Contextual Linear Bandit Problem & Applications

Feng Bi
Joon Sik Kim
Leiya Ma
Pengchuan Zhang
Contents

• Recap of last lecture
  – Multi-armed Bandit Problem / UCB1 Algorithm
• Application to News Article Recommendation
• Contextual Linear Bandit Problem
• LinUCB Algorithm
• Disjoint Model & General Model
• Derivation of LinUCB
• News Article Recommendation Revisited
• Conclusion
Last Lecture : Multi-armed Bandit Problem

• K actions (feature-free)
• Each action has an average reward (unknown): $\mu_k$
• For $t=1,...,T$ (unknown)
  – Choose an action $a_t$ from $\{1, ..., K\}$ actions
  – Observe a random reward $y_t$, where $y_t$ is bounded $[0,1]$
  – $E[y_t] = \mu_{a,t}$ : Expected reward of action $a_t$
• Minimizing Regret: \[
R = \sum_{t=1}^{T} [\mu^* - \mu_{a,t}]
\]

Q. How to choose an action to minimize regret?
Last Lecture: UCB1 Algorithm

- At each iteration, choose the action with highest Upper Confidence Bound (UCB)

\[ a_{t+1} = \operatorname{arg\,max}_a \hat{\mu}_{a,t} + \sqrt{\frac{2 \ln t}{t_a}} \]

- Regret Bound: with high probability, sublinear w.r.t. T

\[ R(T) = O \left( \frac{K}{\Delta} \ln T \right) \]

- Gap between best & 2\(^\text{nd}\) best

\[ \Delta = \mu_1 - \mu_2 \]
Application of Multi-armed Bandit Problem?
News Article Recommendation

• Various algorithms for personalized recommendation
  – Collaborative filtering, Content-based filtering, Hybrid approaches

• But...
  – Web-based contents undergo frequent changes
  – Some users have no previous data to learn from (cold-start problem)

• Exploration vs. Exploitation
  – Need to gather more information about users with more trials
  – Optimize the article selection with past user experience

➤ Use multi-armed bandit setting
Article Recommendation in Feature-Free Bandit Setting

Users $u_1$ with age YOUNG and $u_2$ with age OLD

Retirement planning wishes vs. reality

The Player Wizarding World of Harry Potter ride may conjure a new path for theme park rides

Elon Musk: 198,000 Tesla Model 3 Orders Received in 24 Hours

Not tired yet: Warriors top Spurs for 72nd win, set up date with history
Article Recommendation in Feature-Free Bandit Setting

User $u_1$ with age YOUNG

Retirement planning wishes vs. reality

The Player Wizarding World of Harry Potter ride may conjure a new path for theme park rides

Elon Musk: 198,000 Tesla Model 3 Orders Received in 24 Hours

Not tired yet: Warriors top Spurs for 72nd win, set up date with history
Article Recommendation in Feature-Free Bandit Setting

User $u_1$ with age YOUNG

- Retirement planning wishes vs. reality
  - $0.2$

- The Player: Wizarding World of Harry Potter ride may conjure a new path for theme park rides
  - $0.4$

- Elon Musk: 198,000 Tesla Model 3 Orders Received in 24 Hours
  - $0.5$

- Not tired yet: Warriors top Spurs for 72nd win, set up date with history
  - $0.8$
Article Recommendation in Feature-Free Bandit Setting

User $u_1$ with age YOUNG

- Retirement planning wishes vs. reality
- The Player Wizarding World of Harry Potter ride may conjure a new path for theme park rides
- Elon Musk: 198,000 Tesla Model 3 Orders Received in 24 Hours
- Not tired yet: Warriors top Spurs for 72nd win, set up date with history
Article Recommendation in Feature-Free Bandit Setting

User $u_1$ with age YOUNG

- **Retirement planning wishes vs. reality**
  - The Player: Wizarding World of Harry Potter ride may conjure a new path for theme park rides
  - Elon Musk: 198,000 Tesla Model 3 Orders Received in 24 Hours
  - Not tired yet: Warriors top Spurs for 72nd win, set up date with history
Article Recommendation in Feature-Free Bandit Setting

User $u_2$ with age OLD

0.2 Retirement planning wishes vs. reality

0.4 The Player: Wizarding World of Harry Potter ride may conjure a new path for theme park rides

0.5 Elon Musk: 198,000 Tesla Model 3 Orders Received in 24 Hours

0.8 Not tired yet: Warriors top Spurs for 72nd win, set up date with history
Article Recommendation in Feature-Free Bandit Setting

User $u_2$ with age OLD

- Retirement planning wishes vs. reality
- The Player: Wizarding World of Harry Potter ride may conjure a new path for theme park rides
- Elon Musk: 198,000 Tesla Model 3 Orders Received in 24 Hours
- Not tired yet: Warriors top Spurs for 72nd win, set up date with history
Article Recommendation in Feature-Free Bandit Setting

User $u_2$ with age OLD

---

**Retirement planning wishes vs. reality**

The Player: Wizarding World of Harry Potter ride may conjure a new path for theme park rides

Elon Musk: 198,000 Tesla Model 3 Orders Received in 24 Hours

Not tired yet: Warriors top Spurs for 72nd win, set up date with history
Article Recommendation in Feature-Free Bandit Setting

