

Queueing Matching Bandits with Preference Feedback

Introduction



Motivation examples include

- Ride-hailing platforms where riders are assigned to drivers.
- Online labor service markets where tasks are recommended to freelance workers.

Problem Statement

- There are N agents (queues) and K arms (servers).
- At each time, a job for each agent $n \in [N]$ arrives randomly following a Bernoulli distribution with an unknown arrival rate $\lambda_n \in [0, 1]$.
- At each time, a scheduler assigns agents to arms, $\{S_{k,t}\}_{k \in [K]}$.
- Each agent n has known d-dimensional feature information of $x_n \in \mathbb{R}^d$, and each arm k has latent (unknown) parameter $\theta_k \in \mathbb{R}^d$. Then we adopt the Multi-nomial Logit (MNL) for preference feedback (service rate) as

$$\mu(n|S_{k,t},\theta_k) = \frac{\exp(x_n^{\top}\theta_k)}{1 + \sum_{m \in S_{k,t}} \exp(x_m^{\top}\theta_k)}$$

• Let $Q_n(t)$ be the length of the queue for jobs of agen beginning of time slot t in the system. Queue lengt evolves as

$$Q_n(t+1) = (Q_n(t) + A_n(t) - D_n(t|S_{k_{n,t},t}))^+.$$

Objective Function. Here we provide the goal of this problem. For analyzing the stability of the systems, we define the average queue Jung-hun Kim¹ and Min-hwan Oh¹ ¹Seoul National University

lengths over horizon time T as

$$\mathcal{Q}(T) = \frac{1}{T} \sum_{t \in [T]} \sum_{n \in [N]} \mathbb{E}\left[\right]$$

Definition 1. The systems are denoted to be stable when $\bullet Obs$ $\lim_{T\to\infty}\mathcal{Q}(T)<\infty.$

Assumption 1. (Traffic Slackness) For some traffic slackness Q(T) $0 < \epsilon < 1$, there exists $\{S_k\}_{k \in [K]} \in \mathcal{M}([N])$ such that this set rithm satisfies $\lambda_n + \epsilon \leq \mu(n|S_k, \theta_k)$ for all $n \in S_k$ and $k \in [K]$.

Oracle We denote the oracle assignment

 $\{S_{k,t}^*\}_{k\in[K]} =$ $= \underset{\{S_k\}_{k \in [K]} \in \mathcal{M}(\mathcal{N}_t)}{\operatorname{argmax}} \sum_{k \in [K]} \sum_{n \in S_k} \sum_{k \in [K]} \sum_{k \in [K]} \sum_{n \in S_k} \sum_{k \in [K]} \sum_{k \in S_k} \sum_{k$

Proposition 1. Given the prior knowled the average queue length of MaxWeight is $\mathcal{Q}(T) = O\left(\frac{\min\{N, K\}}{\epsilon}\right)$

$$\overline{k}$$
.
nt $n \in [N]$ at the
of the agent n

Algorithms & Analyses

UCB-based Algorithm (UCB-QMB)

•We define the negative log-likelihood as

$$f_{k,t}(\theta) := -\sum_{n \in S_{k,t} \cup \{n_0\}} y_{n,t} \log \theta$$

where $y_{n,t} \in \{0,1\}$ is observed preference feedback.

• Construct estimator $\hat{\theta}_{k,t}$ from minimizing the negative log-likelihood by applying online newton step as

$$\hat{\theta}_{k,t} = \underset{\theta \in \Theta}{\operatorname{argmin}} \nabla f_{k,t-1} (\hat{\theta}_{k,t-1})^{\top} \theta + \frac{1}{2} \| \theta - \hat{\theta}_{k,t-1} \|_{V_{k,t}}^2.$$

et UCB index for agent $n \in S_k$ as
$$\widetilde{\mu}_t^{UCB} (n | S_k, \hat{\theta}_{k,t}) = \frac{\exp(h_{n,k,t}^{UCB})}{1 + \sum_{m \in S_k} \exp(h_{m,k,t}^{UCB})},$$

• Construct

$$= \underset{\theta \in \Theta}{\operatorname{argmin}} \nabla f_{k,t-1} (\hat{\theta}_{k,t-1})^{\top} \theta + \frac{1}{2} \| \theta - \hat{\theta}_{k,t-1} \|_{V_{k,t}}^{2}.$$

