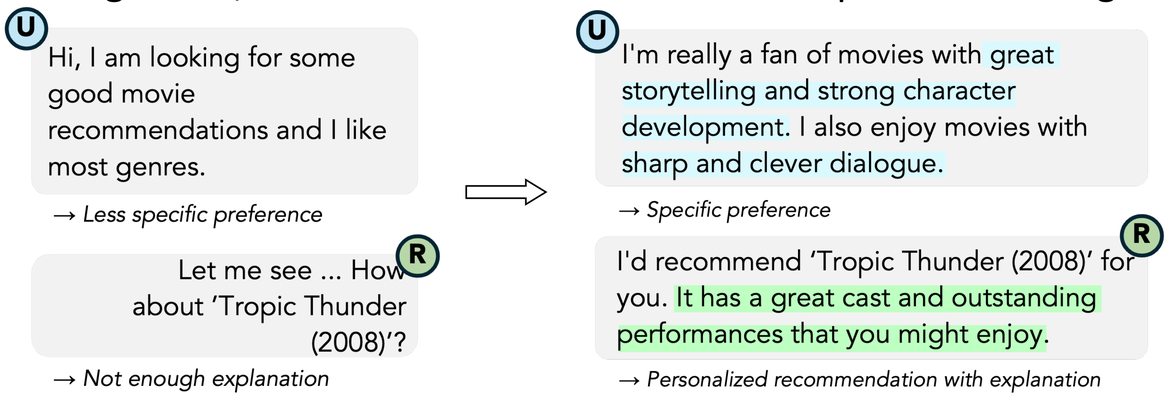
Full paper | Dataset

1 ABSTRACT

- We present a novel conversational recommendation dataset named PEARL, synthesized with persona- and knowledge-augmented LLM simulators.
- We obtain detailed persona and knowledge from real-world reviews and construct a dataset with 57k dialogues.
- We show the quality and utility of PEARL through human and automatic evaluations.

2 MOTIVATION

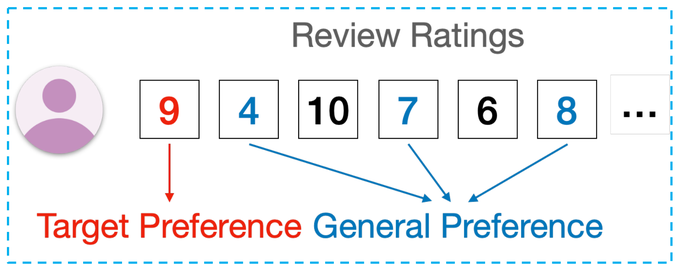
- In existing conversational recommendation datasets, user preferences are too general, and recommendations are not explainable enough.



3 METHOD

1. Construct user-review and movie-review databases with metadata and reviews from IMDb
2. Run persona-augmented user simulator
3. Run knowledge-augmented recommender simulator
4. Complete dialogues via turn-by-turn generation
5. Filter out disqualified or low-quality dialogues

Persona Construction

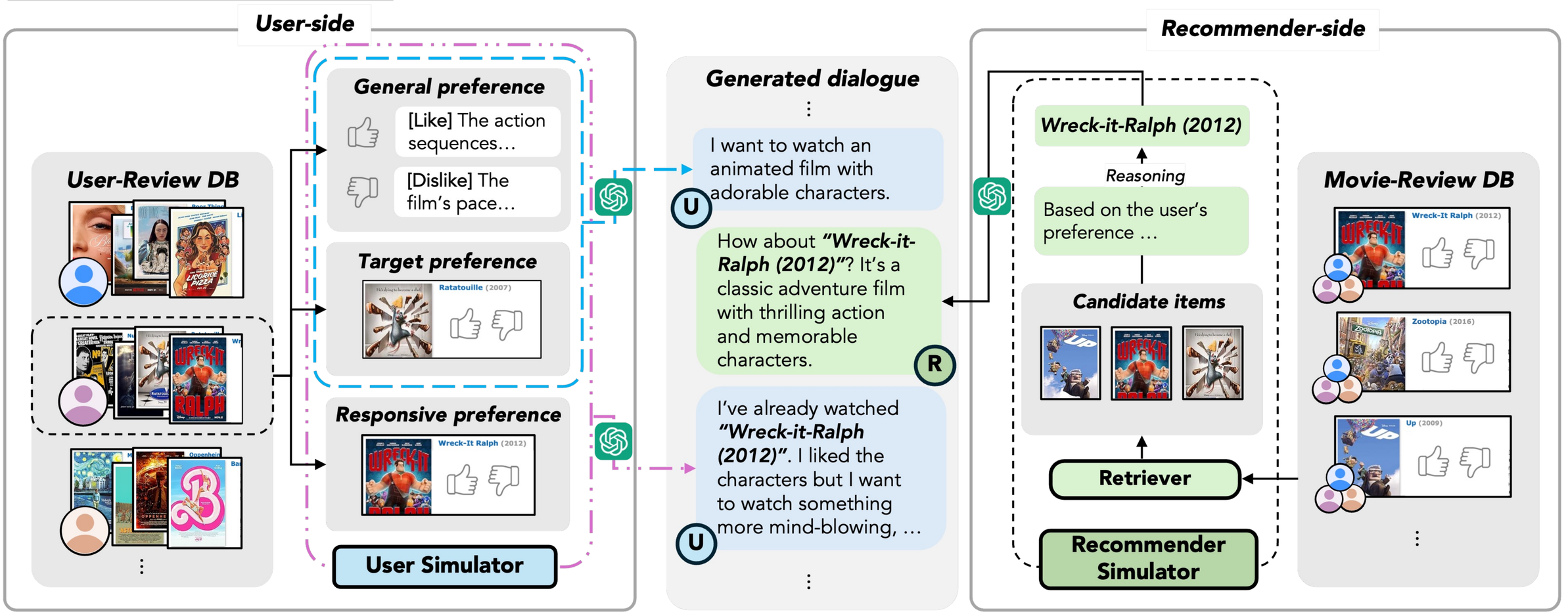


Knowledge Construction

Metadata
(title, genre, director, cast)

