

# Contextual Combinatorial Cascading Bandits

Shuai Li<sup>1</sup>, Baoxiang Wang<sup>1</sup>, Shengyu Zhang<sup>1</sup>, Wei Chen<sup>2</sup>

1 The Chinese University of Hong Kong

2 Microsoft Research

ICML 2016

# Multi-armed Bandit Problem

- A special case of **reinforcement learning**
- There are  $m$  arms (machines)
- Arm  $i$  has an unknown reward distribution with unknown mean  $\mu_i$ 
  - best arm  $\mu^* = \max \mu_i$



# Multi-armed Bandit Problem

- In each round  $t$ , the learning agent selects one arm  $i_t$  to play and observes the reward  $R_t(i_t)$

- Regret after playing  $T$  rounds: Always play the best arm

$$\text{Regret} = T\mu^* - \mathbb{E}[\sum_{t=1}^T R_t(i_t)]$$

- Objective: minimize regret in  $T$  rounds
- **Balancing tradeoff between exploitation and exploration**
  - **Exploration**: try options that have not been tried much before
  - **Exploitation**: try options that yield good results so far



# Multi-armed Bandit Problem

- UCB (Upper Confidence Bound) [Auer, Cesa-Bianchi, Fischer 2002]

- UCB policy: select

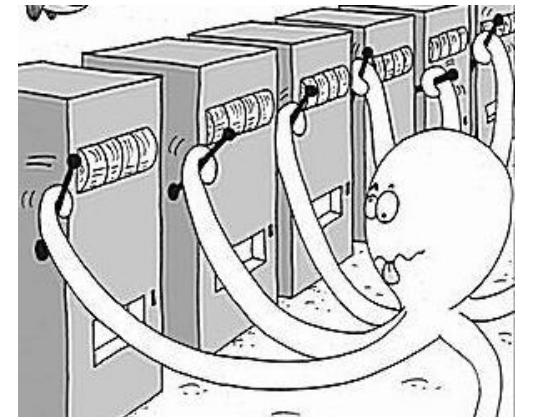
$$i = \operatorname{argmax}_{i \in [m]} \left( \hat{\mu}_i + \sqrt{\frac{2 \ln t}{T_i}} \right)$$

Exploitation

Exploration

where  $T_i$  is the played times of arm  $i$ .

- Gap-dependent bound  $O(\log T \sum_{i: \Delta_i > 0} 1/\Delta_i)$ ,  $\Delta_i = \mu^* - \mu_i$ , match lower bound
- Gap-free bound  $O(\sqrt{mT \log T})$ , tight up to a factor of  $\sqrt{\log T}$



# Combinatorial Multi-Armed Bandit

- Action is combinatorial
  - Selecting a matching, a routing path, a sequence of ads to display, a list of movies to recommend
- May observe some feedback on elements involved (e.g. semi-bandit feedback)
- Challenges
  - Exponential number of actions --- cannot be fully explored
  - Offline optimization may already be hard

The image shows a Google search for "machine learning". The search bar at the top contains the text "machine learning" and the Google logo. Below the search bar, there are tabs for "All", "Images", "News", "Videos", "Books", and "More". The search results show "About 21,400,000 results (0.58 seconds)".

The main search results include:

- Machine Learning Course - coursera.org**: A link to a course on Coursera with a rating of 8.4. The description mentions "Get Certified in Machine Learning Earn a U of Washington Certificate" and "Top Instructors · Flexible Schedule · Top Universities · Learn 24/7".
- Machine Learning Guide - "Statistics & Machine Learning"**: A link to a guide on SAS with a rating of 8.3. The description mentions "Download the Free Paper" and "Statistical Analytics · Predictive Analytics · Business Analytics".

There is also a section titled "In the news" with several articles:

- Machine Learning Is Redefining Business** by Forbes - 2 days ago. Bottom line: "Machine learning is providing new insights into customer behavior, and ..."
- A Gentle Introduction to the Basics of Machine Learning** by miguelgferro.com - 20 hours ago.
- SAPVoice: Welcome to the Machine (Learning)** by Forbes - 2 days ago.

Below the news section is a link for "More news for machine learning".

At the bottom, there is a Wikipedia entry for "Machine learning" with the URL [https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning). The text describes machine learning as a subfield of computer science that evolved from statistical learning theory in artificial intelligence. It mentions that in 1959, Arthur Samuel defined machine learning as a "Field of study that gives computers the ability to learn without being explicitly programmed".

