THE WEB CONFERENCE

Local Clustering in Contextual Multi-Armed Bandits

Yikun Ban and Jingrui He
University of Illinois at Urbana-Champaign
{yikunb2, jingrui}@illinois.edu
www.banyikun.com, www.hejingrui.org

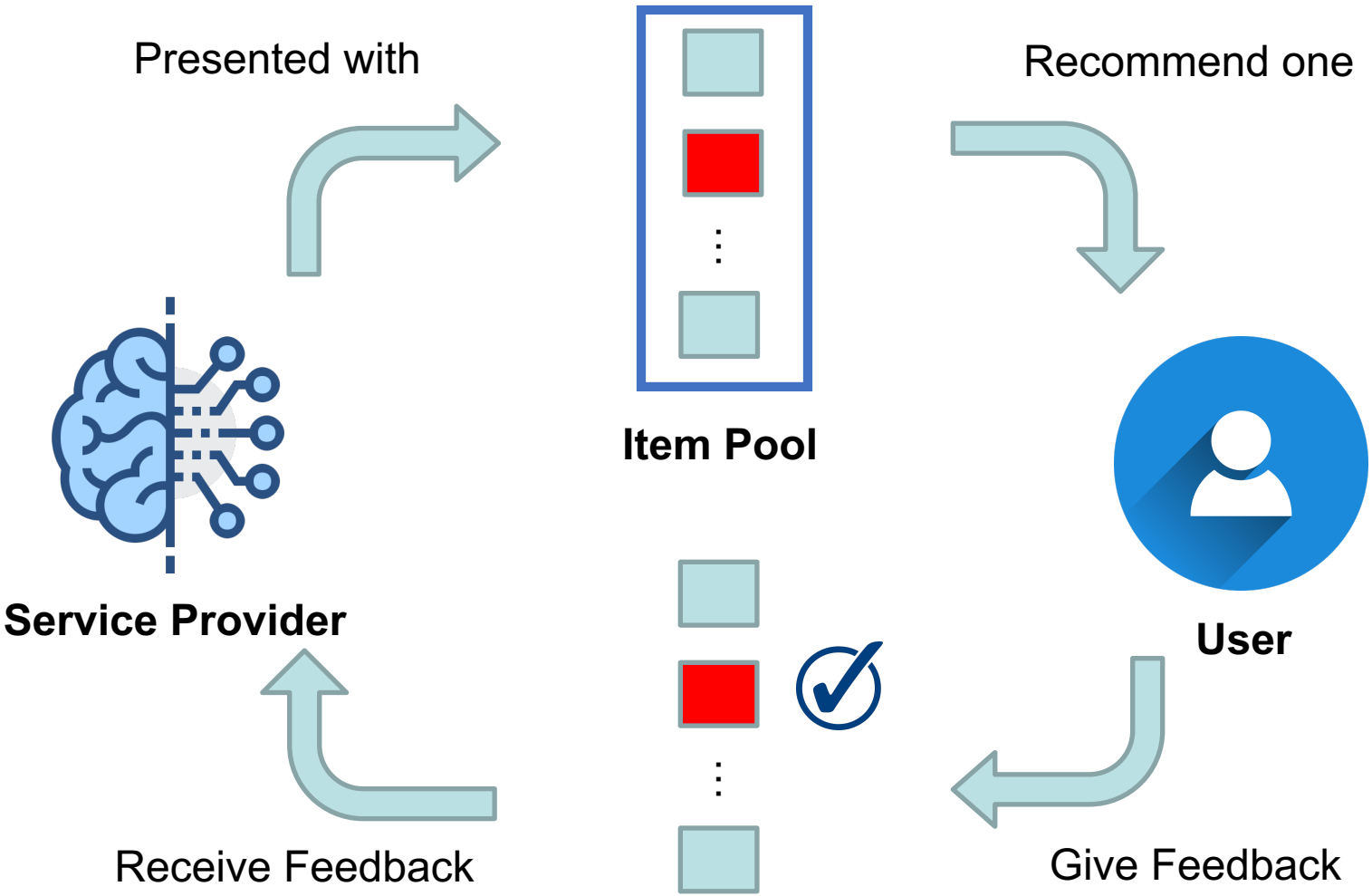
ILLINOIS

Roadmap

- **Background and Motivation** 
- Problem Definition
- Proposed Method
- Theoretical Analysis
- Empirical Performance

Personalized Recommendation

One round: One Recommendation



Personalized Recommendation

- Conventional approaches, e.g., **collaborative and content-based filtering**

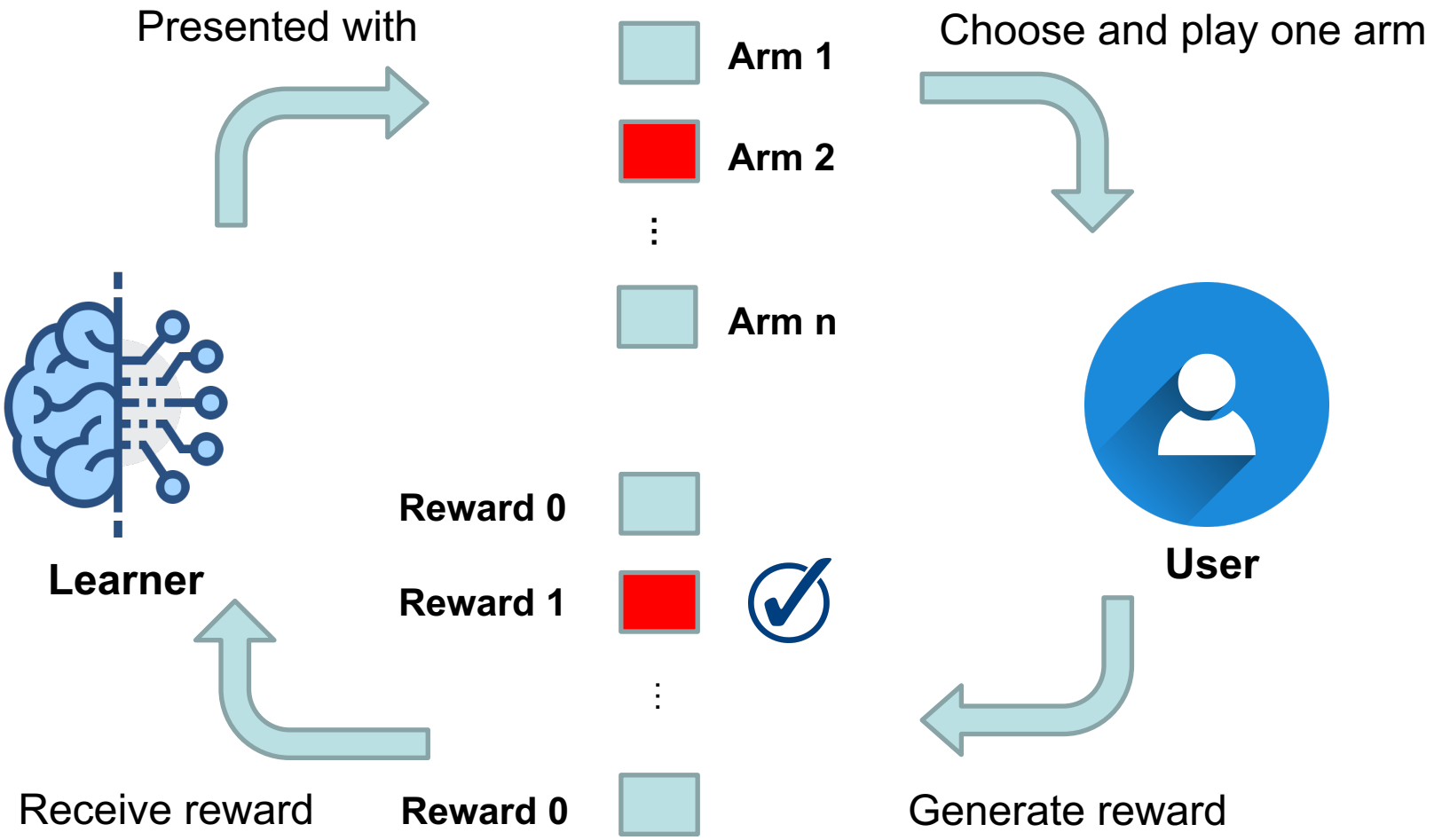
(InCube Group)

