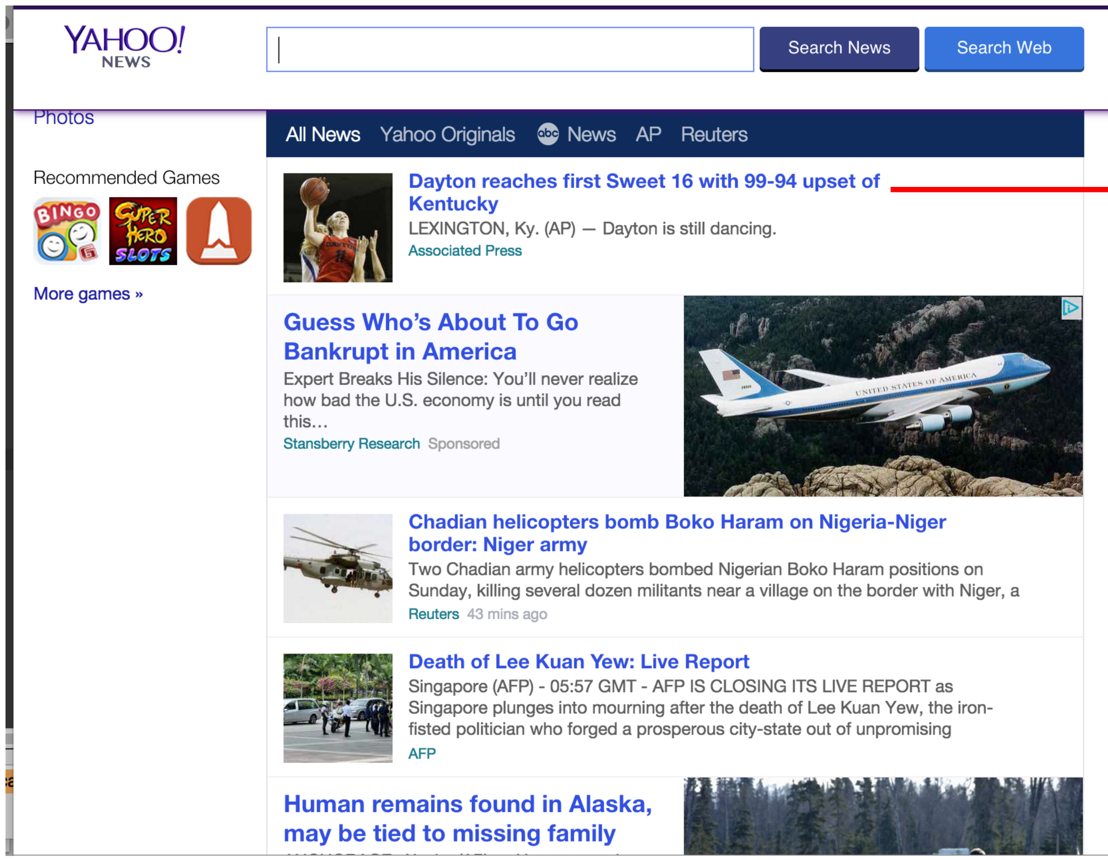
A Contextual-Bandit Approach to Personalized News Article Recommendation

