

Bandit 手法による制約条件下での Web 広告選択

渡邊 僚[†] 中村 篤祥^{††} 工藤 峰一^{††}

[†] 北海道大学工学部 〒060-8628 北海道札幌市北区北 13 条西 8 丁目

^{††} 北海道大学大学院情報科学研究科 〒060-0814 札幌市北区北 14 条西 9 丁目

E-mail: [†]{ryo,atsu,mine}@main.ist.hokudai.ac.jp

あらまし Web 広告配信問題は、クリックを報酬と考えれば多腕 Bandit 問題とみなせる。しかし、広告主と配信数を保証する契約をしている場合には、それにより生ずる制約を満たす必要がある。また、複数のページ (キーワード) に広告を配信する場合には、ページと広告の相性を考慮すべきであるが、それぞれのページビューの回数も制約となる。この問題を定期的なバッチ処理により線形計画問題として解く手法が開発されている。本稿では広告配信スケジュールをオンラインで更新する手法を提案し、シミュレーションにより有効性を検証する。

キーワード Web 広告, オンライン学習, 多腕 Bandit 問題

Constrained Web advertisement selection using Bandit approach

Ryo WATANABE[†], Atsuyoshi NAKAMURA^{††}, and Mineichi KUDO^{††}

[†] School of Engineering, Hokkaido University

Kita 13, Nishi 8, Kita-ku, Sapporo, Hokkaido 060-8628, Japan

^{††} Graduate School of Information Science and Technology, Hokkaido University

Kita 14, Nishi 9, Kita-ku, Sapporo, Hokkaido, 060-0814, Japan

E-mail: [†]{ryo,atsu,mine}@main.ist.hokudai.ac.jp

1. はじめに

インターネットの登場以来 Web 広告は増え続け、今や新聞を抜き、テレビに次ぐ第 2 の広告媒体となった [1]。Web 広告は、検索エンジンサイトやニュースポータルサイトなどといった大規模な Web サイトのみならず、アフィリエイトプログラムに参加している個人の Web サイトにも掲載される。広告を見たサイト訪問者は、その広告をクリックすることで追加の情報を手に入れることができる。この追加情報へのアクセスの容易さも Web 広告の特徴の一つである。

広告主と広告配信会社間での契約形態も様々だが、近年はクリック回数保証や成果報酬型の契約が増えている。そのため、表示できる広告の中でサイト訪問者の興味を引く広告がどれかを学習し、最もクリックされやすい広告や最も購買されそうな商品の広告を表示することができれば、広告収入の増加につながる。

しかし、最適な広告配信を実現するためにはいくつかの問題を克服しなければならない。1 つ目の問題は、知識利用と獲得のトレードオフの問題である。広告のクリック率 (あるいは商品の購買率) は予め知ることのできない知識であり、学習しなけ

ればならない。クリック率推定のために、それまでのクリック率がたとえ低くても配信回数の少ない広告を推定値の信頼度をあげるためにある程度配信しなければならない (知識の獲得)。しかし、過度にそのような広告配信を続けると全体のクリック率が低下し、クリック数を増やすことができない。クリック数を増やすためには、それまでのクリック率が最大のものを配信し続けられればよいように思われるが (知識の利用)、その広告が真に最適でなかった場合、クリック率を上げることができない。

この問題は多腕 Bandit 問題として盛んに研究されてきたものである。広告をスロットマシンのアームと見なし、広告がクリックされたら報酬 1、されなかったら 0 と考えることで、クリック数を最大化する広告配信問題は多腕 Bandit 問題となる。この問題に対しては、Gittins Index [2] や UCB [3] といった対処法が存在する。

もう 1 つの問題は、広告主との契約と広告スペースのアクセス回数からくる制約である。契約については、広告主と直接配信回数で契約することもあるし、クリック数や購買数による契約であってもクリック率、購買率から逆算した配信数を考慮した配信スケジューリングが必要になる。また、1 つの広告スペースにはアクセス数以上に広告を表示することはできない。

Langheinrich ら [4] は、このような制約の下で累計クリック数を最大化するような配信スケジュールを求める問題を線形計画問題として定式化した。

Langheinrich ら [4] は、配信スケジュールの更新を定期的なバッチ処理で行い、知識利用と獲得のトレードオフを解決する方法として、3つの手法を導入していた。

