FedConPE: Efficient Federated Conversational Bandits with Heterogeneous Clients

Zhuohua Li[†], Maoli Liu[†], and John C.S. Lui[*]
The Chinese University of Hong Kong

Introduction

- Recommender systems need to learn and adapt to users' preferences by collecting their feedback, such as click rates.
- Recommender systems can also proactively query users with some questions to obtain additional feedback.

Motivating Example and Main Ideas

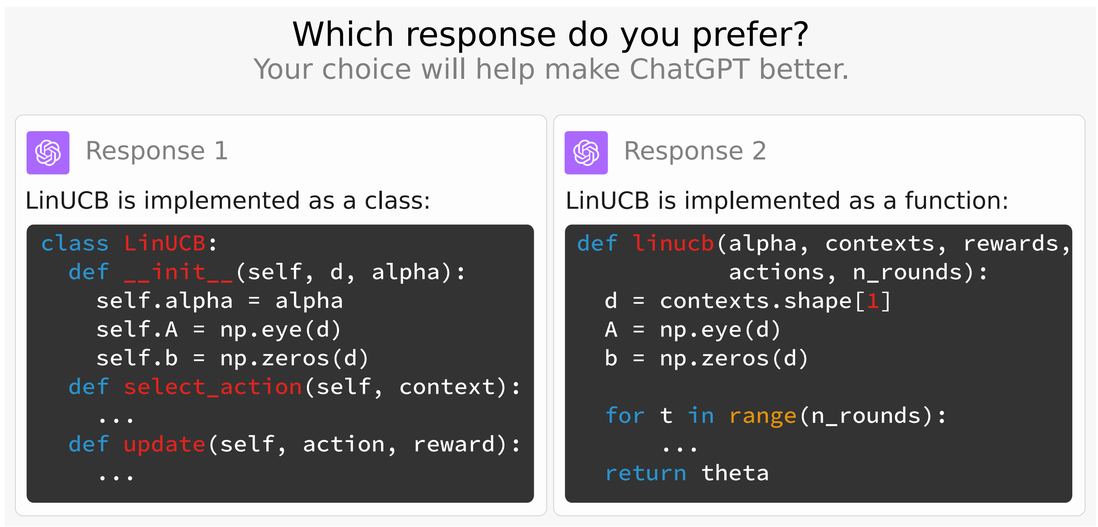


Fig. 1: An example of conversational recommendation.

Main Idea: Query users with questions (i.e., key terms) that are closest to the direction with the maximum uncertainty.

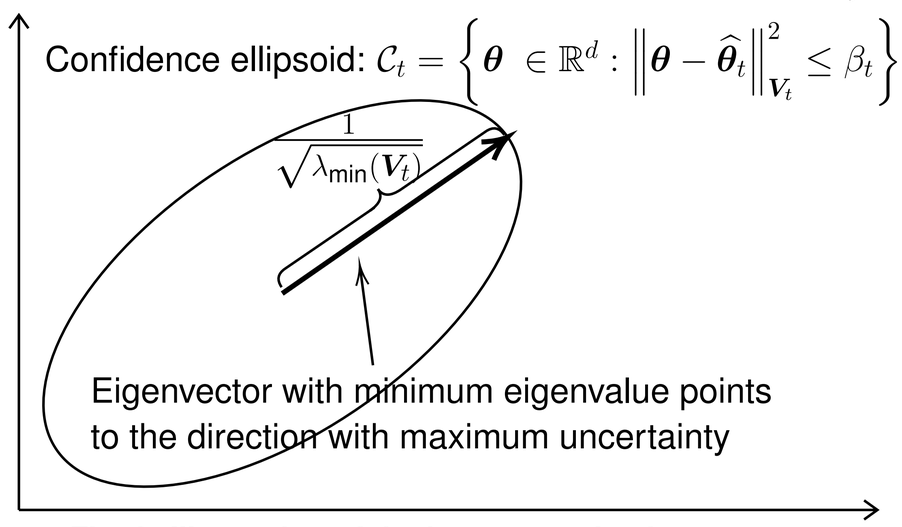


Fig. 2: Illustration of the key term selection strategy.

Contributions

- Propose FedConPE, an algorithm for federated conversational bandits with heterogeneous arm sets.
- Exploit the properties of the conversational setting to enhance efficiency in both computation and communication.
- Theoretical results show that FedConPE is near-optimal.
- Experiments demonstrate that FedConPE achieves lower cumulative regret and uses fewer conversations.

Problem Formulation

- M clients and one central server.
- Each client i has a finite set of recommended items (i.e., arms) $\mathcal{A}_i \subset \mathbb{R}^d$ and $|\mathcal{A}_i| = K$.
- The server has a finite set of questions (i.e., key terms) $\mathcal{K} \subset \mathbb{R}^d$, and we assume that $\mathcal{K}$ is sufficiently rich.

