

ウェブページ最適化問題の定式化と最適化手法の提案

Webpage Optimization Problem and Its Optimization Method

飯塚 修平 *1
Shuhei Iitsuka

東京大学大学院 工学系研究科
Graduate School of Engineering, the University of Tokyo
iitsuka@weblab.t.u-tokyo.ac.jp, <http://tushuhei.com>

松尾 豊
Yutaka Matsuo

(同 上)
matsuo@weblab.t.u-tokyo.ac.jp, <http://ymatsuo.com>

keywords: webpage optimization, optimization problem, A/B testing

Summary

Webpage optimization is an experimental method to make continuous improvements on websites based on users' behavior. This method can be implemented easily but has a drawback that small websites take long time to gather enough data to evaluate the ideas. Although many optimization methods are proposed and conducted so far, there's no mathematical model of this problem. We propose Webpage Optimization Problem and organize existing webpage optimization methods. Combining these methods, we also propose a new webpage optimization method that performs well no matter how many people get to the website. We evaluate the proposal method by simulation experiments and introducing the optimization program to both large and small websites. The results show that our proposal method outperforms existing methods at any size of websites. Webpage Optimization Problem is a framework to create new webpage optimization methods.

1. はじめに

ウェブサービスを成長させる手法として、近年ウェブページ最適化が注目を集めている。ウェブページ最適化とはデザインや機能の一部が異なるウェブページを用意し、ユーザに見せた時の反応の違いを計測することで、広告のクリック率やページ滞在時間などの評価指標を最大化するパターンを発見する手法のことである。このように実験から得られたデータに基づいてウェブページを逐次改善する手法は、ランディングページ最適化やスプリットテストという名前でも親しまれている [Ash 12]。Optimizely*1, Google Website Optimizer*2, planBCD*3 など国内外で様々なウェブページ最適化ツールが生まれ、ウェブページ最適化は一般的な開発手法になりつつある。

ウェブページ上のボタンの文言やページに掲載する写真素材などデザイン上の細かな違いがユーザの行動を左右し、経済的に大きな効果をもたらすことが知られている。たとえば 2008 年アメリカ合衆国大統領選挙では、バ

ック・オバマ氏が公式ウェブサイトでは支援者の登録率を向上させるために、ウェブページ最適化を活用した*4。この実験では 6 種類の写真と 4 種類のボタンからなる 24 種類のパターンに対してテストが行われた (図 1 参照)。実験の結果、最もパフォーマンスの高かったパターンを採用することで、約 6000 万ドルの献金を追加で獲得することに成功した。

しかし、ウェブページ最適化にも問題点はある。そのひとつは、ウェブサイトの規模によって実験結果を得られるまでの時間が異なることである。ユーザ数が多いウェブサイトでは短期間にたくさんのデータを得ることができるため、早期に実験の結果を得ることができるが、ユーザ数が少ないウェブサイトではパターンの評価に十分なデータが集まるまでに時間がかかってしまう。

他にもウェブページ最適化にまつわる問題点はあるが、既存のウェブページ最適化手法が体系化されていないため、問題を解決する新たな最適化手法を生み出す枠組みが存在しないのが現状である。今後のウェブページ最適化の理論および手法の発展のためには、まずウェブページ最適化という問題を定式化することが求められている。

*1 現在は、Google 株式会社に勤務。

*2 Optimizely <https://www.optimizely.com/>

*3 Google Website Optimizer <http://www.google.com/analytics/> 現在は同社が提供するアクセス解析ツール Google Analytics のひとつの機能として統合されている。

*4 planBCD <https://planb.cd>

*4 How Obama Raised \$60 Million by Running a Simple Experiment <http://blog.optimizely.com/2010/11/29/how-obama-raised-60-million-by-running-a-simple-experiment/>



図1 バラック・オバマ氏公式ウェブサイトのトップページでの実験に用いられた写真とボタンのバリエーション

そこで本研究では、ウェブページ最適化を定式化したウェブページ最適化問題を提案する。そして既存のウェブページ最適化手法をウェブページ最適化問題の解法として整理し、これらを組み合わせることで新しい最適化手法を生み出すことが可能になることを示す。今回はウェブサイトのユーザ数によって最適化に要する時間が変化するという問題に着目し、ユーザ数の大小に関わらず最適化を行うことができるウェブページ最適化手法を提案する。

提案手法は解空間全体から線形モデルによって最適解を予測する手法と、観測数を有望な解に集中させることで最適解の探索を高速化する手法を組み合わせたものである。本研究では、筆者が設定した仮想状況におけるシミュレーション実験、大規模ウェブサイトから得られたデータに基づいたシミュレーション実験および中規模ウェブサイトへの最適化プログラムの導入実験を通して提案手法の有効性を評価する。

本研究の貢献は以下のとおりである。

- ウェブページ最適化問題を定式化し、問題を捉える枠組みを提案した。
- 既存のウェブページ最適化手法を整理し、新しいウェブページ最適化手法を生み出す基礎を作った。
- ウェブサイトのユーザ数の大小に関わらずウェブページ最適化を行うことができる手法を提案した。

本論文の構成は以下のとおりである。2章ではウェブページ最適化の関連研究および既存のウェブページ最適化手法について説明する。3章では問題の定式化および既存手法の整理を行う。4章ではユーザ数の大小に関わらず最適化を行うことができる手法を提案する。5章では評価実験について説明する。6章で考察を行い、7章を本論文のまとめとする。

