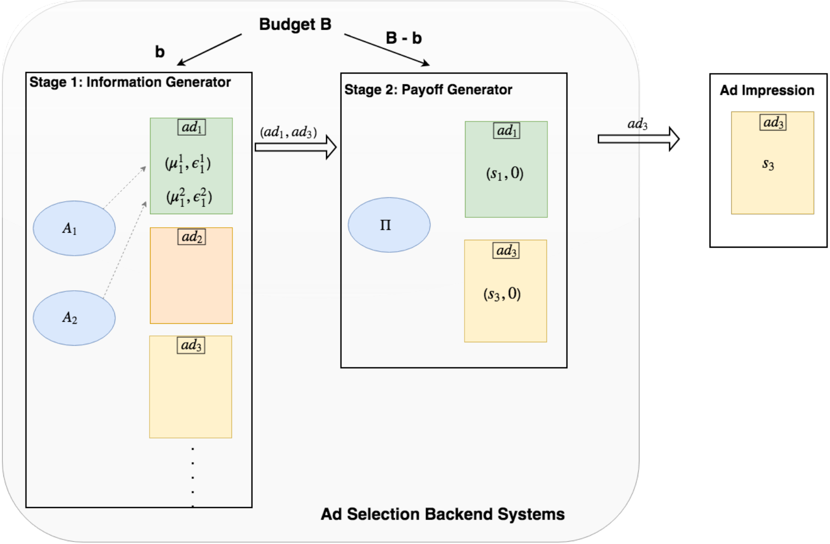
Information Production and Analysis in Large-Scale Budget-Constrained Systems

Ramnik Arora and **Hemant K. Bhargava**
Facebook and U.C. Davis

December 26, 2018

Real-time ad selection system



Some notation

$\boldsymbol{B}$: Total budget

$\boldsymbol{s_i}$: The value from displaying ad$_i$

$\boldsymbol{\Omega}$ **:** Candidate set of ads ($|\Omega| = N$)

$\boldsymbol{\mathcal{I}_0}$: Initial information about Ω ($\mathcal{I}_0 = F_0^N$, F_0 is the prior cdf)

$\boldsymbol{A(i, \lambda, \epsilon_i)}$: The *information generator* applied to ad candidate i with cost λ and error ϵ_i [1]

$\boldsymbol{\Pi(\Omega')}$: The *payoff-generator* or the final stage algorithm with unit cost and zero error evaluating $\Omega' \subset \Omega$

$\boldsymbol{b}$: Budget allocated to the information generator (A can evaluate $\frac{b}{\lambda}$ elements, and $|\Omega'| = B - b$)

[1] A reduces the uncertainty regarding payoff s_i but not perfectly, and the level of reduction is heterogeneous.

Pay-off function

The payoff is the **highest order statistic** amongst the candidates evaluated by Π.

In a single slot auction:

$$\Pi(\Omega^{'}) = \max_{w \in \Omega^{'}} s_i \quad \text{s.t.} \quad w = ad_i$$

Some quick comments and observations

- (Trivial) One element to be evaluated by information generator
- $\lambda \in [0, 1]$ [2]
- The algorithm A produces a point estimate γ_i for ad i, such that:
  - $s_i \in [\gamma_i - \epsilon_i^1, \gamma_i + \epsilon_i^1]$, where ϵ_i captures the inaccuracy of the information generator, and this inaccuracy is heterogeneous across the ads
  - $\mathbb{E}[\gamma_i] = s_i$, or the algorithm A is "calibrated" in expectation.

[2] $\lambda = 0$ is updated prior and $\lambda = 1$ is the information generator

Questions we'll comment on

- **Free evaluation**: Given $I = I_0 \cup A(0, ., .)$ and $|\Omega'| = N - 1$, should $\mathrm{ad}_0 \in \Omega'_*$, the optimal subset to be evaluated by the payoff generator.
- **Single evaluation**: $\mathbb{E}\left[\Pi(\Omega'_*(I_0))\right] \geq \mathbb{E}\left[\Pi(\Omega'_*(\{I_0 \cup A(0, \lambda, .)\})\right]$ where the latter allows only $N - \lambda$ elements to be evaluated.
- **Optimal subset**: Given fixed $B - b$ and I, how does one choose Ω'_* which maximizes $\mathbb{E}\left[\Pi(\Omega')\right]$ over all $B - b$ cardinality subsets.

Questions we'll comment on

- **Budget allocation**: For a given information generator $A(\lambda, \epsilon)$, what should the budget allocation between the information generator(b_*) and the payoff generator($B - b_*$).
- **Learning algorithm**: Given an array of information generators, which one would you choose?

Quick Primer on Order Statistics

For X_1, X_2, ... X_n are iid continuous random variables with pdf f and cdf F, the density of the highest order statistics is:

$$P(X_{(n)} \in [x, x+\epsilon]) = \epsilon \times nf(x)F(x)^{n-1}$$

Order statistics of standard uniforms with n uniform random variables $X_k \sim Beta(k, n-k+1)$ and in particular $\mathbb{E}[X_1] = \frac{n}{n+1}$

Some simple cases

which subset to evaluate?

- Ten $U[0, 1]$ or ten $U[0.85 - \epsilon, 0.85 + \epsilon]$
- Case U[0.4, 0.7] where N = 1 and N = 3
- For the optimal solution to consider all subsets of Ω

Monotone Sub-modular

Let X be a finite set. A function $f : 2^X \to \mathbb{R}$ is:

- **monotone** if for all $S \subset T, f(S) \leq f(T)$
- **submodular** if for all subsets $S \subset T \subset X$ and all $x \in X \setminus T$

$$f(S \cup \{x\}) - f(S) \geq F(T \cup \{x\}) - f(T) \qquad (1)$$

Does this apply to our payoff function?

