

ソーシャルブックマークを利用した ユーザ嗜好に基づくページの抽出

高橋翼[†] 北川博之[‡]

[†] 筑波大学第三学群情報学類

[‡] 筑波大学大学院システム情報工学研究科

1 はじめに

Web が人々に浸透し、様々な情報源から情報が発信されるようになってきた。それにより、情報の信頼度の重要性が強く認識されるようになってきている。一方、Web 上でユーザがブックマーク情報を作成し、興味や関心を持ったページを管理、分類、共有するサービスであるソーシャルブックマーク (以降,SBM) が注目され、普及し始めている。SBM では、ユーザがブックマークするという行為やそれに付与されたタグをユーザの Web ページに対する興味や嗜好を表す指標とみなすことができる。本稿では、SBM 上での Web ページを Authority、ユーザ Hub とみなし、HITS[1] の概念を利用して Web ページの評価値を算出することにより、ユーザの嗜好、信頼度に基づいたページの評価手法を提案する。

本稿の以降の構成は以下のとおりである。2 章では、過去の関連する研究について概観する。3 章では SBM について記述する。4 章では、我々が提案する Web ページの評価手法について述べる。5 章では、提案手法の有効性を測るための評価実験について述べる。最後に、6 章で本稿の結論を述べると共に、今後の課題について述べる。

2 関連研究

近年、SBM に関する研究は盛んになってきている。本研究と密接な関係にある研究として、山家らによる研究があげられる。Yanbe ら [3] は SBM 上でページをブックマークしたユーザの数を SBRank という Web 検索の際の尺度として用い、PageRank[2] との比較実験を行っている。また SBRank と PageRank を統合したランキング手法について提案し、SBRank 値の Web 検索の尺度としての有用性を示している。本研究は、ブックマーク件数を利用した尺度という点は同じだが、ユー

ザの信頼度を測ることを通してページの評価をしているという点と、検索エンジンからは得られないページにまで対象ページを拡大しているという点で Yanbe らの研究とは異なる。

3 ソーシャルブックマーク (SBM)

SBM は近年注目を集めている Web2.0 の概念を持つサービスの一つである。SBM は Web 上でユーザがブックマーク情報を作成し、興味や関心を持ったページを管理、分類、共有するサービスである。ユーザは 1 つのブックマークに 0 個以上の複数のタグの集合を注釈として与えることができる。タグは、ユーザの独自の価値観によって作られた単語やフレーズである。ユーザからページに与えられたタグの集合をブックマークタグ集合 (B タグ集合) とする。様々な価値観を持ったユーザによって評価を与えられ、フィルタリングされた一定の信頼性を有する情報源であるといえる。ある Web ページには複数のユーザがブックマークをしており、何人のユーザがブックマークをしているかという情報は Web ページの信頼度や品質を測る一つの指標と言える。

4 提案手法

HITS の考え方を応用し、SBM から得られる情報を使用した有益なページの抽出手法 S-BITS (Social-Bookmarking Induced Topic Search) について提案する。本研究では、SBM から得られる情報について以下の仮定をする。

- ブックマークするという振舞はページに対して正の評価を与えるということである
- 多くの良きユーザからブックマークされているページは良きページである
- 多くの良きページをブックマークしているユーザは良きユーザである。

上記の仮定は HITS の概念に類似するものであり、HITS の概念を SBM のページとユーザとの関係に拡張したものである。

Extracting Authoritative Pages Based on User Interests Using Social Bookmarks

Tsubasa Takahashi[†] and Hiroyuki Kitagawa[‡]

[†] College of Information Sciences, University of Tsukuba

[‡] Graduate School of Systems and Information Engineering, University of Tsukuba

[†] tsubasa@kde.cs.tsukuba.ac.jp, [‡] kitagawa@cs.tsukuba.ac.jp

4.1 データの収集

まず、評価の対象となる Web ページとユーザの収集を行う。検索語と関連のあるページを収集するために、キーワード集合を検索クエリとして与え、検索エンジンから上位 n 件の Web ページを収集する。この収集したページ集合をルート集合と呼ぶ。ルート集合の Web ページは、従来のリンク構造や tf-idf などの評価手法によって検索語との関連が高いページとして抽出されたものであり、ある程度の信頼度を持っている。ルート集合の各ページにブックマークしているユーザを集め、ユーザ集合を生成する。また、すべてのユーザがルート集合のページに与えている B タグ集合を集め、ルートタギング集合 T を生成する。