Challenges

- **Cold-start** problem (Lack of history data);
- **Rapid change** of recommendation content (New items or users).

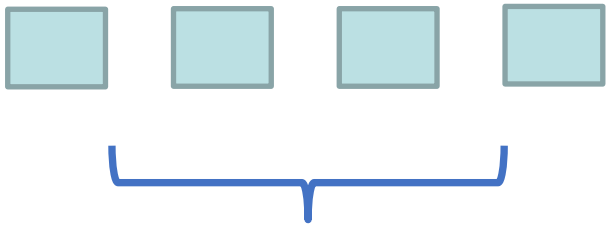
Multi-armed Bandit (MAB)

- Effective tool for **online-decision making.**

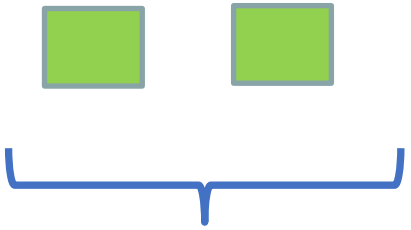


Exploitation VS Exploration

User



Items with historical records

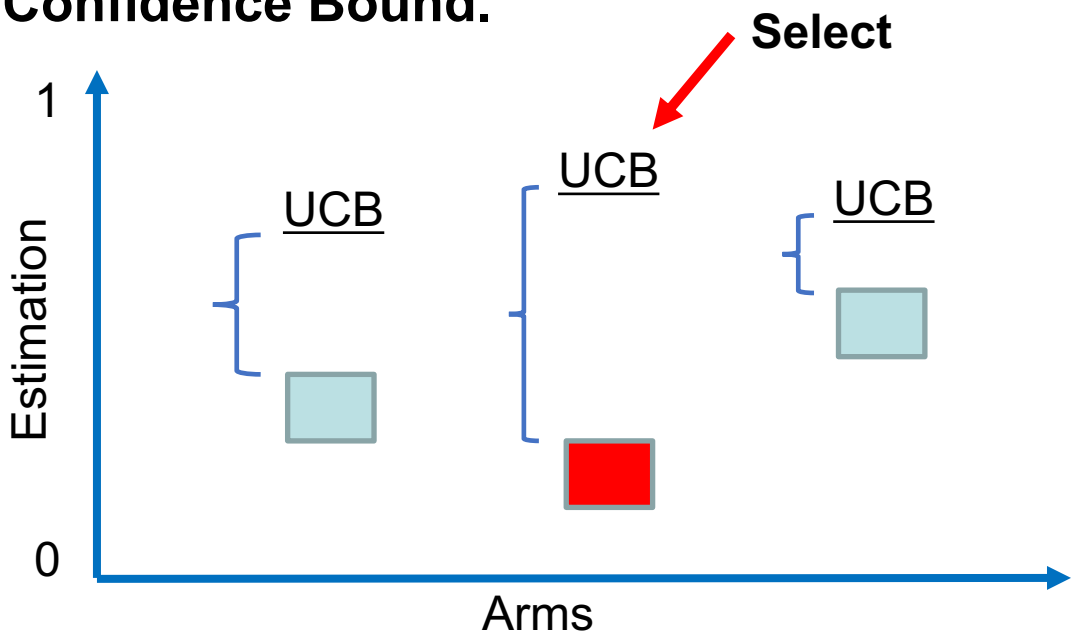


New items

Exploitation **VS** **Exploration**

State-of-the-art in MAB

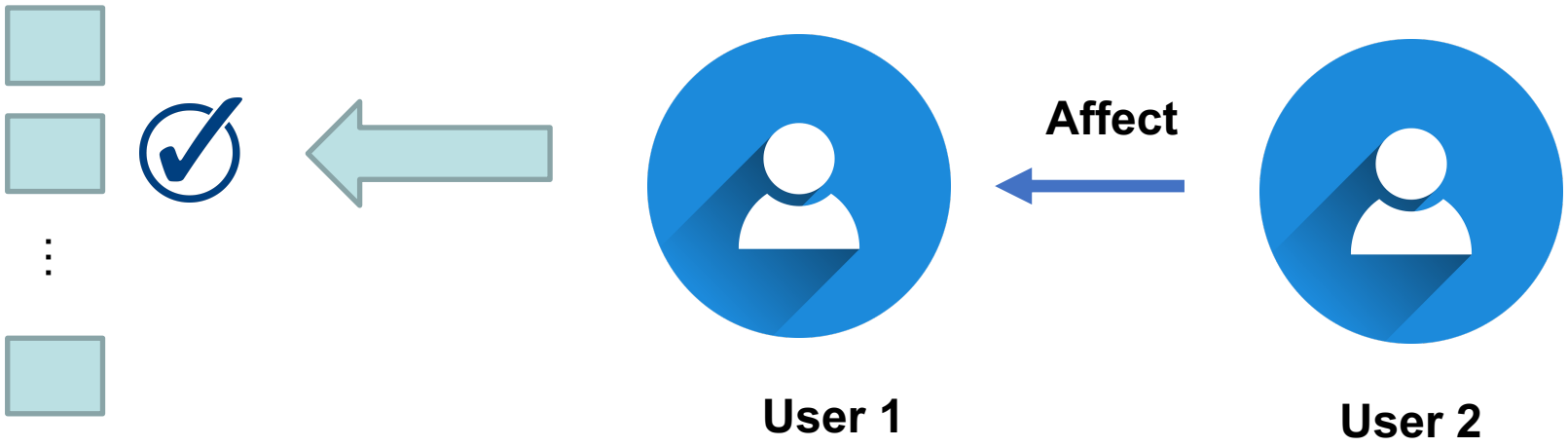
- **Existing algorithms**: ϵ-greedy [3], UCB [1], Tampson Sampling [2], etc.
- **UCB**:
  - Choose the arm with the maximal sum of **estimated reward** and **Upper Confidence Bound.**



