

ソーシャルブックマークにおける 有用なユーザの発見

田代 祐一^{†1} 大石 哲也^{†2} 越村 三幸^{†3}
藤田 博^{†3} 長谷川 隆三^{†3}

Web上の情報が増大するにつれ、ユーザが望む情報を推薦したり、フィルタリングしてくれるような技術の必要性が高まっている。近年、普及しているソーシャルメディアのひとつにソーシャルブックマークがあるが、このソーシャルブックマークにおけるあるユーザのブックマーク情報はユーザの嗜好に基づいてフィルタリングされた情報とみなすことができる。そこで、本研究では特定のトピックについて有用な情報を多くブックマークしているユーザを発見し、推薦する手法を提案する。

Discovery of the useful users in Social Bookmark Service

YUICHI TASHIRO,^{†1} TETSUYA OISHI,^{†2}
MIYUKI KOSHIMURA,^{†3} HIROSHI FUJITA^{†3} and —
RYUZO HASEGAWA^{†3}

As the information on the web increases, recommending or filtering information is needed. Recently, a social bookmark, which is one of social media, is used. The bookmark information of a user on the social bookmark is considered to be the filtered information based on the user's interest. In this paper, we propose a method to discover and recommend users who bookmarked that many useful information of the specific topic.

1. はじめに

ブログやソーシャルネットワークサービスのようなサービスの普及により、Web上で個人がWebコンテンツを生成できる機会が増え、それに伴ってそのようなコンテンツがWeb上で増大している。これらのコンテンツは個人の嗜好や興味を含んだ情報が多い。そういったコンテンツの一つに、ソーシャルブックマークがある。ソーシャルブックマークは元来個人のコンピュータで管理されていたブックマーク情報をWeb上で管理、共有するといったようなサービスである。ブックマークされたページはユーザによって何らかの評価を受けたページであり、また、ユーザがブックマークしているページ群はユーザの嗜好を反映していると考えられる。このため、ソーシャルブックマークのデータを利用し、情報抽出を行った、Web検索結果の改善を行うといった研究が行われている。

ソーシャルブックマークにおけるユーザはブックマークされている情報を評価している一種のフィルタとみなすことができる。あるトピックに関して興味関心をもつユーザがブックマークしたページは、そのトピックに関連している可能性が高く、また、ユーザがそのトピックに関して信頼性の高いユーザである場合、そのユーザによってブックマークされているページも、そのトピックに関して信頼性の高い情報を有していると考えられる。

信頼性の高いページは信頼性の高いユーザに多くブックマークされており、また、信頼性の高いユーザは信頼性の高いページを多くブックマークしていると仮定すると、この関係はHITS¹⁾のHubとAuthorityの関係に類似している。

本研究では、その考えに基づき、あるトピックに関して有用なユーザを提示する手法を提案する。あるトピックに関して有用なユーザは、そのトピックに関するフィルタと考えることが出来、そのユーザがブックマークする情報を得ることにより、トピックに関して有用な情報を得ることが出来る。

以下では、まずソーシャルブックマークにおけるユーザとページの概念を述べ、関連研究について述べた後、今回提案するユーザの推薦手法、及びその実験結果に付いて述べる。最後に、まとめと今後の課題に付いて述べる。

^{†1} 九州大学大学院システム情報科学府
Graduate School of Information Science and Electrical Engineering, Kyushu University
^{†2} 九州大学 情報基盤開発センター
Research Institute for Information Technology, Kyushu University
^{†3} 九州大学大学院システム情報科学研究所
Faculty of Information Science and Electrical Engineering, Kyushu University

2. ソーシャルブックマーク

ソーシャルブックマークは近年普及しているソーシャルメディアの一つであり、代表的な商用サービスとしては delicious³⁾、はてなブックマーク⁴⁾ などがある。これらのソーシャルブックマークは、ユーザが Web 上でブックマークを行う際、主としてタグやコメントを付与することで管理を行う。このタグ付けは Folksonomy と呼ばれる概念に基づいており、ユーザが自由な概念のもとでキーワードによる管理・分類・共有を行うことが出来る。