本稿では、知識利用と獲得のトレードオフを解決する方法として彼らのものとは異なる UCB アルゴリズムの考え方をうい、配信スケジュールをオンライン学習により逐次的に更新する手法を提案する。また、シミュレーションによりその効果を検証する。

2. 問題設定

本稿では、以下のような状況における広告配信問題を考える。

- 選択候補となる広告の数を K 、広告を掲載する Web ページの数を N とする。
- 各ページに掲載できる広告の数は1つのみ。
- 広告料は広告がクリックされた累計回数に比例して支払われる。
- 各広告 $i \in \{1, \dots, K\}$ について、**表示比率**が制約条件として与えられる。

この状況ではページ連動広告しか考えていないが、ページをキーワードに対する検索結果のページとみなすことでキーワード連動広告にも適用できる。また、クリック回数ではなく購買数に応じて広告料が支払われる成果報酬型の契約にも適用できる。

制約条件は K 次元確率ベクトル $\mathbf{p} = (p_1, \dots, p_K)$ で表す。全ての広告表示のうち広告 i が選択された割合が p_i に近づくように広告を選択することで、広告主から要求された制約条件を満たすような広告配信ができるように $\mathbf{p}$ を設定する。

一般に Web 広告が配置されるページは複数存在する。制約がなければそれぞれのページで独立に広告選択をすればよいのだが、広告の表示比率は広告を配信する Web ページ全体で評価される。そのため、1つの広告選択アルゴリズムが全ての Web ページ、広告に関する履歴・統計情報を持ち、各 Web ページに合わせた広告を選択して表示する仕組みとなる。

広告選択のフレームワーク

Web ページへのアクセス $t = 1, \dots$ において、

- (1) アクセスされた Web ページ $J(t) \in \{1, \dots, N\}$ が与えられる。
- (2) $J(t)$ や内部に持っているデータを元に表示する広告 $I(t) \in \{1, \dots, K\}$ を選択し、表示する。
- (3) 表示した広告がクリックされたかどうかのフィードバック $g(t) \in \{0, 1\}$ が与えられる。

この問題設定における最適な広告配信は、全アクセスのうち Web ページ j に広告 i を表示する割合 $x_{i,j}$ ($i = 1, \dots, K, j = 1, \dots, N$) によって表現され、式 (1)-(4) のような線形計画法による最大化問題での解として求めることができる [4]。

$$\text{maximize } \sum_{i=1}^K \sum_{j=1}^N x_{i,j} c_{i,j} \quad (1)$$

$$\text{subject to } \sum_{i=1}^K x_{i,j} = q_j \quad (j = 1, \dots, N) \quad (2)$$

$$\sum_{j=1}^N x_{i,j} = p_i \quad (i = 1, \dots, K) \quad (3)$$

$$x_{i,j} \geq 0 \quad (i = 1, \dots, K, j = 1, \dots, N) \quad (4)$$

目的関数、制約条件に現れる q_j ($j = 1, \dots, N$) と $c_{i,j}$ ($i = 1, \dots, K, j = 1, \dots, N$) はそれぞれ Web ページ j がアクセスされる比率と、Web ページ j に表示した広告 i がクリックされる確率を表している。

広告枠としてはある程度安定したアクセスがあるページを用いると考えられるので、ページアクセス確率 q_j については、履歴データにより十分良い推定値を得ることができる。一方、広告は次々と新しいものに変わるため、クリック確率 $c_{i,j}$ について事前により推定値を得ることはできない。したがって、広告が新しくなるたびにそれぞれの Web ページにおいて各広告を表示して、クリックされる確率を逐次推定する必要がある。

3. 知識利用と獲得のトレードオフ

クリック確率を推定する必要があるのはこれまでに述べたとおりであるが、十分な推定精度を得るためにはどのページにおいてもそれぞれの広告をある程度の回数選択し、表示した広告がクリックされるか探る必要がある (知識獲得)。一方、そのような広告選択方法が続けると、各ページにおいてクリック率の低い広告の配信比率が高まり、全体として累計クリック回数の減少につながるため、これまでのクリック率が高い広告を配信すべきである (知識利用)。この2つのバランスをどうとるか、という問題が**知識利用と獲得のトレードオフ**と呼ばれ、Bandit 問題の特徴となっている。

広告配信問題において、このトレードオフを解決する方法として次のような方法が考えられる [5]。

- 線形計画法による最大化を途中で止め、その時点での解を用いる。
- 信頼度の低い推定値を大きめに見積もった値をクリック率として用い、最適化を行う。
- 配信比率に下限制約を設ける。