[1] Li, L., Chu, W., Langford, J., & Schapire, R. E. (2010, April). A contextual-bandit approach to personalized news article recommendation. *WWW* (pp. 661-670).
[2] Gopalan, A., Mannor, S., & Mansour, Y. (2014, January). Thompson sampling for complex online problems. Icml (pp. 100-108). PMLR.
[3] Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine learning*, *47*(2), 235-256.

Motivation

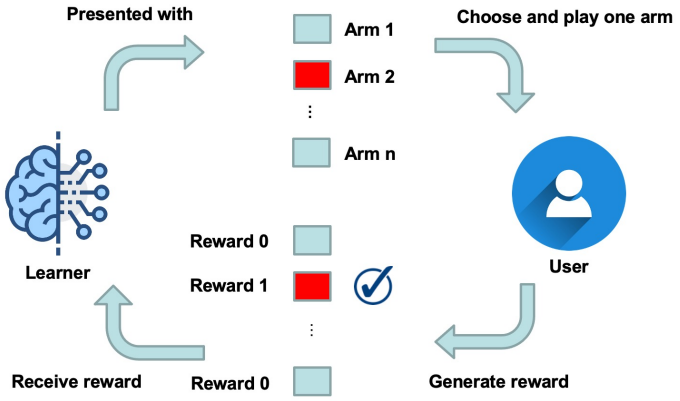
❑ **One user's decision is affected by other users.**



❑ **Utilizing the mutual influence** can improve quality of recommendation.

Motivation

- However, standard MAB algorithms view each user as an individual and do not consider **user dependency**.



- To further improve the recommendation:
  - Objective #1: **Identify user clusters** in MAB;
  - Objective #2: **Exploit the user clusters** to improve the recommendation.

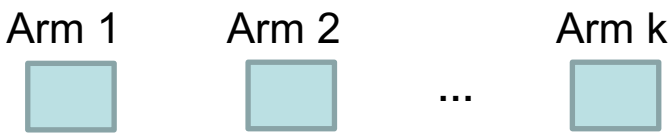
Roadmap

- Background and Motivation
- **Problem Definition** 
- Proposed Method
- Theoretical Analysis
- Empirical Performance

Problem Definition

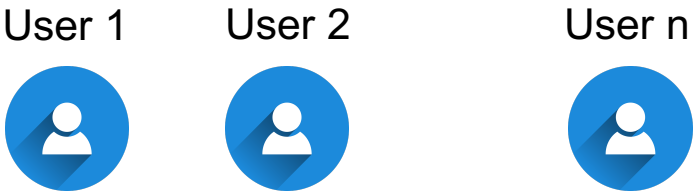
In round t:

- **Arm Pool**:



$\boldsymbol{x_{1,t}}$ $\boldsymbol{x_{2,t}}$ $\boldsymbol{x_{k,t}}$ d-dimensional feature vector

- **User Pool**:



$\boldsymbol{\theta_1}$ $\boldsymbol{\theta_2}$ $\boldsymbol{\theta_n}$ d-dimensional vector (**Unknown**)

- **Reward**:

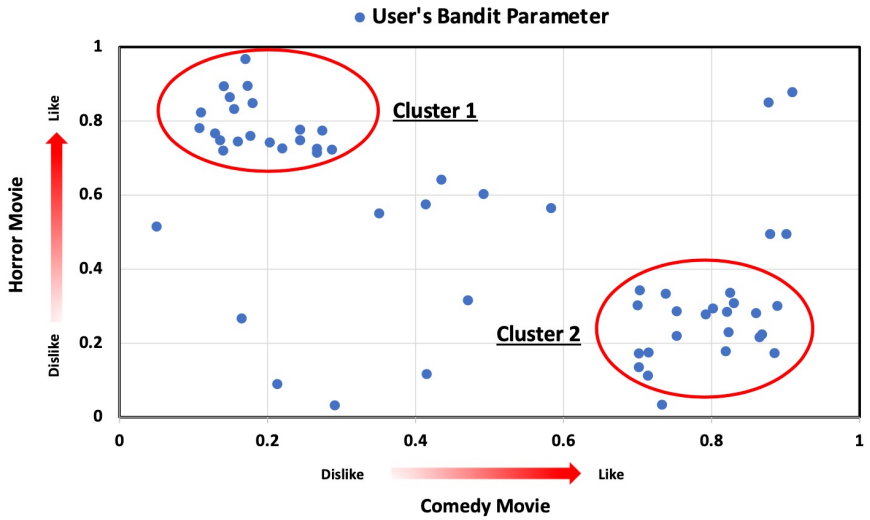
Noise

$$r_t = \boldsymbol{\theta}_{i_t}^{T} \boldsymbol{x}_{a,t} + \eta_t$$

Linear reward function

Problem Definition

- **Related works** [1,2]: Consider the cluster as a set of users with the **same behavior** (θ). However,



- Therefore, we define a cluster as a set of users with **similar behavior**.

DEFINITION 3.1 (γ-CLUSTER). *Given a subset of users $\mathcal{N} \subseteq N$ and a threshold $\gamma > 0$, $\mathcal{N}$ is considered as a γ-Cluster if it satisfies*

$$\forall i, j \in \mathcal{N}, \|\theta_i - \theta_j\| < \gamma.$$

[1] Claudio Gentile, Shuai Li, and Giovanni Zappella. 2014. Online clustering of bandits. *ICML*. 757–765.
[2] Claudio Gentile, Shuai Li, Purushottam Kar, Alexandros Karatzoglou, Giovanni Zappella, and Evans Etrue. 2017. On context-dependent clustering of bandits. *JMLR*. org, 1253–1262.