User u₂ with age OLD

Retirement planning wishes vs. reality

The Player: Wizarding World of Harry Potter ride may conjure a new path for theme park rides

Elon Musk: 198,000 Tesla Model 3 Orders Received in 24 Hours

Not tired yet: Warriors top Spurs for 72nd win, set up date with history
Contextual-Bandit Problem

- For $t=1,\ldots,T$ (unknown)
  - User $u_t$, set $A_t$ of actions (a)
  - Feature vector (context) $x_{t,a}$: summarizes both user $u_t$ and action $a$
  - Based on previous results, choose $a_t$ from $A_t$
  - Receive payoff $r_{t,a_t}$
  - Improve selection strategy with new observation set $(x_{t,a_t}, a_t, r_{t,a_t})$

- Minimizing Regret: $R(T) = \mathbb{E} \left[ \sum_{t=1}^{T} \left( r_{t,a_t}^* - r_{t,a_t} \right) \right]$
Difference?

• Contextual bandit problem becomes K-armed bandit problem when
  – The action set $A_t$ is unchanged and contains K actions for all $t$
  – The user $u_t$ (or the context) is the same for all $t$

• Also called context-free bandit problem
Article Recommendation in Feature-Free Bandit Setting

Users \( u_1 \) with age YOUNG and \( u_2 \) with age OLD

Retirement planning wishes vs. reality

The Player *Wizarding World of Harry Potter* ride may conjure a new path for theme park rides

Elon Musk: 198,000 Tesla Model 3 Orders Received in 24 Hours

Not tired yet: Warriors top Spurs for 72nd win, set up date with history
Article Recommendation in Contextual Linear Bandit Setting

Users $u_1$ with age YOUNG and $u_2$ with age OLD

$\theta_1$  
Retirement planning wishes vs. reality

$\theta_2$  
The Player Wizarding World of Harry Potter ride may conjure a new path for theme park rides

$\theta_3$  
Elon Musk: 198,000 Tesla Model 3 Orders Received in 24 Hours

$\theta_4$  
Not tired yet: Warriors top Spurs for 72nd win, set up date with history

$x_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$x_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
Article Recommendation in Contextual Linear Bandit Setting

Linear Payoff $= x^T \theta$

Users $u_1$ with age YOUNG and $u_2$ with age OLD

$\theta_1$  
Retirement planning wishes vs. reality

$\theta_2$  
The Player Wizarding World of Harry Potter ride may conjure a new path for theme park rides

$\theta_3$  
Elon Musk: 198,000 Tesla Model 3 Orders Received in 24 Hours

$\theta_4$  
Not tired yet: Warriors top Spurs for 72nd win, set up date with history

$x_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$x_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
Article Recommendation in Contextual Linear Bandit Setting

Users $u_1$ with age YOUNG and $u_2$ with age OLD

Linear Payoff $= x^T \theta$

[ 0.1, 0.6 ]

Retirement planning wishes vs. reality

[ 0.5, 0.1 ]

The Player Wizarding World of Harry Potter ride may conjure a new path for theme park rides

[ 0.6, 0.1 ]

Elon Musk: 198,000 Tesla Model 3 Orders Received in 24 Hours

[ 0.9, 0.2 ]

Not tired yet: Warriors top Spurs for 72nd win, set up date with history

$x_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$x_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
Article Recommendation in Contextual Linear Bandit Setting

Users $u_1$ with age YOUNG and $u_2$ with age OLD

$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Linear Payoff $= \mathbf{x}^T \boldsymbol{\theta}$

Retirement planning wishes vs. reality

The Player
- Wizarding World of Harry Potter ride may conjure a new path for theme park rides
- Elon Musk: 198,000 Tesla Model 3 Orders Received in 24 Hours
- Not tired yet: Warriors top Spurs for 72nd win, set up date with history
LinUCB Algorithm

- Assumption: the payoff model is linear
  - Most Intuitive thought: Linear Model
  - Advantage: Confidence interval computed efficiently in closed form

- Tempting to apply UCB on general contextual bandit problems
  - asymptotic optimality
  - strong regret bound

➡ Called LinUCB algorithm.
**Contextual Bandit**

For each trial $t=1,2,3..., T$

1. Observe environment $x_{t,a} \in \mathbb{R}^d$, i.e. user $u_t$ a set of actions $\mathcal{A}_t$ and both their features

2. Choose an arm $a_t \in \mathcal{A}$ based on previous trails an receive payoff $r_{t,a_t}$.

3. Improve arm selection strategy with new observation $(x_{t,a_t}, a_t, r_{t,a_t})$

**Example: News Recommendation**

For each time the news page is loaded $t=1,2,3..., T$

1. Arms or actions are the articles, which can be shown to the user. The environment could be user and article information.