Interaction Protocol

- For each time step $t = 1, 2, \ldots, T$:
  - Each client i chooses an arm $\boldsymbol{a}_{i,t} \in \mathcal{A}_i$, and observes an arm-level reward $x_{i,t} = \boldsymbol{a}_{i,t}^\mathsf{T}\boldsymbol{\theta}^* + \eta_{i,t}$.
  - The server decides whether to initiate queries $\boldsymbol{k}_{i,t} \in \mathcal{K}$ to each client i. If so, the client observes a key term-level reward $\widetilde{x}_{i,t} = \boldsymbol{k}_{i,t}^\mathsf{T}\widetilde{\boldsymbol{\theta}}^* + \widetilde{\eta}_{i,t}$.
- The objective is to minimize the cumulative regret:

$$R_M(T) = \sum_{i=1}^{M}\sum_{t=1}^{T}\left(\max_{\boldsymbol{a}\in\mathcal{A}_i} \boldsymbol{a}^\mathsf{T}\boldsymbol{\theta}^* - \boldsymbol{a}_{i,t}^\mathsf{T}\boldsymbol{\theta}^*\right),$$

while using as few queries as possible.

Theoretical Results

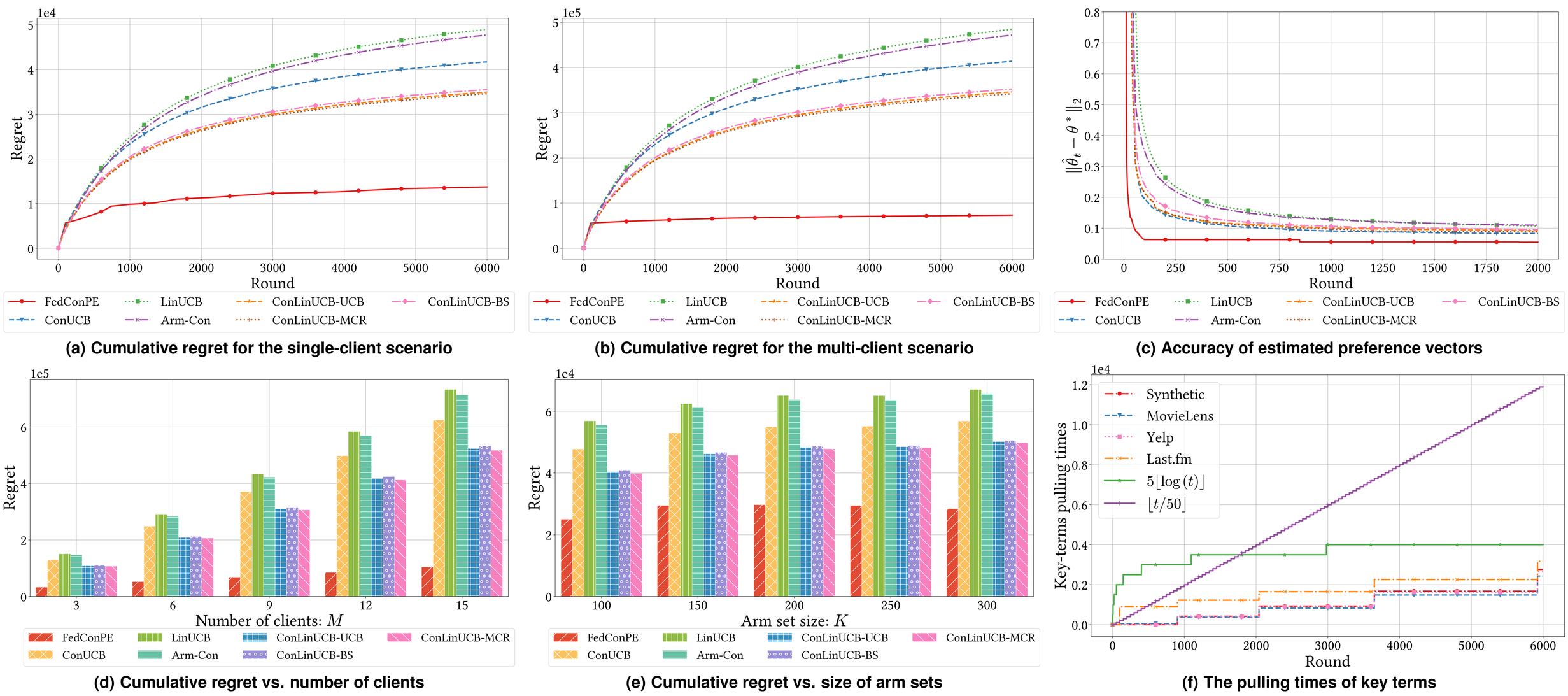
Theorem 1 (Regret upper bound). *With probability at least $1-\delta$, the cumulative regret scales as $\mathcal{O}\left(\sqrt{dMT\log\frac{KM\log T}{\delta}}\right)$.*

Theorem 2 (Regret lower bound). *For any policy that chooses at most one key term at each time step, there exists an instance of the federated conversational bandits such that the policy suffers an expected regret of at least $\Omega\left(\sqrt{dMT}\right)$.*

Theorem 3 (Communication cost). *The number of scalars transmitted between the server and clients scales as $\mathcal{O}\left(d^2M\log T\right)$.*

Theorem 4 (Conversation frequency upper bound). *The fraction of conversations relative to the total time horizon is at most $\frac{1}{NC^2}$.*

Evaluations



(a) Cumulative regret for the single-client scenario

(b) Cumulative regret for the multi-client scenario

(c) Accuracy of estimated preference vectors

(d) Cumulative regret vs. number of clients

(e) Cumulative regret vs. size of arm sets

(f) The pulling times of key terms

Fig. 3: Performance evaluation of FedConPE and comparison with baseline algorithms.

† Zhuohua Li and Maoli Liu contributed equally to this work. ∗ The work of John C.S. Lui is supported in part by the RGC SRFS2122-4S02.