Problem Definition

- **Objective #1**: Identify clusters among users, such that the clusters returned by the proposed algorithm are true γ–clusters with probability at least $1 - \delta$.

- **Objective #2**: Leverage user clusters to improve the quality of recommendation, evaluated by **Regret**.

$$\mathbf{R}_T = \mathbb{E}[\sum_{t=1}^{T} R_t] = \sum_{t=1}^{T} (\boldsymbol{\theta}_{i_t}^{\mathsf{T}} \mathbf{x}_t^* - \boldsymbol{\theta}_{i_t}^{\mathsf{T}} \mathbf{x}_t)$$

Optimal Reward

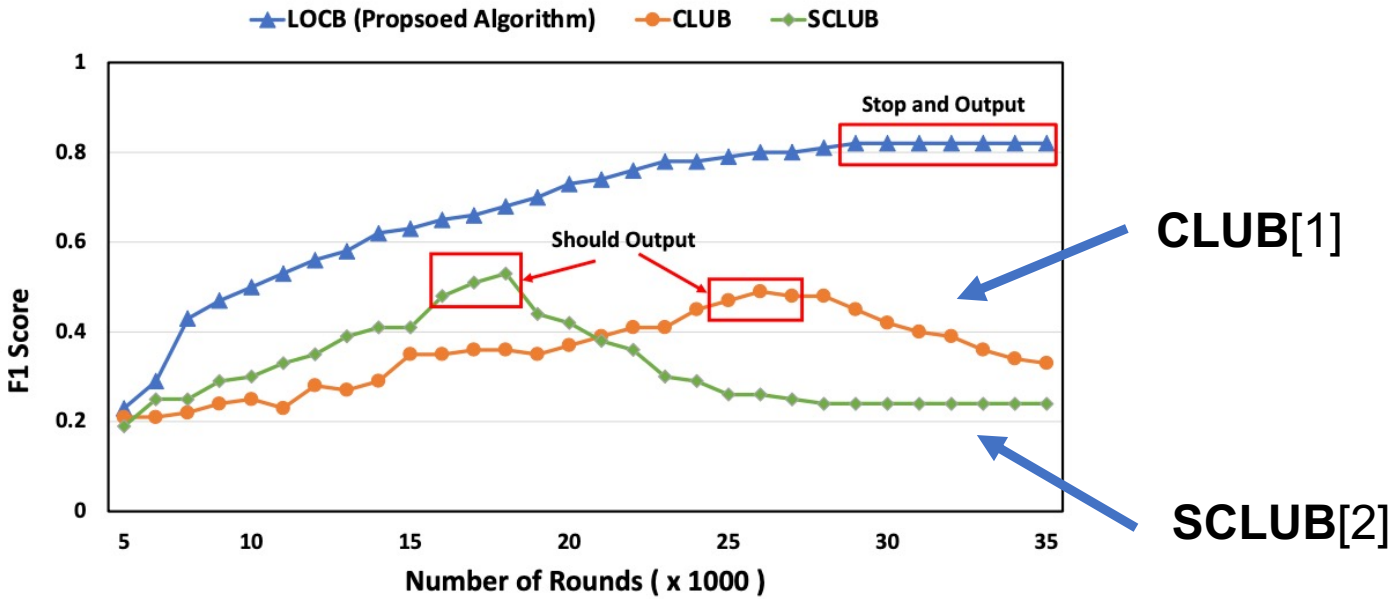
Received Reward

Roadmap

- Background and Motivation
- Problem Definition
- **Proposed Method** 
- Theoretical Analysis
- Empirical Performance

Related Works

- Existing works [1,2] use the top-down hierarchical clustering procedure (**global**).
  - **Limitation #1**: Do not know when to output high-quality clusters.



  - **Limitation #2**: High computational cost.

[1] Claudio Gentile, Shuai Li, and Giovanni Zappella. 2014. Online clustering of bandits. *ICML*. 757–765.
[2] Claudio Gentile, Shuai Li, Purushottam Kar, Alexandros Karatzoglou, Giovanni Zappella, and Evans Etrue. 2017. On context-dependent clustering of bandits. *JMLR*. org, 1253–1262.

Proposed Algorithm: LOCB

- Proposed algorithm: **LOCB**, a novel bandit algorithm embedded with a **local-clustering** procedure.
  - **Clustering module**: Identify K clusters, given K seeds in each round.
  - **Pulling module**: Select an arm based on the clusters provided by Clustering module in each round.

LOCB: Clustering Module

- **Estimated user parameter**: In each round t, LOCB needs to compute the estimation $\widehat{\boldsymbol{\theta}}_{i,t}$ for the user parameter $\boldsymbol{\theta}_i$, based on ridge regression:

$$\hat{\theta}_{i_t,t} = \mathbf{A}_{i_t,t}{}^{-1}\mathbf{b}_{i_t,t}, \quad \mathbf{A}_{i_t,t} = \mathbf{I} + \sum_{t'=1}^{t} \mathbf{x}_{t'}\mathbf{x}_{t'}^{\mathsf{T}}, \quad \mathbf{b}_{i_t,t} = \sum_{t'=1}^{t} \mathbf{x}_{t'} r_{t'},$$

- **Upper Confidence Bound**: Build a confidence interval for $\widehat{\boldsymbol{\theta}}_{i,t}$, for each user, defined as:

$$\mathbb{P}\left(\forall t \in [T], \|\hat{\boldsymbol{\theta}}_{i,t} - \boldsymbol{\theta}_i\| > B_{\boldsymbol{\theta},i}(m_{i,t}, \delta')\right) < \delta',$$

Upper Confidence Bound

LOCB: Clustering Module

- **Neighbor**: Two users are neighbors if they belong to the same γ-cluster.
- **Seed selection**: Randomly choose K users.
- **Potential neighbor**: A user i is considered as as the potential neighbor of seed user s, when:

$$\|\hat{\theta}_{i,t} - \hat{\theta}_{s,t}\| \leq B_{\theta,i}(m_{i,t}, \delta') + B_{\theta,s}(m_{s,t}, \delta').$$

Seed user parameter

- **Cluster**: Seed user + Its potential neighbors.

