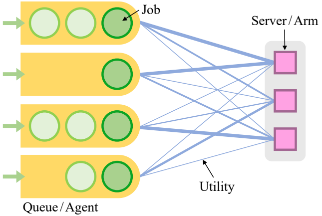
Queueing Matching Bandits with Preference Feedback

Jung-hun Kim[1] and Min-hwan Oh[1]

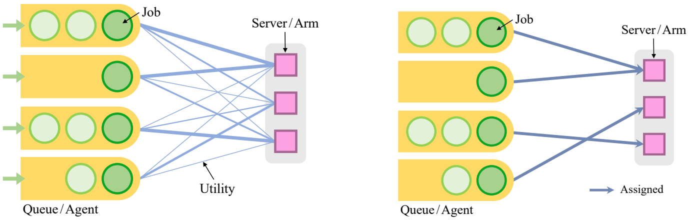
[1]Seoul National University

Queueing Matching Bandit Framework

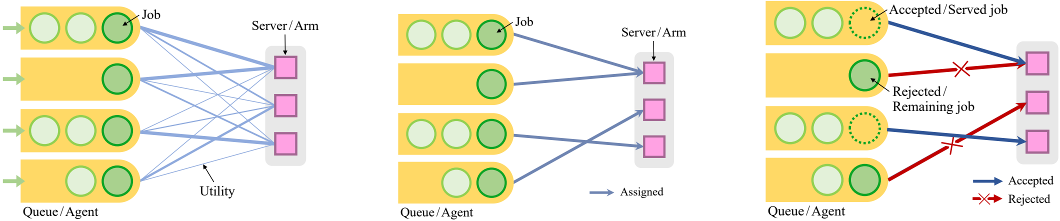
Queueing Matching Bandit Framework



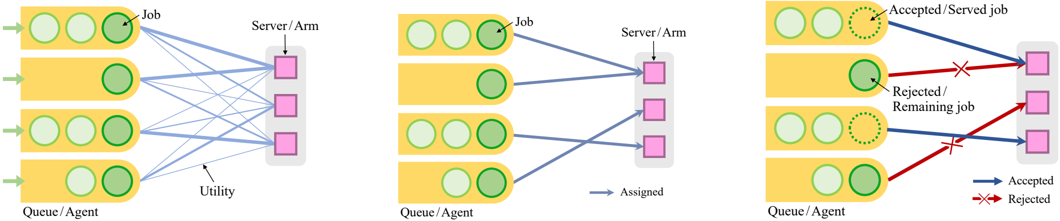
Queueing Matching Bandit Framework



Queueing Matching Bandit Framework



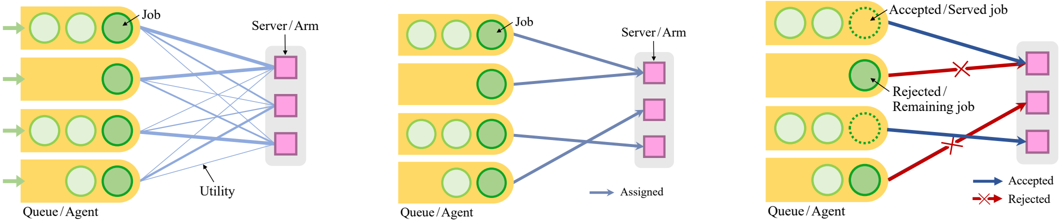
Queueing Matching Bandit Framework



Motivation examples for this framework include

- Ride-hailing platforms where riders are assigned to drivers.

Queueing Matching Bandit Framework



Motivation examples for this framework include

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

Problem Statement - Preference/Service Rate

- For each agent $n \in [N]$, their feature information is known as $x_n \in \mathbb{R}^d$, and each arm $k \in [K]$ is characterized by latent vector $\theta_k \in \mathbb{R}^d$.