2. If the article is clicked $r_{t,a_t} = 1$ otherwise 0.

3. Improve new article selection

Minimize expected regret, i.e

$$R_A(T) = \mathbb{E} \left[ \sum_{t=1}^T r_{t,a_t}^* \right] - \mathbb{E} \left[ \sum_{t=1}^T r_{t,a_t} \right]$$

Lecture 17: The Multi-Armed Bandit Problem
Two Models

• For convenience exposition, first describe simpler form
  – Disjoint linear model

• Then consider the general case
  – hybrid model

⇒ LinUCB is a generic contextual bandit algorithm which applies to applications other than personalized news article recommendation.
Linear Disjoint Model

- We assume the expected payoff of an arm $a$ is linear in its $d$-dimensional feature $x_{t,a}$ with some unknown coefficient vector $\theta_a^*$; namely for all $t$,

$$
E[r_{t,a}|x_{t,a}] = x_{t,a}^T \theta_a^*
$$

- The model is called disjoint because the parameters are not shared among different arms.
“Disjoint” in example

Article Recommendation

Users $u_1$ with age YOUNG and $u_2$ with age OLD

Retirement planning wishes vs. reality

The Player Wizarding World of Harry Potter ride may conjure a new path for theme park rides

Elon Musk: 198,000 Tesla Model 3 Orders Received in 24 Hours

Not tired yet: Warriors top Spurs for 72nd win, set up date with history
Algorithm 1 LinUCB with disjoint linear models.

0: Inputs: $c_t \in \mathbb{R}_+$
1: for $t = 1, 2, 3, \ldots, T$ do
2:   Observe features of all arms $a \in \mathcal{A}_t$: $x_{t,a} \in \mathbb{R}^d$
3:   for all $a \in \mathcal{A}_t$ do
4:     if $a$ is new then
5:       $A_a \leftarrow I_d$ (d-dimensional identity matrix)
6:       $b_a \leftarrow 0_{d \times 1}$ (d-dimensional zero vector)
7:     end if
8:     $\hat{\theta}_a \leftarrow A_a^{-1}b_a$
9:     $p_{t,a} \leftarrow \hat{\theta}_a^T x_{t,a} + c_t \sqrt{x_{t,a}^T A_a^{-1} x_{t,a}}$
10:   end for
11: Choose arm $a_t = \arg \max_{a \in \mathcal{A}_t} p_{t,a}$ with ties broken arbitrarily, and observe a real-valued payoff $r_t$
12: $A_{a_t} \leftarrow A_{a_t} + x_{t,a_t} x_{t,a_t}^T$
13: $b_{a_t} \leftarrow b_{a_t} + r_t x_{t,a_t}$
14: end for
Lecture 17: The Multi-Armed Bandit Problem

Visualization Representation

\[ c_t \sqrt{x_{t,a}^T A_a^{-1} x_{t,a}} \]

\[ \hat{\theta}^T x_{t,a} \]
Feature-free bandit v.s. linear bandit

Feature-free bandit

- $\mathbb{E}[r_{t,a}|x_{t,a}] = \mu_a^*$. 
- $\mu_a^*$ is not known a priori.
- Confidence interval $C_{t,a}$

$$ \{\mu_a : \frac{|\mu_a - \bar{\mu}_{t,a}|}{1/\sqrt{n_{t,a}}} \leq \sqrt{2\log t} \} $$

$$ a_t = \operatorname{arg\,max} \max_{a \in \{1,\ldots,K\}} \mu_a \max_{a \in C_{t,a}} $$

$$ = \operatorname{arg\,max} \bar{\mu}_{t,a} + \sqrt{\frac{2\log t}{n_{t,a}}} $$
Feature-free bandit v.s. linear bandit

Feature-free bandit

- $\mathbb{E}[r_{t,a}|x_{t,a}] = \mu_a^*$.  
- $\mu_a^*$ is not known a priori.
- Confidence interval $C_{t,a}$

- $\{\mu_a : \frac{|\mu_a - \bar{\mu}_{t,a}|}{1/\sqrt{n_{t,a}}} \leq \sqrt{2 \log t}\}$

- $a_t = \arg\max_{a \in \{1,\ldots,K\}} \max_{\mu_a \in C_{t,a}} \mu_a$

  $= \arg\max_{a \in \{1,\ldots,K\}} \bar{\mu}_{t,a} + \sqrt{\frac{2 \log t}{n_{t,a}}}$

Linear bandit

- $\mathbb{E}[r_{t,a}|x_{t,a}] = x_{t,a}^T \theta_a^*$.  
- $\theta_a^*$ is not known a priori.
- Confidence ellipsiod $C_{t,a}$

- $\{\theta_a : ||\theta_a - \hat{\theta}_{t,a}||_{A_{t,a}} \leq c_t\}$

  where $||x||_A \equiv \sqrt{x^T A x}$.