CB index for agent $n \in S_k$ as
 $\widetilde{\mu}_{t}^{UCB}(n|S_k, \hat{\theta}_{k,t}) = \frac{\exp(h_{n,k,t}^{UCB})}{1 + \sum_{m \in S_k} \exp(h_{m,k,t}^{UCB})},$

where $h_{n,k,t}^{UCB} = x_n^{\top} \hat{\theta}_{k,t} + \beta_t ||x_n||_{V_{k,t}^{-1}}$.

 $[Q_n(t)]$.

ts (MaxWeight) as

$$\sum_{S_k} Q_n(t) \mu(n | S_k, \theta_k).$$

End ge of θ_k for all $k \in [I_k]$
is bounded as
 $K_k = 0$.

 $\log \mu(n|S_{k,t},\theta),$

• At each time t, UCB-QMB offers a set of assortments as

$$\begin{split} S_{k,l}\}_{k \in K} &= \operatorname*{argmax}_{\{S_k\}_{k \in [K]} \in \mathcal{M}(\mathcal{N}_l)} \sum_{k \in [K]} \sum_{n \in S_k} Q_n(t) \widetilde{\mu}_l^{UCB}(n|S_k, \widehat{\theta}_{k,l}). \end{split}$$
we preference feedback $y_{n,l} \in \{0, 1\}$ for all $n \in S_{k,l}$ and $k \in [K]$.
em 1. The average queue length of UCB-QMB is bounded as
 $&= O\left(\frac{\min\{N,K\}}{\epsilon} + \frac{d^2N^2K^2 \operatorname{polylog}(T)}{\kappa^{1/2}}\right), which implies that the algo-
achieves stability as
 $&\lim_{T \to \infty} \mathcal{Q}(T) = O\left(\frac{\min\{N,K\}}{\epsilon}\right). \end{aligned}$
em 2. The policy π of UCB-QMB achieves a regret bound of
 $\mathcal{R}^{\pi}(T) = \widetilde{O}\left(\min\left\{\frac{d}{\kappa}\sqrt{KT}Q_{\max}, N\left(\frac{dK}{\kappa^{2}\epsilon^{3}}\right)^{1/4}T^{3/4}\right\}\right). \end{aligned}$
ased Algorithm (TS-QMB) We utilize Thompson Sam-
ith MaxWeight as
 $S_{k,l}\}_{k \in [K]} \leftarrow \operatorname*{argmax}_{\{S_k\}_{k \in [K]}} \sum_{n \in S_k} Q_n(t) \widetilde{\mu}_l^{TS}(n|S_k, \{\widetilde{\theta}_{k,l}^{(l)}\}_{l \in [M]}). \end{aligned}$
em 3. The average queue length of TS-QMB is bounded as
 $= O\left(\frac{\min\{N,K\}}{\epsilon} + \frac{d^{N^2K^2 \text{polylog}(T)}}{\kappa^{1/6}}\right), \text{ which implies that the algo-
echieves stability as
 $\sum_{T \to \infty} Q(T) = O\left(\frac{\min\{N,K\}}{\epsilon}\right).$
em 4. The policy π of TS-QMB achieves a regret bound of
 $\pi(T) = \widetilde{O}\left(\min\left\{\frac{d^{3/2}}{\kappa}\sqrt{KT}Q_{\max}, N\left(\frac{d^2K}{\kappa^{2}\epsilon^{3}}\right)^{1/4}T^{3/4}\right\}\right).$$$

