

# ソーシャルブックマーク情報を用いた Web 検索結果の再ランキング手法とその評価

山家 雄介 中村 聡史 アダム ヤトフト 田中 克己

京都大学大学院 情報学研究科 社会情報学専攻

## 1 はじめに

ブログなどの普及により情報発信の裾野が広がるにつれて、Web 検索結果から有用なページを発見するのは困難になる一方である。最近ではユーザーのブックマーク行動を集約することによって価値のあるページを抽出する、ソーシャルブックマークのような取り組みが盛んになりつつある。我々は以前、ソーシャルブックマークにおけるページのブックマーク数などの情報を用いて、Web 検索結果のページの内より有用なものを上位に提示する再ランキング手法を提案した [1]。本稿では、提案手法を多数のクエリに対して適用し、検索結果に含まれるページの順位変動率や、ページの種類などを調査・分類し、どのような検索目的に本手法が有効なのかを明らかにした。

## 2 SBRank と PageRank による統合検索

本章では、両ランキング尺度を組み合わせた統合型 Web 検索手法の有効性を調査する。統合したランキング尺度を求めるにあたっては次の式を利用する。

$$NewRank = \alpha \cdot SBRank + (1 - \alpha) \cdot PageRank \quad (1)$$

ここで、SBRank はソーシャルブックマークサービスにおけるページに対するブックマーク数であり、PageRank はページに対する Google の PageRank 値である。 $\alpha$  は両尺度に対する重み変数であり、新しい順位は NewRank を降順にソートすることで行われる。

### 2.1 実際の検索結果における提案手法の振る舞い

提案手法は、既存の検索エンジンにおける検索結果のうち、多くのページがソーシャルブックマークにおいてブックマークされていればいる程効果を発揮する。

#### Re-ranking Method for Web Search Based on Social Bookmark Analysis

Yusuke Yanbe, Satoshi Nakamura, Adam Jatowt, Katsumi Tanaka

Department of Social Informatics, Graduate School of Informatics, Kyoto University

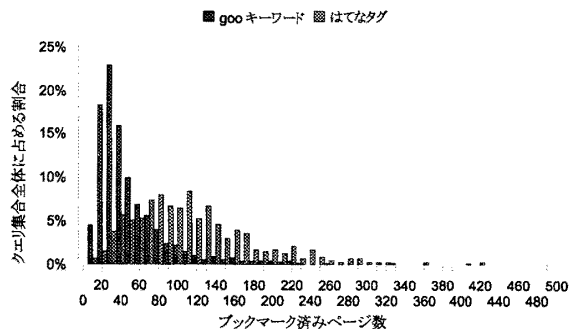


図 1: 両クエリ集合における上位 500 件のブックマーク数の分布

またブックマーク数と検索結果の順位に乖離がある場合、従来とは異なる新たな指標でランキング出来ることになる。これらの特性について、異なる特徴をもつ 2 つのクエリ集合を用いて比較調査を行った。通常のクエリ集合として、Web 検索エンジンの goo が公開している、“goo において 1ヶ月単位に使用頻度の急上昇したキーワードの上位 50 件<sup>1)</sup>”を 2006 年 1 月分から 2007 年 10 月分まで、重複を取り除いた上で計 806 件取得した。またもう一つのクエリ集合として、“はてなブックマークで最近頻りに利用されているタグ<sup>2)</sup>”を 531 件取得した。

次に、2 つのクエリ集合において Yahoo! の検索結果を検索クエリごとに 500 件ずつ取得し、さらに検索結果の各ページについて、はてなブックマークにおけるブックマーク数をおよび PageRank 値を取得した。以降はこの 2 つのデータセットの特性と、それについての考察を述べる。

図 1 は Yahoo! で Web 検索を行った際に、検索結果の上位 500 件のうち、ブックマークされているページ数の分布を示している。横軸が検索結果の上位 500 件のうちブックマークをもつページ数の区間を 10 ページきざみにとったもの、縦軸が各区間に該当する検索クエリの、クエリ集合全体における割合を示したもの

<sup>1)</sup><http://ranking.goo.ne.jp/service/001/>

<sup>2)</sup><http://b.hatena.ne.jp/t>

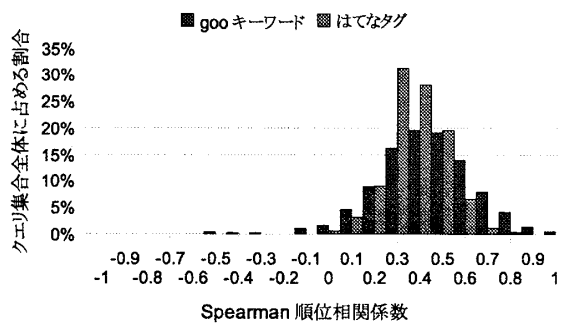


図 2: 提案手法適用後の Spearman 相関係数の分布

である。図 1 によると、goo キーワードでは比較的狭い範囲に収まっているのに対して、はてなタグでは比較的幅広く分布していることがわかる。検索結果に対して被ブックマークページが多い状態、つまり図 1 において分布が右側に偏るほど、提案手法によって検索結果下位のページが上位に再ランクされる可能性が増加することになり好ましい。2つのクエリ集合を比較すると、本手法ははてなタグに基づいたものの方が適していることが見てとれる。