Nakamura ら [5] は2つ目のクリック率推定における解決方法として Gittins Index を用いた方法を提案したが、本稿では UCB アルゴリズム [3] の手法を用い、標準偏差の定数倍を加えることで推定値を大きめに見積もる方法を採用する。

4. アフィン変換法

アフィン変換法は線形計画法の最大化問題の解法の一つで、解空間の中で制約条件がなす凸包の中に初期値をとり、徐々に目的関数の値が大きくなる方向に点を移動させて解を求める内点法と呼ばれるアルゴリズムに分類される [6]。

特にアフィン変換法は、式 (5) のような標準形で表現できる、

m 本の制約条件の下で n 個の変数を持つ目的関数の最小化 (最大化) 問題を対象としたものである。

$$\begin{aligned} & \text{maximize or minimize } \mathbf{c}^T \mathbf{x} \\ & \text{subject to } A\mathbf{x} = \mathbf{b} \end{aligned} \quad (5)$$

ただし, A はサイズ $m \times n$ でフルランクの行列, $\mathbf{b}$ と $\mathbf{c}$ はそれぞれ m, n 次元実数ベクトルとする。

いま扱っている広告配信問題はこのアフィン変換法の標準形で表現でき, $\mathbf{b}$ は広告配信比率と Web ページアクセス確率を並べたもの, $\mathbf{c}$ は各ページにおけるそれぞれの広告のクリック確率 (実際にはその推定値), A は規則的に 1 と 0 が配置された疎行列となる。

具体的な計算方法については割愛するが, 図 1 にアフィン変換法による最大化アルゴリズムを掲載する^(注1)。

アルゴリズム Affine

input $A \in \mathbb{R}^{n \times m}, \mathbf{b} \in \mathbb{R}^m, \mathbf{c} \in \mathbb{R}^n$

output $\mathbf{x}^* \in (\mathbb{R}^+)^n$

parameter $h \in [0, 1], \varepsilon > 0$

- (1) $A\mathbf{x}^{(0)} = \mathbf{b}$ を満たすような初期値 $\mathbf{x}^{(0)}$ を探す。
 $k := 0$.
- (2) $A, \mathbf{x}^{(k)}, \mathbf{c}$ を用いて目的関数の最急上昇方向ベクトル $\mathbf{d}^{(k)}$ を求める。
- (3) $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + h(\mathbf{d}^{(k)} / |\mathbf{d}^{(k)}|)$
- (4) $\mathbf{c}^T \mathbf{x}^{(k+1)} - \mathbf{c}^T \mathbf{x}^{(k)} < \varepsilon$ なら $\mathbf{x}^{(k+1)} = \mathbf{x}^*$ として終了。
- (5) $k := k + 1$ とし手順 3 へもどる。

図 1 アフィン変換法

このアフィン変換法による最大化を十分に行わず途中でなし 1 回で打ち切れば, 準最適解である $\mathbf{x}$ が得られる [5]。

本稿では, アフィン変換法を知識利用と獲得のトレードオフのために用いるのではなく, 前時刻の配信スケジュールに更新された推定クリック率を用いて, 1 回の更新ステップを適用し次時刻の配信スケジュールを求めるオンライン手法として用いる。

5. クリック確率の推定

Web ページ j に表示した広告 i がクリックされたとき 1, されなかった場合 0 という値をとるような確率変数を考える。この確率変数はクリック確率をパラメータとしたベルヌーイ分布 $B(c_{i,j})$ に従う。いま, このパラメータ $c_{i,j}$ を推定することを考える。

パラメータ $c_{i,j}$ の事前分布をベータ分布 $\text{Be}(\alpha_0, \beta_0)$ と仮定する。ベータ分布はベルヌーイ分布の共役事前分布であるので, ページ j に広告 i を n 回配信して m 回クリックされたとする ($n - m$ 回クリックされなかったとする) と, $c_{i,j}$ の事後分布は $\text{Be}(\alpha_0 + n, \beta_0 + (m - n))$ に従う。