Theo

$$\begin{split} & \underset{\{S_k\}_{k\in[K]}\in\mathcal{M}(\mathcal{N}_l)}{\sum}\sum_{k\in[K]}\sum_{n\in S_k}Q_n(t)\widetilde{\mu}_l^{UCB}(n|S_k,\widehat{\theta}_{k,t}).\\ & \text{e feedback } y_{n,t}\in\{0,1\} \text{ for all } n\in S_{k,t} \text{ and } k\in[K].\\ & average queue length of \textit{UCB-QMB} is bounded as \\ & \underbrace{G_{k}^{1}+\frac{d^2N^2K^2\mathrm{polylog}(T)}{\kappa^{1/6}}}_{K^{1/6}}\right), \text{ which implies that the algobility as}\\ & \lim_{T\to\infty}\mathcal{Q}(T)=O\left(\frac{\min\{N,K\}}{\epsilon}\right).\\ & \text{policy } \pi \text{ of UCB-QMB} \text{ achieves a regret bound of}\\ & \left(\min\left\{\frac{d}{\kappa}\sqrt{KT}Q_{\max},N\left(\frac{dK}{\kappa^{2}\epsilon^{3}}\right)^{1/4}T^{3/4}\right\}\right).\\ & \text{orithm (TS-QMB)} \quad \text{We utilize Thompson Sam-ght as}\\ & \underset{\{S_k\}_{k\in[K]}}{\operatorname{average}} \max\sum_{k\in[K]}\sum_{n\in S_k}Q_n(t)\widetilde{\mu}_l^{TS}(n|S_k,\{\widetilde{\theta}_{k,t}^{(i)}\}_{i\in[M]}).\\ & average queue \ length \ of \ TS-QMB \ is \ bounded \ as \\ & \underbrace{G_{k}^{1}+\frac{d^4N^2K^2\mathrm{polylog}(T)}{T}\right), \ which \ implies \ that \ the \ algobility \ as}\\ & \lim_{T\to\infty}\mathcal{Q}(T)=O\left(\frac{\min\{N,K\}}{\epsilon}\right).\\ & policy \ \pi \ of \ TS-QMB \ achieves \ a \ regret \ bounded \ as \\ & \underbrace{G_{k}^{1}+\frac{d^4N^2K^2\mathrm{polylog}(T)}{K^4}, \ which \ implies \ that \ the \ algobility \ as \\ & \lim_{T\to\infty}\mathcal{Q}(T)=O\left(\frac{\min\{N,K\}}{\epsilon}\right).\\ & policy \ \pi \ of \ TS-QMB \ achieves \ a \ regret \ bounded \ as \ (M_{k,k}^{2}\epsilon^{3})^{1/4}T^{3/4}\\ & \lim_{T\to\infty}\mathcal{Q}(T)=O\left(\frac{\min\{N,K\}}{\epsilon}\right).\\ & policy \ \pi \ of \ TS-QMB \ achieves \ a \ regret \ bound \ of \ (M_{k,k}^{2}\epsilon^{3})^{1/4}T^{3/4}\\ & \lim_{T\to\infty}\mathcal{Q}(T)=O\left(\frac{\min\{N,K\}}{\epsilon^{2}\epsilon^{3}}\right).\\ & policy \ \pi \ of \ TS-QMB \ achieves \ a \ regret \ bound \ of \ (M_{k,k}^{2}\epsilon^{3})^{1/4}T^{3/4}\\ & \lim_{T\to\infty}\mathcal{Q}(T)=O\left(\frac{\min\{N,K\}}{\epsilon^{2}\epsilon^{3}}\right)^{1/4}T^{3/4}\\ & \lim_{T\to\infty}\mathcal{Q}(T)=O\left(\frac{\max\{N,K\}}{\epsilon^$$

Theo

$$\{S_{k,t}\}_{k\in[K]} = \underset{\{S_k\}_{k\in[K]}\in\mathcal{M}(\mathcal{N}_{t})}{\operatorname{argmax}} \sum_{k\in[K]} \sum_{n\in S_k} Q_n(t)\widetilde{\mu}_{t}^{UCB}(n|S_k, \widehat{\theta}_{k,t}).$$
erve preference feedback $y_{n,t} \in \{0, 1\}$ for all $n \in S_{k,t}$ and $k \in [K]$.
rem 1. The average queue length of UCB-QMB is bounded as

$$= O\left(\frac{\min\{N,K\}}{\epsilon} + \frac{d^{2}N^{2}K^{2}\operatorname{polytog}(T)}{\pi}\right), \text{ which implies that the algo-achieves stability as}$$