次に、提案手法で  $\alpha=1$  を設定した状態、つまりページのブックマーク数が大きい順に再ランキングを行ったときの、元の検索結果順位からの変化の度合いを、2つの順位データの相関関係を調べるための指標である Spearman 順位相関係数を用いて評価した。Spearman 順位相関係数は、2つの順位データが完全に同一である場合に 1、正反対である場合に -1 をとる。図 2 は各クエリにおいて被ブックマークが 1 つ以上あるページを対象に、提案手法で再ランキングを行った際の Spearman 相関順位係数の分布である。図 2 から、ランキングの順位変動の分布は、両クエリ集合とも 0.4 を中心として同じような傾向があることが分かる。つまりブックマークされているページのみを対象にした場合、元のランキングとは正反対になるような大幅な順位の変動するクエリや、まったく順位変動しないようなクエリはごく一部であり、多くの検索クエリではある程度ランキングが入れ替わるといえる。

## 2.2 Spearman 相関順位係数の値別の特徴

次に、Spearman 順位相関係数が低いクエリと高い検索クエリがそれぞれどのような特徴をもっているのかを調査した。Broder[2] による検索質問の分類に基づく、検索クエリ集合はユーザの意図に応じて調査型クエリと誘導型クエリに分類することができる。ここで両クエリ集合において、Spearman 順位相関係数が最も低かったグループと最も高かったグループ、つま

		Spearman 係数	
		低い	高い
goo	ブログ、検索、株、NEWS、旅行、SEO、ライブ、沖縄、サッカー、日本語、ライブカメラ、mixi、YOUTUBE、ニコニコ動画、Google Earth、民主党、任天堂、Wii、Windows Vista、PSP	寒中見舞い、飯島西、MNMI、日本女子プロゴルフ、ナルト、河合塾、三宅恵美、有馬記念、パロマ、どろろ、ファーストファイバー、柿崎順一、春のワルツ、DADA、FLORA、滝浪愛	
はてな	サイト、api、japan、ブログ、インターネット、ネット、日本、css、ゲーム、検索エンジン、無料、flash、会社、情報、ruby、rails、google、はてなブックマーク、はてな、twitter、gmail、ニコニコ動画	economy、asia、copyright、sex、いい話、branding、おやつ、年金、ベトナムアムド、scalability、brush、amf、ベトナム株、男女、politics、food、military、プラン、pipes、はてなハイク、mtos	

図 3: 両クエリ集合の分類

り図 2 の分布の両端部分に位置する検索クエリをそれぞれ抽出したものが、図 3 である。図 3 において、まず goo のクエリ集合に着目すると、Spearman 順位相関係数が低いグループでは、高いグループに比べて調査型クエリの割合が高くなる。具体的には、ブログ、検索、株、旅行、SEO、沖縄、サッカーなどである。また、Spearman 順位相関係数が高いグループで多く見られる人名やイベント名が含まれないという特徴がある。この傾向は、はてなのクエリ集合も同様である。

以上をまとめると、Spearman 順位相関係数が低い、提案手法が強く作用する検索クエリ集合は、同係数が高いグループに比べ調査型クエリの割合が高く、誘導型クエリの割合が低いという特徴が見出せる。つまり我々のシステムは調査型クエリにおいて効果を発揮できる可能性を秘めているといえる。

## 3 おわりに

本稿では、Web 検索結果に対してソーシャルブックマークのブックマーク数を用いて再ランキングを行った場合、どの程度作用し、またどのようなクエリに対して効果的に作用するのかを確かめた。今後の課題は、提案手法による再ランキングを複数のユーザに実際に使用してもらうことでユーザ評価を行うことである。

## 4 謝辞

本研究は一部、計画研究「情報爆発時代に対応するコンテンツ融合と操作環境融合に関する研究」(課題番号 18049041) および、京都大学グローバル COE プログラム「知識循環社会のための情報学教育研究拠点」によるものです。ここに記して謝意を表するものとします。

## 参考文献

- [1] Y. Yanbe, A. Jatowt, S. Nakamura, K. Tanaka: Towards Improving Web Search by Utilizing Social Bookmarks, Proc. 7th ICWE, LNCS, Vol.4607, pp. 343-357, July 2007.
- [2] A. Broder, A taxonomy of web search. ACM SIGIR Forum, Vol. 36, No. 2, pp. 3