ベータ分布 $\text{Be}(\alpha, \beta)$ に従う確率変数 X の期待値と分散は式

(6) と (7) で与えられる。

$$E[X] = \frac{\alpha}{\alpha + \beta} \quad (6)$$

$$\text{Var}[X] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \quad (7)$$

分散については表示回数の増加に従い小さくなっていくことに注意されたい。

多腕 Bandit で用いられる UCB アルゴリズムにおいては, 信頼区間の考え方に基づき期待値に標準偏差の定数倍を加えた値をアームの評価値として用いている [3], [7]。これらに倣い, ページ j に広告 i を表示してクリックされる確率の時刻 t における推定値 $\hat{c}_{i,j}(t)$ を, 式 (8) で求める。

$$\hat{c}_{i,j}(t) = \mu_{i,j}(t) + \gamma\sqrt{V_{i,j}(t)} \quad (8)$$

ただし, $\mu_{i,j}(t), V_{i,j}(t)$ はベータ分布 $\text{Be}(\alpha_{i,j}(t), \beta_{i,j}(t))$ の期待値と分散を表す。ここで, $\alpha_{i,j}(t), \beta_{i,j}(t)$ はそれぞれ時刻 t までにページ j に広告 i を表示した回数 $n_{i,j}(t)$, クリックされた回数 $m_{i,j}(t)$ により式 (9), (10) で定義される。 α_0, β_0 は $c_{i,j}$ の事前分布を決定するパラメータである。

$$\alpha_{i,j}(t) = \alpha_0 + m_{i,j}(t) \quad (9)$$

$$\beta_{i,j}(t) = \beta_0 + (n_{i,j}(t) - m_{i,j}(t)) \quad (10)$$

6. 提案手法

6.1 配信広告決定方法

求められた配信スケジュールを元に配信する広告を決定する方法を二つ提案する。

6.1.1 ルーレット選択

1 つは, 求めた広告表示比率に従ってルーレット選択を行う方法である。ここで注意すべき点は, 広告表示比率 $x_{i,j}(t)$ の Web ページ j に関する成分の和は,

$$\sum_{i=0}^K x_{i,j}(t) = q_j$$

となっている点である。そのため, Web ページ $J(t)$ に表示する広告を決定するルーレットは, 各広告が選択される確率 $P(I(t) = i)$ を

$$P(I(t) = i) = \frac{x_{i,J(t)}(t)}{q_{J(t)}}$$

と定めなければならない。

6.1.2 決定的に広告を選択する手法

$x_{i,j}(t)$ は全アクセスに対する Web ページ j に広告 i を表示すべき回数の割合の目標値であるといえる。一方, 時刻 t において広告 i を選択した場合に, 実際の広告選択における全アクセスに対する Web ページ j に広告 i を表示した割合は,

$$\frac{n_{i,j}(t) + 1}{t}$$

(注1): $\mathbb{R}^+$ は非負実数を表すものとする。

と求めることができる。この値と $x_{i,j}(t)$ を比較し表示すべき割合に対して実際に表示された回数が最も不足している広告を時刻 t での選択とする。この手法により、時刻 t で選択する広告 $I(t)$ は、式 (11) で表すことができる。

$$I(t) = \arg \max_{i \in \{1, \dots, K\}} \left(x_{i,J(t)}(t) - \frac{n_{i,J(t)}(t) + 1}{t} \right) \quad (11)$$

6.2 広告選択アルゴリズム

UCB と内点法を組み合わせた Web 広告選択アルゴリズムは、図 2 のようになる。

広告選択アルゴリズム

parameter $h \in [0, 1], \alpha_0 \in \mathbb{R}^+, \beta_0 \in \mathbb{R}^+, \gamma \in \mathbb{R}^+.$

initialize $n_{i,j}(0) = 0, m_{i,j}(0) = 0, x_{i,j}(0) = p_i q_j.$

- (1) t 番めのアクセスされたページ $J(t)$ が与えられる。
- (2) **すべての広告、Web ページの組**について式 (8) によりクリック確率の推定値 $\hat{c}_{i,j}(t)$ を求める。
- (3) $\hat{\mathbf{c}}(t) = (\hat{c}_{1,1}(t), \dots, \hat{c}_{K,N}(t))$ と $\mathbf{x}(t-1)$ を用いてアフィン変換法を 1 ステップだけ進めて、 $\mathbf{x}(t)$ を求める。
- (4) $x_{i,J(t)}(t)$ や $n_{i,J(t)}(t)$ などを元に $I(t)$ を定める。