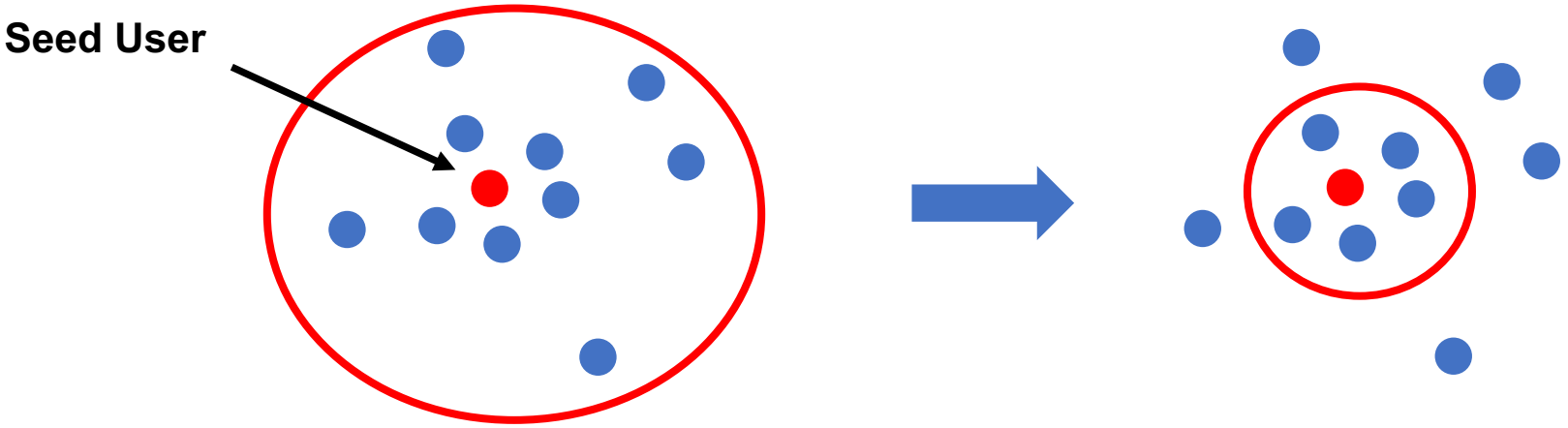
LOCB: Clustering Module

- **Evolving of potential neighbor**:

$$\|\hat{\theta}_{i,t} - \hat{\theta}_{s,t}\| \leq B_{\theta,i}(m_{i,t}, \delta') + B_{\theta,s}(m_{s,t}, \delta').$$

UCB is shrinking as more and more rounds are played for this user.

- **Evolving of cluster**:



LOCB: Clustering Module

- **Termination Status**: Given a cluster $\mathcal{N}_{s,t}$, Clustering module outputs this cluster when

$$\sup\{B_{\theta,i}(m_{i,t},\delta') : i \in \mathcal{N}_{s,t}\} < \frac{\gamma}{8}$$

procedure CLUSTERING(i_t, S)
  $\hat{\theta}_{i_t,t} \leftarrow \mathbf{A}_{i_t,t}^{-1}\mathbf{b}_{i_t,t}$ ← **Serving user**
  for each $s \in S$ **do**
    $\hat{\theta}_{s,t} \leftarrow \mathbf{A}_{s,t}^{-1}\mathbf{b}_{s,t}$ ← **Seed user**
    if $\|\hat{\theta}_{i_t,t} - \hat{\theta}_{s,t}\| > B_{\theta,i_t}(m_{i_t,t},\delta') + B_{\theta,s}(m_{s,t},\delta')$ **then**
      $\mathcal{N}_{s,t} \leftarrow \mathcal{N}_{s,t-1} - \{i_t\}$ # Remove i_t from s's potential neighbors
    else
      $\mathcal{N}_{s,t} \leftarrow \mathcal{N}_{s,t-1} \cup \{i_t\}$ # To ensure i_t is in $\mathcal{N}_{s,t}$
    if $\sup\{B_{\theta,i}(m_{i,t},\delta') : i \in \mathcal{N}_{s,t}\} < \frac{\gamma}{8} \cdot \tau$ **then**
      $S \leftarrow S - \{s\}$ # $\mathcal{N}_{s,t}$ is ready to return and remove s from S

Update membership for each cluster

Terminate and output the cluster

- Input K seeds, output K (**overlapped**) clusters.

LOCB: Pulling Module

- **Individual UCB**: For each user, build a confidence interval for the estimated reward $\hat{\theta}_{i,t}^{\mathsf{T}}\mathbf{x}_{a,t}$ of $\theta_i^{\mathsf{T}}\mathbf{x}_{a,t}$, defined as:

$$\mathbb{P}\left(\forall t \in [T], |\hat{\theta}_{i,t}^{\mathsf{T}}\mathbf{x}_{a,t} - \theta_i^{\mathsf{T}}\mathbf{x}_{a,t}| > CB_{r,i}\right) < \delta',$$

UCB

- **Cluster parameter:** Given a cluster $\mathcal{N}_{s,t}$, we define its cluster parameter as:

$$\theta_{\mathcal{N}_{s,t}} = \frac{1}{|\mathcal{N}_{s,t}|}\sum_{i\in\mathcal{N}_{s,t}}\theta_i.$$

Represent the behavior of this cluster

estimation

$$\hat{\theta}_{\mathcal{N}_{s,t}} = \frac{1}{|\mathcal{N}_{s,t}|}\sum_{i\in\mathcal{N}_{s,t}}\hat{\theta}_{i,t}.$$

LOCB: Pulling Module

- **Cluster UCB**: For each cluster, build a confidence interval for $\hat{\theta}_{\mathcal{N}_{s,t}}^{\mathsf{T}} \mathbf{x}_{a,t}$ such that

$$\mathbb{P}\left(\forall t \in [T], |\hat{\boldsymbol{\theta}}_{\mathcal{N}_{s,t}}^{\mathsf{T}} \mathbf{x}_{a,t} - \boldsymbol{\theta}_{\mathcal{N}_{s,t}}^{\mathsf{T}} \mathbf{x}_{a,t}| > CB_{r,\mathcal{N}_{s,t}}\right) < \delta',$$

UCB

- Inspired by UCB algorithm, Pulling module selects one arm by:

$$\mathbf{x}_t = \arg \max_{\mathbf{x}_{a,t} \in \mathbf{X}_t} \hat{\boldsymbol{\theta}}_{\mathcal{N}_{s,t}}^{\mathsf{T}} \mathbf{x}_{a,t} + CB_{r,\mathcal{N}_{s,t}}.$$

