

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

Minjin Kim*1, Minju Kim*1, Hana Kim1, Beong-woo Kwak1, SeongKu Kang², Youngjae Yu¹, Jinyoung Yeo¹, Dongha Lee¹

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





Full paper Dataset

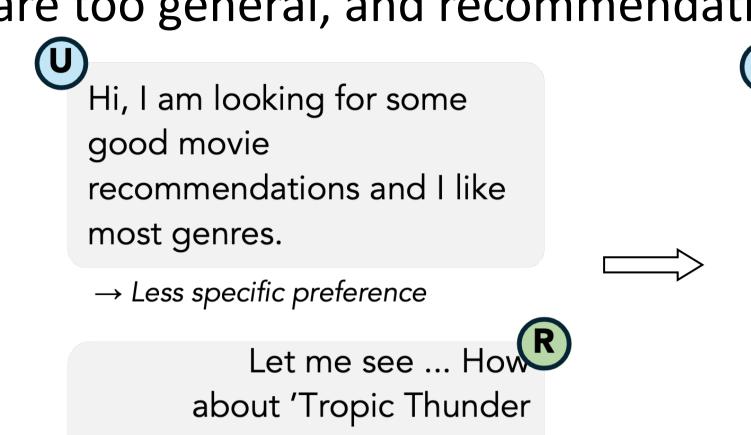


ABSTRACT

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

MOTIVATION

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



(2008)'?

I'm really a fan of movies with great storytelling and strong character development. I also enjoy movies with sharp and clever dialogue.

→ Specific preference

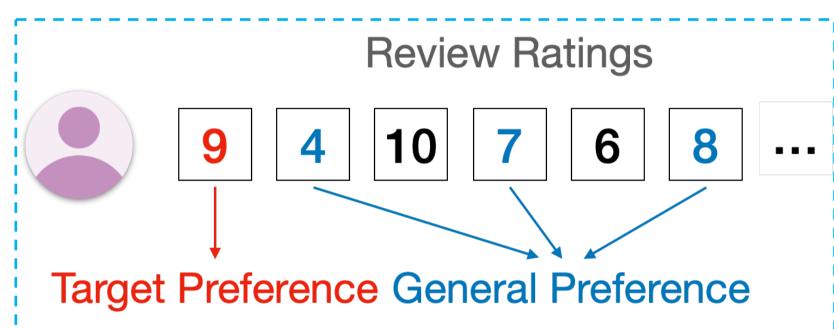
I'd recommend 'Tropic Thunder (2008)' for you. It has a great cast and outstanding performances that you might enjoy.

