

高速なウェブページ最適化手法の提案

Fast Webpage Optimization Method

飯塚 修平

Shuhei Iitsuka

松尾 豊

Yutaka Matsuo

東京大学大学院 工学系研究科 技術経営戦略学専攻

Department of Technology Management for Innovation, The University of Tokyo

Webpage optimization is an experimental method to make continuous improvements by finding high-performing variation 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. In this research, we propose a fast webpage optimization method that outperforms naive A/B testing.

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のパターンを複数用意して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ており、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、バラック・オバマ氏が公式ウェブサイト上で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}。この実験は公式ウェブサイトのトップページで 6 種類の写真と 4 種類のボタンからなる 24 種類のパターンをユーザにランダムに表示するというもので、それぞれのパターンを表示した時のユーザの登録率を計測した (図 1 参照)。実験の結果、最も登録率が高かったパターンをサイト全体に採用することで、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、ユーザが少ない小規模なウェブサイトでは、それぞれのパターンを評価するのに十分なデータを集めるのに時間がかかってしまうという問題点がある。実験に時間がかかることはウェブサイトの成長を鈍化させてしまうだけでなく、季節性による外部要因が混入する余地が大きくなるため実験上好ま

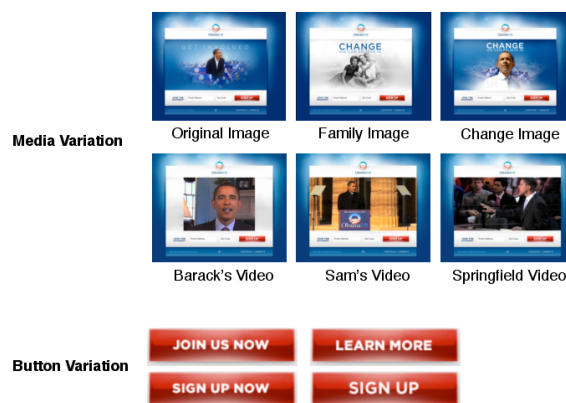


図 1: バラック・オバマ氏公式ウェブサイトのトップページで実験に用いられた写真とボタンのバリエーション

しくない。小規模ウェブサイトでも短期間で実験結果を得ることができるウェブページ最適化手法が求められている。

そこで本研究では、既存のウェブページ最適化手法を整理して組み合わせることで、小規模なウェブサイトでも高速に最適なパターンを発見することができるウェブページ最適化手法を提案する。また、提案手法をウェブページ最適化プログラムとして実装し、実際の壁紙画像検索サイト「Imagerous(以下**実サイト A**と呼ぶ)^{*5}」に導入することで有効性を評価する。

2. 関連研究

効率的なウェブページ最適化を実現するための研究は以前から行われている。評価指標に応じて実験に必要なユーザ数をあらかじめ算出する研究 [Kohavi 07]、より当てはまりの良い確率分布を評価指標に仮定することで、より少ないユーザ数でパターンの比較を可能にする研究 [Rosset 11] などが行われている。実際のウェブサイト上でウェブページ最適化を行なった研究もあり、実験の上で注意すべき点がケーススタディとしてまとめられている [Crook 09, Kohavi 09]。また、従来の最適化問題で用いられている探索手法を応用した研究もあり、遺

連絡先: 飯塚修平, 東京大学工学系研究科技術経営戦略学専攻, 東京都文京区弥生 2-11-16, 03-5841-7672, iitsuka@weblab.t.u-tokyo.ac.jp

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

*2 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。

*3 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/>

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

伝的アルゴリズムをウェブページに表示する要素の最適化に適用した研究がある [Asllani 07].

効率的な最適化を行うための知見は蓄積されているものの、そこから新たに最適化手法を生み出そうとする試みは少ない。本研究ではウェブページ最適化問題にまつわる問題の中から、ユーザ数が少ない小規模ウェブサイトでは実験に時間がかかってしまう問題に着目して、新たなウェブページ最適化手法を構築する。

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る。そして各パターンでのユーザの反応を比較することで、最適なパターンを決定する手法である。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。最適なパターンを決定する際には、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て反応を比較する。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 [Ash 12].