- $a_t = \arg\max_{a \in \{1,\ldots,K\}} \max_{\theta_a \in C_{t,a}} x_{t,a}^T \theta_a$

  $= \arg\max_{a \in \{1,\ldots,K\}} x_{t,a}^T \hat{\theta}_{t,a} + c_t \sqrt{x_{t,a}^T A_{t,a}^{-1} x_{t,a}}$
Confidence ellipsoid $C_{t,a} = \{ \theta_a : \| \theta_a - \hat{\theta}_{t,a} \|_{A_{t,a}} \leq c_t \}$

$$x_{t,a}^T \hat{\theta}_{t,a} + c_t \sqrt{x_{t,a}^T A_{t,a}^{-1} x_{t,a}} = \max_{\theta_a} x_{t,a}^T \theta_a$$

s.t. $$(\theta_a - \hat{\theta}_{t,a})^T A_{t,a} (\theta_a - \hat{\theta}_{t,a}) \leq c_t$$
Feature-free bandit $= \text{Linear bandit with } x_{t,a} \equiv 1, \theta_a = \mu_a$

Feature-free bandit
- $\mathbb{E}[r_{t,a}|x_{t,a}] = 1^T \mu^*_a$.
- $\mu^*_a$ is not known a priori.
- Confidence interval $C_{t,a}$
  
  \[
  \{\mu_a : \|\mu_a - \bar{\mu}_{t,a}\|_{n_{t,a}} \leq \sqrt{2 \log t}\}
  \]

  where $\|\mu\|_n \equiv \sqrt{\mu^T n \mu}$.

  $a_t = \arg \max_{a \in \{1, \ldots, K\}} \max_{\mu_a \in C_{t,a}} 1^T \mu_a$

  $= \arg \max_{a \in \{1, \ldots, K\}} 1^T \bar{\mu}_{t,a} + \sqrt{\frac{2 \log t}{n_{t,a}}}$

Linear bandit
- $\mathbb{E}[r_{t,a}|x_{t,a}] = x_{t,a}^T \theta^*_a$.
- $\theta^*_a$ is not known a priori.
- Confidence ellipsiod $C_{t,a}$
  
  \[
  \{\theta_a : \|\theta_a - \hat{\theta}_{t,a}\|_{A_{t,a}} \leq c_t\}
  \]

  where $\|x\|_A \equiv \sqrt{x^T A x}$.

  $a_t = \arg \max_{a \in \{1, \ldots, K\}} \max_{\theta_a \in C_{t,a}} x_{t,a}^T \theta_a$

  $= \arg \max_{a \in \{1, \ldots, K\}} x_{t,a}^T \hat{\theta}_{t,a} + c_t \sqrt{x_{t,a}^T A_{t,a}^{-1} x_{t,a}}$
A Bayesian approach to derive $\hat{\theta}_{t,a}$ and $A_{t,a}^{-1}$

- Gaussian prior $p_0(\theta_a) \sim \mathcal{N}(0, \lambda I_d)$. 

The reward $x^T_t \theta_a \sim \mathcal{N}(x^T_t \hat{\theta}_{t,a}, x^T_t A_{t,a}^{-1} x_t, a)$. The upper confidence bound (UCB) is $x^T_t \hat{\theta}_{t,a} + c t \sqrt{x^T_t A_{t,a}^{-1} x_t, a}$. 
A Bayesian approach to derive $\hat{\theta}_{t,a}$ and $A_{t,a}^{-1}$

- Gaussian prior $p_0(\theta_a) \sim \mathcal{N}(0, \lambda I_d)$.

- $n_{t,a}$ noisy measurements: $y_{t,a} \sim \mathcal{N}(D_{t,a} \theta_a, I_{n_{t,a}})$.

\[
\begin{bmatrix}
\vdots \\
y_{t,a}(i)
\end{bmatrix} = \begin{bmatrix} \vdots \\ x_{i,a}^T \end{bmatrix} \theta_a + \begin{bmatrix} \vdots \\ \eta_{i,a} \end{bmatrix}
\]
A Bayesian approach to derive $\hat{\theta}_{t,a}$ and $A_{t,a}^{-1}$

- Gaussian prior $p_0(\theta_a) \sim \mathcal{N}(0, \lambda I_d)$.

- $n_{t,a}$ noisy measurements: $y_{t,a} \sim \mathcal{N}(D_{t,a} \theta_a, I_{n_{t,a}})$.

$$
\begin{bmatrix}
\vdots \\
y_{t,a}(i) \\
\vdots
\end{bmatrix}
= 
\begin{bmatrix}
\vdots \\
x_{i,a}^T \\
\vdots
\end{bmatrix}
\theta_a + 
\begin{bmatrix}
\vdots \\
\eta_{i,a} \\
\vdots
\end{bmatrix}
$$

- Posterior distribution $p_{t,a}(\theta_a) \sim \mathcal{N}(\hat{\theta}_{t,a}, A_{t,a}^{-1})$.

$$
\hat{\theta}_{t,a} = (D_{t,a}^T D_{t,a} + \frac{1}{\lambda} I_d)^{-1} D_{t,a}^T y_{t,a},
$$

$$
A_{t,a} = D_{t,a}^T D_{t,a} + \frac{1}{\lambda} I_d.
$$
A Bayesian approach to derive $\hat{\theta}_{t,a}$ and $A_{t,a}^{-1}$

- Gaussian prior $p_0(\theta_a) \sim \mathcal{N}(0, \lambda I_d)$.

