Aligning LLM Agents by Learning Latent Preference from User Edits

https://github.com/gao-g/prelude

Ge Gao*, Alexey Taymanov*, Eduardo Salinas, Paul Mineiro, and Dipendra Misra

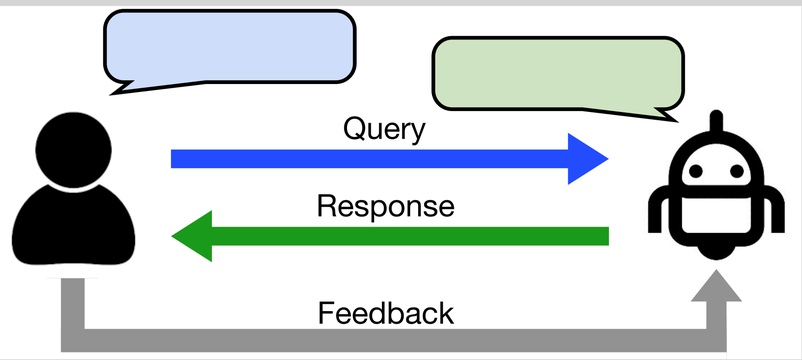
Microsoft Research

Overview

How to learn from user feedback in the form of edits to improve AI writing assistants?

We introduce:

- **PRELUDE** framework: formulates the interaction progress and preference learning as a cost minimization problem
- **CIPHER:** learns a prompt policy to infer a descriptive user preference



Learning Framework: PRELUDE

- **PRE**ference **L**earning from **U**ser's **D**irect **E**dits
- User directly makes edits to the agent response based on a latent preference
- Agent infers a user preference from interaction history, and uses it to generate a response
- Cost minimization to account for the amount of efforts spent by the user on making edits

Round t

1. User provides a context x_t to the LLM agent
2. LLM agent infers a preference f_t based on history given the context
   LLM agent generates a response y_t given the context
3. User edits the agent response, resulting in a revision y_t', according to a latent preference f_t^*

Agent incurs a cost $c_t = \Delta_{\text{edit}}(y_t, y_t')$

Minimize cumulative cost $\sum_{t=1}^{T} c_t$

Method: CIPHER

- **C**onsolidates **I**nduced **P**references based on **H**istorical **E**dits with **R**etrieval
- Leverages LLMs by prompting
- Infers a descriptive user preference from history with retrieval
- Manages a collection of context-dependent preference history
- Computationally efficient
- User friendly
- Interpretable

Round t

1. User provides a context x_t to the LLM agent

   Retrieve top-k examples from history based on $\phi(x_t)$
   Aggregate induced preferences in those retrieved examples $\{\tilde{f}_{z_i}\}_{i=1}^{k}$
2. LLM agent infers a preference f_t based on history given the context
   LLM agent generates a response y_t given the context and inferred preference
3. User edits the agent response, resulting in a revision y_t', according to a latent preference f_t^*

Agent incurs a cost $c_t = \Delta_{\text{edit}}(y_t, y_t')$

History $D \leftarrow D \cup \{(\phi(x_t), \tilde{f}_t)\}$

$c_t \geq \delta$: LLM agent generates a preference $\tilde{f}_t$ to explains the user edits

$c_t < \delta$: $\tilde{f}_t \leftarrow f_t$

Experiments & Analysis

- Writing task: Summarization
- Experiment: 200 rounds of interaction
- User setup:
  - GPT-4 user as a simulation
  - Provide different context documents
  - Show context-dependent preference
- CIPHER setup:
  - GPT-4 as the base LLM
  - MPNet as the context representation function
  - Top 5 retrieval with cosine similarity

Use Case	Latent User Preference
Introduce a political news to kids	targeted to young children, storytelling, short sentences, playful language, interactive, positive
Promote a paper to invoke more attention	tweet style, simple English, inquisitive, skillful foreshadowing, with emojis
Take notes for knowledge from Wikipedia	bullet points, parallel structure, brief
Use online stories for inspirations in writing	second person narrative, brief, show emotions, invoke personal reflection, immersive
Extract main opinions from a move review	question answering style