総当り実験はすべてのパターンを対象に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。それぞれのパターンについてユーザの反応を計測した後は、各要素において評価指標を最大にする値を組み合わせたものを最適なパターンとする。

3. ウェブページ最適化の定式化

3.1 問題の定式化

ウェブページ最適化は、デザインや機能の一部が異なるウェブページのパター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^{*6}。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る。それぞれの要素を変数として捉えれば、パターンは変数を組み合わせた解として表現することができる。したがって、ウェブページ最適化は評価指標を最大化する変数値の組み合わせを探索する **組み合わせ最適化問題** の一種である。

ウェブページのパターン $\mathbf{x} = (x_1, \dots, x_m)$ 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 $\mathbf{x}$ 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ことになる。

表 1: 従来のウェブページ最適化手法の整理

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

ウェブサイト開発者は解空間 X の中から評価指標 $f(\mathbf{x})$ を最大化するパターン $\mathbf{x}^* \in X$ を求めたい。しかし評価値は直接観測することが出来ず、パターン $\mathbf{x}$ 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 $p(y|\mathbf{x})$ に従って与えられるのみである。ここでは評価指標 $f(\mathbf{x})$ を観測値 y の条件付き期待値 $\mathbb{E}[y|\mathbf{x}]$ によって推定することにする。

以上より、ウェブページ最適化は式 1 のように定式化することができる。本研究ではこの問題を **ウェブページ最適化問題** と呼ぶ。

ウェブページ最適化問題

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

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

3.2 既存のウェブページ最適化手法の整理

ウェブページ最適化問題に対する解法として従来のウェブページ最適化手法を整理したものを表 1 に示す。A/B テストはオリジナルのウェブページのパターンを暫定解 $\mathbf{x}$ として局所探索を行うことに相当する。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 $\mathbf{x}'$ 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変数値の組み合わせを最適解として推定している点で、下式に示す線形モデルを仮定した最適解の推定を行なっている。

$$y = a_0 + \sum_{i=1}^{l_1-1} a_{1i}x_1^i + \dots + \sum_{i=1}^{l_m-1} a_{mi}x_m^i$$

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。

4. 提案手法

小規模ウェブサイトでは観測値を得るのに時間がかかるため、すべての実行可能解を評価する総当り探索を導入することは難しい。総当り探索と異なり最適性の保証はないが、高速に近似解を与える局所探索が小規模ウェブサイトにおけるウェブページ最適化と相性が良い解法であると考えられる。しかし、局所探索は初期解によって探索のパフォーマンスが大きく変化することが知られている [古川 12]。そこで提案手法では、局所探索の初期解の決定に線形モデルによる最適解の予測を用いることで優れた解から局所探索を開始し、高速に最適解を探索

*6 ここでは最適化問題を最大化問題として考える。

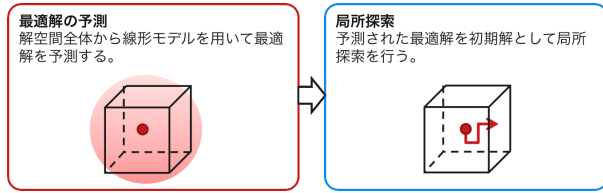


図 2: 提案手法の概要図

することを目的とする。提案手法の概要図を図 2 に、アルゴリズムを Algorithm 1 に示す。

Algorithm 1 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: X as the set of solutions.

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

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

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

$\mathbf{x}^*, n \leftarrow Initialization(X, N_1)$

repeat

$\mathbf{x}^*, n \leftarrow Move(\mathbf{x}^*, X, N_2)$

until exit condition

return $\mathbf{x}^*$ as the optimal solution.

提案手法による最適化に必要なものは、有意水準 α 、検出力 $1 - \beta$ 、効果量 Δ 、解空間 X の 4 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。

線形モデルによる最適解の予測を行うのに要するサンプルサイズ N_1 には、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。線形モデルによる最適解の予測のためには、各変数がとりうるすべての値について観測値を得られればよいので、ここでは分散分析する群の数を $k = \sum_{i=1}^m l_i$ とする。ただし m は変数の数、 l_i は変数 x_i が取りうる値の数である。局所探索の解の評価に要するサンプルサイズ N_2 には、対応のない 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 $\mathbf{x}$ のある変数 x_i について値を入れ替えて近傍解 $\mathbf{x}'$ を生成し、お互いに 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それぞれの解の評価値の期待値 $E[y|\mathbf{x}]$ を比較し、期待値が大きい方の解で暫定解を更新する。終了条件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。観測数 n は累計の観測数を表している。