- $n_{t,a}$ noisy measurements: $y_{t,a} \sim \mathcal{N}(D_{t,a} \theta_a, I_{n_{t,a}})$.

$$
\begin{bmatrix}
\vdots \\
y_{t,a}(i)
\end{bmatrix} = \begin{bmatrix}
\vdots \\
\end{bmatrix} x_{i,a}^T \theta_a + \begin{bmatrix}
\vdots \\
\end{bmatrix} \eta_{i,a}
$$

- Posterior distribution $p_{t,a}(\theta_a) \sim \mathcal{N}(\hat{\theta}_{t,a}, A_{t,a}^{-1})$.

$$
\hat{\theta}_{t,a} = (D_{t,a}^T D_{t,a} + \frac{1}{\lambda} I_d)^{-1} D_{t,a}^T y_{t,a},
$$

$$
A_{t,a} = D_{t,a}^T D_{t,a} + \frac{1}{\lambda} I_d.
$$

- The reward $x_{t,a}^T \theta_a \sim \mathcal{N}(x_{t,a}^T \hat{\theta}_{t,a}, x_{t,a}^T A_{t,a}^{-1} x_{t,a})$. The upper confidence bound (UCB) is

$$
x_{t,a}^T \hat{\theta}_{t,a} + c_t \sqrt{x_{t,a}^T A_{t,a}^{-1} x_{t,a}}.
$$
A general form of linear bandit

Disjoint linear model
- $\mathbb{E}[r_{t,a}|x_{t,a}] = x_{t,a}^T \theta_a^*.$

- $a_t = \arg \max_{a \in \{1, \ldots, K\}} \max_{\theta_a \in \mathcal{C}_{1,a}} x_{t,a}^T \theta_a$

A general linear model
- $\mathbb{E}[r_t|x_t] = x_t^T \theta^*.$

- $x_t = \arg \max_{x \in \mathcal{A}_t} \max_{\theta \in \mathcal{C}_t} x^T \theta$
A general form of linear bandit

Disjoint linear model

\[ \mathbb{E}[r_{t,a}|x_{t,a}] = x_{t,a}^T \theta^*_a. \]

\[ a_t = \arg \max_{a \in \{1,\ldots,K\}} \max_{\theta_a \in C_{t,a}} x_{t,a}^T \theta_a \]

A general linear model

\[ \mathbb{E}[r_t|x_t] = x_t^T \theta^*. \]

\[ x_t = \arg \max_{x \in \mathcal{A}_t} \max_{\theta \in C_t} x^T \theta \]

\[ \theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_a \\ \vdots \\ \theta_K \end{bmatrix}, \quad \mathcal{A}_t = \left\{ \begin{bmatrix} \vdots \\ 0 \\ x_{t,a} \\ 0 \\ \vdots \end{bmatrix} : a = 1, 2, \ldots, K \right\} \]
A hybrid linear model

\[ \mathbb{E}[r_{t,a}|z_{t,a}, x_{t,a}] = z_{t,a}^T \beta^* + x_{t,a}^T \theta_a^*. \]

\[ a_t = \arg \max_{a \in \{1, \ldots, K\}} \max_{\beta, \theta_a \in C_t} z_{t,a}^T \beta + x_{t,a}^T \theta_a \]

A general linear model

\[ \mathbb{E}[r_t|x_t] = x_t^T \theta^*. \]

\[ x_t = \arg \max_{x \in \mathcal{A}_t} \max_{\theta \in C_t} x^T \theta \]

\[ \theta = \begin{bmatrix} \beta \\ \theta_1 \\ \vdots \\ \theta_a \\ \vdots \\ \theta_K \end{bmatrix}, \quad \mathcal{A}_t = \begin{bmatrix} z_{t,a} \\ \vdots \\ 0 \\ x_{t,a} \\ 0 \\ \vdots \end{bmatrix} : a = 1, 2, \ldots, K \]
A general form of linear bandit, continued

Disjoint linear model

\[ a_t = \arg \max_{a \in \{1, \ldots, K\}} \max_{\theta_a \in C_{t,a}} x_{t,a}^T \theta_a \]

\[ C_{t,a} = \{ \theta_a : \| \theta_a - \hat{\theta}_{t,a} \|_{A_{t,a}} \leq c_t \} \]

\[ \hat{\theta}_{t,a} = (D_{t,a}^T D_{t,a} + \frac{1}{\lambda} I_d)^{-1} D_{t,a}^T y_{t,a}, \]

\[ A_{t,a} = D_{t,a}^T D_{t,a} + \frac{1}{\lambda} I_d. \]
A general form of linear bandit, continued

Disjoint linear model

\[ a_t = \arg \max_{a \in \{1, \ldots, K\}} \max_{\theta_a \in C_{t,a}} x_{t,a}^T \theta_a \]

\[ C_{t,a} = \{ \theta_a : \| \theta_a - \hat{\theta}_{t,a} \|_{A_{t,a}} \leq c_t \} \]

\[ \hat{\theta}_{t,a} = (D_{t,a}^T D_{t,a} + \frac{1}{\lambda} I_d)^{-1} D_{t,a}^T y_{t,a} , \]

\[ A_{t,a} = D_{t,a}^T D_{t,a} + \frac{1}{\lambda} I_d . \]

A general linear model

\[ x_t = \arg \max_{x \in \mathcal{A}_t} \max_{\theta \in C_t} x^T \theta \]

\[ C_t = \{ \theta : \| \theta - \hat{\theta}_t \|_{A_t} \leq c_t \} \]

\[ \hat{\theta}_t = (D_t^T D_t + \frac{1}{\lambda} I_d)^{-1} D_t^T y_t , \]

\[ A_t = D_t^T D_t + \frac{1}{\lambda} I_d . \]
An $O(d \sqrt{T})$ regret bound

Theorem (Theorem 2 + Theorem 3 in APS 2011)

Assume that

1. The measurement noise $\eta_t$ is independent of everything and is $\sigma$-sub-Gaussian for some $\sigma > 0$, i.e., $\mathbb{E}[e^{\lambda \eta_t}] \leq \exp\left(\frac{\lambda^2 \sigma^2}{2}\right)$ for all $\lambda \in \mathbb{R}$.