2. 関連研究

ユーザの行動を左右するウェブページの要素を特定し、ウェブサイトの改善につなげる研究は以前から行われている。たとえばEコマースサイトをレイアウト、デザイ

ン、情報の種類などの属性によって表現し、ユーザヒアリングの結果から各属性の影響度を算出する研究が行われている [Gehrke 99]。同様にオークションサイトを属性で表現し、ユーザの入札意思から各属性の影響度を算出する研究も行われている [Kwon 02]。これらは一般的な属性の影響度を算出したものであるが、個々のウェブサイトにおいて重要な要素を発見する手法を述べたものではない。それに対し、本研究ではあらゆるウェブサイトを導入することができる最適化手法およびその枠組みを提案する。

ウェブページ最適化を実際のウェブサイトで行なった研究もある [Crook 09, Kohavi 09]。実験上の注意点や各最適化手法の特徴がケーススタディとしてまとめられているものの、既存のウェブページ最適化手法を体系だてて整理したものは少ない。

従来のウェブページ最適化手法の代表的なものにはA/Bテストがある。A/BテストではオリジナルのパターンをA、操作を施したパターンをBとしてユーザをランダムに振り分け、ユーザの反応を比較することで最適なパターンを探し出す。効果を正しく検証するために、操作を加える要素はひとつに絞るのが原則である。また、ばらつきがある結果に対して統計的検定を用いることで、有意差を持って判断を下すことができる。

A/Bテストは一度にひとつの要素について最適化を行うのに対して、複数の要素の最適化を同時に行うのが多変量テスト (Multivariate Testing) である。多変量テストでは、まず各要素が取りうる値のすべての組み合わせを生成する。そして、その中から評価するパターンを選択してユーザを振り分け、反応を比較する [Ash 12]。多変量テストには主に総当り実験 (Full Factorial Test Design) と直交配列実験 (Fractional Factorial Test Design) の二つの方法がある。

総当り実験はすべてのパターンを対象として対照実験を行う方法である。厳密に最適なパターンを見つけることができるが、要素の数が多くなると評価すべきパターンの数が膨大になるという欠点がある。

直交配列実験は一部の要素間の高次の相乗効果は十分に小さいため無視できるという仮定を置くことで、評価するパターンを一部に絞って対照実験を行う方法である。それぞれの要素が取りうる値を同じ回数だけ評価できるように工夫して、予め評価するパターンを決定する。

近年ではウェブページのパターンを解、表示するパターンを決定するプログラムをエージェントとみなすことで、バンディットアルゴリズムがウェブページ最適化に応用されている [White 12]。バンディットアルゴリズムは、強化学習においてエージェントが効率的な学習を行うために用いられるアルゴリズムである。エージェントが取る行動には、それまでの計測値から算出される期待値が最も大きい解を選択する活用と、期待値が最も大きい解以外を選択することで未知の解の評価値を求める探求のふ

たつがある．バンディットアルゴリズムはこの活用と探求のトレードオフの中でエージェントが最良の解を速く見つけ出すための解法である [Auer 02] ．

3. ウェブページ最適化問題の定式化

3.1 問題の定式化

ウェブページ最適化とは、デザインや機能の一部が異なるウェブページのバリエーションの中からクリック率やページ滞在時間などの評価指標を最大化するパターンを探索することである*5．それぞれのパターンはウェブページを構成するボタンやラベルなどの要素の組み合わせによって構成されているため、ウェブページ最適化は組合せ最適化問題のひとつとして捉えることができる．また、ユーザにパターンを表示することではじめて反応を計測できるため、逐次的に評価値が与えられる環境下での最適化問題となる．

ウェブページのパターン $x = (x_1, \dots, x_m)$ はウェブページを構成する m 個の要素の組み合わせによって表現することができる．それぞれの要素 x_i は離散値であり、その取りうる値の集合 $V_i = \{v_{i1}, \dots, v_{il_i}\}$ の中からひとつ選ばれるものとする．ただし l_i は要素 x_i が取りうる値の数である．たとえば、ある食料品を紹介するウェブページ x が商品の写真 x_1 、キャッチコピー x_2 、購入ボタンの色 x_3 という 3 つの要素の組み合わせによって表現されるとき、購入ボタンの色 x_3 は $V_3 = \{\text{赤}, \text{青}, \text{緑}, \text{黄}\}$ のいずれかひとつの値を取るようになる．

ウェブサイト開発者は評価指標 $f(x)$ を最大化するパターン x^* を求めたい．しかし評価指標 $f(x)$ は直接観測することが出来ず、パターン x をユーザに表示した時の反応から計測される観測値 y がある確率分布 $p(y|x)$ に従って与えられるのみである．ここでは評価指標 $f(x)$ を観測値 y の条件付き期待値 $\mathbb{E}[y|x]$ によって推定する．

ウェブページ最適化では実験に割くことができる時間が限られていることが多く、その場合は限られたユーザ数の中で最適化を行う．最適化に用いることができる観測値の数を限界観測数 N とする．また、ユーザに表示したパターンと観測値の対の集合 $\{(x_1, y_1), \dots, (x_N, y_N)\}$ を観測データ Y と呼ぶことにする．