ソーシャルブックマークにおけるユーザ集合を U 、ページ集合を P 、タグ集合を T とするとき、ユーザ $u_i \in U$ がページ $p_j \in P$ をタグ集合 $UT_{ij} = \{t_1^{ij}, \dots, t_n^{ij}\}$ を用いて時刻 d にブックマークしていることを $Bookmark(u_i, p_j, UT_{ij}, d)$ で表すこととする。 $u_k^{ij} \in UT_{ij}$ はユーザ u_i によって自由に付与されたタグであり、単語やフレーズで構成される。また、 UT_{ij} は 0 個以上の要素を持つ。また、ページ $p_i \in P$ に付与されているタグ集合を PT_i とする。

3. 関連研究

ソーシャルブックマークに関して、これまでに多くの研究が行われている。

Golder ら⁵⁾ はソーシャルブックマークのユーザやタグ、ブックマークの性質について分析を行ない、Web ページに付与されるタグの出現頻度が一定値に収束することを示している。 Paul ら⁶⁾ はソーシャルブックマークのデータが Web 検索において大きな変革を起こせるかを検討しており、タグの揺らぎの問題や、Web 空間全体に対するソーシャルブックマーク上のデータ量の少なさの問題から現時点では難しいが、今後データ量が増えることがあればその可能性があるとの結論に至っている。

Yanbe ら⁷⁾ はソーシャルブックマークにおけるブックマーク数を指標とした SBRank を提案し、PageRank との統合を行うことで Web 検索結果の精度向上を行っている。 Takahashi ら⁸⁾ はソーシャルブックマークのネットワーク構造に注目し、ユーザと Web ページ間に HITS アルゴリズムを適用した S-BITS を提案し、Web ページのランキングを行っている。

白土ら⁹⁾ はソーシャルブックマークにおけるブックマーク情報から興味の近いユーザを推薦し、コミュニティ構造を表示するシステムを開発している。 大力はソーシャルブックマークにおけるイノベータ、すなわち率先して情報をブックマークしていくようなユーザに着目して情報推薦を行っている。 本研究はソーシャルブックマークにおける有用なユーザを推薦するという点でこれらの研究に近いが、これらの研究は推薦を受けるユーザがソーシャ

ルブックマークを利用しているユーザに限られているのに対し、本研究ではキーワードベースの推薦であり、ソーシャルブックマークを利用していないようなユーザでも推薦可能といった点で異なる。

4. 提案手法

本研究における提案手法について述べる。

4.1 概要

本研究では、ソーシャルブックマークにおける有用なユーザを発見し、推薦することを目的としている。ここで、ソーシャルブックマークにおける有用なユーザとは、あるトピックに関して有用であるような Web ページをより多く、早くブックマークしているようなユーザとする。提案手法では、次の手順で有用なユーザを発見する。

- (1) ユーザはトピックを表すキーワード K を与える。この K はソーシャルブックマーク上に存在するタグ T に等しいものとする。
- (2) タグ T が付与されているページのうち、より最近ブックマークされていてかつブックマーク数が m 以上のものをソーシャルブックマークのデータから上位 l 件取得し、これをページ集合 P とする。
- (3) ページ集合 P をブックマークしているユーザ集合 U を取得し、 P と U の各要素から構成されるグラフ間において HITS に基づいたランキングを行い、上位ユーザを有用なユーザとする。

以下では各ステップの詳細な方法について述べる。

4.2 トピックに関するキーワード

ユーザは得たいと思うトピックに関するキーワード K を与える。このとき、 K は対象とするソーシャルブックマークに存在するタグ T と等しいものとする。タグ T はソーシャルブックマークにおいて T に関するページに対して付与されているため、トピックに関連するページとみなすことが出来る。