Lihong li, Wei Chu, John Langford, Rebort E. Schapire

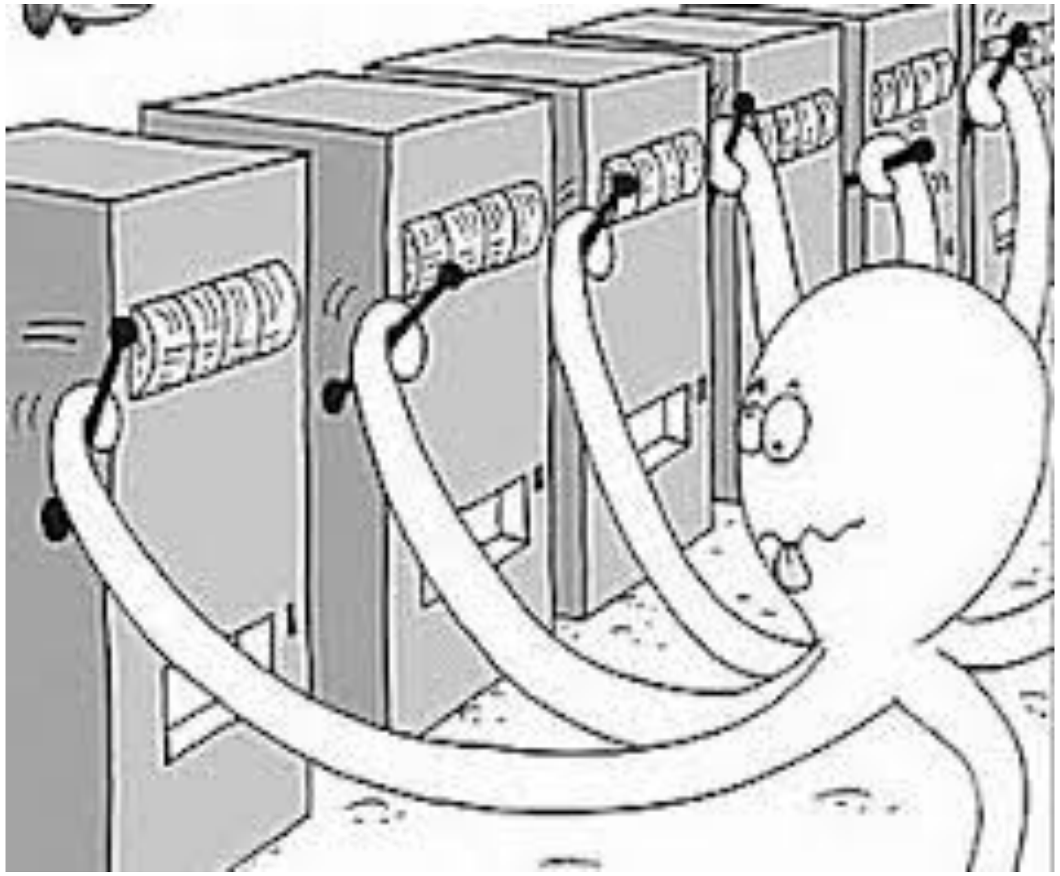
Presentator: Qingyun Wu

News Recommendation Cycle



A K-armed Bandit Formulation

- A gambler must decide which of the K non-identical slot machines(we called them arms) to play in a sequence of trails in order to maximize total reward.



News Website <—> gambler
Candidate news articles <—> arms
User Click <—> Reward

How to pull arms to maximize reward?

A K-armed Bandit formulation

- Setting
  - Set of K choices(arms)
  - Each choice a is associate with an unknown probability distribution p_a supported in [0,1]
  - play the game for T rounds

- In each round t

  (1)we pick article j

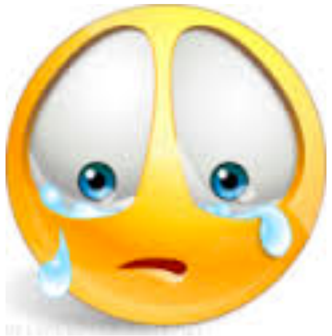
  (2)we observe random sample X_t from p_j

Our Goal: maximize $\sum_{t=1}^{T} X_t$

Ideal Solution

Pick $\arg\max_{a} \mu_a$

But we DO NOT know the mean.



Feasible Solution

Choices	X_1	X_2	X_3	X_4	X_5	X_6	...
a_1					1	1	
a_2	0		1	0			
...							
a_k		0					

Time

Every time we pull an arm we learn a bit more about the distribution.

Exploitation VS. Exploration

Exploitation: pull an arm for which we current have the highest estimate of mean of reward

Exploration: Pull an arm we never pulled before

Extreme examples:

Greedy Strategy:
Take the arm with the highest average reward

Random Strategy:
Randomly choose an arm

Too confident

Too unconfident

How to make trade off

Don’t just look at the mean(that’s the expected reward), but also the confidence!