以上より、ウェブページ最適化問題は式 (1) のように定式化することができる．

ウェブページ最適化問題

解 $x \in X$ に対する観測値 y が確率分布 $p(y|x)$ に従って与えられるとき、下式を満たす解 x^* を求めよ．

$$x^* = \arg \max_{x \in X} \mathbb{E}[y|x] \text{ s.t. } y \sim p(y|x) \quad (1)$$

*5 ここでは最適化問題を最大化問題として考える．

表 1 既存のウェブページ最適化手法とウェブページ最適化問題の解法の対応付け

最適化手法	解法	工夫
A/B テスト	局所探索	なし
総当り実験	総当り	なし
直交配列実験	総当り	線形モデル予測
バンディット	総当り	観測数の集中

3.2 局所探索法を用いた解法

ウェブページ最適化問題は一種の組合せ最適化問題であるため、一般的な探索手法は導入することができる．ウェブページ最適化問題には観測回数に対して強い制約が与えられているため、最適性の保証は無いが高速に近似解を得ることができるヒューリスティクスを用いた探索手法が相性が良いと考えられる．特に組合せ最適化問題では良い解同士は似通った構造を持っているという仮定が成り立つことが多く、局所探索法が広く用いられている [金澤 02] ．ここではウェブページ最適化問題の解法として局所探索法を用いた場合について説明する．

局所探索法のアルゴリズムを Algorithm 1 に示す．局所探索法では、まず適当に初期解 $x \in X$ を設定する．この初期解を暫定解とし、暫定解から近傍解 X' を生成する．近傍解の中に暫定解より良い解があればその解に移動する．局所探索法ではこの近傍解の生成と暫定解の移動を終了条件を満たすまで繰り返すことで、最適解を求める [野々部 08] ．

Algorithm 1 局所探索法のアルゴリズム

Choose $x \in X$.

repeat

$X' \leftarrow \text{Neighbors}(x)$ ▷ 近傍解の生成

$x \leftarrow \text{Move}(x, X')$ ▷ 暫定解の移動

until 終了条件

return x

近傍解とは、暫定解に小さな操作を加える事で得ることのできる暫定解と似た解のことである．ウェブページ最適化問題における解 x は独立な変数の組み合わせによって表現されるため、ある変数 x_i の値を変数 x_i が取りうる他の値と入れ替えることによって、近傍解 X' を生成することができる．

暫定解の移動とは、近傍解に暫定解より評価値が高い解があれば、その解で暫定解を更新することである．ウェブページ最適化問題では解 x の真の評価値 $f(x)$ を直接観測することが出来ないため、解 x の観測値の条件つき期待値 $\mathbb{E}[y|x]$ によって解の評価を行う．

3.3 既存のウェブページ最適化手法の整理

既存のウェブページ最適化手法をウェブページ最適化問題の解法として対応付けたものを表 1 にまとめる．

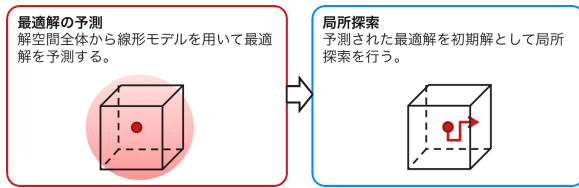


図2 提案手法の概要図

A/B テストを繰り返すことは、無作為に選んだ解を初期解 x_0 として局所探索することに相当している。テストを行う変数 x_i を選び、その変数が取りうる値の集合 V_i から変数値をひとつ選んで入れ替えて近傍解 x' を生成する。近傍解が暫定解より評価値が高ければ、その近傍解で暫定解を更新することでひとつの A/B テストが完了する。その後新たな変数 x_j を取り出して再び A/B テストを行うことで、探索を繰り返すことができる。

総当り実験は実行可能解すべてを観測して最適解を探索することに相当する。直交配列実験も同様に総当り探索であるが、データの収集後に各変数の最適な値を組み合わせ合わせたものを最適解として推定しているという点で、下式で表される線形モデルで評価関数を仮定し、最適解の予測を行なっていると考えることができる。

$$f(x) = g_1(x_1) + \dots + g_m(x_m) + a$$

ただし $g_i(x_i)$ は変数 x_i の評価関数 $f(x)$ への寄与を表しており、 a は定数である。本研究ではこの工夫のことを線形モデルによる最適解の予測と呼ぶ。

バンディットアルゴリズムでは条件付き期待値 $\mathbb{E}[y|x]$ に応じて、各解に割り当てる観測数を変化させている。評価値が劣る解の観測数を少なくし、有望な解に多くの観測数を割くことで、効率的な探索を実現している。このような工夫を本研究では観測数の集中と呼ぶ。

4. 提案手法

本研究では既存のウェブページ最適化手法を組み合わせることで、ユーザ数の大小に関わらず適用することができるウェブページ最適化手法を提案する。多くのユーザ数が得られる大規模ウェブサイトではすべての解について十分な観測数を得ることができるが、ユーザ数が少ない小規模ウェブサイトではすべての解について十分な観測数を得ることが難しい。そのため、評価関数の線形性を仮定してでも、できるだけ少ない観測数で解空間全体から最適解を得ることが望ましい。そこで提案手法では、線形モデルによる最適解の予測と局所探索を組み合わせた探索を行う。

提案手法の概要図を図2に、アルゴリズムを Algorithm 2 に示す。提案手法による最適化に必要なものは、解空間 X 、限界観測数 N 、有意水準 α 、検出力 $1 - \beta$ 、効果量 Δ の5つである。有意水準 α は帰無仮説を棄却する

正確さを表す。検出力 $1 - \beta$ は比較する解の評価値の差を検出することができる確率を表す。効果量 Δ は解の評価値の差を標準偏差を用いて標準化したものである。

Algorithm 2 提案アルゴリズムの全体

Require: α as the significant level.