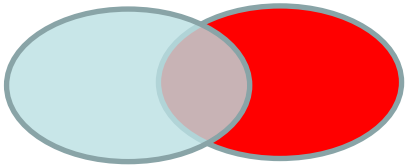
Cluster behavior

Instead of using $\hat{\theta}_{i,t}^{\mathsf{T}} \mathbf{x}_{a,t}$

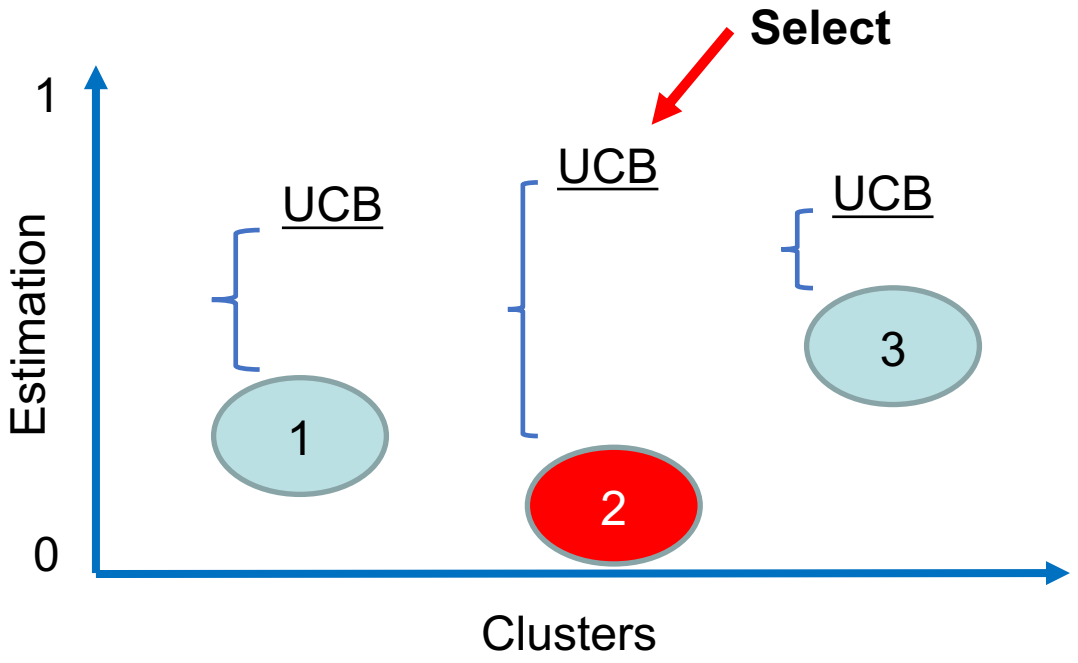
Personal behavior

LOCB: Pulling Module

- However, a user may belong to multiple clusters (overlapped), which is usually the case in real-world problems.

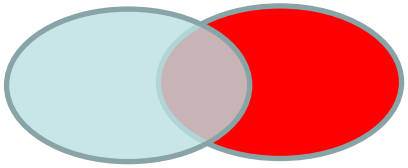


- **Cluster selection**: For a serving user, given multiple clusters returned by Clustering module, Pulling module chooses the cluster by:



LOCB: Pulling Module

- However, a user may belong to multiple clusters (overlapped), which is usually the case in real-world problems.



- **Cluster selection**: For a serving user, given multiple clusters returned by Clustering module, Pulling module chooses the cluster by:

$$\mathbf{x}_t = \arg \max_{\mathbf{x}_{a,t} \in \mathbf{X}_t} \max_{s \in S_t(i_t)} \left(\hat{\theta}^{\mathsf{T}}_{\mathcal{N}_{s,t}} \mathbf{x}_{a,t} + CB_{r,\mathcal{N}_{s,t}} \right).$$

Arm set Cluster set

  - **Choose the best arm with the best cluster.**

LOCB: Workflow

- In round t = 1, 2, …, T,
  - Receive the serving user i and a set of arms;
  - Obtain the clusters containing i, returned by Clustering module;
  - Pulling module selects the best arm with the best cluster
  - Receive reward and update parameters;
  - Clustering module updates each cluster;

LOCB: Seed Selection

- **Challenge**: The number of seeds affects the performance of LOCB.
- **Intuition**: The more seeds LOCB is given, the higher chance it has to find good clusters.
- **Solution**: It is encouraged to **choose a large number of seeds** (Simply, set each user as a seed).
  - Clustering module finds overlapped clusters.
  - Pulling module chooses the best cluster among candidate clusters.

Roadmap

- Background and Motivation
- Problem Definition
- Proposed Method
- **Theoretical Analysis** 
- Empirical Performance

Theoretical Analysis

- **Correctness:** Each cluster returned by LOCB is a true γ-cluster with probability $1 - \delta$.

THEOREM 5.1 (CORRECTNESS). *Given a threshold γ and a set of seeds $S \subseteq N$, for each $s \in S$, let $\mathcal{N}_s$ represent the cluster output by LOCB with respect to s. The terminate criterion of Clustering module is defined as:*

$$\sup\{B_{\boldsymbol{\theta},i}(m_{i,t}, \delta') : i \in \mathcal{N}_{s,t}\} < \frac{\gamma}{8}.$$

Then, with probability at least $1 - \delta$, after the Clustering module terminates, for each $s \in S$, it has

$$\forall i, j \in \mathcal{N}_s, \|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\| < \gamma.$$

- With this guarantee, the learner is confidence about quality of each overlapped cluster, in order to make item recommendation or friend recommendation.

Theoretical Analysis

- **Efficiency**: The number of rounds for LOCB to return the clusters is upper bounded by $O(n \log n)$.

THEOREM 5.2. *Suppose each user is evenly served and* $m_{i,t} \geq \frac{2\times 32^2}{\lambda^2} \log\left(\frac{2nd}{\delta'}\right) \log\left(\frac{32^2}{\lambda^2} \log\left(\frac{2nd}{\delta'}\right)\right)$ *for any* $i \in N$. *Then, with probability at least* $1-\delta$, *the number of rounds* $\hat{T}$ *needed for the Clustering module to terminate is upper bounded by*

$$\hat{T} < \frac{2nd}{C} \log \frac{nd}{C} + \frac{2n}{C}\left(\log\left(\frac{2^{(d+1)}n}{\delta}\right) - \frac{\gamma^2 - 256}{512\sigma^2}\right) + n.$$

where $C = \frac{\lambda\gamma^2}{16^3\sigma^2}$.