図 3: 実サイト A のスクリーンショット

5. 実験

小規模ウェブサイトにおいて提案アルゴリズムが有効であることを評価するために、一日の訪問者数が約 300 人の実サイト A を対象に実験を行なった。実サイト A 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閲覧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 (図 3 参照)。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、ユーザをウェブサイト内に留める力が強いウェブページを構築することを目指した。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に導入してウェブページ最適化を行うことで、提案手法の有効性を評価した。また、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ようなロボットとみられる動きをするものもあったため、評価指標には上限値を設定した。

今回は写真一覧ページを構成する要素として表 2 に示す変数と取りうる値を設定し、これらの組み合わせを解空間とした。ここでは各変数の値は 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。たとえば、 $\mathbf{x} = (0, 2, 1, 0)$ 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、無作為に初期解を決定して局所探索を行うベースラインアルゴリズムと、線形モデルによる最適解の推定によって初期解を決定して局所探索を行う提案アルゴリズムを比較した。提案アルゴリズムは N_1 の観測数を費やして初期解を決定する一方、ベースラインアルゴリズムは初期解の決定に観測数を費やさず、解空間から無作為にひとつ解を選択して初期解とする。同期間で二つのアルゴリズムを用いて最適化を行い、終了時点で到達した暫定解の評価値の期待値を比較して、

表 2: 実サイト A を構成する変数とその取りうる値

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

表 3: 各アルゴリズムにおける暫定解と評価値の期待値の推移

手法	観測数	暫定解	期待値	標準誤差
ベースライン	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	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

アルゴリズムの有効性を評価した。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。ここから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 A で行なった実験の結果を表 3 に示す。提案アルゴリズムが到達した最適解の評価値の期待値は、ベースラインアルゴリズムのそれを上回った。 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。ベースラインアルゴリズム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、提案アルゴリズムでは期待値の高い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

提案手法は解空間 X が大きい場合でも適用することができる手法である。解空間のサイズは変数の数 m に対して指数関数オーダー $O(c^m)$ で増えていくのに対し、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 は線形関数オーダー $O(m)$ でゆるやかに増加する。そのため、提案手法は変数の増加による解空間の拡大に強い手法である。

また、提案手法は大域的にみて評価値が高いと予測される解から探索を開始することができる。予測モデルによっては評価関数への当てはまりが悪く、評価値の高い解から探索を行うことが出来ない可能性もあるが、その後局所探索を行うことで最終的には評価値の高い解に到達することができる。加えて、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ることで大域的に最適解を探索する手法も考えられる。

提案手法は、ウェブページ最適化問題と同様の形の問題であればウェブ以外の分野でも適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

あり、得られる観測値が少ない場合には提案手法が有効であると考えられる。

7. まとめ

本研究では、ウェブページ最適化において小規模なウェブサイトでは実験に時間がかかってしまう問題に着目し、高速にウェブページ最適化を行う手法を提案した。まずウェブページ最適化を一種の組み合わせ最適化問題として捉え、ウェブページ最適化問題を定式化した。その上で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に対する解法として整理した。直交配列実験で用いられている線形モデルによって最適解を予測する工夫を局所探索の初期解の決定に用いることで、高速に最適解を探索する手法を提案した。提案手法をウェブページ最適化プログラムとして実装し、実際の小規模ウェブサイトを導入して実験した。その結果、無作為に初期解を選択して局所探索による最適化を行う場合に比べて、提案手法は同期間でより評価値の期待値が高い解に到達できることがわかった。ばらつきのある観測値を元に最適組み合わせを探索する問題であれば、提案手法はウェブ以外の分野にも応用できると考えている。

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)
- [Asllani 07] Asllani, A. and Lari, A.: Using genetic algorithm for dynamic and multiple criteria web-site optimizations, *European journal of operational research*, Vol. 176, No. 3, pp. 1767–1777 (2007)
- [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 ACM (2009)
- [Kohavi 07] Kohavi, R., Henne, R. M., and Sommerfield, D.: Practical guide to controlled experiments on the web: listen to your customers not to the hippo, in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 959–967 ACM (2007)
- [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)
- [Rosset 11] Rosset, S. and Borodovsky, S.: A/B Testing Using the Negative Binomial Distribution in an Internet Search Application (2011)
- [古川 12] 古川 正志, 川上 敬, 渡辺 美知子, 木下 正博, 山本 雅人, 鈴木 育男: メタヒューリスティクスとナチュラコンピューティング, コロナ社 (2012)