Require: $1 - \beta$ as the power.

Require: Δ as the effect size.

Require: N as the maximum number of observations.

Require: X as the set of solutions.

Set Y as an empty set for observed data.

Set $n \leftarrow 0$ as the number of total observations.

Set $N_1 \leftarrow N_{ANOVA}(\alpha, \beta, \Delta, X)$

Set $N_2 \leftarrow N_{T-TEST}(\alpha, \beta, \Delta)$

$x^*, Y, n \leftarrow Initialization(X, Y, N_1, n)$

while $n < N$ **do**

$x^*, Y, n \leftarrow Move(x^*, X, Y, N_2, n)$

return x^* as the optimal solution.

線形モデルによって最適解を予測するのに要するサンプルサイズ N_1 には、各変数にとる変数値によって評価値が変化することを検出するのに十分な大きさが求められる。ここでは一元配置分散分析に必要なサンプルサイズ N_{ANOVA} を用いることにする。予測のためには各変数のすべての値について観測値を得られればいいので、ここでは群の数 $k = \sum_{i=1}^m l_i$ とする。局所探索の解の評価に要するサンプルサイズ N_2 には t 検定に要するサンプルサイズ N_{T-TEST} を用いることにする。サンプルサイズの算出には、いずれも [Cohen 88] の方法を用いる。

提案手法ではまず *Initialization* にて局所探索の初期解を決定する。ここでは解空間 X 全体から解を N_1 回無作為に抽出して観測値を収集し、線形モデルによる最適解の予測を行う。解の移動 *Move* では、*Initialization* で得られた初期解 x のある変数 x_i について値を入れ替えて近傍解 x' を生成し、それぞれ N_2 回観測値を収集する。観測値の収集が終了した時点で評価値の期待値 $\mathbb{E}[y|x]$ を比較し、期待値が大きい方の解で暫定解を更新する。累計の観測数 n が限界観測数 N に達するまで解の移動を繰り返して局所探索を行うことで、線形モデルによる予測では見逃されてしまった最適解を探索する。

ここに観測数の集中を導入することで探索を高速化することができる。解の期待値に応じて観測数の配分を変化させるためにバンディットアルゴリズムを用いることもできるが、ここでは区間推定を応用したレーシングアルゴリズムを用いることにする。レーシングアルゴリズムは観測値が得られるたびに解 x の評価値 $f(x)$ の信頼区間 $[\mu_{Lx}, \mu_{Ux}]$ を算出し、有意に評価値が良い解があれば採用し、有意に評価値が悪い解があれば観測対象から除くことで探索を高速化する工夫である。強化学習におけるエージェントの方策やメタヒューリスティクスのパ

ラメータ設定の優劣を比較するために用いられることが多い [Birattari 02, Heidrich-Meisner 09] . μ_{Lx} を解 x の下側信頼限界, μ_{Ux} を解 x の上側信頼限界と呼ぶ .

解 x_w の下側信頼限界 μ_{Lx_w} が他のすべての解の上側信頼限界を上回る, すなわち

$$\mu_{Lx_w} > \max_{x \in X \setminus \{x_w\}} \mu_{Ux}$$

が成り立つとき, 解 x_w を最良の近傍解として採用することにして観測を中断し, 暫定解の更新を行うことで観測数を節約できる . 本研究ではこの工夫のことを観測の短縮と呼ぶ . 一方, 解 x_l の上側信頼限界 μ_{Ux_l} を上回る下側信頼限界を持つ解が存在する, すなわち

$$\mu_{Ux_l} < \max_{x \in X \setminus \{x_l\}} \mu_{Lx}$$

が成り立つとき, 解 x_l が近傍解のなかで最良解となる可能性は無いとみなして観測対象から外すことで, 有望な解に多くの観測数を割くことができる . 本研究ではこの工夫のことを敗者解のカットオフと呼ぶ .

提案アルゴリズムにこれらの工夫を導入することで, さらに効率的に探索を行うことができる . 提案アルゴリズムに観測の短縮を適用することで, *Initialization* において $n = N_1$ に到達するのを待つことなく初期解を決定することができる . *Move* では有意に良い近傍解があれば観測数 N_2 を消費する前に解の移動を行うことができる . ここでは提案アルゴリズム (提案 A と呼ぶ) に観測の短縮を導入したものを提案 B と呼ぶことにする .

さらに, 提案アルゴリズムに敗者解のカットオフを導入することで *Initialization* において有意に悪い変数値を持つ解の探索を省略し, 最適解の予測を高速化することができる . *Move* では有意に劣る近傍解の探索を省略し, 有望な近傍解へ観測数を集中することができる . ここでは提案 B に敗者解のカットオフを導入したものを提案 C と呼ぶことにする .