4.2 対象ページの拡張

対象ページの拡張は、検索クエリと類似した内容を持つページに対してのみ行う。ユーザはタグの集合をページに与え、管理、分類しており、タグ集合がページの識別に用いられている。そこで、 T の頻出アイテム集合を抽出し、検索語と関連があり、ルート集合を代表するタグ集合の部分集合 T' を得る。本研究では、 T' を T の極大頻出アイテム集合として抽出する。ここでのアイテムとは、1つのタグである。また、 T の要素であるタグ集合をトランザクションとして扱う。極大頻出アイテム集合は、全トランザクション中にある一定以上の割合で含まれているアイテム集合の中で、それ以上の超集合がないアイテム集合である。 T' は最大公約数的に抽出された検索クエリとの関連の高いタグ集合を表すと考えられる。

ユーザによってブックマークされたページ p に B タグ集合 t_p が与えられているとき、タグ集合 $t \in T'$ が $t \subset t_p$ であるページ p を SBM から収集し、ルート集合と合わせてベース集合とする。

4.3 評価値の計算

HITS の概念を拡張し、SBM におけるページを Authority、ユーザを Hub とみなすことで、ページ-ユーザ間の関係を相互に強化することを通して、ページの評価値を算出する(式 1)。ページ、ユーザの両スコアが平衡を迎えるまで繰り返し計算を行い、よりスコアの高いユーザからの評価がページの評価値へと反映される。これによって、単純な被ブックマーク数という指標だけでなく、ユーザの嗜好・信頼度を基にしたページの評価が可能になる。

$$p_i = \sum_{(j,i) \in E} u_j \quad u_i = \sum_{(i,j) \in E} p_j \quad (1)$$

5 評価実験

提案した手法の有用性を測るために評価実験を行った。Yahoo!, SBRank 値、ルート集合、ベース集合を対

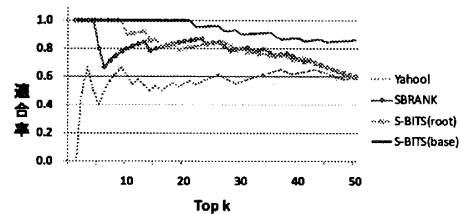


図 1: Top k 件の適合率曲線

象とした S-BITS の 4 つのランキングの妥当性を比較する。以下のような環境において実装し、実験を行った。

- 検索エンジン API: Yahoo Search API[5]
- 対象 SBM: はてなブックマーク [4]

SBRank, S-BITS は Yahoo! Search API から上位 200 件を取得し、そこから SBM に登録されているページを抽出した。それぞれの評価手法でスコアを算出し、それぞれのランキングを作成した。

各手法の評価は、検索クエリに対するページの妥当性を被験者に手作業による判別することで行った。手作業による判定には作業量的限界があるため、上位 50 件のみを対象とした。評価結果は適合率曲線(図 1)で示す。提案手法は、クエリ“open social”において、高い適合率を示している。

6 まとめ

SBM におけるページとユーザの関係に HITS の Authority, Hub の概念を取り入れたページの評価・抽出手法 S-BITS を提案した。今後は精度の向上や、より汎用性のある評価手法とするために、様々な要素を考慮した手法について検討していきたい。

謝辞

本研究の一部は科学研究費補助金特定領域研究(#19024006)による。

参考文献

- [1] J. Kleinberg. Authoritative Sources in a Hyperlinked Environment. In Proc. of the 9th ACM SIAM Symposium on Discrete Algorithms (SODA'98).
- [2] L. Page, S. Brin, R. Motwani and T. Winograd. The pagerank citation ranking: Bringing order to the Web. Technical report, Stanford Digital Library, 1998.
- [3] Yanbe, Y., Jatowt, A., Nakamura, S. and Tanaka, K. Can social bookmarking enhance search in the web?. ACM IEEE JCDL, 2007.
- [4] はてなブックマーク . <http://b.hatena.ne.jp/>
- [5] Yahoo! Search Web Services. <http://developer.yahoo.co.jp/search/web/V1/webSearch.html>