On the right side of the search results, there is a "Recommended For You" section with several movie and TV series recommendations:

- Wei zhuang zhe (TV series 2015- )**: Drama, 8.4 rating, starring Kai Wang, Dong Jin, Ge Hu.
- The Terminator (1984)**: Action | Sci-Fi, 1 h 47 min, 8.1 rating, Metascore 83, starring Arnold Schwarzenegger, Linda Hamilton, Michael Biehn.
- Back to the Future (1985)**: Adventure | Comed..., 1 h 56 min, 8.5 rating, Metascore 86, starring Michael J. Fox, Christopher Lloyd, Lea Thompson.
- Breaking Bad (TV series 2008-2013)**: Crime | Drama | Thriller, 49 min, 9.5 rating, starring Bryan Cranston, Aaron Paul, Anna...
- Argo (2012)**: Drama | History, 2 h, 8.8 rating, starring Ben Affleck, Bryan Cranston, Alan Arkin.

# Motivation of Cascading Bandit

- Websites search results
- Recommended movies
- Etc.
- All are sequential lists
  - Users are likely to go through the list **from top down**
  - Stop at the first satisfactory item
  - Click as the feedback
  - Online feedback helps improving list quality

The image shows a Google search for "machine learning". The search bar at the top contains the text "machine learning" and the Google logo. Below the search bar, there are tabs for "All", "Images", "News", "Videos", "Books", and "More". The search results show "About 21,400,000 results (0.58 seconds)".

The main search results include:

- Machine Learning Course - coursera.org**: An advertisement for a course on Coursera, with a link to [www.coursera.org/machine-learning](http://www.coursera.org/machine-learning). The ad text says: "Get Certified in Machine Learning Earn a U of Washington Certif Top Instructors · Flexible Schedule · Top Universities · Learn 24/7".
- Machine Learning Guide - "Statistics & Machine Learning"**: An advertisement for a guide on SAS, with a link to [www.sas.com/machine-learning](http://www.sas.com/machine-learning). The ad text says: "Download the Free Paper. Statistical Analytics - Predictive Analytics - Business Analytics All Analytics Products - Buy SAS Analytics - Analytics White Paper".

Below the main results is a section titled "In the news" with a small thumbnail image. The text reads: "Machine Learning Is Redefining Business" by Forbes - 2 days ago. The bottom line says: "Machine learning is providing new insights into customer behavior, and ...".

There are also links to "A Gentle Introduction to the Basics of Machine Learning" by miguelgferro.com - 20 hours ago, and "SAPVoice: Welcome to the Machine (Learning)" by Forbes - 2 days ago.

At the bottom of the search results, there is a link to "More news for machine learning" and a Wikipedia entry for "Machine learning - Wikipedia, the free encyclopedia". The Wikipedia snippet reads: "Machine learning is a subfield of computer science that evolved from statistical learning theory in artificial intelligence. In 1959, Arthur Samuel defined machine learning as a "Field of study that gives computers the ability to learn without being explicitly programmed". List of machine learning ... - Supervised learning - Computational Intelligence and Data Science".

On the right side of the search results, there is a "Recommended For You" sidebar. It contains several movie and TV series recommendations:

- Wei zhuang zhe (TV series 2015- )**: Drama, 8.4 rating, starring Kai Wang, Dong Jin, Ge Hu.
- The Terminator (1984)**: Action | Sci-Fi, 1 h 47 min, 8.1 rating, Metascore 83, starring Arnold Schwarzenegger, Linda Hamilton, Michael Biehn.
- Back to the Future (1985)**: Adventure | Comed..., 1 h 56 min, 8.5 rating, Metascore 86, starring Michael J. Fox, Christopher Lloyd, Lea Thompson.
- Breaking Bad (TV series 2008-2013)**: Crime | Drama | Thriller, 49 min, 9.5 rating, starring Bryan Cranston, Aaron Paul, Anna...
- Argo (2012)**: Drama | History, 2 h, R rating, starring Ben Affleck, Bryan Cranston, John Goodman.