5. 実験

提案手法によってウェブサイトのユーザ数の大小に関わらずウェブページ最適化ができることを検証する . まず様々なウェブサイトの状況を想定したシミュレーション実験を行う . 次に実際の大規模ウェブサイトと中規模ウェブサイトに提案手法を適用することで, ユーザ数の大小に関わらず提案手法が有効であることを示す .

5.1 比較検討するアルゴリズム

今回の実験で比較検討するアルゴリズムを表 2 にまとめる . ベースライン A は従来の総当り実験に相当する最適化手法であり, 無作為に解を選択して観測を行う . ベースライン B は従来の A/B テストに相当する最適化手法であり, 山登り法による探索を行う . ベースライン B の

表 2 比較検討するアルゴリズム

	<i>Initialization</i>	<i>Move</i>
ベースライン A	ランダム	総当り
ベースライン B	ランダム	山登り
提案 A	線形モデル予測	山登り
提案 B	線形モデル予測 +観測の短縮	山登り +観測の短縮
提案 C	線形モデル予測 +観測の短縮 +カットオフ	山登り +観測の短縮 +カットオフ

表 3 シミュレーションの問題設定

問題	評価関数	限界観測数 N
1	線形関数 $f_1(x)$	400 ($N < N_1$)
2	線形関数 $f_1(x)$	1000 ($N > N_1$)
3	非線形関数 $f_2(x)$	400 ($N < N_1$)
4	非線形関数 $f_2(x)$	1000 ($N > N_1$)

近傍解の評価にはサンプルサイズ N_2 の観測数を割くものとする . 提案 A は解空間全体から N_1 回観測値を収集し, 線形モデルによる最適解の予測の結果を初期解として山登り法による探索を行う . 提案 B は提案 A に観測の短縮を導入したものである . 提案 C は提案 B に敗者解のカットオフを導入したものである .

5.2 仮想の評価関数によるシミュレーション実験

仮想状況におけるシミュレーション実験では, 様々なウェブサイトの状況を想定して表 3 に記す問題を独自に設定し, それぞれの問題に対してベースラインアルゴリズムと提案アルゴリズムを適用した . 問題を設定する際には, 評価関数が線形関数 $f_1(x)$ のときと非線形関数 $f_2(x)$ のときで場合分けをした . ここでいう非線形性とは, $g_1(x_1)g_2(x_2)$ のように複数の変数の相乗効果をもたらす非線形項が評価関数に含まれることを指す . さらに, 限界観測数 N が線形モデルによる最適解の予測に要するサンプルサイズ N_1 より大きい場合か否かで場合分けをした . ここでは変数の数 $m = 6$, それぞれの変数 x_i が $V = \{0, 1, 2\}$ のいずれかの値を取るものとして解空間および評価関数を設定した*6 . このとき解候補の数は $3^6 (= 729)$ である . この解候補 X の中から限界観測数 N の中で最適解 x^* を探索することが解くべき問題となる . 統計的検定にま

*6 評価関数は

$$f_1(x) = x_1 + x_2 + x_3 - x_4 - x_5 - x_6 + N(0, 1)$$

$$f_2(x) = x_1 + x_2 + x_3 - x_4 - x_5 - x_6 - x_1x_2 + N(0, 1)$$

とした . $f_1(x)$ では各変数にかかる符号が同回数現れるように工夫し, $x_i = 0$ が最適解となる変数と $x_i = 2$ が最適解となる変数を同数含むようにした . $f_2(x)$ では同じ符号を持つ変数をふたつ取り出して掛けあわせ, 逆の符号をつけることで負の相乗効果をもたらす非線形項を加えた . ただし $x_i \in \{0, 1, 2\}$ であり, $N(0, 1)$ は平均 $\mu = 0$, 標準偏差 $\sigma = 1$ の正規分布に従う確率変数である .

表4 各アルゴリズムの正確度

問題	ベースライン		提案		
	A	B	A	B	C
1	0.28	0.00	1.00	1.00	1.00
2	0.34	0.01	1.00	1.00	1.00
3	0.37	0.02	0.18	0.19	0.34
4	0.32	0.01	0.29	0.91	0.93

つわるパラメータは $\alpha = 0.05, 1 - \beta = 0.8, \Delta = 0.2$ と設定してサンプルサイズを算出し、 $N_1 = 550, N_2 = 400$ を得た。

最適化手法の評価には最適解からの平均誤差、暫定解が最適解である確率（正確度と呼ぶ）や損失の総和などが用いられることが多い[Auer 02]。今回の実験では各アルゴリズムが限界観測数 N の時点で最適解に到達できていることを評価するために、正確度を用いて各アルゴリズムのパフォーマンスを比較した。

各問題に対して比較検討するアルゴリズムを適用した結果を表4に示す。問題1ではベースラインAは観測値にばらつきがあるために正確度を上昇させることができず、ベースラインBも十分な解の移動を行うことができないため低い正確度にとどまっていることがわかる。それに対し提案アルゴリズムでは線形モデルを仮定して最適解の予測を行なうため、高い精度で最適解を予測することができている。問題2についても問題1と同様の傾向が見られ、評価関数に非線形性が無い限りは線形モデルによる予測によって高い正確度で探索できることがわかる。

