Personalized LLM Decoding via Contrasting Personal Preference

EMNLP 2025 Suzhou, China | 中国苏州

YONSEI UNIVERSITY

Project Page Code

Hyungjune Bu[1], Chanjoo Jung[1], Minjae Kang[2], Jaehyung Kim[1]

[1]Yonsei University, [2]Opt-AI Inc.

Overview

Background & Motivation

- LLM personalization is essential for personal assistants.
- Prompt-based: simple but limited
- Training-based: effective but costly & forgetful.
- PEFT helps, but decoding-time personalization remains underexplored.

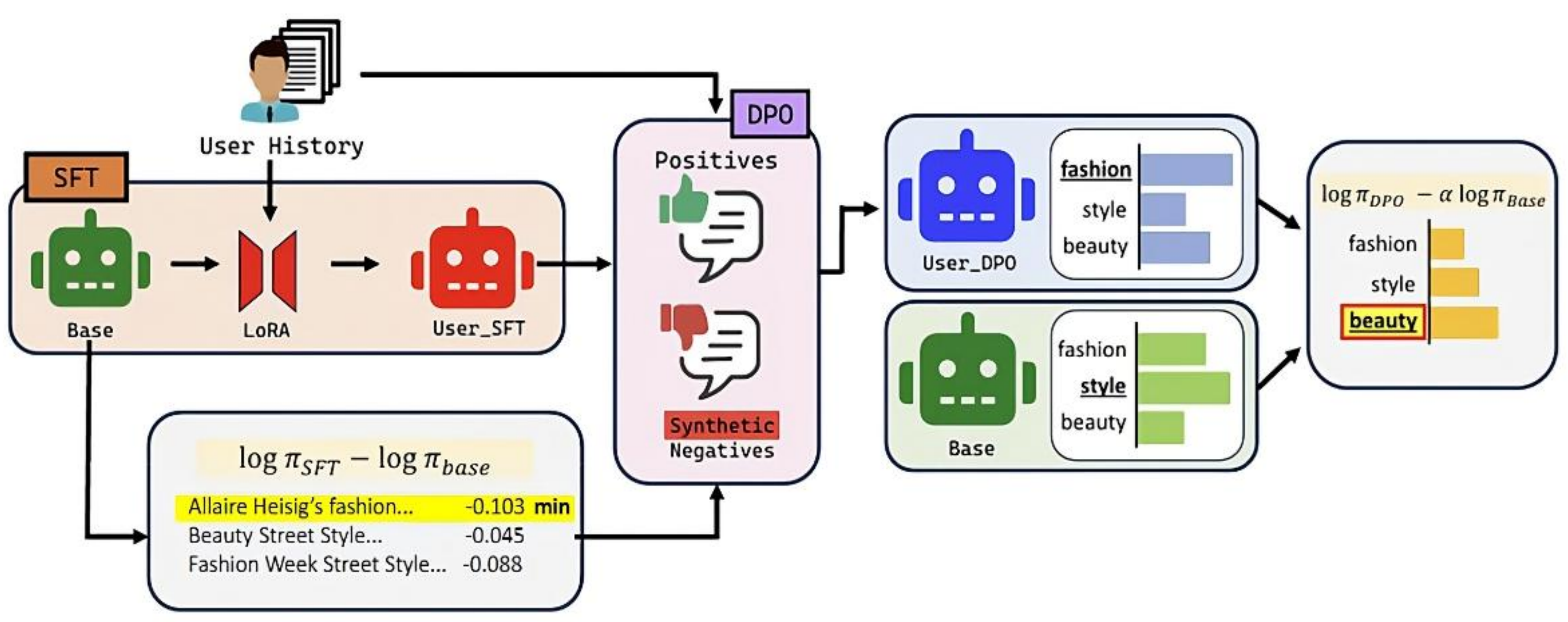
Our Approach: CoPe

- Reward-guided decoding after PEFT using implicit reward (personalized vs. base model).
- Enhanced with DPO and synthetic negatives.

Key Contributions

- First decoding-based LLM personalization without external reward models.
- **Unified pipeline** that integrates PEFT, synthetic negatives, and DPO to maximize implicit reward.
- Implicit reward maximization via contrastive decoding
- **Model-agnostic**, compatible with various LLMs (LLaMA, Gemma, Qwen).
- **Average gain** of +10.57% ROUGE-L across 5 personalized generation tasks from LaMP and LongLaMP benchmarks.