Problem Statement - Preference/Service Rate

- For each agent $n \in [N]$, their feature information is known as $x_n \in \mathbb{R}^d$, and each arm $k \in [K]$ is characterized by latent vector $\theta_k \in \mathbb{R}^d$.
- Given $S_{k,t}$ (set of agents assigned to arm k, 'assortment'), the service rate for arm k to serve a job of agent $n \in S_{k,t}$ follows Multi-nomial Logit (MNL) model as follows

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

Problem Statement - Preference/Service Rate

- For each agent $n \in [N]$, their feature information is known as $x_n \in \mathbb{R}^d$, and each arm $k \in [K]$ is characterized by latent vector $\theta_k \in \mathbb{R}^d$.
- Given $S_{k,t}$ (set of agents assigned to arm k, 'assortment'), the service rate for arm k to serve a job of agent $n \in S_{k,t}$ follows Multi-nomial Logit (MNL) model as follows

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

- It is allowed to have an outside option (n_0), or not to choose any, for each arm k according to MNL with probability as $\mu(n_0|S_{k,t}, \theta_k) = \frac{1}{1+\sum_{m \in S_{k,t}} \exp(x_m^\top \theta_k)}$.

Problem Statement - Objective Function

- For the stability of the systems, we define the average queue lengths over horizon time T as

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

Problem Statement - Objective Function

- For the stability of the systems, we define the average queue lengths over horizon time T as
$$\mathcal{Q}(T) = \frac{1}{T} \sum_{t \in [T]} \sum_{n \in [N]} \mathbb{E}[Q_n(t)].$$
- By following previous queueing bandit literature [Sentenac et al. 2021; Freund et al. 2022; Yang et al. 2023; Huang et al. 2024], we define the system stability as follows:

Problem Statement - Objective Function

- For the stability of the systems, we define the average queue lengths over horizon time T as

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

- By following previous queueing bandit literature [Sentenac et al. 2021; Freund et al. 2022; Yang et al. 2023; Huang et al. 2024], we define the system stability as follows:

Definition 1 (System Stability)

The systems are denoted to be stable if $\lim_{T \to \infty} \mathcal{Q}(T) < \infty$.

Summary of Our Contributions

- As a preliminary study, we analyze an oracle strategy of MaxWeight under known service rates. It is shown to achieve $\mathcal{Q}(T) = O\left(\frac{\min\{N,K\}}{\epsilon}\right) < \infty$, where ϵ is a traffic slackness.

Summary of Our Contributions

- As a preliminary study, we analyze an oracle strategy of MaxWeight under known service rates. It is shown to achieve $\mathcal{Q}(T) = O\left(\frac{\min\{N,K\}}{\epsilon}\right) < \infty$, where ϵ is a traffic slackness.
- Under unknown service rates, we propose UCB-based and Thompson Sampling-based MaxWeight algorithms.

Summary of Our Contributions

- As a preliminary study, we analyze an oracle strategy of MaxWeight under known service rates. It is shown to achieve $\mathcal{Q}(T) = O\left(\frac{\min\{N,K\}}{\epsilon}\right) < \infty$, where ϵ is a traffic slackness.
- Under unknown service rates, we propose UCB-based and Thompson Sampling-based MaxWeight algorithms.
- Our algorithms achieve stability with $\lim_{T\to\infty} \mathcal{Q}(T) = O\left(\frac{\min\{N,K\}}{\epsilon}\right)$.

Summary of Our Contributions

- As a preliminary study, we analyze an oracle strategy of MaxWeight under known service rates. It is shown to achieve $\mathcal{Q}(T) = O\left(\frac{\min\{N,K\}}{\epsilon}\right) < \infty$, where ϵ is a traffic slackness.
- Under unknown service rates, we propose UCB-based and Thompson Sampling-based MaxWeight algorithms.
- Our algorithms achieve stability with $\lim_{T\to\infty} \mathcal{Q}(T) = O\left(\frac{\min\{N,K\}}{\epsilon}\right)$.
- Our algorithms achieve regret bounds of $\widetilde{O}\left(\min\{\sqrt{T}Q_{\max}, T^{3/4}\}\right)$.

Summary of Our Contributions

- As a preliminary study, we analyze an oracle strategy of MaxWeight under known service rates. It is shown to achieve $\mathcal{Q}(T) = O\left(\frac{\min\{N,K\}}{\epsilon}\right) < \infty$, where ϵ is a traffic slackness.
- Under unknown service rates, we propose UCB-based and Thompson Sampling-based MaxWeight algorithms.
- Our algorithms achieve stability with $\lim_{T\to\infty} \mathcal{Q}(T) = O\left(\frac{\min\{N,K\}}{\epsilon}\right)$.
- Our algorithms achieve regret bounds of $\widetilde{O}\left(\min\{\sqrt{T}Q_{\max}, T^{3/4}\}\right)$.
- We demonstrate the effectiveness of our algorithms using synthetic datasets.