

インターネットにおける広告戦略モデル

有坂 昌樹

指導教員 合原 一幸 教授

2007年2月5日

1 はじめに

インターネット広告の世界において、検索エンジンで検索されたキーワードに合わせてテキストなどを表示される広告形態である検索連動型広告が急激な広がりを見せている。

検索連動型広告の研究にはサービス側の視点からの研究と広告主側からの研究がある。広告主側の研究は広告主の収益最大化が大きな目的である。この研究分野の目的の一つはキーワード集合を選択することによる最適化である。

本研究は、広告主の収益が上がるようにキーワード集合を適切に選択する方法を提案する。既存の研究 [1] では、過去の観測結果を元に利益対コストの比が大きい順にキーワードを選択していくアルゴリズムが提案されている。本研究ではこれと同様に、選択するキーワードの個数を改良することにより、理論的に近似率を向上させることに成功した。また、数値実験を行った結果、収益を 12.7% 上昇させることに成功した。

2 目的と定式化

本研究では、広告主が検索連動型広告に一定期間広告を出稿することを考える。まず広告として入札できるキーワードは N 個あるものとし、 $\{1, \dots, N\}$ を入札可能なキーワード集合とする。本研究の目的は、各期間 $t = 1, 2, \dots, T$ で入札するキーワード集合 $A_t \subseteq \{1, \dots, N\}$ を適切に選択することによって、広告主の収益 Z_{A_t} の期待値の総和 $\sum_{t=1}^T E[Z_{A_t}]$ を最大化することである。

1 回の検索でどのキーワードが検索されるかという確率を λ_i で表す。また、各キーワード $1 \leq i \leq N$ に対して、 $0 \leq p_i \leq 1, c_i > 0, \pi_i \geq 0$ をそれぞれクリック率、1 クリックあたりの広告料、1 クリックあたりの期待収益とする。 λ_i は検索連動型広告会社のサービスで各キーワードの月間検索回数を調べられるので推定できる。また、広告料は広告主が設定するので既知として扱う。

π_i は一般には未知であるが、本研究では既知として扱う。 p_i も一般には未知である。本研究ではこの p_i をまずは既知の場合として扱い、近似アルゴリズムを提案する。そして、この既知の場合の結果を用いて、 p_i が未知の場合におけるアルゴリズムを提案する。各パラメータは t に関して一定とする。また、本研究において、一期間の予算を 1 として c_i と π_i を標準化する。

先行研究 [1] において、 p_i が既知の場合は以下のように定式化されている。

$$E[Z_A] = \sum_{i \in A} \pi_i \lambda_i p_i \sum_{r=1}^{\infty} E[\mathbf{1}(B_r^A \geq c_i, S \geq r)]. \quad (1)$$

ここで B_r^A は $A \subseteq \{1, \dots, N\}$ に入札したときに、 r 番目の検索が行われたときの残りの予算を表し、 S は一期間の総検索回数である。この S は平均 μ の確率分布に従うものとする。各 i で $p_i = 1$ のとき式 (1) が Stochastic Knapsack Problem に帰着でき、NP 完全であることが示されている。

3 研究結果

[1] では、この問題に対して以下の仮定を導入している。

- (a) キーワードの添え字は $\pi_1/c_1 \geq \pi_2/c_2 \geq \dots \geq \pi_N/c_N$ の順に並んでいるものとする。
- (b) $\mu \sum_{i=1}^N c_i p_i \lambda_i > 1$.
- (c) 全ての i に対して $c_i \leq 1/k$ かつ $\lambda_i \mu \leq k^\alpha$ を満たすようなある $k \geq 1$ と $0 \leq \alpha < 1$ が存在する。

この仮定の下、キーワード集合 $P \subseteq \{1, \dots, N\}$ を提案されている近似アルゴリズムによって選択することで、 Z^* を最適収益とすれば、 $\rho \left(1 - \frac{1}{k^{(1+2\alpha)/3}} - \frac{2}{k^{(1-\alpha)/3}}\right) Z^* \leq E[Z_P] \leq Z^*$ となることが示されている。但し $\rho = E[\min\{S, \mu\}]/\mu$ であり、 S がポアソン分布に従うならば $1 - 3/\sqrt{\mu} \leq \rho \leq 1$ となる。

これに対して、本研究では、[1] の仮定 (c) を

- (c') 全ての i に対して $c_i \leq 1/k$ かつ $\lambda_i \mu \leq k^\alpha$ を満たすようなある $k \geq 1$ と $0 \leq \alpha < 2/3$ が存在する。

と変更し, (a),(b) は同様とした. これは [1] が調査した実データでも適応でき, また予算を十分に多く用意することで満たされる条件である.

我々は p_i が既知の場合において以下の結果を示した. これは先行研究における近似率の $\alpha = 0$ の場合であり, 下界の上限を示したことになっている.

定理 1. k, α は我々の仮定を満たすとす. p_i が既知であり, さらに $1/k + 1/k^{1-\alpha} + 1/k^{1/3} \leq 1$ とす. また, Z^* を最適収益とし, $\rho = E[\min\{S, \mu\}]/\mu, P = \{1, 2, \dots, \mathcal{I}_U\}$ とす. そして $\mathcal{I}_U \equiv \max\{\ell : \mu \sum_{u=1}^{\ell} c_u \lambda_u p_u \leq 1 - 1/k - 1/k^{1/3}\}$ とするとき,

$$\rho \left(1 - \frac{3}{k^{1/3}}\right) Z^* \leq E[Z_P] \leq Z^*.$$

□

また, クリック率が未知の場合において我々は以下のアルゴリズムを提案した.