図 2 UCB と内点法を用いた広告選択アルゴリズム

7. 実験

7.1 実験データの生成

提案手法の性能を検証するためにシミュレーションを行った。シミュレーションでは

- 広告数 $K = 10$, Web ページ数 $N = 20$
- 試行 (アクセス) 回数 $T = 1,000,000$
- パラメータ $h = 0.03, \alpha_0 = 0.2, \beta_0 = 9.8, \gamma = 2$

とした。シミュレーションのために生成しなければならないデータは

- 広告表示比率 p_i ($i = 1, \dots, K$)
- Web ページアクセス確率 q_j ($j = 1, \dots, N$)。
- クリック確率の真値 $c_{i,j}$ ($i = 1, \dots, K, j = 1, \dots, N$)。

であるが、これらに関しては乱数を用いて以下のように生成した。

広告表示比率については、まず一様分布 $U(0.1, 1)$ に従う乱数を K 個生成し、和が 1 になるよう正規化したのちそれぞれを $p_1, \dots, p_K$ の値とした。このように生成することで表示比率が最も少ない広告と多い広告の比率の差が高々 10 倍で抑えられる。Web ページアクセス確率も同様の手法で与えた。

最後に、クリック確率の真値 $c_{i,j}$ はすべて一様分布 $U(0.001, 0.01)$ に従う乱数を生成し、そのままクリック確率とした。この値域については、実際に Web 広告がクリックされる割合はおおよそ 0.7% 程度であるという報告を参考とした [8]。どの手法についても 5 回ずつ実験を行った。

7.2 実験結果

7.2.1 従来手法との比較

図 3 は提案手法 (+ 広告のルーレット選択) と定期的なバッチ

処理により配信スケジュールを更新するオフライン更新手法 [5] の累積クリック確率の変化を示したグラフである。実線 (UCB LP det.) は提案手法、二重破線 (offline update) はオフライン更新手法、破線 (optimal schedule) は今回用いた実験データにおける真の最適解に従って広告を表示した場合のクリック率の期待値を、点線 (random schedule) は常に制約条件である広告配信比率 $\mathbf{p} = (p_1, \dots, p_K)$ に従った単純なルーレット選択による期待値^(注2)を示す。配信スケジュールをバッチ更新する既存手法について、更新間隔 (Learning interval) を設定する必要があるが、今回は 3,125[回] としている。

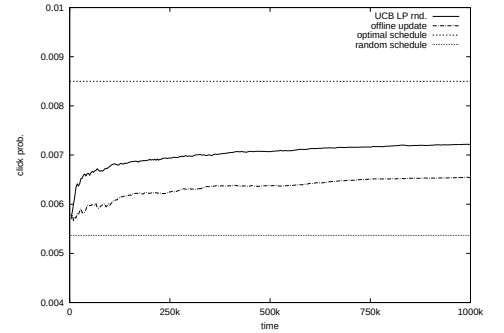


図 3 累積クリック確率の推移

最終的には提案手法による累計クリック確率がやや大きいという結果が得られた。しかし両手法とも学習が進み少しずつ最適値に近づいていることは分かるが、十分に学習が進んでいるとは言い難い。

7.2.2 配信広告決定方法の比較

図 4 は第 6.1 節で提案した二つの配信広告決定方法のそれぞれによる累積クリック確率の推移を示したグラフである。決定的な選択ではルーレット選択より最適解からより遠い位置で収束している。

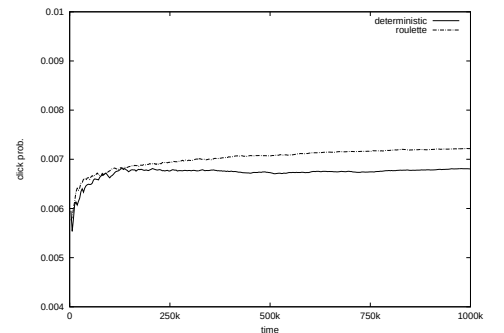


図 4 累積クリック確率による配信広告決定方法の比較

最後に、各手法で得られたクリック数を表 1 に掲載する。

(注2)：この値はは一切の学習をせずとも得られる値である。

表 1 各手法によるクリック数

	オフライン	ルーレット	決定的
平均	6,537	7,217	6,805
最大	6,639	7,805	6,940
最小	6,443	6,646	6,546

7.3 制約充足度の評価

最終的に制約条件として与えられた広告配信比率に近い広告配信を行えたかを評価するために、式 (12) のような評価式を考えた。

$$E = \sum_{i=1}^K \left| \left(\sum_{j=1}^N n_{i,j}(T) \right) - T p_i \right| \quad (12)$$