4.3 ブックマーク情報の取得

与えられた T を用い、ソーシャルブックマークからページ集合 P の収集を行う。収集の対象となるページは、ページに付与されたタグに T を含み、ブックマークしているユーザ数が m 人以上で、収集時点でブックマークが最初に登録された日付 d が最近のものから上位 l 件とする。 T がタグとして用いられていることからキーワード K に関連するページである。また、ブックマークしているユーザが一定数いることからある程度の評価がなされて

おり、また、日付を最近のものとするこゝで、よりアクティブなユーザを発見することにつながる。

4.4 ユーザの取得とランキング

得られた P をブックマークしているユーザ集合 U を得る。 P と U はブックマークと被ブックマークというリンク構造を有している。この関係から、ユーザをスコアリングしてランキングを行う。ユーザのランキングは HITS アルゴリズムの考え方に基づいて行っていく。

HITS とは、PageRank²⁾ と並んでよく知られるページのリンク構造解析の手法の一つである。HITS では Hub と Authority といった考えに基づいてページを評価する。Hub とは多くの out-link(参照)をもつページのことであり、多くの有用なページを参照しているページ、といった概念に基づいている。一方 Authority は多くの in-link(非参照)をもつページであり、特定のトピックにおいて多くの人に信頼されており、多く参照されているようなページ、といった概念に基づいている。また、良き Hub は多くの良き Authority を参照しており、良き Authority は多くの良き Hub から参照されているといった相互関係を持っている。

これをソーシャルブックマークにおけるブックマーク関係に当てはめると、Hub はユーザであり、Authority はブックマークされている Web ページと考えることが出来る。すなわち、良きユーザは良きページを多くブックマークしており、良きページは良きユーザから多くブックマークされている、といった関係性が仮定できる。これにより、HITS における Hub スコア、Authority スコアを求めるアルゴリズムを適応することで、ユーザ及びページのスコアが算出できる。

具体的なアルゴリズムについては図 1 に示す。ユーザ集合 U 、ページ集合 P の各ユーザ、各ページに関してスコアを計算していく。それぞれのスコアは繰り返しのたびに収束していき、収束条件の閾値 ϵ_p 、 ϵ_u (十分の小さい値) 以下になるまで計算を繰り返す。 α 回目の繰り返しは $\alpha - 1$ 回目のスコアの影響を受けるため、繰り返しのたびにページの信頼度が重み付けされていき、ユーザの評価値が決定されていく。

5. 実 験

提案手法の有効性を評価するための実験を行う。今回、対象とするソーシャルブックマークははてなブックマークとした。また、実験に用いるトピックははてなブックマークのタ

Algorithm

$$p^0 = \{1, 1, 1, \dots, 1\};$$

$$u^0 = \{1, 1, 1, \dots, 1\};$$

$$\alpha = 1;$$

Repeat

foreach $p_i \in P$

$$p_i^\alpha = \sum_{(j,i) \in E} u_j^{\alpha-1}$$

foreach $u_i \in U$

$$u_i^\alpha = \sum_{(i,j) \in E} p_j^{\alpha-1}$$

// normalization

$$p^\alpha = p^\alpha / \|p^\alpha\|_1$$

$$u^\alpha = u^\alpha / \|u^\alpha\|_1$$

until $\|p^\alpha - p^{\alpha-1}\|_1 < \epsilon_p$ and $\|u^\alpha - u^{\alpha-1}\|_1 < \epsilon_u$

return p^α and u^α

end

図 1 ユーザ評価値算出のアルゴリズム

Fig. 1 The Algorithm about computing score of users on social bookmark

グー覧^{*1}から 10 個のタグ T_{test} (akb48, food, game, iPhone, music, ruby, sns, Web サービス, カメラ, 就職) を選択した。これらのタグがつけられており、2010 年 9 月 15 日以前にブックマークされたページ P を取得し、評価値を算出した。尚、取得するページ数は $l = 200$ とし、ブックマーク件数は $m = 3$ 以上のものを収集した。