UCB(Upper Confidence Bound) algorithm

Pick $$\arg\max_a(\hat{\mu}_a + \alpha * Varance)$$

$$\downarrow$$

Pick $$\arg\max_a(\hat{\mu}_a + \alpha * UCB)$$

Confidence Interval is a range of values within which we are sure the mean lies with a certain probability

UCB1 $$\arg\max_a(\hat{\mu}_a + \sqrt{\frac{2\ln T}{n_a}})$$

Reference: Finite-Analysis of the Multi-armed Bandit Problem, Peter Auer, Nicolo Cesa-Bianchi, Paul Fischer
http://homes.di.unimi.it/~cesabian/Pubblicazioni/ml-02.pdf

Make use of Contextual Information

- User feature: demographic information, geographic features, behavioral categories

- Article feature:URL categories, topic categories

Assumption about the reward:

The expected reward of an arm a is linear in its d-dimensional feature $x_{t,a}$, with some unknown coefficient vector θ_a^*, namely, for all t ,

$$E(r_{t,a} \mid x_{t,a}) = x_{t,a}^T \theta_a^*$$

UCB(Upper Confidence Bound) algorithm

Assumption $$E(r_{t,a} \mid x_{t,a}) = x_{t,a}^T \theta_a^*$$

Parameter Estimation $$\hat{\theta}_a = (D_a^T D_a + I_d)^{-1} D_a^T c_a$$ (Ridge Regression)

Bound of the variance $$\left| x_{t,a}^T \hat{\theta}_a - E(r_{t,a} \mid x_{t,a}) \right| \le \alpha \sqrt{x_{t,a}^T (D_a^T D_a + I_d)^{-1} x_{t,a}}$$

Bound we need!!!

Pick $$\arg\max_a \left(x_{t,a}^T \hat{\theta}_a + \alpha \sqrt{x_{t,a}^T (D_a^T D_a + I_d) x_{t,a}} \right)$$

Performance Evaluation

algorithm	size = 100%		size = 30%		size = 20%		size = 10%		size = 5%		size = 1%	
	deploy	learn	deploy	learn	deploy	learn	deploy	learn	deploy	learn	deploy	learn
ϵ-greedy	1.596 0%	1.326 0%	1.541 0%	1.326 0%	1.549 0%	1.273 0%	1.465 0%	1.326 0%	1.409 0%	1.292 0%	1.234 0%	1.139 0%
ucb	1.594 0%	1.569 18.3%	1.582 2.7%	1.535 15.8%	1.569 1.3%	1.488 16.9%	1.541 5.2%	1.446 9%	1.541 9.4%	1.465 13.4%	1.354 9.7%	1.22 7.1%
ϵ-greedy (seg)	1.742 9.1%	1.446 9%	1.652 7.2%	1.46 10.1%	1.585 2.3%	1.119 −12%	1.474 0.6%	1.284 −3.1%	1.407 0%	1.281 −0.8%	1.245 0.9%	1.072 −5.8%
ucb (seg)	1.781 11.6%	1.677 26.5%	1.742 13%	1.555 17.3%	1.689 9%	1.446 13.6%	1.636 11.7%	1.529 15.3%	1.532 8.7%	1.32 2.2%	1.398 13.3%	1.25 9.7%
ϵ-greedy (disjoint)	1.769 10.8%	1.309 −1.2%	1.686 9.4%	1.337 0.8%	1.624 4.8%	1.529 20.1%	1.529 4.4%	1.451 9.4%	1.432 1.6%	1.345 4.1%	1.262 2.3%	1.183 3.9%
linucb (disjoint)	1.795 12.5%	1.647 24.2%	1.719 11.6%	1.507 13.7%	1.714 10.7%	1.384 8.7%	1.655 13%	1.387 4.6%	1.574 11.7%	1.245 −3.5%	1.382 12%	1.197 5.1%
ϵ-greedy (hybrid)	1.739 9%	1.521 14.7%	1.68 9%	1.345 1.4%	1.636 5.6%	1.449 13.8%	1.58 7.8%	1.348 1.7%	1.465 4%	1.415 9.5%	1.342 8.8%	1.2 5.4%
linucb (hybrid)	1.73 8.4%	1.663 25.4%	1.691 9.7%	1.591 20%	1.708 10.3%	1.619 27.2%	1.675 14.3%	1.535 15.8%	1.588 12.7%	1.507 16.6%	1.482 20.1%	1.446 27%

Table 1: Performance evaluation: CTRs of all algorithms on the one-week evaluation dataset in the deployment and learning bucket (denoted by "deploy" and "learn" in the table, respectively). The numbers with a percentage is the CTR lift compared to ϵ-greedy

Summary

- Model news recommendation as a K-armed Bandit Problem
- UCB-type Algorithm
- Take Contextual Information in to consideration

Q&A