E の値は小さいほど制約を満たすような広告配信が実現されている (制約充足度の高い広告配信手法である) といえる。この評価式に基づき、オフライン更新手法と提案手法の決定的な配信広告選択、ルーレット選択のそれぞれの配信回数の制約充足度を評価した値を、表 2 にまとめた。オフライン更新手法ではほぼ制約条件に忠実な広告配信が実現されているのに対し、提案手法についてはどちらの配信広告選択方法についても E の値が非常に大きく、やや制約を外れがちになることが見てとれる。

表 2 各手法の制約充足度

手法	オフライン	決定的	ルーレット
E	105.64	69,386	49,826

7.4 計算効率

オフライン更新手法と提案手法のそれぞれの計算量について考える。オフライン更新手法の時刻 t における広告選択は既に計算してある広告表示比率 $x_{i,J(t)}(t)$ を元にルーレット選択するのみであるため、一度の広告選択にかかる計算量は $O(1)$ である。一方、提案手法においては広告の選択自体は $O(K)$ で可能なのだが、前処理としてアフィン変換法を 1 ステップ進める必要がある。アフィン変換法の 1 ステップ中で律速段階となるのは行列の積演算で、行列のサイズは $(K + N - 1) \times KN$ となっている。よって、提案手法では 1 回の広告選択にかかる計算量は $O(K^3 N^3)$ となる。

表 3 は提案手法およびオフライン更新手法のシミュレーション全体で要した実計算時間の表である。

表 3 各手法のシミュレーションに要した実計算時間

手法	従来法	提案手法
オンライン処理	ほぼ 0	14 [ms]
オフライン処理	1700 [ms]	0

8. 考察とまとめ

オフライン更新手法との比較によると、提案手法は累計クリック確率の改善が見られた。広告配信期間の初めにおいて高い学習効率を示していたことが要因であると考えられるが、最適な配信スケジュールに近づく前にクリック率の上昇が鈍っている。

この原因を探り、最適な配信スケジュールに近づけるように改善する必要がある。

配信履歴とスケジュールのずれに着目した決定的な配信広告選択法ではルーレット選択に比べて平均的な性能が良くなかった。

提案手法の次なる問題点として、その高い計算コストがあげられる。実験では問題サイズを 10×20 としたが、Nakamura らの実験 [5] では問題サイズを 32×128 としている。この問題サイズにおいて提案手法による 1 回の広告選択の実計算時間を概算すると、およそ 15 秒程度となる。これは実用的な数値とはいえないため、計算コストの削減が必要である。手がかりとして、制約条件の係数を表すサイズ $(K + N - 1) \times KN$ の行列 A が規則的な構造を持った疎行列となっていることが利用できるのではないかと考えられる。

文 献

- [1] 電通. 2010 年の日本の広告費は 5 兆 8,427 億円. 前年比 1.3%減. NEWS RELEASE, 2011. <http://www.dentsu.co.jp/news/release/2011/pdf/2011019-0223.pdf>.
- [2] J. Gittins, R. Weber and K. Glazebrook, *Multi-armed bandit allocation indices*. Wiley Online Library, 1989.
- [3] P. Auer, N. Cesa-Bianchi and P. Fischer, Finite-time analysis of the multiarmed bandit problem. *Machine learning*, **47**, 2(2002), 235–256.
- [4] M. Langheinrich, A. Nakamura, N. Abe, T. Kamba and Y. Koseki, Unintrusive customization techniques for Web advertising. *Computer Networks*, **31**, 11-16(1999), 1259–1272.
- [5] A. Nakamura and N. Abe, Improvements to the linear programming based scheduling of web advertisements. *Electronic Commerce Research*, **5**, 1(2005), 75–98.
- [6] 茨木俊秀・福島雅夫, 最適化の手法. 共立出版, 1993.
- [7] L. Li, W. Chu, J. Langford and R. Schapire, A contextual-bandit approach to personalized news article recommendation. *Proceedings of the 19th international conference on World wide web*, 2010, 661–670.
- [8] X. Dreze and F. Hussherr, Internet advertising: Is anybody watching?. *Journal of interactive marketing*, **17**, 4(2003), 8–23.