# Contextual Combinatorial Cascading Bandit

- Contexts

- User profiles, search keywords
- Important for search, recommendations

- Combinatorial

- Action is selection of a sequence
- May have other combinatorial constraints (children movies)

The image shows a Google search interface for the query "machine learning". The search results page includes a search bar with the Google logo, the search term "machine learning", and a search button. Below the search bar, there are navigation tabs for "All", "Images", "News", "Videos", "Books", and "More". The main search results area shows "About 21,400,000 results (0.58 seconds)". The first result is an advertisement for "Machine Learning Course - coursera.org" with a link to "www.coursera.org/machine-learning". The second result is another advertisement for "Machine Learning Guide - 'Statistics & Machine Learning'" with a link to "www.sas.com/machine-learning". Below the advertisements, there is a section titled "In the news" with a snippet from Forbes: "Machine Learning Is Redefining... Bottom line: Machine learning is providing... applications, and...". There are also links to "A Gentle Introduction to the Basics of Machine Learning" and "SAPVoice: Welcome to the Machine (Learning)". A "More news for machine learning" link is also present. The bottom of the search results shows a Wikipedia entry for "Machine learning - Wikipedia, the free encyclopedia".

The "Recommended For You" sidebar on the right side of the page features several recommendations:

- Wei zhuang zhe** (TV series 2015- )  
Drama  
★ 8.4  
Kai Wang, Dong Jin, Ge Hu
- The Terminator (1984)**  
1 h 47 min - Action | Sci-Fi  
★ 8.1 83 Metascore  
Arnold Schwarzenegger, Linda Hamilton, Michael Biehn
- Back to the Future (1985)**  
PG 1 h 56 min - Adventure | Comed...  
★ 8.5 86 Metascore  
Michael J. Fox, Christopher Lloyd, Lea Thompson
- Breaking Bad** (TV series 2008-2013)  
49 min - Crime | Drama | Thriller  
★ 9.5  
Bryan Cranston, Aaron Paul, Anna...
- Argo (2012)**  
1 h 26 min - Biography | Drama | History

# Our Contribution

- Formulate the **C**ontextual **C**ombinatorial **C**ascading **B**andits problem
- Proposed **C<sup>3</sup>-UCB** algorithm, handles
  - contextual information
  - cascading feedback
  - position discount (top positions may be more important)
  - general reward function

	context	cascading	Position discount	General reward
Combinatorial UCB <sup>1</sup>	No	Yes	No	Yes
Contextual Combinatorial UCB <sup>2</sup>	Yes	No	No	Yes
Comb-Cascade <sup>3</sup>	No	Yes	No	No
C <sup>3</sup> -UCB(ours)	Yes	Yes	Yes	Yes

- Theoretical analysis and empirical evaluation

1 Chen et al. 2013

2 Qin et al. 2014

3 Kveton et al. 2015



# Setting & Algorithms

# Setting of C<sup>3</sup>B

- $E = \{1, \dots, L\}$ : set of base arms
- Action  $A = (a_1, \dots, a_k)$ : a **sequence** of base arms
  - There is a feasible action set  $\mathcal{S}$ .
- At each time  $t \geq 1$ 
  - set of contexts  $\{x_{t,a}\}_{a \in E}$  are given (e.g. user/keyword features)
  - learning agent selects a feasible action  $\mathbf{A}_t = (\mathbf{a}_1^t, \dots, \mathbf{a}_{|\mathbf{A}_t|}^t)$
  - The user checks from the first item and stops at  $\mathbf{O}_t$ -th item.
  - Feedback: observe weights of first  $\mathbf{O}_t$  items,  $\mathbf{R}_t(\mathbf{a}_k^t), k \leq \mathbf{O}_t$ .

$$\mathbb{E}[\mathbf{R}_t(a)] = \theta_*^\top \cdot x_{t,a} = w_{t,a}$$

Fixed but unknown

# Setting of C<sup>3</sup>B

- Assume the expected reward of an action  $A$  is a function of  $w_t = \{w_{t,a}\}_{a \in E}$  of each base arm,  $f(A, w_t)$ .