2. For all $t$ and all $x \in A_t$, $x^T \theta^* \in [-1, 1]$.

Then, for any $\delta > 0$, with probability at least $1 - \delta$, for all $t \geq 0$,

1. $\theta^*$ lies in the confidence ellipsoid

$$C_t = \left\{ \theta : \|\theta - \hat{\theta}_t\|_{A_t} \leq c_t := \sigma \sqrt{\log \det A_t + d \log \lambda + 2 \log \frac{1}{\delta} + \frac{\|\theta^*\|}{\sqrt{\lambda}}} \right\}$$

2. The regret of the linUCB algorithm satisfies

$$R_t = \sqrt{8t} \sqrt{\log \det A_t + d \log \lambda} \left( \sigma \sqrt{\log \det A_t + d \log \lambda + 2 \log \frac{1}{\delta} + \frac{\|\theta^*\|}{\sqrt{\lambda}}} \right)_{	ext{III}, c_t}$$
An $O(d \sqrt{T})$ regret bound

Theorem (Theorem 2 + Theorem 3 in APS 2011)

Assume that

1. The measurement noise $\eta_t$ is independent of everything and is $\sigma$-sub-Gaussian for some $\sigma > 0$, i.e., $\mathbb{E}[e^{\lambda \eta_t}] \leq \exp\left(\frac{\lambda^2 \sigma^2}{2}\right)$ for all $\lambda \in \mathbb{R}$.
2. For all $t$ and all $x \in A_t$, $x^T \theta^* \in [-1, 1]$.

Then, for any $\delta > 0$, with probability at least $1 - \delta$, for all $t \geq 0$,

1. $\theta^*$ lies in the confidence ellipsoid

   $$ C_t = \left\{ \theta : \|\theta - \hat{\theta}_t\|_{A_t} \leq c_t := \sigma \sqrt{\log \det A_t + d \log \lambda + 2 \log \frac{1}{\delta} + \frac{\|\theta^*\|}{\sqrt{\lambda}}} \right\} $$

2. The regret of the linUCB algorithm satisfies

   $$ R_t = \sqrt{8t} \left\{ \sqrt{\log \det A_t + d \log \lambda} \left( \sigma \sqrt{\log \det A_t + d \log \lambda + 2 \log \frac{1}{\delta} + \frac{\|\theta^*\|}{\sqrt{\lambda}}} \right) \right\}_{\text{III, } c_t} $$

Lemma (Determinant-Trace Inequality, Lemma 10 in APS 2011)

If for all $t \geq 0$, $\|x_t\|_2 \leq L$ then

$$ \log \det A_t \leq d \log \left( \frac{1}{\lambda} + \frac{tL^2}{d} \right) $$
The ♥ of the proof

We consider the high probability event $\theta^* \in C_t$ for all $t \geq 0$.

\[
\begin{align*}
  r_t &= \langle x_t^*, \theta^* \rangle - \langle x_t, \theta^* \rangle \\
  x_t, \tilde{\theta}_t &= \arg \max_{x \in A_t} \max_{\theta \in C_t} \langle x, \theta \rangle \\
  &\leq \langle x_t, \tilde{\theta}_t \rangle - \langle x_t, \theta^* \rangle \\
  &\leq \langle x_t, \tilde{\theta}_t - \theta^* \rangle \\
  &\leq \langle x_t, \hat{\theta}_t - \theta^* \rangle + \langle x_t, \tilde{\theta}_t - \hat{\theta}_t \rangle \\
  &\leq \|x_t\|_{A_t^{-1}} \|\hat{\theta}_t - \theta^*\|_{A_t} + \|x_t\|_{A_t^{-1}} \|	ilde{\theta}_t - \hat{\theta}_t\|_{A_t} \\
  &\leq 2c_t \|x_t\|_{A_t^{-1}} \\
  \theta^*, \tilde{\theta}_t \in C_t &= \{\theta : \|\theta - \hat{\theta}_t\|_{A_t} \leq c_t\}
\end{align*}
\]
The heart of the proof

We consider the high probability event $\theta^* \in C_t$ for all $t \geq 0$.