$$\lim_{T\to\infty} Q(T) = O\left(\frac{\min\{N,K\}}{\epsilon}\right).$$
rem 2. The policy π of UCB-QMB achieves a regret bound of
 $\mathcal{R}^{\pi}(T) = \widetilde{O}\left(\min\left\{\frac{d}{\kappa}\sqrt{KT}Q_{\max}, N\left(\frac{dK}{\kappa^{2}\epsilon^{3}}\right)^{1/4}T^{3/4}\right\}\right).$
pased Algorithm (TS-QMB) We utilize Thompson Samwith MaxWeight as

$$[S_{k,t}]_{k\in[K]} \leftarrow \underset{\{S_k\}_{k\in[K]}}{\operatorname{argmax}} \sum_{k\in[K]} \sum_{n\in S_k} Q_n(t)\widetilde{\mu}_{t}^{TS}(n|S_k,\{\widetilde{\theta}_{k,t}^{(i)}\}_{i\in[M]}).$$
rem 3. The average queue length of TS-QMB is bounded as

$$= O\left(\frac{\min\{N,K\}}{\epsilon} + \frac{d^{3}N^{2}K^{2}\operatorname{polytog}(T)}{\pi^{4}\epsilon^{6}}\right), \text{ which implies that the algo-achieves stability as}$$

$$\lim_{T\to\infty} Q(T) = O\left(\frac{\min\{N,K\}}{\epsilon}\right).$$
rem 4. The policy π of TS-QMB achieves a regret bound of

$$\mathcal{R}^{\pi}(T) = \widetilde{O}\left(\min\left\{\frac{d^{3/2}}{\kappa}\sqrt{KT}Q_{\max}, N\left(\frac{d^{2}K}{\kappa^{2}c^{3}}\right)^{1/4}T^{3/4}\right\}\right).$$

$\subset [K], \mathbf{TS-b}$

pling

$$\{S_{k,t}\}_{k\in[K]} = \underset{\{S_k\}_{k\in[K]}\subset\mathcal{M}(\mathcal{N}_t)}{\operatorname{argmax}} \sum_{k\in[K]} \sum_{n\in S_k} Q_n(t) \widetilde{\mu}_t^{UCB}(n|S_k; \widehat{\theta}_{k,t}).$$
we reperference feedback $y_{n,t} \in \{0, 1\}$ for all $n \in S_{k,t}$ and $k \in [K]$.
prem 1. The average queue length of UCB-QMB is bounded as
$$= O\left(\frac{\min\{N,K\}}{\epsilon} + \frac{d^2N^2K^2\operatorname{polylog}(T)}{\kappa^2\epsilon^3}\right), \text{ which implies that the algo-
to achieves stability as
$$\lim_{T\to\infty} \mathcal{Q}(T) = O\left(\frac{\min\{N,K\}}{\epsilon}\right).$$
orem 2. The policy π of UCB-QMB achieves a regret bound of
 $\mathcal{R}^{\pi}(T) = \widetilde{O}\left(\min\left\{\frac{d}{\kappa}\sqrt{KT}Q_{\max}, N\left(\frac{dK}{\kappa^2\epsilon^3}\right)^{1/4}T^{3/4}\right\}\right).$
based Algorithm (TS-QMB) We utilize Thompson Sam-
with MaxWeight as
$$\{S_{k,t}\}_{k\in[K]} \leftarrow \operatorname*{argmax}_{\{S_k\}_{k\in[K]}} \sum_{k\in[K]} \sum_{n\in S_k} Q_n(t)\widetilde{\mu}_t^{TS}(n|S_k, \{\widetilde{\theta}_{k,t}^{(i)}\}_{i\in[M]}).$$
orem 3. The average queue length of TS-QMB is bounded as
$$= O\left(\frac{\min\{N,K\}}{\epsilon} + \frac{d^{N^2K^2}\operatorname{polylog}(T)}{\kappa^2\epsilon^3}\right), \text{ which implies that the algo-
based achieves stability as
$$= O\left(\frac{\min\{N,K\}}{\epsilon} + \frac{d^{N^2K^2}\operatorname{polylog}(T)}{\kappa^2\epsilon^3}\right), \text{ which implies that the algo-
based achieves stability as
$$= O\left(\frac{\min\{N,K\}}{\epsilon} + \frac{d^{N^2K^2}\operatorname{polylog}(T)}{\kappa^2\epsilon^3}\right), \text{ which implies that the algo-
based achieves a regret bound of TS and $T \to 0$

