

ソーシャルブックマークにおける タグ付けを利用した Web ページ評価手法の検討

根本 幸人[†], 後藤 慎弥[†], 金井 敦[†]

[†]法政大学大学院 工学研究科

あらまし

近年, 多様に変化し, 膨大に増加し続ける Web ページの中で, 価値ある情報が埋没する問題が出てきている. その中で, Web ページ間の link 構造を利用した従来の Web ページ評価手法だけでは必ずしも欲しい情報が得られなくなっている. そこで本稿では, Web ページ評価のために, Web ページ間の link 構造だけでなく, ユーザ, Web ページ及びそれらの関連情報を考慮したモデルを提案する. また, 本モデルを用いて, 実際のソーシャルブックマーク上で Web ページに対して付加されているタグを利用してユーザおよび Web ページの評価値を算出し, 本モデルの性質を明確化する.

A Web Page Evaluation Methodology Using Tags in Social Bookmark

Yukihito Nemoto[†], Shinya Goto[†], Atsushi Kanai[†]

[†] Graduate School of Engineering, Hosei University

Abstract

Useful Web page is buried under huge number of Web pages which keep increasing explosively in recent years. Evaluation methodologies of Web pages, which are used at present based on link structure between Web pages, unfortunately do not work very well. Therefore, social bookmark services where bookmarks are shared and grouped using tags become popular. In this paper, a Web page evaluation model is proposed. The model consists of users of Web and related information as well as link information. Then, evaluation values of Web pages based on the model are calculated using actual Web pages and characteristics of this model are clarified and discussed.

1. はじめに

近年の情報爆発に伴い, インターネット上の膨大な Web ページの中から有益な情報を選択することが困難になってきている. 様々な情報を選択する有効な手段として, Google や Yahoo! などの Web 検索エンジンがあり, 多くのユーザが Web 検索エンジンを利用している. Google では, Web ページの評価手法として Web ページ間同士の link 構造を利用した Page Rank[6]を使用しており, 現在でも Google が提供する Web 検索エンジンの重要な位置を示している. 一方, Blog, Wiki や SNS のような CGM の普及が急速に進んでいる. そのようなサイトでは, Web ページ作成者とユーザの区別はなくユ

ーザが Web ページのコンテンツを作成する. また, ユーザ同士でのコメントやトラックバックを簡単に書き込めるため, Web ページ間の link 構造の中には Web ページの作成者の意図とは異なる状態になる場合がある. さらに, スパムトラックバックやコメントスパムなど, Web ページ間の link 構造を巧みに利用したスパム行為が増加しており, 現状のような Web ページの評価手法では, ユーザの意志に反したランキングになってしまうこともある.

そのような背景から, ソーシャルブックマーク[5]を利用した新しい Web ページの評価手法が提案されている. ソーシャルブックマークサイトでランキングとして利用されている被ブックマーク数を SBRank 値とし

て評価し, SBRank と Page Rank を統合する試み[1]やソーシャルブックマークサイトにおいて, ユーザがブックマークするという行為はその Web ページに対して正の評価を与えると仮定し, Web ページを Authority, ユーザを Hub と見なして, HITS の概念を適応した S-BITS という Web ページの評価手法[2]などがその例である。また, ソーシャルブックマークと Folksonomy[5]を利用し, インターネット全体を対象とした Web ページ推薦システム[3]やユーザが付けているタグ情報を用いて, 似た嗜好を推測し関連 Web ページを推薦する方式[4]など, ソーシャルブックマークにおけるタグを利用した研究も盛んに行われている。

本稿では, ソーシャルブックマークを利用する点は共通だが, Web ページに付けられたタグを用いてユーザおよび Web ページを評価するという点で上記の研究と異なる。具体的には, ユーザの意思を反映させるため, Web ページ間の link 構造だけでなくユーザ, Web ページとそれらの関連情報の三者間構造に拡張した Social Knowledge Network モデルを提案する。また, 三者間構造を用いた評価はまだ少なく発展途上と言える。そこで, 本モデルを用いた Web ページの評価手法である SKRank モデルを使い, 実際のソーシャルブックマーク上で Web ページに対して付加されているタグを利用して, ユーザおよび Web ページの評価値を算出し, 本モデルの性質を明確化する。