$$r_t = \langle x_t^*, \theta^* \rangle - \langle x_t, \theta^* \rangle$$

$$x_t, \tilde{\theta}_t = \arg \max_{x \in \mathcal{A}_t} \max_{\theta \in C_t} \langle x, \theta \rangle$$

$$\leq \langle x_t, \tilde{\theta}_t \rangle - \langle x_t, \theta^* \rangle$$

$$= \langle x_t, \tilde{\theta}_t - \theta^* \rangle$$

$$= \langle x_t, \hat{\theta}_t - \theta^* \rangle + \langle x_t, \tilde{\theta}_t - \hat{\theta}_t \rangle$$

$$\leq \|x_t\|_{A^{-1}_t} \|\hat{\theta}_t - \theta^*\|_{A_t} + \|x_t\|_{A^{-1}_t} \|\tilde{\theta}_t - \hat{\theta}_t\|_{A_t}$$

Cauchy-Schwarz

$$\leq 2c_t \|x_t\|_{A^{-1}_t} \theta^*, \tilde{\theta}_t \in C_t = \{\theta : \|\theta - \hat{\theta}_t\|_{A_t} \leq c_t\}$$

Since $x^T \theta^* \in [-1, 1]$ for all $x \in \mathcal{A}_t$, then we have $r_t \leq 2$. Therefore,

$$r_t \leq \min \{2c_t \|x_t\|_{A^{-1}_t}, 2\} \leq 2c_t \min \{\|x_t\|_{A^{-1}_t}, 1\}$$
The heart of the proof, continued

\[
r_t^2 \leq 4c_t^2 \min\{\|x_t\|_{A_t^{-1}}^2, 1\}
\]  

(1)
The heart of the proof, continued

\[ r_t^2 \leq 4c_t^2 \min\{\|x_t\|_A^{-1}, 1\} \]  \hspace{1cm} (1)

Consider the regret \( R_T \equiv \sum_{t=1}^{T} r_t \),

\[
R_T \leq \sqrt{T \sum_{t=1}^{T} r_t^2} \leq \sqrt{T \sum_{t=1}^{T} 4c_t^2 \min\{\|x_t\|_A^{-1}, 1\}} \\
\leq 2 \sqrt{T} \sum_{t=1}^{T} \min\{\|x_t\|_A^{-1}, 1\} \quad c_t \text{ is monotonically increasing.}
\]
The heart of the proof, continued

\[ r_t^2 \leq 4c_t^2 \min\{\|x_t\|_{A_t^{-1}}^2, 1\} \quad (1) \]

Consider the regret \( R_T \equiv \sum_{t=1}^{T} r_t \),

\[
R_T \leq \sqrt{T \sum_{t=1}^{T} r_t^2} \leq \sqrt{T \sum_{t=1}^{T} 4c_t^2 \min\{\|x_t\|_{A_t^{-1}}^2, 1\}} \\
\leq 2 \sqrt{T} c_T \sum_{t=1}^{T} \min\{\|x_t\|_{A_t^{-1}}^2, 1\} \\
\leq 2 \log(1 + \|x_T\|_{A_T^{-1}}) = 2(\log \det A_t + d \log \lambda).
\]

Since \( x \leq 2 \log(1 + x) \) for \( x \in [0, 1] \), we have

The last equality is proved in Lemma 11 in APS_2011.
Algorithm Evaluation

How do we evaluate the performance of a recommendation algorithm?

• Can we just run the algorithm on “live” data?

• Build a simulator to model the bandit process, evaluate the algorithm based on the simulated data?
Algorithm Evaluation

How do we evaluate the performance of a recommendation algorithm?

• Can we just run the algorithm on “live” data? Difficult logistically.
• Build a simulator to model the bandit process, evaluate the algorithm based on the simulated data? May introduce bias from the simulator.

• Yahoo! Today Module! (random article)
Algorithm Evaluation

0: Inputs: $T > 0$, algorithm $\pi$, stream of events
1: $h_0 = 0$, $R_0 = 0$
2: for $t = 1, 2, \ldots, T$ do
3: \hspace{1em} repeat
4: \hspace{2em} Get next event $(x_1, \ldots, x_K, a, r_a)$
5: \hspace{2em} until $\pi(h_{t-1}, (x_1, \ldots, x_K)) = a$
6: \hspace{2em} $h_t \leftarrow \text{add}(h_{t-1}, (x_1, \ldots, x_K, a, r_a))$
7: \hspace{2em} $R_t \leftarrow R_{t-1} + r_a$
8: end for
9: Output $R_t / T$
Algorithm Evaluation

0 : Inputs: $T > 0$, algorithm $\pi$, stream of events
1 : $h_0 = 0$, $R_0 = 0$
2 : for $t = 1, 2, ..., T$ do
3 :   repeat
4 :     Get next event $(x_1, ..., x_K, a, r_a)$
5 :     until $\pi(h_{t-1}, (x_1, ..., x_K)) = a$
6 :     $h_t \leftarrow \text{add}(h_{t-1}, (x_1, ..., x_K, a, r_a))$
7 :     $R_t \leftarrow R_{t-1} + r_a$
8 : end for
9 : Output $R_t / T$

No bias! About $T$ events are accepted from TK trails.
Algorithm Evaluation (data collection)

Yahoo News ...

Randomly shoot user an article $a$ as highlighted news.
Algorithm Evaluation (data collection)

Yahoo News ...

Randomly shoot user an article $a$ as highlighted news.

This event contains article $a$, feature vector $X$, and response $r = 1/0$.

Accept this event iff algorithm predicts the same article $a$. 
Algorithm Evaluation (construct features)

In Yahoo’s data base, either a user or article is depicted by hundreds raw features.

Need to reduce the feature dimensions.
Algorithm Evaluation (construct features)

In Yahoo’s data base, either a user or article is depicted by hundreds raw features.

Need to reduce the feature dimensions.

\[
\begin{array}{c|c|c|c|c|c|c}
\text{Article} & \text{Long} & \text{Domestic} & \text{Tech} & \text{Politics} & \ldots \\
\phi_a & 1 & 1 & 0 & 1 & \ldots \\
\end{array}
\]

\[
\begin{array}{c|c|c|c|c|c|c}
\text{User} & \text{Gender} & \text{Age}>20 & \text{Age}>40 & \text{Student} & \ldots \\
\phi_u & 1 & 1 & 0 & 1 & \ldots \\
\end{array}
\]
Algorithm Evaluation (construct features)

In Yahoo’s data base, either a user or article is depicted by hundreds raw features.

Need to reduce the feature dimensions.

| Raw | $\phi_a$ | Article | Long | Domestic | Tech | Politics | ...
|-----|---------|---------|------|----------|------|----------|------
|     | $\phi_a$ | 1       | 1    | 0        | 1    | ...      |

| Raw | $\phi_u$ | User | Gender | Age>20 | Age>40 | Student | ...
|-----|---------|------|--------|--------|--------|---------|------
|     | $\phi_u$ | 1    | 1      | 0      | 1      | ...      |

Suppose there’s a weight matrix $W$, st. the probability of user clicking on article $a$ is:

$$ P = \phi_u^T W \phi_a $$
Algorithm Evaluation (construct features)

\[ P = \phi_u^T W \phi_a \]

Logistic Regression to get \( W \)

K-means method to find clusters/groups of the users.

\[ \psi_u \equiv \phi_u^T W \quad \text{(Cluster this projected feature vector.)} \]
Algorithm Evaluation (construct features)

\[ P = \phi_u^T W \phi_a \]

Logistic Regression to get \( W \)

K-means method to find clusters/groups of the users.

\[ \psi_u \equiv \phi_u^T W \]  (Cluster this projected feature vector.)

- The constructed feature vector for a user would be the possibilities of being in different groups. Denote as: \( x_{t,a} \)

- The same procedure can be applied to the article and get the constructed feature vector.

- **Disjointed LinUCB**, use \( x_{t,a} \) as input data.
- **Hybrid model**, the outer product of constructed user and article feature vectors is also included as global features.
Algorithm Evaluation (construct features)

For example, we have binary raw feature vectors:

$$\phi_a = (1, 1, 0, 1, 1, ...) \quad \phi_u = (1, 1, 0, 1, 0, ...)$$
Algorithm Evaluation (construct features)

For example, we have binary raw feature vectors for user and article:
\[ \phi_a = (1, 1, 0, 1, 1, ...) \quad \phi_u = (1, 1, 0, 1, 0, ...) \]

- After clustering:

<table>
<thead>
<tr>
<th>Group</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Membership</td>
<td>0.2</td>
<td>0.1</td>
<td>0.35</td>
<td>0.3</td>
<td>0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Membership</td>
<td>0.7</td>
<td>0.05</td>
<td>0.15</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Algorithm Evaluation (construct features)

For example, we have binary raw feature vectors for user and article:

\[
\phi_a = (1, 1, 0, 1, 1, \ldots) \quad \phi_u = (1, 1, 0, 1, 0, \ldots)
\]

- After clustering:

<table>
<thead>
<tr>
<th>Group</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Membership</td>
<td>0.2</td>
<td>0.1</td>
<td>0.35</td>
<td>0.3</td>
<td>0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Membership</td>
<td>0.7</td>
<td>0.05</td>
<td>0.15</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>

- The outer product of the two constructed feature vectors is:

\[
z_{t,a} \quad (25 \text{ dimensional here})
\]

- Hybrid model:

\[
E[r_{t,a} | x_{t,a}] = z_{t,a}^T \beta^* + x_{t,a}^T \theta_a^*
\]

Remove this term will get back to disjointed LinUCB
Algorithm Evaluation

Without utilizing features:
• Purely Random
• E-greedy
• UCB
• Omniscient

Algorithms with features:
• E-greedy (run on different clusters)
• E-greedy (hybrid, epoch greedy)
• UCB (cluster)
• LinUCB (disjoint)
• LinUCB (hybrid)
Algorithm Evaluation

(CTR normalized by the random recommendation CTR)
Algorithm Evaluation

(CTR normalized by the random recommendation CTR)
Conclusion

For multi-armed bandits problem

• UCB algorithm without feature has regret bound:
  \[ R_T = O\left(\frac{K}{\epsilon} \ln T\right) \]

• LinUCB using feature vectors has regret bound:
  \[ R_T = O(D\sqrt{T}) \]

• Evaluate using Yahoo Front Page Today Module data.
• Introducing contextual information (features) to the recommendation algorithm, the CTR (reward) has been improved by about 10%.
References