Exploitation and Exploration in a Performance based Contextual Advertising System

Wei Li, Xuerui Wang, Ruofei Zhang, Ying Cui

Jianchang Mao and Rong Jin

Performance-based online advertising

- Performance based online advertising
  - A form of online advertising that has a direct measurable financial goal associated
  - CPC is a predominant pricing model for sponsored search and content match
  - Click prediction (CTR) is critical in CPC-based auction
  - eCPM prediction shares the same challenge as CTR prediction

- Challenges in CTR measurement in performance based advertising system
  - The marketplace is very dynamic. New ads come in at hourly basis.
  - Historical CTR measurement is often sparse and unreliable
  - User fatigue. Same ads repeatedly to the same users
  - Low economic efficiency. Over-exposing for some advertisers and under-exposing for some others

Exploitation vs. exploration (EE) trade-off

- Short term: maximize revenue from current knowledge (exploitation)
- Long term: learn from unknown ad space, and improve performance in the future (exploration)
- Primarily studied in reinforcement learning
  - Ad placement: an extension to multi-armed bandit problem
    - Select a small number of ads as action
    - Receive a user feedback (click or non-click) as reward
    - Maximize accumulated reward
- No systematic study of short-term and long-term effects of EE algorithms in the context of dynamic performance based online advertising

Two common EE strategies

- ε-greedy
  - With probability 1-ε, choose the best action
  - With probability ε, choose any other action u formly
  - Optimal ε is difficult to find
  - ε needs to be adjus due to f o e advertising marketplace
- ε-decreasing
  - ε decrea
  - Focus on exp xploitat later
  - Not desirab
    - Old ads expire and ds em ge

Balance EE tradeoff adaptively in response to the data dynamics

Our approach (1)
ε -greedy with EG update

- Extend ε-greedy by updating ε dynamically
  - Not necessarily decreasing
  - Sample ε from a finite set of candidates at each iteration
  - Candidate probabilities updated via Exponetiated Gradient (EG)
    - Increase its probability if a candidate leads to click

Algorithm 2 The ϵ-greedy algorithm with the Exponentiated Gradient update

1: $p_k \Leftarrow 1/T$ and $w_k \Leftarrow 1$, $k = 1, \cdots, T$
2: **for** $i = 1$ to N **do**
3: Sample d from Discrete$(p_1, \cdots, p_T)$
4: Run Algorithm 1 with ϵ_d
5: Receive a click feedback c_i from the user
6: $w_k \Leftarrow w_k \exp\left(\frac{\tau\,[c_i I(k=d)+\beta]}{p_k}\right)$, $k = 1, \cdots, T$
7: $p_k \Leftarrow (1-\kappa)\frac{w_k}{\sum_{j=1}^{T} w_j} + \frac{\kappa}{T}$, $k = 1, \cdots, T$
8: **end for**

Our approach (2) confidence-based exploration

- Improve random exploration aspect of ε-greedy
  - Not desirable to explore ads randomly
    - Do not waste opportunities on the established bad performers
    - Focus on ads that might lead to higher revenue in the long run
  - Introduce confidence metric of performance measure
    - Decide which ads need to be explored.
    - Serve as a dynamic switch between exploitation and exploration.
    - When the confidence increases to a certain level, some exploration budget will be automatically shifted to exploitation.

Algorithm 3 The Confidence-based EE alg. for advertising

$F \Leftarrow \{a_1, \cdots, a_r\}$ {the final ranking list F}
$P \Leftarrow \{\}$ {the promotional queue P}
for $i = r + 1$ to n **do**
    if $x_i < w$ **then**
        $p_i \Leftarrow 1 - \tanh(x_i/b)$
    else
        $p_i \Leftarrow 0$
    end if
end for
Sample q ads from $\{a_{r+1}, \cdots, a_n\}$ with probabilities $\{p_{r+1}, \cdots, p_n\}$ and append the q ads to P
repeat
    Sample z from Bernoulli(ϵ)
    if $z = 0$ **then**
        $a = \mathrm{POP}(A)$, $F \Leftarrow F \cup \{a\}$ if $a \notin F$
    else
        $a = \mathrm{POP}(P)$, $F \Leftarrow F \cup \{a\}$ if $a \notin F$
    end if
until P or A is empty
$F \Leftarrow F \cup P \cup A$