2. Social Knowledge Network モデル

Web ページおよびそれを参照する人 (ユーザ) を評価するためのモデルとして, Web ページ, ユーザおよびそれらの関連情報の三者に着目したモデルを提案し, 以後, Social Knowledge Network モデル (以下, SKN モデル) と呼ぶ。図 1 に本モデルにおける情報提供モデルを示す。ユーザは Web ページに対して, 参照, ブックマーク, タグ付けなど様々な行為を行うと同時に Web ページの情報を提供する。さらに, ユーザはオークションサイトのユーザ相互評価のようにユーザに関する情報も提供する。また, Web ページはハイパーリンクやコメント表示などにより Web ページの情報を提供し, Web ページの参照や掲示板記入などのユーザの行為によりユーザの情報を提供する。これらの関係を表 1 に示す。

Web ページが Web ページの情報を提供する情報の例としては, Web ページ同士の link 構造があげられる。こ

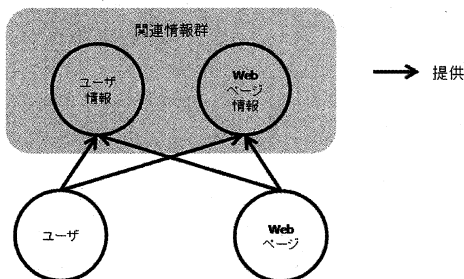


図 1 SKN モデルの情報提供モデル

表 1 提供される 4 種類の情報

ユーザ	ユーザによるユーザの情報
	ユーザによるWebページの情報
Webページ	Webページによるユーザの情報
	WebページによるWebページの情報

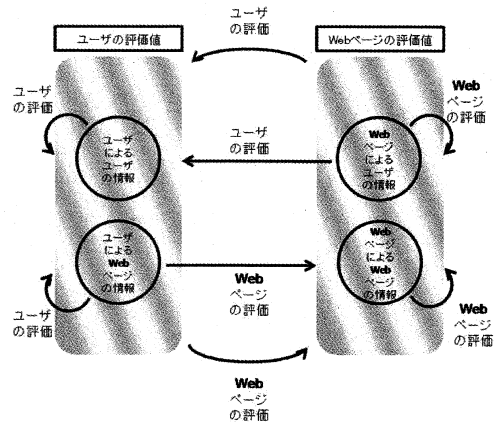


図 2 SKN モデルの評価モデル

れらの情報を利用した Web ページ評価としては Google の Page Rank など従来から様々な手法が提案されている。ユーザがユーザの情報を提供する例としては, ネットオークションにおける取引相手評価が典型的であり, オークションサイトにおいてユーザ評価手法として定着している。SKN モデルを用いた評価は, ユーザの評価と Web ページの評価がある。これらの評価は, 表 1 に示した情報を元に行う。図 2 に表 1 の情報を用いて, ユーザと Web ページが評価し合う評価の構造を示す。網掛けの部分が SKN モデルにおけるユーザ及び Web ページの評価値であり, ユーザと Web ページから提供された情報と各々の評価値を元に, ユーザと Web ページを相互に評価し合う評価構造となっている。次章では, この評価モデルを用いて, Web ページを評価する手法について述べる。

3. SKRank (Social Knowledge Rank) モデル

3.1. 概要

SKRank とは, SKN モデルに基づいて Web ページの友人度合を定量的に評価する計算モデルである。ユーザ及び Web ページに対して評価値をもち, SKRank モデルにおけるユーザ i の評価値を U_i , Web ページ j の評価値を P_j とする。今回は, P_j が SKRank 値となる。