Theoretical Analysis

- **Effectiveness**: The regret of LOCB is upper bounded by

$$O(\sqrt{T}) + \gamma O(T - n \log n)$$

Similar to linear bandit

Caused by the deviation between cluster center and serving user

THEOREM 5.3. *Suppose that each user is evenly served. Given γ and a set of seeds S, after $T > \hat{T}$ rounds, the accumulated regret of LOCB can be upper bounded as follows:*

$$\mathbf{R}_T \leq \left[\sqrt{T} \cdot \sqrt{2\log(1+T)} + O\left(nd \log nd\right)\right] \cdot O\left(\sqrt{d \log\left(Tn/\delta\right)}\right) + \left(T - O\left(nd \log nd\right)\right)\gamma.$$

Roadmap

- Background and Motivation
- Problem Definition
- Proposed Method
- Theoretical Analysis
- **Empirical Performance** 

Experiments (1) : Clustering Accuracy

- **Data sets**: Synthetic, Yelp, MovieLens, and Yahoo.
- **Baselines**:
  - (1) N-CLUB [1]: Terminates it when clusters have not changed in the last consecutive $10/\delta$ rounds.
  - (2) ST-CLUB [1]: Same termination status with LOCB
  - (3) ST-SCLUB [2]: Same termination status with LOCB
  - (4) N-LCOB: LOCB terminates when clusters have not changed in the last consecutive $10/\delta$ rounds.
- **Metric**: F1 score.

[1] Claudio Gentile, Shuai Li, and Giovanni Zappella. 2014. Online clustering of bandits. *ICML*. 757–765.
[2] Li, S., Chen, W., & Leung, K. S. (2019). Improved algorithm on online clustering of bandits. AAAI.

Experiments (1): Clustering Accuracy

	Synthetic			Yelp		
	F1	Pre	Recall	F1	Pre	Recall
N-CLUB	0.390	0.246	0.943	0.484	0.334	0.884
ST-CLUB	0.578	0.549	0.612	0.626	0.593	0.663
ST-SCLUB	0.714	0.745	0.687	0.768	0.863	0.693
N-LOCB	0.662	0.618	0.714	0.675	0.620	0.743
LOCB	**0.880**	**0.913**	**0.856**	**0.879**	**0.908**	**0.853**

	MovieLens			Yahoo		
	F1	Pre	Recall	F1	Pre	Recall
N-CLUB	0.417	0.286	0.773	0.454	0.334	0.709
ST-CLUB	0.520	0.429	0.663	0.528	0.385	0.841
ST-SCLUB	0.538	0.739	0.424	0.632	0.781	0.532
N-LOCB	0.472	0.432	0.524	0.615	0.553	0.692
LOCB	**0.814**	**0.892**	**0.749**	**0.869**	**0.935**	**0.813**

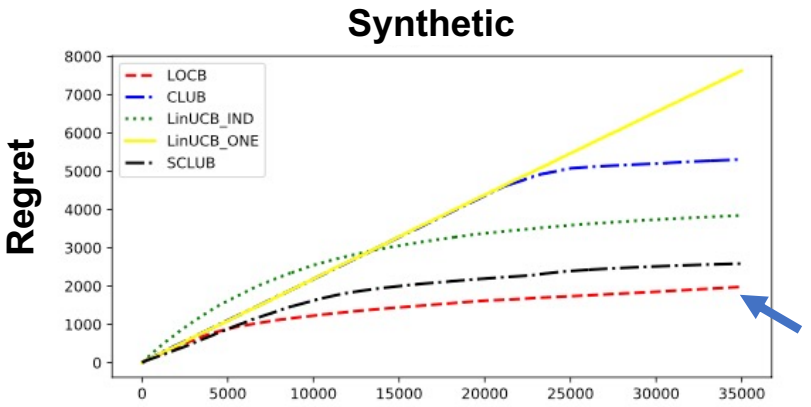
- **Advantages of LOCB**:
  - Effective to recover multiple clusters (multiple clustering centers);
  - Use sets to represent clusters instead of connected components.

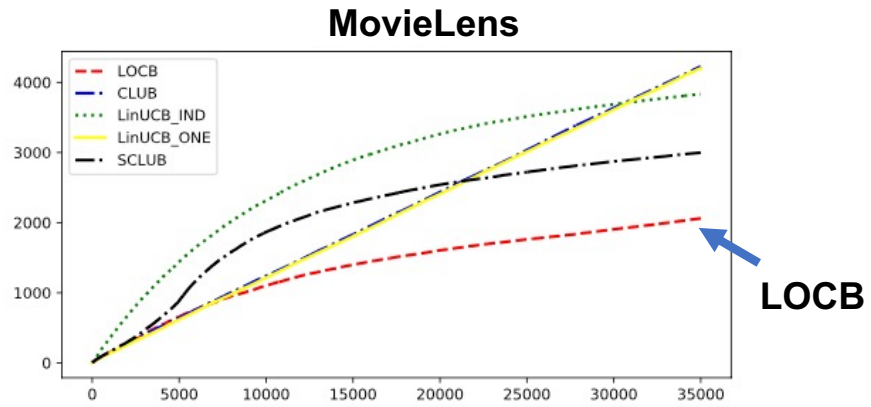
Experiments (2) : Regret comparison

- **Data sets**: Synthetic, Yelp, MovieLens, and Yahoo.
- **Baselines**:
  - (1) LinUCB-ONE [3]: Linear UCB where all users share one parameter.
  - (2) LinUCB-IND [3]: Linear UCB where each user has a parameter.
  - (3) CLUB [1]: Top-down global clustering
  - (4) SCLUB [2]: Improved Top-down global clustering.
- **Metric**: Regret.

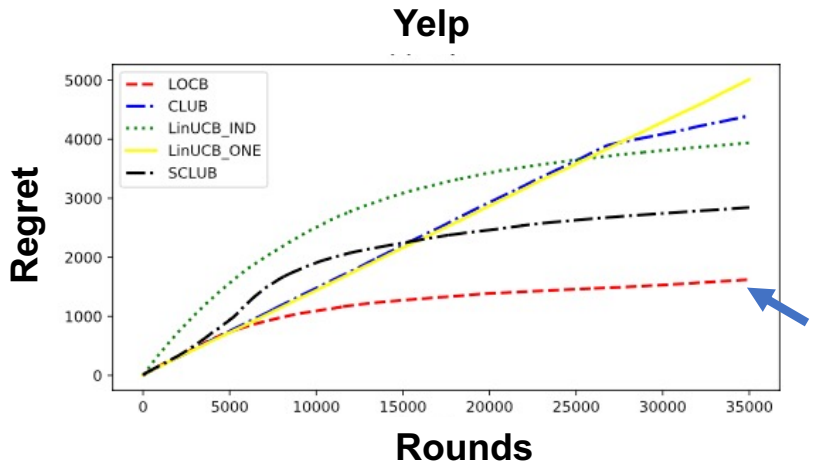
[1] Claudio Gentile, Shuai Li, and Giovanni Zappella. 2014. Online clustering of bandits. *ICML*. 757–765.
[2] Li, S., Chen, W., & Leung, K. S. (2019). Improved algorithm on online clustering of bandits. AAAI.
[3] Li, L., Chu, W., Langford, J., & Schapire, R. E. (2010, April). A contextual-bandit approach to personalized news article recommendation. *WWW* (pp. 661-670).