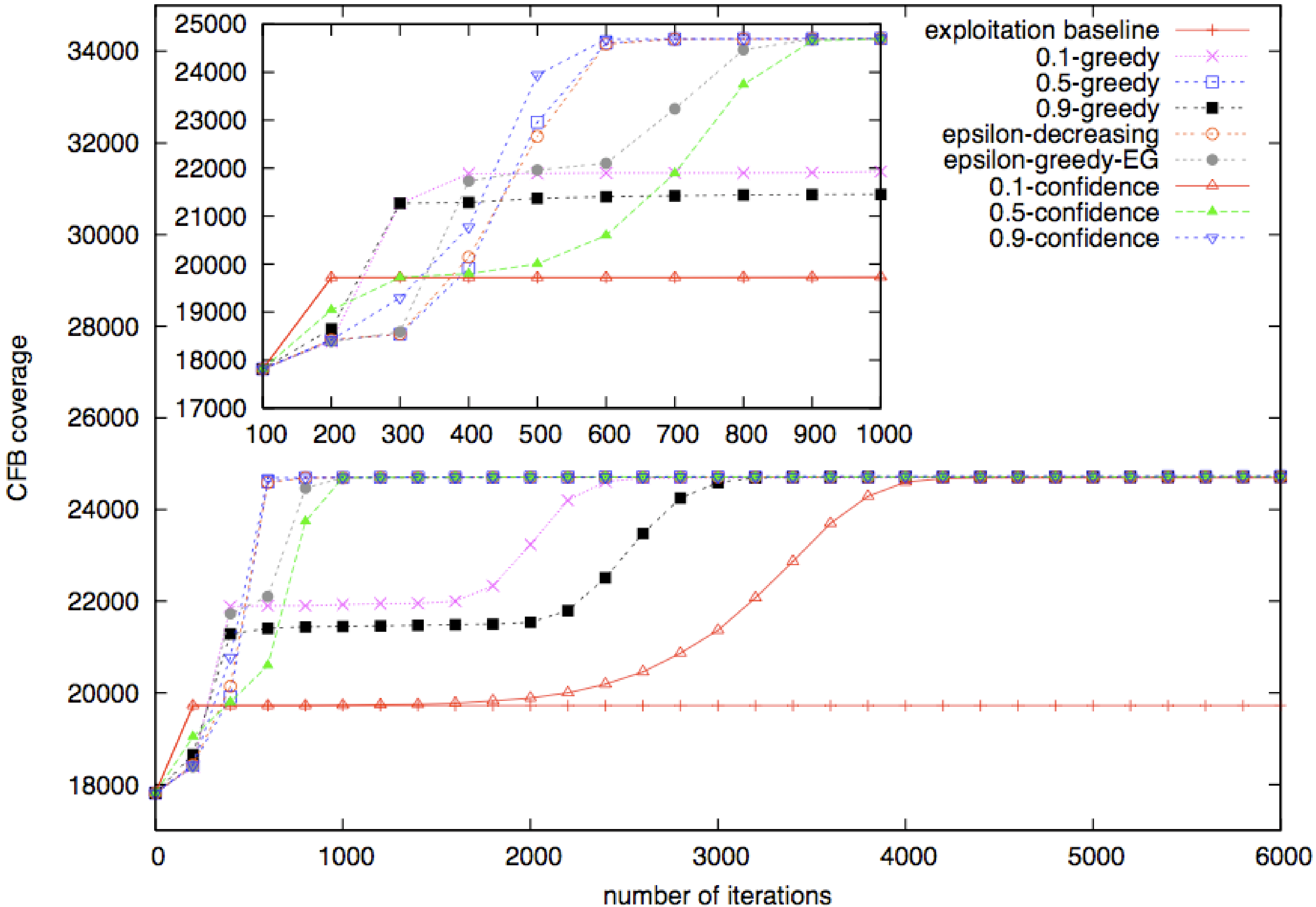
Simulation and experiments

- Evaluate the long-term effects of EE algorithms
  - Take time to explore unknown space and discover new ads with good CTR
  - During such a long time period, CTR may change significantly
  - Speed up iterations to demonstrate the results faster
- Design an offline simulation framework
  - Get more accurate evaluation
  - Mimic emitting the online events using real event logs
  - Set up controlled experiment buckets to perform apple-to-apple comparisons between a pure exploitation baseline and various EE algorithms.