評価のため、2010 年 9 月 16 日から 2010 年 10 月 16 日の期間にタグ $t \in T_{test}$ を用いたユーザ U_{test} をランダムに 100 人取得する。タグ t を用いているユーザは t に興味があるユーザと仮定できる。提案手法によってランキングされたユーザが同期間にブックマークしているページのうち、上位から積集合を取っていった結果と、 U_{test} がブックマークしているページを比較し、適合率を見ることで提案手法の評価を行う。

比較のため、ページ $p \in P$ をブックマークしているユーザを P 内でブックマークしている数が多い順にランキングし、同様に評価を行う。

*1 <http://b.hatena.ne.jp/t>

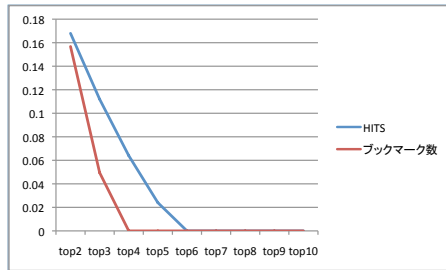


図 2 タグ [AKB48] に関する適合率
Fig. 2 The Precision about Tag[AKB48]

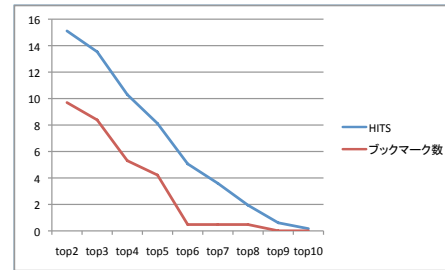


図 3 タグ [food] に関する適合率
Fig. 3 The Precision about Tag[food]

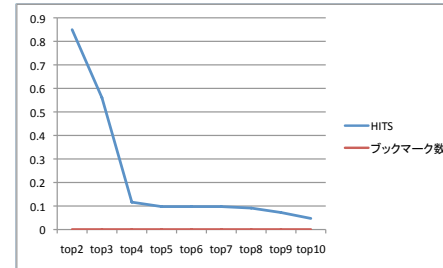


図 6 タグ [music] に関する適合率
Fig. 6 The Precision about Tag[music]

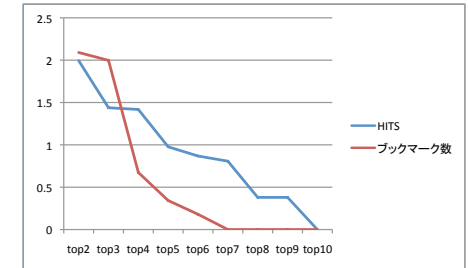


図 7 タグ [ruby] に関する適合率
Fig. 7 The Precision about Tag[ruby]

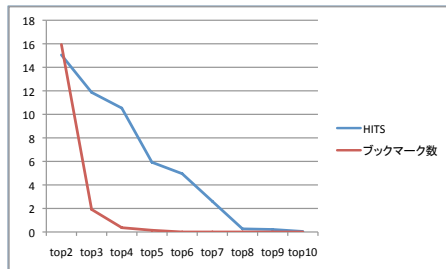


図 4 タグ [game] に関する適合率
Fig. 4 The Precision about Tag[game]

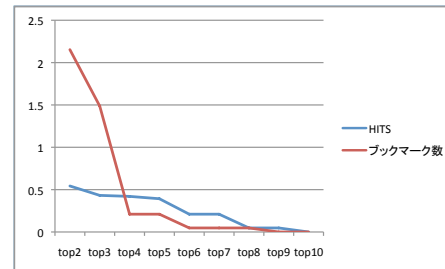


図 5 タグ [iPhone] に関する適合率
Fig. 5 The Precision about Tag[iPhone]

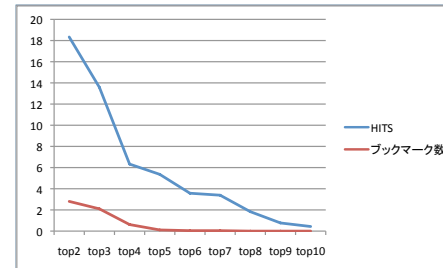


図 8 タグ [sns] に関する適合率
Fig. 8 The Precision about Tag[sns]