$$= O\left(\frac{\min\{N,K\}}{\epsilon} + \frac{d^{N^2K^2}\operatorname{polylog}(T)}{\kappa^2\epsilon^3}\right), \text{ which implies that the algo-
based achieves a regret bound of TS and $T \to 0$

$$= O\left(\frac{\min\{N,K\}}{\epsilon} + \frac{d^{N^2K^2}\operatorname{polylog}(T)}{\kappa^2\epsilon^3}\right), \text{ which implies that the algo-
based achieves a regret bound of $T \to 0$

$$= O\left(\min\{N,K\}\right).$$

$$= O\left(\min\{N,K\}\right) = O\left(\min\{N,K\}\right).$$$$$$$$$$$$$$

Theo Q(T)rithm

$$= \underset{\{S_k\}_{k \in [K]} \in \mathcal{M}(\mathcal{N}_t)}{\operatorname{argmax}} \sum_{k \in [K]} \sum_{n \in S_k} Q_n(t) \widetilde{\mu}_t^{UCB}(n|S_k, \widehat{\theta}_{k,t}).$$
we feedback $y_{n,t} \in \{0, 1\}$ for all $n \in S_{k,t}$ and $k \in [K]$.
 $average queue length of UCB-QMB is bounded as$

$$\frac{K!}{k + \frac{d^2N^2K^2 \operatorname{polylog}(T)}{\kappa^4 \epsilon^6}}, which implies that the algo-bility as$$

$$\lim_{T \to \infty} \mathcal{Q}(T) = O\left(\frac{\min\{N, K\}}{\epsilon}\right).$$
 $policy \pi \text{ of UCB-QMB achieves a regret bound of}$

$$\left(\min\left\{\frac{d}{\kappa}\sqrt{KT}Q_{\max}, N\left(\frac{dK}{\kappa^2 \epsilon^3}\right)^{1/4}T^{3/4}\right\}\right).$$
 $orithm (TS-QMB)$ We utilize Thompson Samight as
$$\underset{\{S_k\}_{k \in [K]}}{\operatorname{argmax}} \sum_{k \in [K]} \sum_{n \in S_k} Q_n(t) \widetilde{\mu}_t^{TS}(n|S_k, \{\widetilde{\theta}_{k,t}^{(i)}\}_{i \in [M]}).$$
 $e average queue length of TS-QMB is bounded as$

$$\underset{\{S_k\}_{k \in [K]}}{\operatorname{argmax}} \mathcal{Q}(T) = O\left(\frac{\min\{N, K\}}{\epsilon}\right), which implies that the algo-bility as$$
 $\underset{\{S_k\}_{k \in [K]}}{\operatorname{argmax}} \sum_{k \in [K]} \sum_{n \in S_k} Q_n(t) \widetilde{\mu}_t^{TS}(n|S_k, \{\widetilde{\theta}_{k,t}^{(i)}\}_{i \in [M]}).$
 $\underset{\{S_k\}_{k \in [K]}}{\operatorname{argmax}} \sum_{k \in [K]} \sum_{n \in S_k} Q_n(t) \widetilde{\mu}_t^{TS}(n|S_k, \{\widetilde{\theta}_{k,t}^{(i)}\}_{i \in [M]}).$
 $\underset{\{S_k\}_{k \in [K]}}{\operatorname{argmax}} \sum_{k \in [K]} \sum_{n \in S_k} Q_n(t) \widetilde{\mu}_t^{TS}(n|S_k, \{\widetilde{\theta}_{k,t}^{(i)}\}_{i \in [M]}).$
 $\underset{\{S_k\}_{k \in [K]}}{\operatorname{argmax}} \sum_{k \in [K]} \sum_{n \in S_k} Q_n(t) \widetilde{\mu}_t^{TS}(n|S_k, \{\widetilde{\theta}_{k,t}^{(i)}\}_{i \in [M]}).$
 $\underset{\{S_k\}_{k \in [K]}}{\operatorname{argmax}} \sum_{k \in [K]} \sum_{n \in S_k} Q_n(t) \widetilde{\mu}_t^{TS}(n|S_k, \{\widetilde{\theta}_{k,t}^{(i)}\}_{i \in [M]}).$
 $\underset{\{S_k\}_{k \in [K]}}{\operatorname{argmax}} \sum_{k \in [K]} \sum_{n \in S_k} Q_n(t) \widetilde{\mu}_t^{TS}(n|S_k, \{\widetilde{\theta}_{k,t}^{(i)}\}_{i \in [M]}).$
 $\underset{\{S_k\}_{k \in [K]}}{\operatorname{argmax}} \sum_{k \in [K]} \sum_{n \in S_k} Q_n(t) \widetilde{\mu}_t^{TS}(n|S_k, \{\widetilde{\theta}_{k,t}^{(i)}\}_{i \in [M]}).$
 $\underset{\{S_k\}_{k \in [K]}}{\operatorname{argmax}} \sum_{k \in [K]} \sum_{n \in S_k} Q_n(t) \widetilde{\mu}_t^{TS}(n|S_k, \{\widetilde{\theta}_{k,t}^{(i)}\}_{i \in [M]}).$
 $\underset{\{S_k\}_{k \in [K]}} \sum_{n \in S_k} Q_n(t) \widetilde{\mu}_t^{TS}(n|S_k, \{\widetilde{\theta}_{k,t}^{(i)}\}_{i \in [M]}).$
 $\underset{\{S_k\}_{k \in [K]}} \sum_{n \in S_k} Q_n(t) \widetilde{\mu}_t^{TS}(n|S_k, \{\widetilde{\theta}_{k,t}^{(i)}\}_{i \in [M]}).$

