

組合せ最適化手法を用いたウェブサイト内部リンク構造の最適化

菅沼孝二[†] 西村直樹[†] 酒井亮平^{††} 観山正道[‡] 大関真之^{‡‡}

[†]株式会社リクルートライフスタイル ^{††}株式会社リクルート [‡]東北大学 ^{‡‡}東京工業大学

1 はじめに

検索エンジン最適化 (Search Engine Optimization, SEO) [1] は, ウェブサイトのブランディングや集客, 収益性の向上に大きく寄与する重要な取り組みである. その取り組みの中でも, 同一ウェブサイト内のページ間に貼られる内部リンクによって形成される内部リンク構造を適切に設計することは重要であると考えられている [1, 2]. しかし内部リンク構造設計において, 一般に以下のような課題がある.

- 無数の内部リンクの候補がある中で, 定量的な評価基準に基づいた内部リンクの選定ができていない
- 内部リンク構造の分布を考慮していないため, 特定のページに内部リンクが偏って貼られる場合がある

そこで本研究では, 内部リンク構造の分布を考慮し, 取得可能なデータから算出した推定値に基づいて内部リンク構造を最適化する方法を提案する. 提案手法の有効性は, 実運用中のウェブサイト導入実験によって検証する.

本論文の構成は次の通りである. 2章では, 適切な内部リンク構造を決定するために必要な考え方を導入する. 3章では, 本研究における問題設定と提案手法について述べる. 4章では, 本研究で行った数値実験について述べる.

2 内部リンク構造最適化

内部リンクが貼られるページをリンク元ページ, 内部リンクの遷移先ページをリンク先ページと呼ぶものとする. 内部リンク構造は, ウェブサイトの運営者によって任意の構造に決定することができる. 適切な内部リンク構造を設計すると, ユーザや検索エンジンの

クローラーが, リンク先ページに素早く辿り着けたり, 関連性が高いリンク先ページを簡単に見つけられるようになるため, ユーザビリティやクローラビリティが向上すると考えられる. そのようなリンク先ページは, 検索エンジンから高く評価されやすく, 掲載順位の改善に繋がる.

上記を背景に, 本研究における内部リンク構造最適化では, ユーザビリティ及びクローラビリティの高い内部リンク構造を設計することを目的とする. 内部リンク構造の決定は, 以下の性質を仮定して行う.

- 質の高いリンク元ページに貼る内部リンクほど, リンク先ページに対する効果が高い
- 関連性が高いページ間の内部リンクほど, リンク先ページに対する効果が高い
- 各リンク元ページに貼ることができるリンク本数は上限がある

3 提案手法

ウェブサイトにおける全てのページのうち, リンク元ページ及びリンク先ページの対象となるページ集合をそれぞれ $S = \{s_1, s_2, \dots, s_M\}$, $D = \{d_1, d_2, \dots, d_N\}$ で表す. ただし M, N はページ集合のサイズを表す.

任意のリンク元ページ s_i 及びリンク先ページ d_j の組合せ $(s_i, d_j) \in S \times D$ に対して, ページ間の内部リンクの有無を表現する決定変数を $x_{i,j} \in \{0, 1\}$ とし, リンク元ページ s_i にリンク先ページ d_j のリンクを貼ることを $x_{i,j} = 1$ で表し, そうでない場合 $x_{i,j} = 0$ で表すものとする. また, リンク元ページ s_i に貼ることができるリンク本数を L_i とする. 提案手法では, 2章で述べた内部リンク構造の決定に関する性質を考慮し, リンク本数 L_i に関する制約を満たす中で, $x_{i,j}$ を求める組合せ最適化問題を解くことを目的とする. 提案手法は以下の通りである.

1. リンク先ページの優先度の推定
2. 内部リンク効果の推定
3. 内部リンク構造決定のための目的関数
4. 組合せ最適化手法による求解

Optimizing the internal link structure of a website using combinatorial optimization methods

Koji SUGNAUMA[†], Naoki NISHIMURA[†], Ryohei SAKAI^{††},

Masamichi MIYAMA[‡] and Masayuki OHZEKI^{‡‡}

[†]Recruit Lifestyle Co., Ltd. ^{††}Recruit Co., Ltd.