Method: CoPe



Task-adapted base model (TAM)

- Base LLM π_{base} is adapted to the target task via PEFT (LoRA).
- Output: Task-aware but **non-personalized** model.

User-specific personalization (OPPU)

- Apply LoRA fine-tuning on π_{base} with user history $H_{\texttt{user}}$

$$\pi_{\texttt{user}} = \pi_{\texttt{base}} + \Delta_{\texttt{user}}$$

- Output: Personalized model capturing the user's style and preferences.

Generate synthetic negatives

- Sample K outputs from the π_{base} for each input.
- Select the **lowest-reward** output:

$$\tilde{y}^{i,*} = \arg \min_{y \in \{\tilde{y}^{i,1}, \dots, \tilde{y}^{i,K}\}} \sum_t r_{\texttt{user}}(y_t),$$

where

$$r_{\texttt{user}}(y_t) = \log \frac{\pi_{\texttt{user}}(y_t \mid y_{<t})}{\pi_{\texttt{base}}(y_t \mid y_{<t})^{\alpha}},$$

Direct Preference Optimization (DPO)

- Train $\pi_{\texttt{user}}$ to prefer user-aligned y^{pos} over negative y^{neg}

$$\mathcal{L}_{\text{DPO}} = -\log \sigma\left(\beta \cdot [r_{\text{user}}(y^{pos}) - r_{\text{user}}(y^{neg})]\right)$$

Reward-guided decoding (CoPe)

- At inference, select next token maximizing $r_{\texttt{user}}$ among plausible candidates:

$$y_t^* = \arg \max_{y_t \in \mathcal{V}_{\text{head}}^t} r_{\texttt{user}}(y_t).$$

- Ensures outputs align with implicit user reward without external reward models.

Empirical Validations

Main results

Methods	Abstract Generation		Review Writing		Topic Writing		News Headline		Scholarly Title	
	ROUGE-1	ROUGE-L	ROUGE-1	ROUGE-L	ROUGE-1	ROUGE-L	ROUGE-1	ROUGE-L	ROUGE-1	ROUGE-L
Base	0.341	0.186	0.287	0.126	0.246	0.105	0.119	0.105	0.409	0.324
RAG	0.347	0.205	0.272	0.128	0.243	0.115	0.141	0.124	0.425	0.347
PAG	0.344	0.186	0.256	0.125	0.262	0.107	0.118	0.102	0.372	0.289
TAM	0.357	0.204	0.289	0.122	0.253	0.107	0.200	0.179	0.514	0.456
OPPU	0.378	0.218	0.319	0.134	0.278	0.112	0.203	0.182	0.510	0.454
CoPE (Ours)	**0.392**	**0.239**	**0.335**	**0.146**	**0.281**	**0.120**	**0.205**	**0.184**	**0.519**	**0.461**

Qualitative example

News Headline	
Query	Generate a headline for the following article: When we first saw Michael H. Rohde's photography series \"Below The Floor\" on Design-Milk.com, we were floored. (Pun intended
User Answer	Michael H. Rohde, German Photographer, Shoots Breathtaking Series 'Below The Floor' (PHOTOS)
TAM	Craft Of The Day: Create A Floating Photo Gallery With This DIY
OPPU	'Below The Floor' Photography Series Reveals The Hidden Beauty In The Places We Walk On
CoPe (ours)	'Below The Floor' Photography Series By Michael H. Rohde Is A Whole New Perspective On The World (PHOTOS)

Ablation study

	CD	DPO	Abstract Generation		News Headline	
			ROUGE-1	ROUGE-L	ROUGE-1	ROUGE-L
OPPU	✗	✗	0.378	0.218	0.203	0.181
	✓	✗	0.385	0.232	0.204	0.183
	✗	✓	0.386	0.230	0.203	0.182
CoPE (Ours)	✓	✓	0.392	0.239	0.205	0.184

Compatibility of CoPe

Methods	LLaMA 3.1-8B	Gemma 3-4B	Qwen 2.5-1.5B
Base	0.172	0.135	0.130
RAG	0.183	0.170	0.128
PAG	0.183	0.169	0.130
TAM	0.198	0.181	0.150
OPPU	0.202	0.194	0.163
CoPE (Ours)	**0.261**	**0.237**	**0.233**