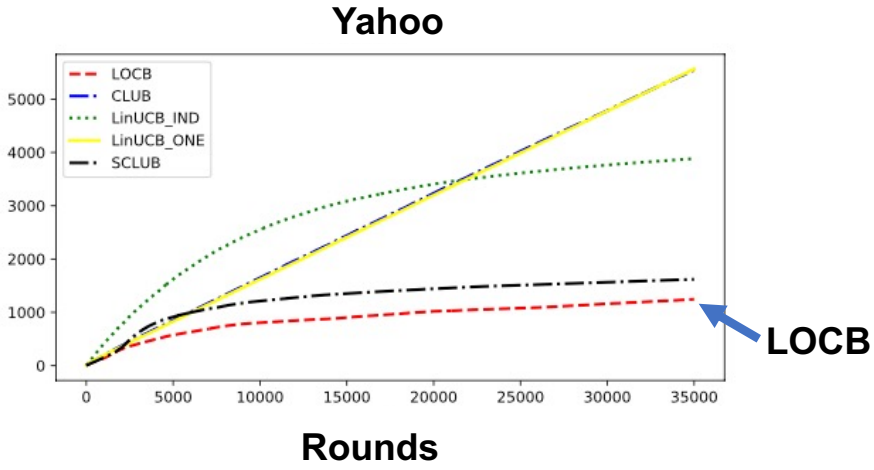
Experiments (2) : Regret analysis





Better

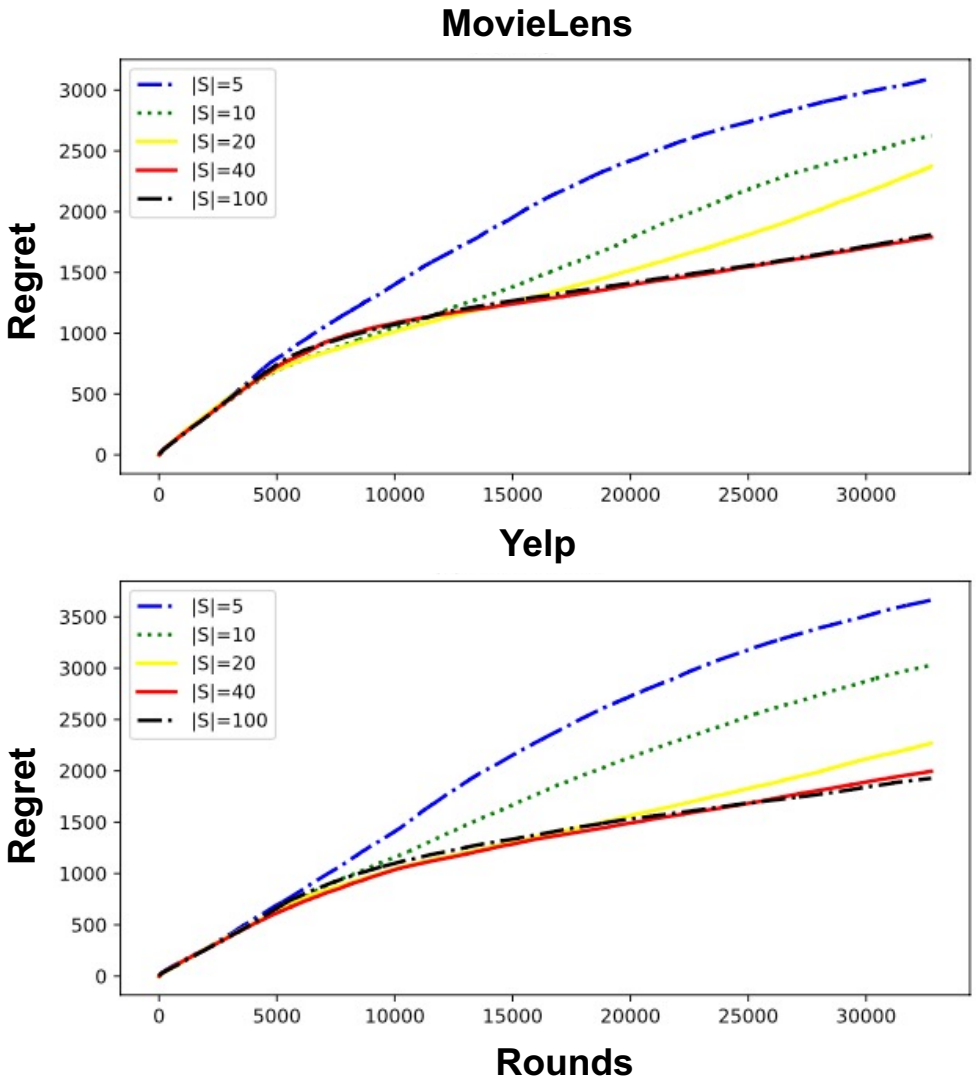




❑ Advantages of LOCB:

- ➢ Provide multiple optional clusters for a user.
- ➢ Choose the best cluster to make recommendation.

Experiments (3) : Effect of Parameters



- Number of users: 100
- Seeds 0-40: improve significantly
- Seeds >40: little improvement

Conclusion

❑ Contributions

- (1) **Problem definition**: We introduce a user clustering problem in the contextual MAB, which needs weaker assumptions and is applicable to many real-world scenarios.

- (2) **Algorithm**: We propose a bandit algorithm, LOCB, embedded with a local clustering procedure. Different from global online clustering, it is more scalable and effective. And we first study the overlapped clusters in MAB.

- (3) **Theoretical Analysis**: We provide three main theorems regarding the correctness and efficiency of clustering and the effectiveness of recommendation.

- (4) **Evaluation**: We evaluate LOCB from various aspects on four data sets including clustering accuracy, regret comparison, and effect of parameters.

THE WEB CONFERENCE

Thank You!

Local Clustering in Contextual Multi-Armed Bandits

Yikun Ban and Jingrui He

University of Illinois at Urbana-Champaign

www.banyikun.com, www.hejingrui.org

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN