Recommendation Algorithm: online clustering of contextual cascading bandits

May 1, 2022

1 main idea

In our model, the learning agent uses exploration exploitation techniques to learn the preferences of users over items by interactions with users. At each time step, the learning agent recommends a list of posts to the current user, observes the click feedback, and receives a reward of 1 if the user clicks on an item (and receives reward 0 otherwise). The learning agent aims to maximize its cumulative rewards after n rounds.[Gentile et al., 2014]

In this problem, there are u users, denoted by set $[u] = \{1, \ldots, u\}$. At each time step t, a user i_t comes to be served with contents and the learning agent receives the user index with a finite feasible content set $D_t \subset \mathbb{R}^{d\times 1}$, where $\|x\|_2 \leq 1$ for all $x \in D_t$, the feature of posts is captured by the content vector.[Gentile et al., 2017] Then the learning agent recommends a ranked list of *distinct* K items $X_t = (x_1, ..., x_K) \in \Pi^K(D_t)$ to the user. The user checks the items in the order from the first one to the last one, clicks the first attractive item, and stops checking after the click. We use the Bernoulli random variable $\boldsymbol{y}_{t,k}$ to indicate whether the item $x_{t,k}$ has been clicked or not. The learning agent receives the feedback of the index of the clicked item, that is[Li and Zhang, 2018]

$$\boldsymbol{C}_t = \inf\{k : \boldsymbol{y}_{t,k} = 1\}. \tag{1}$$

Note that $\inf(\emptyset) = \infty$ and $\boldsymbol{C}_t = \infty$ represents that the user does not click on any given item. Let $\boldsymbol{K}_t = \min\{\boldsymbol{C}_t, K\}$. The user checks the first $\boldsymbol{K}_t$ items and the learning agent receives the feedback $\{\boldsymbol{y}_{t,k}, k = 1, \ldots, \boldsymbol{K}_t\}$.

Let $\mathcal{H}_t$ be the entire history information until the end of round t. Then the action $\boldsymbol{X}_t$ is $\mathcal{H}_{t-1}$-adaptive. We will write $\mathbb{E}_t[\cdot]$ for $\mathbb{E}[\cdot|\mathcal{H}_{t-1}]$ for convenience of notation, use the boldface symbols to denote random variables, and denote $[m] = \{1, \ldots, m\}$.

We assume the probability of clicking on an item to be a linear function of item feature vector. Specifically there exists a vector $\theta_{i_t} \in \mathbb{R}^{d\times 1}, \|\theta_{i_t}\|_2 \leq 1$ for user i_t, such that the expectation of the binary click feedback $\boldsymbol{y}$ on the checking item x is given by the inner product of x with θ_{i_t}, i.e.,

$$\mathbb{E}_t[\boldsymbol{y}|x] = \theta_{i_t}^\top x, \tag{2}$$

independently of any other given item.

We also use the idea of clustering to describe the overall feature between different users in order to accumulate the establish of users' portrait. Besides, when new user comes, we dynamically add it to the existing graph and initialize it.[Li et al., 2019]

2 algorithm

Algorithm 1 Recommendation algorithm: CLUB-cascade

```
Input: $\lambda, \alpha, \beta > 0$, all of these are dynamically adjustable parameters
Initialize: $G_0 = ([u], E_0)$ is a complete graph over all users, $\boldsymbol{S}_{i,0} = 0_{d\times d}, \boldsymbol{b}_{i,0} = 0_{d\times 1}, T_{i,0} = 0$ for all $i \in [u]$.
for all $t = 1, 2, \ldots, T$ do
    Receive user index $i_t$, and the feasible context set $D_t \subset \mathbb{R}^{d\times 1}$;
    Find the connected component $V_t$ for user $i_t$ in the current graph $G_{t-1} = ([u], E_{t-1})$, and compute
        $\boldsymbol{M}_{V_t,t-1} = \lambda I + \sum_{i\in V_t} \boldsymbol{S}_{i,t-1}, \quad \boldsymbol{b}_{V_t,t-1} = \sum_{i\in V_t} \boldsymbol{b}_{i,t-1},$
        $\hat{\boldsymbol{\theta}}_{V_t,t-1} = \boldsymbol{M}_t^{-1}\boldsymbol{b}_t;$
    For all $x \in D_t$, compute
        $\boldsymbol{U}_t(x) = \min\{\hat{\boldsymbol{\theta}}_{V_t,t-1}^{\top} x + \beta\sqrt{x^{\top}\boldsymbol{M}_{V_t,t-1}^{-1}x}, 1\};$    (3)
    Recommend a list of $K$ items $\boldsymbol{X}_t = (\boldsymbol{x}_{t,1}, \ldots, \boldsymbol{x}_{t,K})$ with largest $\boldsymbol{U}_t(\cdot)$ values and receive feedback $\boldsymbol{C}_t \in \{1, \ldots, K, \infty\}$;
    Update statistics
        $\boldsymbol{S}_{i_t,t} = \boldsymbol{S}_{i_t,t-1} + \sum_{k=1}^{\boldsymbol{K}_t} \boldsymbol{x}_{t,k}\boldsymbol{x}_{t,k}^{\top},$
        $\boldsymbol{b}_{i_t,t} = \boldsymbol{b}_{i_t,t-1} + \sum_{k=1}^{\boldsymbol{K}_t} \boldsymbol{x}_{t,k}\mathbf{1}\{\boldsymbol{C}_t = k\},$
        $T_{i_t,t} = T_{i_t,t-1} + \boldsymbol{K}_t,$
    where $\boldsymbol{K}_t = \min\{\boldsymbol{C}_t, K\}$ and update
        $\hat{\boldsymbol{\theta}}_{i_t,t} = (\lambda I + \boldsymbol{S}_{i_t,t})^{-1}\boldsymbol{b}_{i_t,t};$
    Update
        $\boldsymbol{S}_{\ell,t} = \boldsymbol{S}_{\ell,t-1}, \boldsymbol{b}_{\ell,t} = \boldsymbol{b}_{\ell,t-1}, T_{\ell,t} = T_{\ell,t-1},$
        $\hat{\boldsymbol{\theta}}_{\ell,t} = \hat{\boldsymbol{\theta}}_{\ell,t-1}$
    for all $\ell \neq i_t$;
    Delete the edge $(i_t, \ell) \in E_{t-1}$, if
        $\left\|\hat{\boldsymbol{\theta}}_{i_t,t} - \hat{\boldsymbol{\theta}}_{\ell,t}\right\|_2$
        $\quad \geq \alpha\left(\sqrt{\frac{1+\ln(1+T_{i_t,t})}{1+T_{i_t,t}}} + \sqrt{\frac{1+\ln(1+T_{\ell,t})}{1+T_{\ell,t}}}\right)$
    and obtain a new graph $G_t = ([u], E_t)$;
end for $t$
```

References

[Gentile et al., 2017] Gentile, C., Li, S., Kar, P., Karatzoglou, A., Zappella, G., and Etrue, E. (2017). On context-dependent clustering of bandits. In *International Conference on Machine Learning*, pages 1253–1262. PMLR.

[Gentile et al., 2014] Gentile, C., Li, S., and Zappella, G. (2014). Online clustering of bandits. In *International Conference on Machine Learning*, pages 757–765. PMLR.

[Li et al., 2019] Li, S., Chen, W., and Leung, K.-S. (2019). Improved algorithm on online clustering of bandits. *arXiv preprint arXiv:1902.09162*.

[Li and Zhang, 2018] Li, S. and Zhang, S. (2018). Online clustering of contextual cascading bandits.