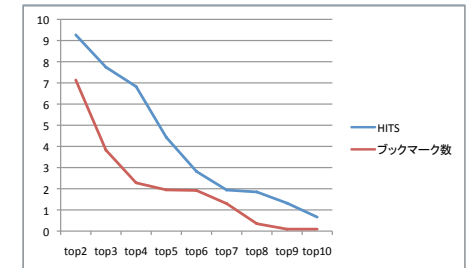


図 9 タグ [Web サービス] に関する適合率
Fig. 9 The Precision about Tag[Web サービス]

図 2 から図 11 にそれぞれのキーワードに関する結果を示す。また、表 1 にタグ'sns'、表 2 にタグ'game'に関して、それぞれ提案手法とブックマーク数によるランキング上位 10 名のユーザを示す。

それぞれの図は、'HITS'を用いた今回の提案手法、ブックマーク数によるランキングによって得た結果を比較したものであり、縦軸が適合率(%), 横軸が上位 n 人のブックマークしたページの積集合を取った時の結果を示す。 n が増えるに連れて、積集合のページ、すなわち共通してブックマークされているページが減少してため、適合率も減少していく。それぞれの結果に関して、適合率は高い数値を示しているとは言えない。これは、比較的網羅的にブックマークを行うようなユーザが上位にきており、求めるトピックとは関係ないようなページも多数ブックマークしていることが多かったためと考えられる。表 1 と表 2 にみられるように、あるトピックで上位に出てきたユーザが、別のトピックでもまた上位にラ

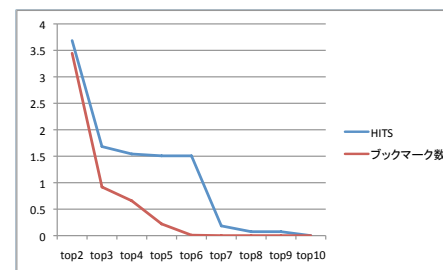


図 10 タグ [カメラ] に関する適合率
Fig. 10 The Precision about Tag[カメラ]

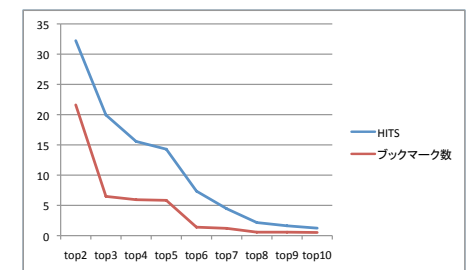


図 11 タグ [就職] に関する適合率
Fig. 11 The Precision about Tag[就職]

表 1 タグ [sns] に関するランキング上位ユーザ
 Table 1 The high-ranking users about Tag[sns]

	HITS	ブックマーク数
1	minazuki6	minazuki6
2	A410	hide7096
3	harax	giffie11
4	raitu	andsoatlast
5	repunit	mouri_yuji
6	beautymaquia	iriso
7	bojoys	mure
8	nkoz	raitu
9	came8244	advblog
10	w03www	Naruhodius

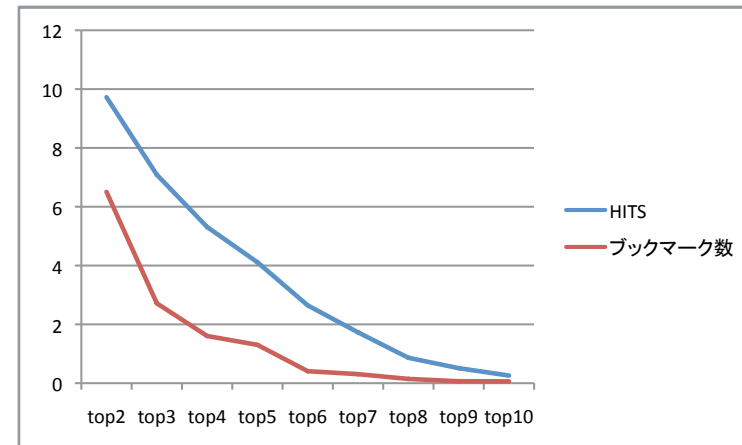


図 12 全タグの適合率の平均
 Fig. 12 The Average of Precision about All Tags

表 2 タグ [game] に関するランキング上位ユーザ
 Table 2 The high-ranking users about Tag[game]