- Regret in  $T$  rounds

$$\text{Regret} = \sum_{t=1}^T f_t^* - \mathbb{E} \left[ \sum_{t=1}^T f(A_t, w_t) \right]$$

Best cumulative reward











- $f_t^*$ : **max** expected reward in round  $t$

# Example – movie recommendation

- Each movie  $i$  has a feature vector  $m_i$
- At time  $t$ ,
  - A random user comes with feature vector  $u_t$
  - Use  $x_{i,t} = g(m_i, u_t)$ , a function of  $m_i$  and  $u_t$ , (e.g. direct sum, outer-product) as context
  - The learning agent recommends a list of movies  $A_t$
  - The user checks from the first movie and stops at the attractive one.
  - The learning agent receives reward  $\gamma_k$  if the user stops at position  $k$ .

$$1 = \gamma_k \geq \dots \geq \gamma_k \geq 0$$

Recommended For You

	<p>Wei zhuang zhe (TV series 2015- ) Drama ★ 8.4 Kai Wang, Dong Jin, Ge Hu</p>	
	<p>The Terminator (1984) R 1 h 47 min - Action   Sci-Fi ★ 8.1 <span>83</span> Metascore Arnold Schwarzenegger, Linda Hamilton, Michael Biehn</p>	
	<p>Back to the Future (1985) PG 1 h 56 min - Adventure   Comed... ★ 8.5 <span>86</span> Metascore Michael J. Fox, Christopher Lloyd, Lea Thompson</p>	
	<p>Breaking Bad (TV series 2008-2013) 49 min - Crime   Drama   Thriller ★ 9.5 Bryan Cranston, Aaron Paul, Anna...</p>	
	<p>Argo (2012) IIB - 2 h - Biography   Drama   History</p>	

# C<sup>3</sup>-UCB Algorithm

- For round  $t = 1, 2, \dots, T$

- obtain context:  $\{x_{t,a}\}_{a \in E}$

- From  $\mathbb{E}[\mathbf{R}(a)] = \theta_*^\top x_a = w_a$ , we get an estimate  $\hat{\theta}_{t-1}$  of  $\theta_*$ .  
(use linear regression, details omitted.)

With high probability

$$w_{t,a} \in (\hat{\theta}_{t-1}^\top x_{t,a} - \beta_{t-1} \|x_{t,a}\|_{V_{t-1}^{-1}}, \hat{\theta}_{t-1}^\top x_{t,a} + \beta_{t-1} \|x_{t,a}\|_{V_{t-1}^{-1}})$$

- The upper confidence bound (UCB) of base arms:

$$U_t(a) = \min \left\{ \hat{\theta}_{t-1}^\top x_{t,a} + \beta_{t-1} \|x_{t,a}\|_{V_{t-1}^{-1}}, 1 \right\}$$

- use offline oracle to find the best action for UCB:  $A_t = \mathcal{O}_S(U_t)$
- play action  $A_t$ , observe prefix feedback  $\mathbf{R}_t(a_k^t), j \leq O_t$
- update observations (details omitted)

# Result

- Regret bound in  $T$  rounds:

$$\text{Regret} = O\left(\frac{d}{p^*} \sqrt{TK} \ln(T)\right)$$

- $d$ : dimension of latent and feature vectors
  - $p^*$ : minimum probability of triggering all arms in a sequence
  - $K$ : largest length of the sequence
- Regret bound of disjunctive objective in  $T$  rounds:

$$\text{Regret} = O\left(\frac{d}{1 - f^*} \sqrt{TK} \ln(T)\right)$$

- $f^* = \max f_t^*$ : the maximal expected reward in  $T$  rounds.

# Result

	context	cascading	Position discount	General reward	Regret bound
Combinatorial UCB <sup>1</sup>	No	Yes	No	Yes	$O(m\sqrt{mT \log T})$
Contextual Combinatorial UCB <sup>2</sup>	Yes	No	No	Yes	$O(d\sqrt{T} \log T)$
Comb-Cascade <sup>3</sup>	No	Yes	No	No	$O\left(\sqrt{\frac{KLT \log T}{f^*}}\right)$
C <sup>3</sup> -UCB (ours)	yes	Yes	Yes	Yes	$O\left(\frac{dB}{p^*} \sqrt{TK} \ln(T)\right)$