• 初期化:

- y_i, x_i をそれぞれキーワード i の累積表示回数と累積クリック回数とする. 全てのキーワード i に対して, $y_i = x_i = 0$, クリック率の推定値 $\hat{p}_i^0 = 1$ とす.
- 時刻 t において, ランダムにキーワード集合を選ぶ確率を $(\gamma_t = t^{-2} : t \geq 1)$ とす.

• アルゴリズム: $t = 1, 2, \dots$ に対して,

- $\mu \sum_{u=1}^{\ell_t} c_u \lambda_u \hat{p}_u^{t-1} \leq 1 - \frac{1}{k} - \frac{1}{k^{1/3}} - \frac{1}{k^{(1-\alpha)/3}} \leq \mu \sum_{u=1}^{\ell_t+1} c_u \lambda_u \hat{p}_u^{t-1}$ を満たすように ℓ_t を定める.
- θ_t を $\{1, 2, \dots, N\}$ の中から一様乱数を用いてランダムに選ぶ.
- キーワード集合の決定変数 g_t を

$$g_t = \begin{cases} \ell_t & \text{確率 } 1 - \gamma_t \\ \theta_t & \text{otherwise} \end{cases}$$

とする.

- $G_t = \{1, 2, \dots, g_t\}$ とし, これらのキーワードを入札する.
- 更新: V_i^t, W_i^t をそれぞれ期間 t におけるキーワード i の表示回数とクリック回数とする. このとき, 全てのキーワード i に対して, $y_i \leftarrow y_i + V_i^t, x_i \leftarrow x_i + W_i^t$ として

$$\hat{p}_i^t = \begin{cases} 1, & \text{if } y_i = 0 \\ \frac{x_i}{y_i}, & \text{if } y_i > 0. \end{cases}$$

• 出力: キーワード集合列 $(G_t : t \geq 1)$.

このアルゴリズムの収束率を以下に示す.

定理 2. $(G_t : t \geq 1)$ を提案したアルゴリズムによって生成されたキーワード集合列とし, $\rho = E[\min\{S, \mu\}]/\mu$ とす. そして, $1/k + 1/k^{1-\alpha} + 1/k^{1/3} + 1/k^{(1-\alpha)/3} \leq 1$ と仮定する. このとき, 我々の仮定の下で全ての $T \geq 1$ に対して

$$\frac{\sum_{t=1}^T E[Z_{G_t}]}{TZ^*} \geq \rho \left(1 - \frac{3}{k^{1/3}} - \frac{2}{k^{(1-\alpha)/3}}\right) - \frac{\sum_{i=1}^T \gamma_i}{T} - \frac{M}{T},$$

$$\text{ここで, } M = \frac{32\mu\mathcal{I}_U}{k\rho(1-1/k^{1/3})} + \frac{12\mathcal{I}_U^2 \sum_{i=1}^{\mathcal{I}_U} \lambda_i \mu / (1-e^{-\lambda_i \mu})}{k}.$$

□

[1] で示されている収束率は $\frac{\sum_{t=1}^T E[Z_{G_t}]}{TZ^*} \geq \rho \left(1 - \frac{1}{k^{(1+2\alpha)/3}} - \frac{4}{k^{(1-\alpha)/3}}\right) - \frac{\sum_{i=1}^T \gamma_i}{T} - \frac{M}{T}$, であり, 収束率は改善されており, $\alpha = 0$ のときに一致する. 図 1 に数値実験の結果を示す.

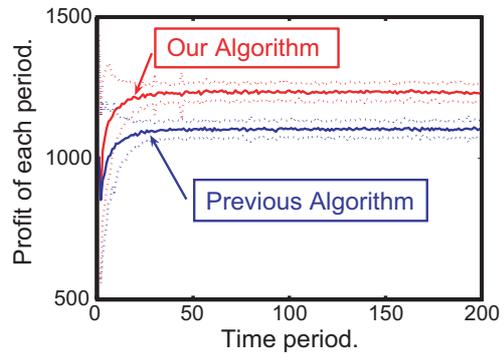


図 1 数値実験による期待収益の比較. 条件は検索数を平均 40,000 のポアソン分布に従わせ, キーワードを 8,000 個, 予算を 400, 期間を 200 としたものを 100 回繰り返した. 平均を実線で, 95% 信頼区間を破線で示している. 上側が我々の結果で, 下側が [1] の結果.

4 今後の課題

本研究では c_i, π_i を一定として扱った. しかし, 実際, これらは変化するパラメータである. これらは掲載順位によって変動するし, また競争相手を考慮すると同じ掲載順位においても変動する. また π_i を既知として扱ったが, 初めて広告を出稿する際には, 一般的には π_i は未知である. 従ってこれらのことを考慮したモデルを構築することが今後の課題となる.

参考文献

[1] P. Rusmevichientong and D. P. Williamson, *Proceedings of the 7th ACM Conference on Electronic Commerce*, 2006.