一方、評価関数に非線形性が含まれる問題3では、線形モデルを仮定した最適解の予測の精度が落ちてしまうため、提案アルゴリズムの正確度はベースラインと同程度またはそれ以下まで落ちてしまっている。しかし、評価関数に非線形性が含まれる場合でも問題4のように大きな限界観測数 N が与えられる場合には、その後局所探索を行うことで正確度を向上することができている。特に提案B、Cでは探索の効率化のための工夫を導入しているため、高い正確度で最適解に到達している。

5.3 大規模ウェブサイトでのシミュレーション実験

次に、実際の大規模ウェブサイトで得られたデータを元にシミュレーションを行い、提案手法が有効であることを検証した。今回は一日のユーザ数が数万規模の大規模な人物検索サイト「あのひと検索スパイシー^{*7}（以下実サイトAと呼ぶ）」で得たログデータをもとに評価関数を設定し、シミュレーションを行なった。実サイトAは人物のプロフィールを紹介する「プロフィールページ」と人物同士のつながりを紹介する「関連図ページ」から構成されるシンプルなウェブサイトである。そのため評

表5 実サイトAにおける各施策の効果

施策名	表示回数	クリック率
施策A	9,739	7.57%
施策B	1,4923	5.73%
施策C	5,869	5.06%
オリジナル	71457	6.40%

価指標が設定しやすいと判断し、本サイトを採用した。

今回の実験では、関連図ページにおける広告のクリック率を評価指標として設定した。2013年5月14日から18日にかけて実サイトAで行なった実験で、ある施策Aを関連図ページに施すことによって広告のクリック率を1.17%引き上げることに成功した。同時期に他に施策B、Cも行なっていたが、これらの施策は逆にクリック率を減少させる方向に働いた。各施策による効果を表5にまとめる。オリジナルパターンのクリック率と各施策を施したパターンのクリック率の間でカイ二乗検定を行なった結果、有意水準0.01でパターン間の広告クリック率に有意差があることがわかった。

ここで、施策 i を行うか否かの二値 (0, 1) をとる変数 x_i を用いると、ある解 x を訪問者に表示した時に広告がクリックされる確率は

$$q(x) = 0.0640 + 0.0117x_A - 0.0067x_B - 0.0134x_C$$

と表すことができる。今回の実験では実際のユーザの反応によるデータを収集しているうえに、大規模ウェブサイトの場合は施策の成否がウェブサイト全体の収益に大きな影響を与えるため、実験を繰り返し行うことが難しい。そのため、ここでは各施策が線形に効果を表すものと仮定する。評価関数 $f(x)$ は解 x に対して確率 $q(x)$ で1を、確率 $1 - q(x)$ で0を返すものとする。統計的検定にまつわる各パラメータは $\alpha = 0.05, 1 - \beta = 0.8, \Delta = 0.05$ と設定した。ここから算出される各サンプルサイズは $N_1 = 5500, N_2 = 6500$ であった。

それぞれのアルゴリズムを適用した結果を図3に示す。ベースラインAは観測回数を経るごとに観測データが増え、着実に正確度を向上することができている。ベースラインBは解の比較を繰り返すには与えられた限界観測数が不十分であり、他のアルゴリズムに比べると正確度が低くなっている。提案手法はいずれもベースラインを上回る結果となった。

今回のシミュレーションで仮定した評価関数は施策を行う・行わないの二値変数の線形結合であるため、提案アルゴリズムはいずれも観測数 $n = 5500$ の初期解決時点で高い正確度を示している。局所探索では、提案BおよびCが観測の短縮および敗者解のカットオフによって頻繁な解移動を行っていることがわかる。ただしその変化はばらつきが大きく、評価値が低い解に移動してしまう様子も見て取れる。

*7 あのひと検索スパイシー <http://spysee.jp/>

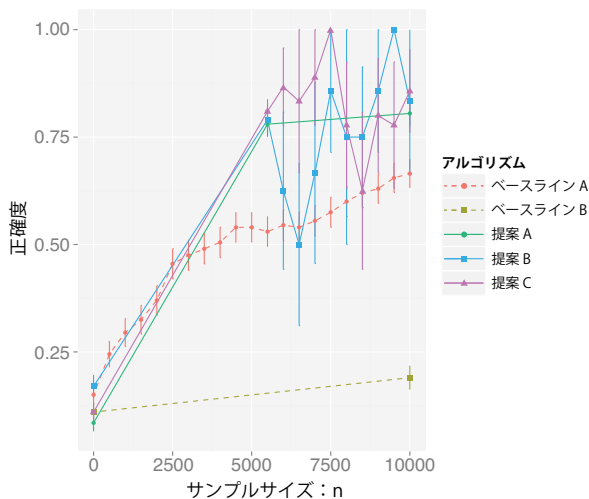


図3 実サイト A から得た評価関数における各アルゴリズムの正確度の推移

表6 実サイト B を構成する変数と値

変数	対応する要素	取りうる値
x_1	画像の枠線の太さ	0px, 5px
x_2	画像の間隔	0px, 5px, 10px
x_3	画像のサイズ	100px, 200px, 300px
x_4	画像の切り抜き	正方形, 正円