ユーザ i によるユーザ i' の情報を $K_{ii'}$, ユーザ i による Web ページ j の情報を K_{ij} , Web ページ j によるユーザ i の情報を W_{ji} と Web ページ j による Web ページ j' の情報を $W_{jj'}$ とする。この 4 種類の情報と各評価値を使い, SKRank の評価モデルを表現すると図 3 のようになり, (1) (2) 式のように定義できる。

$$U_i = U_i\{i, j, K_{ij}', K_{ij}, W_{ji}, P_j\} \quad (1)$$

$$P_j = P_j\{i, j, K_{ij}, W_{ji}, W_{jj}', U_i\} \quad (2)$$

本モデルを用いた評価は理想的にはすべての情報を用いて、(1) (2)式のように評価すべきであるが、本稿では、特に、これまで評価事例が少ないユーザが提供する Web ページの情報を用いた評価手法が有効かどうかに着目するため、Web ページ間の情報 (link 構造など) は用いない。また、ユーザ間の評価はオークション以外では実際には情報収集が困難であること、Web ページによるユーザ情報の提供も現時点では現実的でないことから省略する。さらに、(1) (2)式では、Web ページの評価結果がユーザの評価に反映され、そのユーザ評価の結果が Web ページに反映されるという再起構造になっている。従って、それぞれの評価を行うには、再帰的に数値計算を行わなければならない、結果の安定性等様々な問題があり、現時点では現実的でない。そこで今回は、再起構造の原因となる Web ページ評価によるユーザ評価部分は省略する。

以上を考慮し、(1) (2)式を以下のように縮退させる。また、縮退後の SKRank 評価モデルは図 3 の実線部のみ関係となる。

$$U_i = U_i\{i, j, K_{ij}\} \quad (3)$$

$$P_j = P_j\{i, j, K_{ij}, U_i\} \quad (4)$$

理想的にはユーザはインターネットを利用するすべてのユーザ、ユーザによる Web ページの情報はインターネット上にあるすべての情報、Web ページはインターネット上にあるすべての Web ページとなる。ユーザによる Web ページの情報は、2 章で述べたように数多くものが挙げられるが、今回はソーシャルブックマークサイトのタグを利用する。ソーシャルブックマークにおいて、ユーザがブックマークする際、Web ページを閲覧または推測し、その Web ページに関する内容や感想をタグとして表現しているため、タグにはユーザ各々の嗜好が出やすく、Web ページの玄人度合を評価する関連情報としては良い情報であると言える。

そこで、(3) (4)式において、ユーザ*i*による Web ページ*j*の情報 K_{ij} をソーシャルブックマークサイトのタグ及

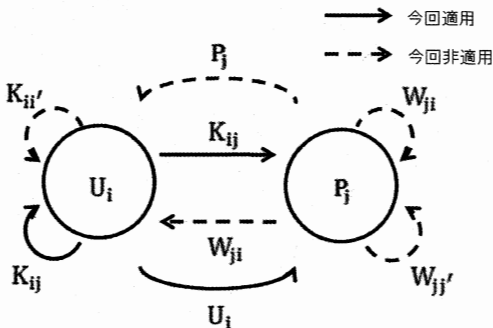


図 1 SKRank 評価モデル

びブックマークなどの情報 T_{ij} として定義すると、(5) (6)式のようになる。

$$U_i = U_i\{i, j, T_{ij}\} \quad (5)$$

$$P_j = P_j\{i, j, T_{ij}, U_i\} \quad (6)$$

また、ユーザ*i*をソーシャルブックマークサイトに登録しているユーザ及び Web ページ*j*はブックブックマークされた Web ページとして、SKRank モデルに適用する。

3.2. User Rank: $U(i)$

ユーザの評価値である U_i に相当する部分である。ユーザ*i*が各ページに対して付けた総タグ付与数を $T(i)$ とする。情報を提供してくれるユーザは信頼度があるユーザと仮定すれば、タグを多く付けるユーザつまり総タグ付与数 $T(i)$ が多いユーザは信頼あるユーザとなる。ここで、総タグ付与数 $T(i)$ を正規化したものをユーザ*i*の信頼度を $U(i)$ とすると、式(7)のようになる。

$$U(i) = \frac{T(i)}{\max_{i:1 \leq i \leq N} [T(i)]} \quad (7)$$

また、ここで N はユーザの総数、ユーザ*i*が Web ページ*j*に付与したタグ数を $T(i, j)$ 、ユーザ*i*の総ブックマーク数を $B(i)$ とすることで、 $T(i)$ は(8)式のように定義できる。