Greedy Algorithms are good enough?

Theorem: Let f be a monotone, submodular, non-negative function on X. The greedy algorithm, which starts with S as the empty set and at every step picks an element x which maximizes the marginal benefit $(f(S \cup \{x\}) - f(S))$ provides a set S that achieves a $(1 - \frac{1}{e})$ approximation of the optimum.[3]

[3] M. Fischer, G. Nemhauser, L. Wolsey. An analysis of approximations for maximizing submodular set functions.

Budget allocation

If the "value of the information" generated by A is monotonic in the b, then the optimal budget allocation should be until:

$$\mathbb{E}\left[\Pi[I_0 \cup A(b+\lambda)]\right] < \mathbb{E}\left[\Pi[I_0 \cup A(b)]\right]$$

where the cardinality of the former set to the payoff generator is $B - b - \lambda$ while $B - b$ for the later set.

But... how do you compute the "marginal value of additional information?"

Moving to uniform distributions

- At the start of the evaluation process the s_i are treated as independent random variables, identically distributed over an interval $U[0, 1]$ with uniform density.
- Given ad_i, A produces an interval $[\alpha_i; \beta_i]$ over which s_i is uniformly distributed. In particular:
  - $\mathbb{E}[s_i] = \frac{\alpha_i + \beta_i}{2}$ or the algorithm is *calibrated*

When would you include the evaluated point?

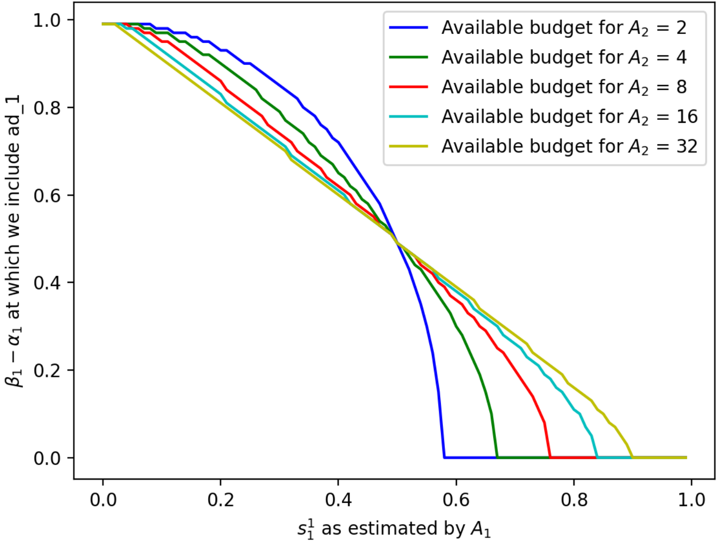
If a budget B is available for Π after A generates $[\alpha_i, \beta_i]$ for ad i, and if $\mathcal{I}' = U[0,1]^B \cup \{U[\alpha_i, \beta_i]\}$, then

$$\mathbb{E}\left[\Pi(B, \mathcal{I}')\right] = \underbrace{\frac{B-1}{B}}_{\mathbb{E}[\Pi] \text{ for } B-1 \text{ ads}} + \underbrace{\frac{1}{B(B+1)}}_{\text{loss from 1 less ad}} \times \underbrace{\frac{\beta_i^{B+1} - \alpha_i^{B+1}}{\beta_i - \alpha_i}}_{\text{gain due to i}}$$

and the condition for including ad i is

$$\mathbb{E}\left[\Pi(\mathcal{B}, \mathcal{I}_0)\right] > \mathbb{E}\left[\Pi(\mathcal{B}, \mathcal{I}')\right] \iff \frac{\beta_i^{B+1} - \alpha_i^{B+1}}{\beta_i - \alpha_i} > 1.$$

When would you include the evaluated point?



Decision tree for one element?

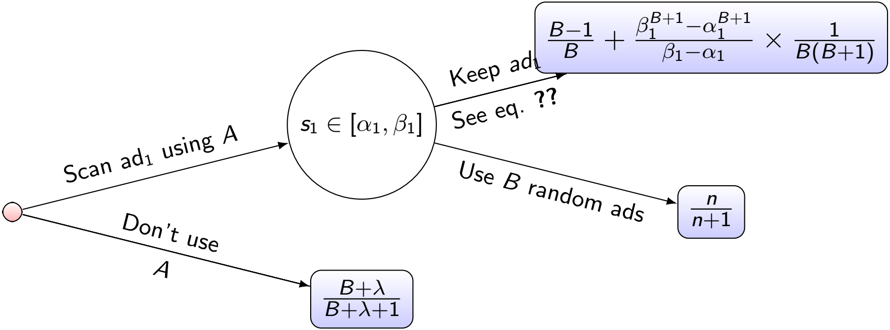


Figure: Decision tree for assigning a single unit of computation to A, and then deciding how to leverage the new information $[\alpha_i, \beta_i]$.

Summary

- Cost / accuracy trade-off in real-time ad selection systems.
- Simply picking elements with highest expected value might be naive, sub-optimal.
- There exist approximate algorithms even in the setting when ad candidates are processed as a stream.
- The budget allocation should balance marginal utility of the information with the marginal cost of the information.

questions and suggestions?