	HITS	ブックマーク数
1	minazuki6	minazuki6
2	A410	repunit
3	repunit	makamaka_at_donzoko
4	beautymaquia	orbis
5	harax	otsune
6	zenibuta	zenibuta
7	Nao_u	Naruhodius
8	came8244	snow113
9	gyu-tang	silverscythe
10	makamaka_at_donzoko	Nachbar

ンクインするということが多々見られた。また、すべての結果において、ブックマーク数順にランキングした結果と比べ、 $n > 4$ のとき HITS を用いた提案手法による適合率の方が上回っている。 $n = 2, 3$ のときもタグ 'iPhone', 'ruby' の場合以外は全て上回っている。表 1 や、表 2 にみられるように、2 つの手法の両方にランクインしているようなユーザがいるが、提案手法の結果の方が上回っていることから、ランキングが適切に行われているということが考えられる。また、図 12 に全タグの適合率の平均を示す。

図 12 の結果は全ての n において提案手法の方が良い結果を示している。これは、トピック、すなわちあるタグ t に興味をもつようなユーザが、良いと評価する、すなわちブックマークするようなページをより多くブックマークしているようなユーザが上位にきていと考えられる。

6. ま と め

ソーシャルブックマークにおいて、ユーザがページをブックマークするという振る舞いは、ページに正の評価を与えるものと仮定することが出来る。そこで、ソーシャルブックマークにおけるユーザとページ間の関係に対して、HITS の考え方を適用し、あるトピックに関して有用である情報を持つと思われるようなユーザを提示する手法の提案を行った。本手法

を用いた結果, あるトピックに関してより有用であるようなページを多くブックマークする
ようなユーザを上位に提示することができた. しかし, 適合率についてはあまり良い数値
を示すことはできなかった. これは, 多くのトピックに関して多くのブックマークを行うよ
うなユーザが上位に来てしまったことが理由と考えられる. このため, 今後の課題として,
トピックに関するページに限ってなるべく多くのページをブックマークしているようなユー
ザを取り出す手法の提案を行っていきたい.

謝 辞

本研究は科研費 (21500102) の助成を受けたものである.

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), pp.668-677, 1998.
- 2) L.Page, S.Brin, R.Motwani and T.Winograd. The pagerank citation ranking: Bringing order to the Web. Technical report, Stanford Digital Library Technologies Project, 1998.
- 3) del.icio.us. <http://del.icio.us/>
- 4) はてなブックマーク. <http://b.hatena.ne.jp/>
- 5) S.A.Golder and B.A. Huberman, The structure of collaborative tagging system. In Information Dynamics Laboratory, HP Labs, 2008.
- 6) Paul Heymann, Georgia Koutrika, and Hector Garcia-Molina. Can Social Bookmarking improve web search? In WSDM' 08, 2008.
- 7) Yusuke Yanbe, Adam Jatowt, Satoshi Nakamura, and Katsumi Tanaka. Can Social bookmarking enhance search in the web? In Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2007), pp.107-116, 2007.
- 8) Tsubasa Takahashi and Hiroyuki Kitagawa, S-BITS: Social Bookmaking Induced Topic Search. In Proc. of the 9th International Conference on Web-Age Information Management (WAIM2008), pp.25-30, 2008.
- 9) 白土慧, 吉井真一郎, 古川正志. ソーシャルブックマークを利用した情報レコメンデーション. 情報処理学会研究報告. ICS, 知能と複雑系, pp.15-20, 2006.
- 10) 大力慶祐, 大向一輝, 武田英明, ソーシャルブックマークにおけるイノベータに注目した情報推薦手法の提案. 人工知能学会第 22 回全国大会 (JSAI2008), 2008.