5.4 中規模ウェブサイトにおける導入実験

一日のユーザ数が数百と少ない中規模ウェブサイトにおいても提案アルゴリズムが有効であることを評価するために、壁紙画像検索サービス「Imagerous^{*8} (以下実サイト B と呼ぶ)」を対象に実験を行なった。本実験では提案手法を実装した最適化プログラムをウェブサイト導入して有効性を評価した。実サイト B はパソコンやスマートフォンの壁紙に使える画像を検索したり、アイドルや女優のグラビア画像を検索したりして楽しむことを目的としたウェブサイトである。写真一覧ページと写真詳細ページからなるシンプルなウェブサイトであるため、評価指標の設定および測定が容易と判断して採用した。今回は訪問あたり閲覧ページ数を評価指標として設定した。ユーザの中には一回の訪問で数百ものページを閲覧する、ロボットとみられる動きをするものもあったため、評価指標に上限値を設定した。

今回はこのページを構成する要素として表 6 に示す変数と値を設定し、これらの組み合わせを候補とした。ここでは各変数の値を 0 以上の整数で表すこととし、整数の組み合わせによって解を表現することにする。たとえば、 $x = (0, 2, 1, 0)$ は { 画像の枠線の太さ: 0px, 画像の間隔: 10px, 画像のサイズ: 200px, 画像の切り抜き: 正方形 } という設定を表している。

今回の実験ではベースライン B と提案 A の 2 つのア

表7 暫定解と期待値の推移

手法	観測数	暫定解	期待値	標準誤差
ベース B	0	(1, 1, 2, 1)	N/A	N/A
	288	(1, 2, 2, 1)	2.560	0.188
	593	(1, 2, 2, 1)	2.470	0.136
提案 A	0	N/A	N/A	N/A
	122	(1, 0, 0, 0)	7.000	Inf.
	376	(0, 0, 0, 0)	3.760	0.300
	610	(0, 0, 0, 0)	3.873	0.214

ルゴリズムを適用した。中規模サイトでは最適化に用いることができるユーザ数が少数であるため、多くのアルゴリズムを同時に試そうとすると時間がかかり、時間変化による外的要因が実験結果に入り込んでしまう。そのため、今回は提案アルゴリズムのうち高速化の工夫を導入していない提案 A のみを取り上げ、同じ期間でベースラインアルゴリズムよりも良い解を見つけることができることを示す。提案 B, C は高速化のための工夫を導入しているため、提案 A 以上のパフォーマンスを示すことが予想される。今回の実験では有意水準 $\alpha = 0.05$ 、検出力 $1 - \beta = 0.8$ 、効果量 $\Delta = 0.3$ とした。各サンプルサイズは $N_1 = 120, N_2 = 240$ と算出された。

2014 年 2 月 4 日から 14 日にかけて実サイト B で行なった実験の結果を表 7 に示す。提案 A が到達した最適解の評価値の期待値は、ベースライン B のそれを上回った。t 検定の結果、二つの期待値の間には有意水準 0.01 で有意な差が見られた。ベースライン B ではランダムに選ばれた初期解から大きく評価値を改善することができていないのに対して、提案 A は有望な初期解から局所探索を開始し、評価値の期待値が高い解に到達することができている。

6. 考 察

ウェブサイトによっては、得られるサンプルサイズに比べて非常に小さい限界観測数 N の中で最適化を行わなければならない状況もありうるが、提案手法は少ない観測数でも最適化を行うことができる。解空間のサイズ $|X| = \prod_{i=1}^m l_i$ は変数の数 m に対して指数関数オーダーで増加するのに対し、線形モデルによる最適解の予測を行うのに必要なサンプルサイズ $N_1 = \sum_{i=1}^m l_i$ は線形関数オーダーで増加する。したがって、提案手法は得られるサンプルサイズよりも解空間が大きい場合でも適用することが可能であり、サンプルサイズの制約に強い最適化手法だということができる。

変数間の相乗効果が大きく、線形モデルによる予測では評価値の高い解から探索を開始できない可能性もあるが、提案手法はその後局所探索を行うことで最終的に評価値の高い解に到達することができる。提案手法の発展とし

*8 Imagerous <http://imagero.us/>

て、擬似焼きなまし法やタブーサーチなどのメタヒューリスティクスを用いて大域的に最適解を探索する手法も考えられる。

逐次的に評価値が与えられる環境下での最適化問題に対する解法として、他にも強化学習やメタヒューリスティクスを用いた探索などが考えられる。他の最適化手法と比較して提案手法が有利な点の一つは、パラメータを設定する必要がないことである。これらの手法ではパラメータによってパフォーマンスが大きく変化することが多く、何度か施行して試行錯誤する中で最適なパラメータを探し出す必要がある。それに対して、提案手法に必要なパラメータはサンプルサイズの算出に要する有意水準、検定力、効果量のみであり、[Cohen 88] で目安となる値も示されている。このことから、提案手法は未知のウェブページに対して適用しても安定したパフォーマンスを発揮することができる手法であると考えられる。一方、提案手法は他の最適化手法と異なり評価関数の線形性を仮定する必要がある。評価関数の特性によっては評価値の低い解から探索を開始することになり、探索に時間がかかったり局所最適解に収束したりする可能性がある。

本研究では大規模実サイト A、中規模実サイト B の 2 つを取り上げ、ユーザ数に関係なく提案手法が有効であることを示した。また、実サイト A ではクリック率を評価関数に設定したのに対し、実サイト B では訪問あたりのページビュー数を評価関数として実験を行なった。このことから、様々な確率分布の評価関数について提案手法が有効であることが示されたと考えている。