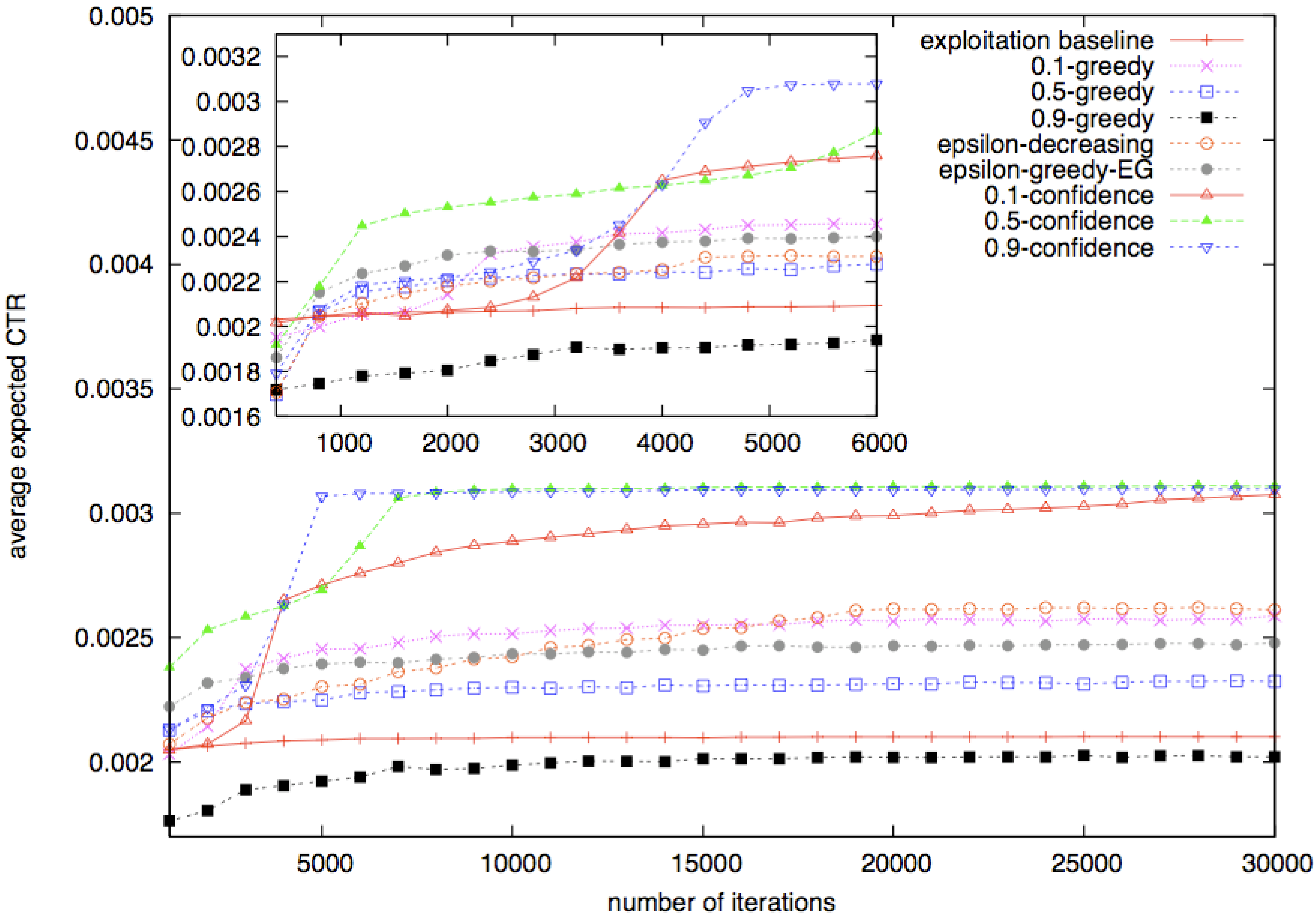
Evaluation metrics

- Ad reach of performance history
  - # of (page,ad) pairs in the statistics table with sufficient impressions
  - Positively correlated with CTR estimation accuracy
  - With increased coverage, we select and rank ads from a larger pool
- Average expected CTR
  - Actual CTR changes abruptly over iterations for random click feedback
  - The ratio between # of *expected* clicks and # of impressions
  - Calculate the average *expected* CTR over every 100 iterations

Experiment results: ad reach



Experiment results: avg expected CTR



Summary

- Proposed two approaches adaptively balance EE trade-off
  - ε-greedy with Expoinitated Gradient update
  - Confidence-based exploration

- Designed an offline simulation framework
  - A controlled environment mimicking the online ads selection and click feedbacks
  - Compared different EE strategies. Our approaches perform superiorly in ad reach and expected CTR measures

- Several findings of short-term/long-term effects of EE algorithms.
  - The ad reach convergence rate is un-sensitive to ε in ε-greedy due to data sparsity
  - The converged CTR increases as ε decreases in ε-greedy
  - The ε-greedy-EG has faster convergence rate and the higher CTR than ε-greedy when the ad space is under-discovered.