1 Chen et al. 2013

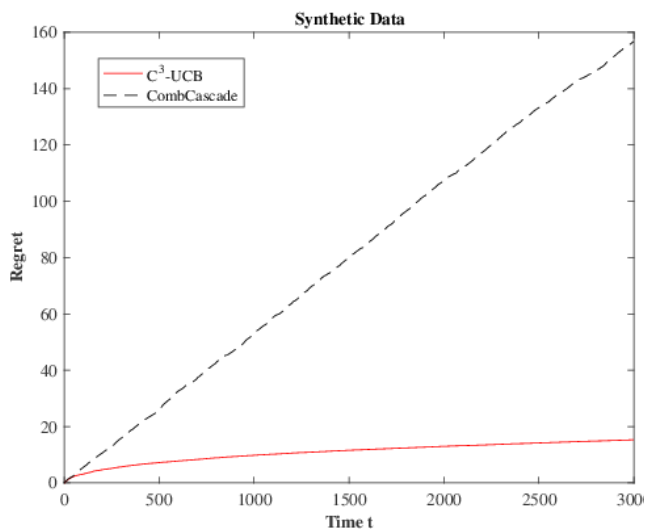
2 Qin et al. 2014

3 Kveton et al. 2015

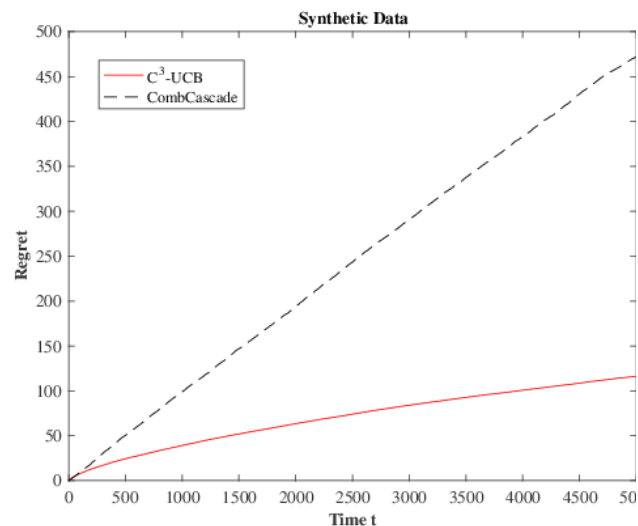
# Experimental Results



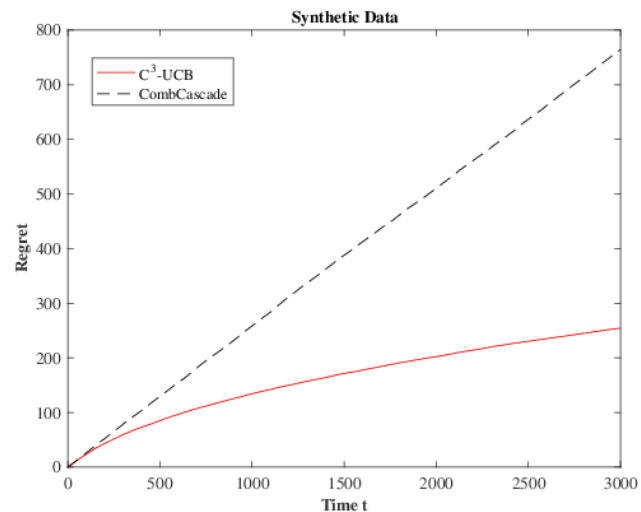
# Regret comparisons in Synthetic Data



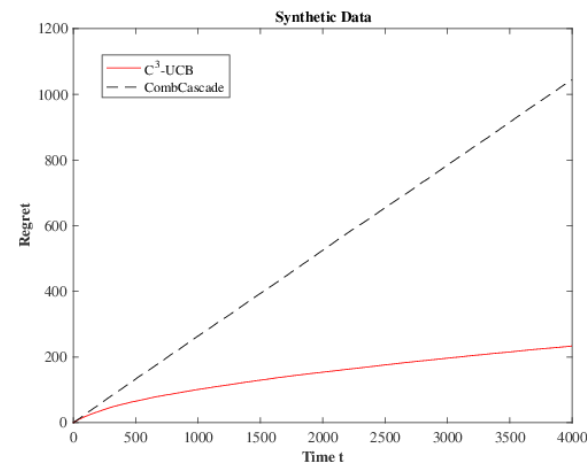
Disjunctive,  $\gamma_k = 1$   
9.77%



Disjunctive,  $\gamma_k = 0.9^{k-1}$   
24.6%



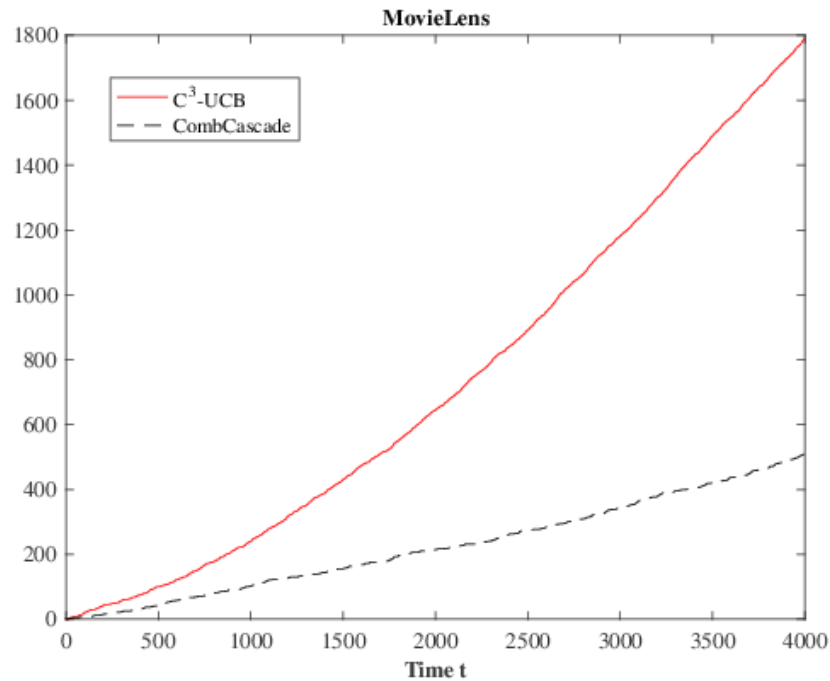
Conjunctive,  $\gamma_k = 1$   
33.3%



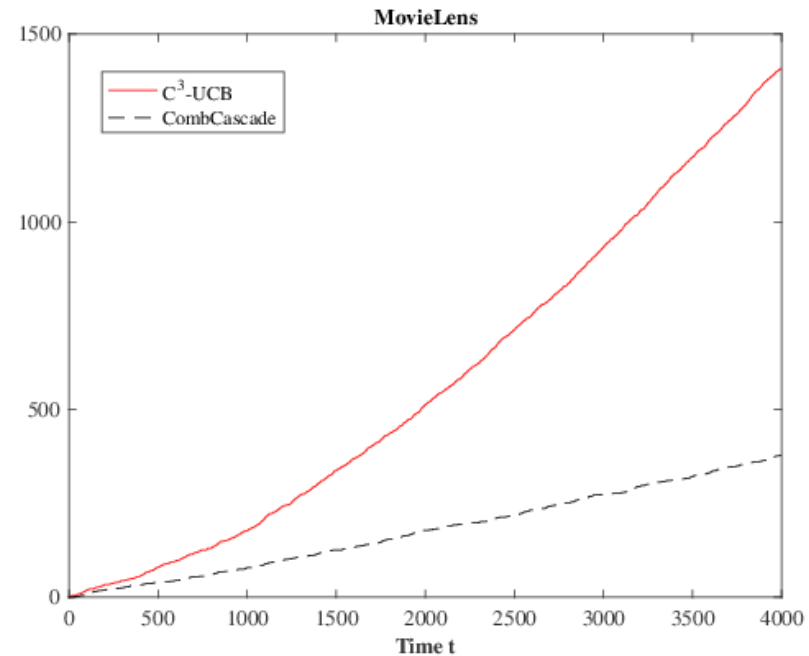
Conjunctive,  $\gamma_k = 0.9^{k-1}$   
22.4%

100 items, select 4 items  
latent and feature vector dimension = 4

# Reward comparisons in MovieLens



$\gamma_k = 1$   
3.52 times



$\gamma_k = 0.9^{k-1}$   
3.74 times

MovieLens dataset, 200 movies, select 4 items  
d= 400 (By SVD decomposition)

# Conclusions

- Incorporating contextual information to cascading bandit
- Advancing the research in combinatorial online learning
- Application potential
  - Any sequential list recommendation (search, ads, mobile recommendations)
    - Need online (real-time) feedback
- Future work
  - Theoretical lower bounds
  - Other non-sequential click models

Thank you!

Q & A