今回提案したウェブページ最適化問題はウェブ以外の分野にも応用することができる。たとえば近年の 3D プリンタの登場によって、ものづくりの世界でもウェブと同様に簡単に製品を生成することができるようになった。ここで製品をデザインや機能の組み合わせによって表現できる解とし、製品への評判や利用状況に関するログを観測値とみなせば、ものづくりもウェブページ最適化問題のひとつとして捉えることができる。ウェブページ最適化問題は、ばらつきのある観測値を元に最適解を探索するあらゆる分野に共通するものである。

7. ま と め

本研究ではウェブページ最適化問題の定式化、既存のウェブページ最適化手法の整理およびウェブページのユーザ数の大小に関わらず最適化を行うことができるウェブページ最適化手法の提案を行なった。ウェブページ最適化問題は解の評価値がばらつきを持った値として与えられ、観測値に制約がある中で最適解を探索する組み合わせ最適化問題である。今回はウェブページのユーザ数の大小によってウェブページ最適化に要する時間が変化してしまう問題に着目し、ユーザ数が少なくても適用することができる最適化手法を提案した。提案手法はウェブ

ページ最適化問題を定式化したことによって整理された既存の最適化手法を組み合わせたものである。最初に線形モデルによる最適解の予測を行い、それを初期解とした局所探索を行う。提案手法を様々な状況を想定したシミュレーション実験および大規模サイトと中規模サイトでの実験によって評価した結果、提案手法の有効性が示された。今回提案したウェブページ最適化問題および最適化手法は、ばらつきのある観測値を元に最適解を探索する問題であれば、ウェブ以外の分野にも応用できるものである。

◇ 参 考 文 献 ◇

- [Ash 12] Ash, T., Ginty, M., and Page, R.: *Landing Page Optimization: The Definitive Guide to Testing and Tuning for Conversions*, ITPro collection, Wiley (2012)
- [Auer 02] Auer, P., Cesa-Bianchi, N., and Fischer, P.: Finite-time analysis of the multiarmed bandit problem, *Machine learning*, Vol. 47, No. 2-3, pp. 235–256 (2002)
- [Birattari 02] Birattari, M., Stützle, T., Paquete, L., and Varrentrapp, K.: A Racing Algorithm for Configuring Metaheuristics, in *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 11–18 (2002)
- [Cohen 88] Cohen, J.: *Statistical power analysis for the behavioral sciences*, Psychology Press (1988)
- [Crook 09] Crook, T., Frasca, B., Kohavi, R., and Longbotham, R.: Seven Pitfalls to Avoid when Running Controlled Experiments on the Web, in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1105–1114 (2009)
- [Gehrke 99] Gehrke, D.: Determinants of Successful Website Design: Relative Importance and Recommendations for Effectiveness, in *Proceedings of the Thirty-second Annual Hawaii International Conference on System Sciences* (1999)
- [Heidrich-Meisner 09] Heidrich-Meisner, V. and Igel, C.: Hoeffding and Bernstein Races for Selecting Policies in Evolutionary Direct Policy Search, in *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 401–408 (2009)
- [Kohavi 09] Kohavi, R., Longbotham, R., Sommerfield, D., and Henne, R. M.: Controlled experiments on the web: survey and practical guide, *Data Mining and Knowledge Discovery*, Vol. 18, No. 1, pp. 140–181 (2009)
- [Kwon 02] Kwon, O. B., Kim, C.-R., and Lee, E. J.: Impact of website information design factors on consumer ratings of web-based auction sites, *Behaviour & information technology*, Vol. 21, No. 6, pp. 387–402 (2002)
- [White 12] White, J.: *Bandit Algorithms for Website Optimization*, O'Reilly (2012)
- [金澤 02] 金澤 貴彦, 安田 恵一郎: 組合せ最適化問題における Proximate Optimality Principle を考慮した Tabu Search の探索効率改善, 最適化シンポジウム講演論文集, Vol. 2002, No. 5, pp. 219–224 (2002)
- [野々部 08] 野々部 宏司, 柳浦 睦憲: 局所探索法とその拡張 タブー探索法を中心として, 計測と制御, Vol. 47, No. 6, pp. 493–499 (2008)

〔担当委員：大原 剛三〕

2014 年 3 月 15 日 受理

著 者 紹 介

**飯塚 修平**

2012 年 3 月東京大学工学部システム創成学科知能社会システムコース卒業，2014 年 3 月同大学院工学系研究科技術経営戦略学専攻修士課程修了．2014 年 Google 株式会社ウェブデベロッパー．

**松尾 豊(正会員)**

1997 年 東京大学工学部卒業．2002 年同大学院博士課程修了．博士（工学）．同年より，産業技術総合研究所研究員．2005 年 10 月よりスタンフォード大学客員研究員．2007 年 10 月より，東京大学大学院工学系研究科技術経営戦略学専攻 准教授．2007 年より国際 WWW 会議プログラム委員，2013 年より Web Mining 部門のチェア．2012 年より人工知能学会理事・編集委員長．人工知能学会論文賞，情報処理学会長尾真記念特別賞，ドコモモバイルサイエンス賞など受賞．専門は，Web 工学，人工知能．