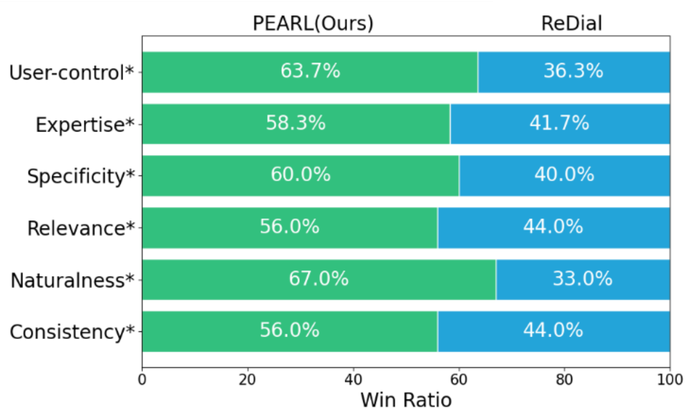
Preferred & unpreferred attributes
(summarized from IMDb's top-featured reviews)

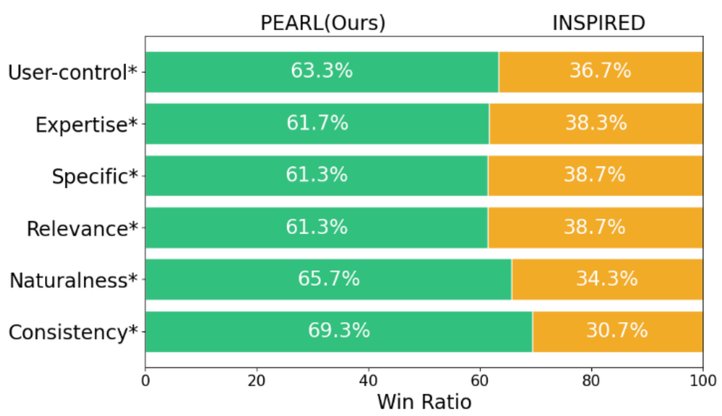
PEARL Construction Overview



4 EXPERIMENTS

Evaluation on Dialogue Quality

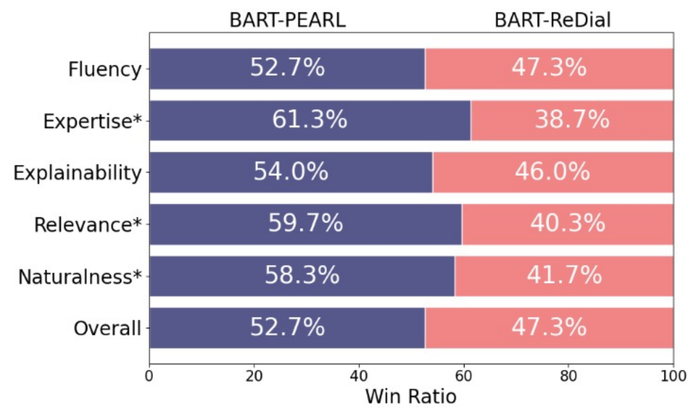




	ReDial	INSPIRED	PEARL
# of dialogues	10,006	1,001	57,277
# of utterances	182,150	35,811	548,061
2-gram specificity	65.44	119.56	141.79
3-gram specificity	65.97	123.01	149.75
4-gram specificity	65.37	122.81	153.00

- Dialogues of PEARL shows superiority in expertise, specificity, etc. over those of baseline datasets.
- PEARL is superior to previous datasets in scalability and specificity in user preferences.

Evaluation on Dialogue Utility



- Models trained on PEARL generate better responses than those trained on a baseline dataset.

Model	Dist-3	Dist-4
BART-ReDial	0.6220	0.5057
BART-PEARL	**0.9241**	**0.8861**
UniCRS-ReDial	0.5413	0.3667
UniCRS-PEARL	**0.9338**	**0.9007**
PECRS-ReDial	0.6798	0.5906
PECRS-PEARL	**0.9132**	**0.8947**
GPT-3.5	0.9256	0.8910

- PEARL improves the diversity of responses generated by downstream models.

Model	R@1	R@10	R@50
BERT-PEARL	0.0018	0.0208	0.0736
UniCRS-PEARL	**0.0310**	**0.0697**	**0.1202**
PECRS-PEARL	0.0151	0.0339	0.0798
GPT-3.5	0.0071	0.0355	0.0709

- All models, including GPT-3.5, show low performances.
- Future research on utilizing PEARL, built based on real-world data rather than on parametric knowledge, is necessary.