→ Personalized recommendation with explanation

3 METHOD

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

Persona Construction

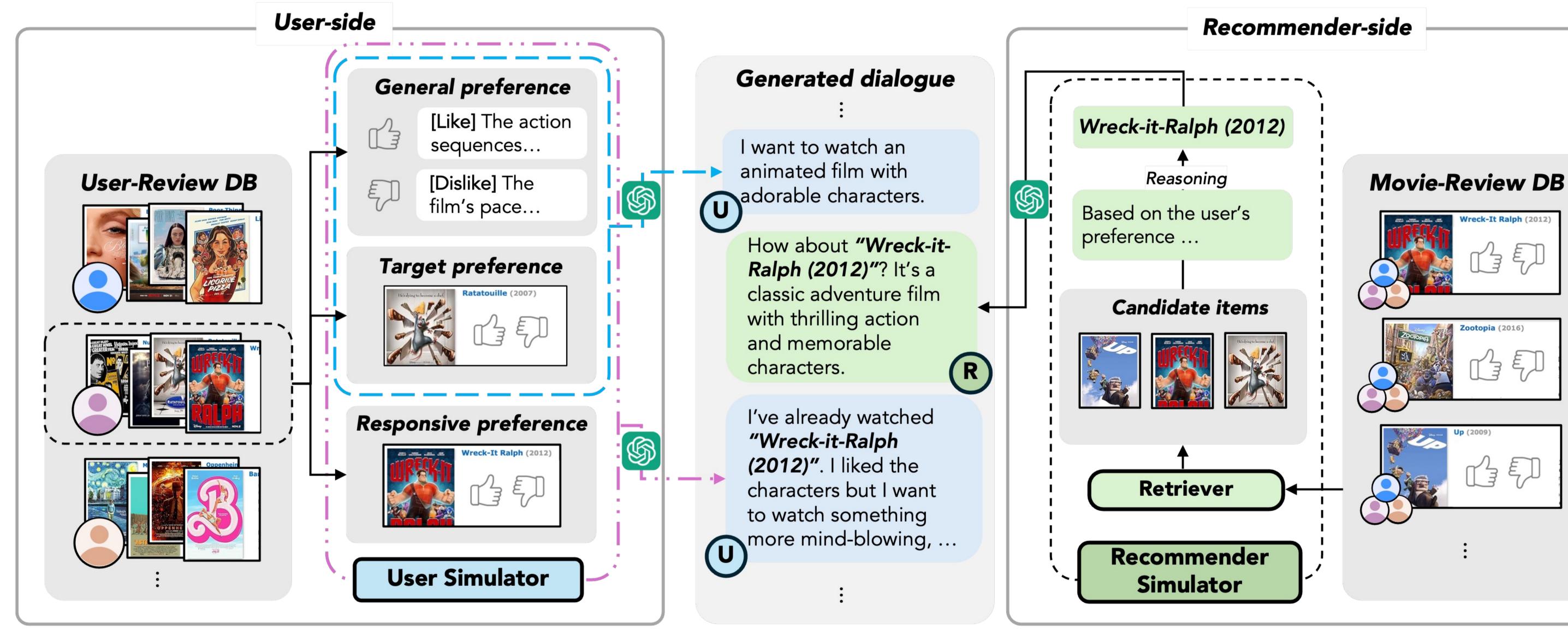


→ Not enough explanation

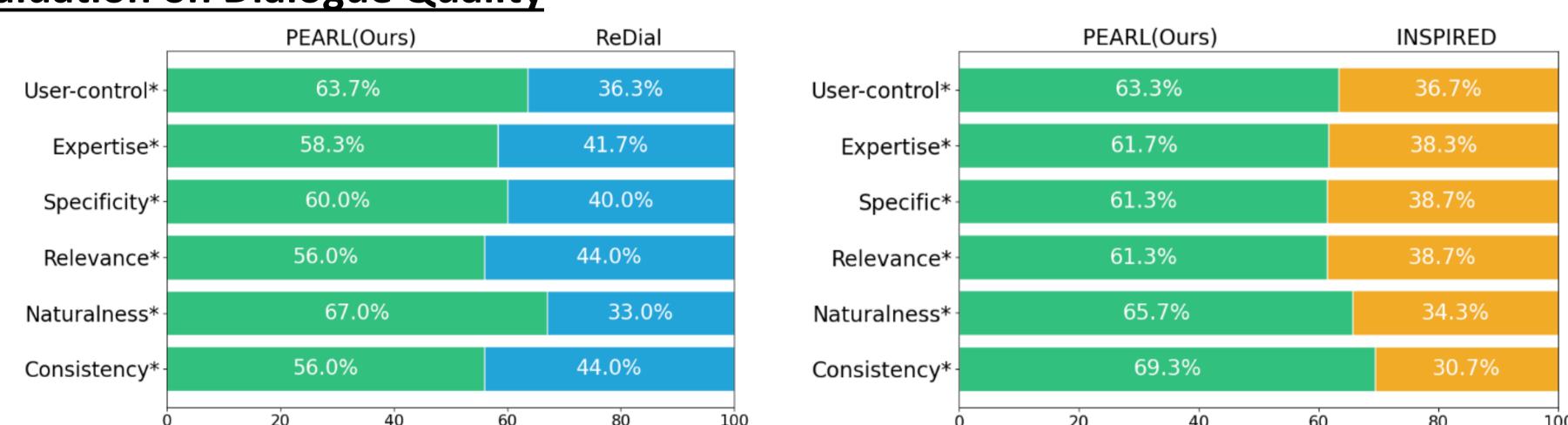
Knowledge Construction

Metadata (title, genre, director, cast) **Preferred & unpreferred attributes** (summarized from

PEARL Construction Overview



EXPERIMENTS



IMDb's top-featured reviews)

Evaluation on Dialogue Quality

	PEARL(Ours)	ReDial	1 -	PEARL(Ours)	INSPIRED
User-control*	63.7%	36.3%	User-control*	63.3%	36.7%
Expertise*	58.3%	41.7%	Expertise*	61.7%	38.3%
Specificity*	60.0%	40.0%	Specific*	61.3%	38.7%
Relevance*	56.0%	44.0%	Relevance*	61.3%	38.7%
Naturalness*	67.0%	33.0%	Naturalness*	65.7%	34.3%
Consistency*	56.0%	44.0%	Consistency*	69.3%	30.7%
(20 40 Win Ratio		00 0	20 40 60 Win Ratio	0 80 100

Dialogues of PEARL shows superiority in expertise, specificity, etc. over those of baseline

	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

PEARL is superior to previous datasets in scalability and specificity in user preferences.

Evaluation on Dialogue Utility

	BART-PEARL	BART-ReDial	
Fluency	52.7%	47.3%	
Fluency -	32.770	47.570	
Expertise*	61.3%	38.7%	
Explainability -	54.0%	46.0%	
Relevance*	59.7%	40.3%	
Naturalness*	58.3%	41.7%	
Overall -	52.7%	47.3%	
O.	20 40 Win Ra	60 80 100	

		Ó	20	40	60	80	100
				Win I	Ratio		
•	Models t	raine	d on P	PEARL g	enerat	e bette	er
	rochonco	c tha	n thac	o train	od on a	hacali	ina datacat
	response	5 illa	II UIOS	euam	eu on a	a Dasell	ine dataset.

Model	Dist-3	Dist-4
BART-ReDial BART-PEARL	0.6220 0.9241	0.5057 0.8861
UniCRS-ReDial UniCRS-PEARL	0.5413 0.9338	0.3667 0.9007
PECRS-ReDial PECRS-PEARL	0.6798 0.9132	0.5906 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 UniCRS-PEARL PECRS-PEARL	0.0018 0.0310 0.0151	0.0208 0.0697 0.0339	0.0736 0.1202 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.