Theo

$$\{S_{k,l}\}_{k\in[K]} = \underset{\{S_k\}_{k\in[K]} \in \mathcal{M}(N_l)}{\operatorname{argmax}} \sum_{k\in[K]} \sum_{n\in S_k} Q_n(t) \widetilde{\mu}_t^{UCB}(n|S_k, \widehat{\theta}_{k,l}).$$
serve preference feedback $y_{n,l} \in \{0, 1\}$ for all $n \in S_{k,l}$ and $k \in [K]$.
orem 1. The average queue length of UCB-QMB is bounded as
 $p = O\left(\frac{\min\{N,K\}}{\epsilon} + \frac{d^2N^2K^2\operatorname{polylog}(T)}{\pi^4}\right), \text{ which implies that the algo-
 n achieves stability as
 $\lim_{T \to \infty} Q(T) = O\left(\frac{\min\{N,K\}}{\epsilon}\right).$
orem 2. The policy π of UCB-QMB achieves a regret bound of
 $\mathcal{R}^{\pi}(T) = \widetilde{O}\left(\min\left\{\frac{d}{\kappa}\sqrt{KT}Q_{\max}, N\left(\frac{dK}{\kappa^2\epsilon^3}\right)^{1/4}T^{3/4}\right\}\right).$
based Algorithm (TS-QMB) We utilize Thompson Sam-
with MaxWeight as
 $\{S_{k,l}\}_{k\in[K]} \leftarrow \operatorname*{argmax}_{\{S_k\}_{k\in[K]}} \sum_{k\in[K]} Q_n(t)\widetilde{\mu}_l^{TS}(n|S_k, \{\widetilde{\theta}_{k,l}^{(i)}\}_{i\in[M]}).$
orem 3. The average queue length of TS-QMB is bounded as
 $p = O\left(\frac{\min\{N,K\}}{\epsilon} + \frac{d!N'K^2\operatorname{polylog}(T)}{\kappa^4\epsilon^0}\right), \text{ which implies that the algo-
 n achieves stability as
 $\lim_{T \to \infty} Q(T) = O\left(\frac{\min\{N,K\}}{\epsilon}\right).$
orem 4. The policy π of TS-QMB achieves a regret bound of
 $\mathcal{R}^{\pi}(T) = \widetilde{O}\left(\min\left\{\frac{d^{3/2}}{\kappa}\sqrt{KT}Q_{\max}, N\left(\frac{d^2K}{\kappa^2\epsilon^3}\right)^{1/4}T^{3/4}\right\}\right).$$$

Experiments

