Online Clustering of Contextual Cascading Bandits

Shuai Li*1  Shengyu Zhang*1,2 1. The Chinese University of Hong Kong  2. Tencent

Motivation

- Cascading feedback
  - Scenarios: web search results, online recommendation systems, --
  - Model: On an ordered list
    - A user goes through the list from top down,
    - stops at the first satisfactory item, and
    - clicks it.
  - Task: Use this online feedback to help improve future list recommendation
- Contexts: Features for an item or and item-user pair
  - Important for recommendations
- Combinatorial:
  - An action is an ordered sequence of items
- Clustering
  - Users have a clustering structure
  - We only see the user indices
  - Need to learn the user similarities online

Algorithm: CLUB-cascade

1. Parameters: $\lambda, \alpha, \beta > 0$
2. Initialization:
   - $\mathcal{G}$ is a complete graph over users;
   - $S_i = 0_{d \times d}; b_i = 0_{d \times 1}; T_i = 0$ for any user $i$.
3. For all $t = 1, 2, \ldots, n$ do
   1. Obtain user index $I_t$ and item set $D_t \subset \mathbb{R}^{d \times 1}$
   2. Find the connected component $V_i$ of user $I_t$
   3. Compute $\hat{\theta} = M^{-1}b$
   4. For any $x \in D_t$
      - Compute $U_i(x) = -\ln \beta_i x + \beta_i |x|^{\beta_i}$
      - Recommend the $k$ items with largest $U_i$ values and observe $R_i, W_i(x_i), k \leq K_t$
   5. Update statistics
      - $S_i = S_i + \sum_{k=1}^{K_t} x_{i,k} x_{i,k}^T$
      - $b_i = b_i + \sum_{k=1}^{K_t} x_{i,k} W_i(x_{i,k})$
      - $T_i = T_i + k$
      - $b_i = (d_i + S_i)^{-1} b_i$
   6. Delete edge $(i, j)$ if
      - $\|\delta_i - \delta_j\|_2 \geq \alpha \sqrt{T_i + T_j}$
   7. End for $t$

Setting

- $n_u$ of users.
- Each action is an ordered list of $K$ items.
- At time step $t$,
  - User $I_t$ comes to be served with items $D_t \subset \mathbb{R}^{d \times 1}$.
  - Let $\mathcal{H}_t$ denote the history so far.
  - The learning agent recommends $A_t = (x_{t,1}, \ldots, x_{t,K})$ to the user.
  - The user checks from the first item of $A_t$ and stops at $K_t$-th item.
  - The learning agent observes the weights of first $K_t$ base arms in $A_t$.
  - Assume that given $\mathcal{H}_t$, $W_i(x)$'s are mutually independent Bernoulli random variables with $E[W_i(x)] = \theta_i^T x_i$.
  - For some unknown $\theta_i \in \mathbb{R}^{d \times 1}$ with $\|\theta_i\|_2 \leq 1$, $0 \leq \theta_i^T x_i \leq 1$.
  - Cluster regularity: All users in the same cluster have the same $\theta_i$. Users in different clusters have noticeably different $\theta_i$'s.
  - User uniformity: At each time, the user is drawn uniformly from the set of all users, independently over the past.
  - Item regularity: At each time step, the items are drawn independently from a fixed distribution where $E[x_i^T x_j] \geq \gamma > 0$.
- Task: Minimize the cumulative regret of $n$ rounds

$$R(n) = \sum_{t=1}^{n} R(t, A_t)$$

Theoretical analysis

**Theorem 1.** Let $\beta = \sqrt{\ln(1 + n/d)} + 2 \ln 4mK + 8\sqrt{d}$, and $\alpha = 4\sqrt{\ln(1 + n/d)}$. Then the regret of our algorithm, CLUB-cascade, satisfies

$$R(n) = O(d \sqrt{nK \ln n})$$

**Corollary 2.** When the number of clusters $m = 1$, the regret satisfies

$$R(n) = O(d \sqrt{mK \ln n})$$

**Theorem 3.** Consider a general linear reward function

$$\mu(\theta_i^T x_{i,a})$$

where $\mu$ is strictly increasing, continuously differentiable, and Lipschitz with constant $\lambda$. Let $c = \ln \lambda \ln(2 + 2\sqrt{\lambda})$. Then the regret satisfies

$$R(n) = O\left(\frac{cd^2}{\lambda} \sqrt{nK \ln n}\right)$$

Experiments

- **Figure 1.** Experimental results for synthetic data. 40 users, 200 items, $K = 4, d = 20$.
- **Figure 2.** Cumulative rewards on Yelp dataset. $K = 4, d = 20, l = 1$.
- **Figure 3.** Cumulative rewards on MovieLens dataset. $K = 4, d = 20, l = 1$.

Conclusions

- Formulate Online Clustering of Contextual Cascading Bandits problem.
- Propose a CLUB-cascade algorithm that can learn clustering over users and, at the same time, effectively handle
  - contextual information
  - cascading feedback
  - Theoretical analysis
  - Empirical evaluation

References


Contact

Shuai Li
Email: shuaill@cse.cuhk.edu.hk

Shengyu Zhang
Email: syzhang@cse.cuhk.edu.hk