100-6640, 1-9-2, Marunouchi, Chiyoda-ku, Tokyo

[‡]Tohoku University

980-8579, 6-3-09, Aoba, Aramaki, Aoba-ku, Sendai

^{‡‡}Tokyo Institute of Technology

152-8550, 2-12-1, Ookayama, Meguro-ku, Tokyo

suganuma_koji@r.recruit.co.jp

3.1 リンク先ページの優先度の推定

SEOにおける収益性の向上幅は、検索数や現状の掲載順位などによりページごとに異なる。また、現在の掲載順位に応じて掲載順位を改善する容易さが変わることも考慮する。以上を踏まえ、内部リンク構造を設計する上でのリンク先ページ d_j の優先度を P_j とし、以下のように定義する。

$$P_j = imp_j \times \Delta p_{ctr}(r_j) \times prof_j \times p_e(r_j)$$

ただし、右辺の各項は表 1 を参照。

3.2 内部リンク効果の推定

2章で述べた通り、どのリンク先ページのリンクをどのリンク元ページに貼るかによって、期待される内部リンクの効果が変わると考えられている。これを踏まえ、リンク先ページ d_j のリンクをリンク元ページ s_i に貼った時に得られる内部リンクの効果を $E_{i,j}$ とし、以下のように定義する。

$$E_{i,j} = val_i \times sim_{i,j}$$

ただし、右辺の各項は表 1 を参照。

3.3 内部リンク構造決定のための目的関数

以下の戦略に基づいて目的関数を設計する。

- 優先度の高いリンク先ページを、期待効果が高いリンク元ページに優先的に貼る
- 優先度の高い特定のリンク先ページに内部リンクが集中しないように、内部リンクを分散して貼る

以上の点を踏まえ、決定変数 $x_{i,j}$ を用いて、以下のように定式化する。なお、表記の簡単化のために $w_{i,j} = P_j \times E_{i,j}$ とおき、掛け算の記号は省略する。

$$\max. \sum_{j=1}^N \sum_{i=1}^M w_{i,j} x_{i,j} - \lambda \sum_{j=1}^N \left(\sum_{i=1}^M x_{i,j} \right)^2 \quad (1)$$

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

$$x_{i,j} \in \{0, 1\} \quad \forall i, j.$$

これは決定変数が $M \times N$ 個、目的関数が 2 次の 0-1 整数計画問題となり、 $M \times N$ が大きい場合には汎用の最適化ソフトウェアでは求解が困難である。

式 (1) の第 1 項が優先度と内部リンク効果に対応する割り当ての役割を担い、第 2 項が内部リンクの偏りを防ぐ罰則項の役割を担う。 λ は罰則項の影響の程度を調整する係数であり、 λ の値が大きくなるほど内部リンクが分散して貼られるようになる。

表 1: 数式中の各項の説明

項	説明
imp_j	検索エンジンにおける特定のキーワードに対するページ d_j の検索数。
$\Delta p_{ctr}(r_j)$	ページ d_j の現在の掲載順位 r_j から順位が一つ上昇した時の増分クリック率。 r_j に対するクリック率は、各掲載順位とそれに対応するクリック率のデータから推定する。
$prof_j$	ページ d_j の収益性。
$p_e(r_j)$	ページ d_j の現在の掲載順位 r_j から順位を一つ上昇させることの容易さ。これは、各掲載順位とそれに対応する内部リンク効果のデータから推定する。
val_i	ページ s_i の価値。アクセス数やその対数値、ユニークユーザ数などを用いる。
$sim_{i,j}$	ページ s_i, d_j 間の類似度。テキストやアクセスログの類似性計算により推定する。

3.4 組合せ最適化手法による求解

前項の目的関数及び制約条件に対して、組合せ最適化手法を利用し決定変数 $x_{i,j}$ を求める。

本研究では、決定変数 $x_{i,j}$ の初期解の決定には貪欲法を、初期解からの最適化には 2-flip 近傍探索、及び 4-flip 近傍探索を利用したヒューリスティックな局所探索法を用いた。また、局所解への収束を回避するために、近傍探索にはランダムサンプリングを導入した。探索の終了条件は、目的関数の値が収束するか、一定時間経過した場合とする。詳細については当日報告する。

4 実験結果

提案手法を利用して、人間ドック予約サイトの「人間ドックのここカラダ」[3] のデータにおける提案手法の求解性能の検証と、実際のウェブサービスでの適用による効果検証を行った。その結果、提案手法導入前と比較して、(i) 1 セッションあたりの予約完了率、(ii) 病院名の平均検索順位、(iii) 検索エンジン経由の想定売上高について、いずれも改善することを確認した。実験結果の詳細については当日報告する。

参考文献

- [1] Google: Search Engine Optimization (SEO) Starter Guide <https://support.google.com/webmasters/answer/7451184?hl=en>
- [2] Sonya, Z. and Neal, C.: Search Engine Optimization: Comparison of Link Building and Social Sharing, JCIS, Vol.57, No.2, pp.148-159 (2017) .
- [3] <https://dock.cocokarada.jp/>