$$T(i) = \sum_{j=1}^{B(i)} T(i, j) \quad (8)$$

3.3. SKRank: $P(j)$

Web ページの評価値である P_j に相当する部分であり、また、今回の場合は $P(j)$ が SKRank 値となる。

ユーザ*i*がタグをある Web ページ*j*に対して、そのユーザが付与する平均タグ付与数 $\overline{T(i)}$ よりも多い Web ページを信頼が高い Web ページ、少ない Web ページを信頼が低いページと仮定する。このとき、ユーザの信頼度である User Rank と組み合わせることで信頼あるユーザが評価した場合には重みが大きくなり、信頼ある Web ページほど高い値となる。 $T(i, j)$ と $\overline{T(i)}$ との差分を $\alpha(i, j)$ とすると、 $\alpha(i, j)$ は(9)式のようになる。

$$\alpha(i, j) = T(i, j) - \overline{T(i)} \quad (9)$$

そして、この差分と User Rank を掛けたものが $P(i, j)$ となり、 $P(i, j)$ を Web ページに登録しているユーザについて、足し合わせたものが $P(j)$ となり、 $P(i, j)$ 及び $P(j)$ は(10)、(11)式のようになる。

$$P(i, j) = \alpha(i, j) \cdot U(i) \quad (10)$$

$$P(j) = \sum_{i=1}^{N(j)} P(i, j) \quad (11)$$

ただし、 $N(j)$ は Web ページ j に登録しているユーザ数である。また、 j は $1 \leq j \leq M$ であり、 M がソーシャルブックマークサイトに登録された総 Web ページ数となる。

4. SKRank の評価

4.1. データセット

実際のソーシャルブックマークサイトを用いて

SKRank 値の算出を行った。ソーシャルブックマークサイトは Delicious[7] やはてなブックマーク[8] などが有名だが、今回のデータセットには、Delicious を用いた。検索クエリを“カメラ”とし、約 6800 件の検索結果の内、取得可能な 2000 件のデータの中から、被ブックマーク数が 10 以上である Web ページを 37 件のデータを収集し、SKRank 値を算出した。また、他の評価手法との比較のため、被ブックマーク数や総タグ数の他に、Web ページの Page Rank を Google Toolbar を使って取得した。37 件の Web ページから、Page Rank が 3 の場合と被ブックマーク数が近いものを抽出し、SKRank 値に基づいて並び替えたものを表 1, 2 に示す。

表 2 Page Rank が 3 の被ブックマーク数、総タグ数、SKRank 値

Web ページ名	被ブックマーク数	総タグ数	SKRank 値
ITmedia Biz. ID:Flickrと携帯カメラで記憶のスナップショット	105	367	11.81
痛いニュース(ﾉﾌﾟ):車載カメラで当て逃げの瞬間を撮影して、警察に持って行くも警察は動かず... 2ちゃんねるで当て逃げ動画公開	20	55	4.26
かつて音楽と呼ばれたもの “ Blog Archive ” 赤外線カメラの作り方	28	93	2.05
トイカメラ風にデジカメ画像を変換するPhotoshop用プラグイン : Gizmodo Japan(ギズモード・ジャパン)	40	112	0.37
カメラマンWeb - 読者と作る, カメラと写真のコミュニティ	15	37	0.23
カメラ★ライフ	11	24	0.11
カメラのまつばら光機 HASSELBLAD, ハッセルブラッド, 舶来カメラ, Leica, ライカ, Rollei, ローライ, M42, Linhof, Sinar, etc.	10	21	0.020
MouRa デジタルカメラ, 家族旅行 デジタルカメラこども撮影術 超簡単「魔法のテクニック」	14	30	0.011
【楽天市場】マップカメラ:東京新宿にある新品・中古カメラ専門店です。購入後のケアも万全.	10	15	-0.012
沖縄のライブカメラ 那覇港 三重城からライブ中継	12	15	-0.12
トイカメラ - ギズモショップ (世界のユニークなトイカメラのご紹介)	11	30	-0.12
Rauru Blog “ Blog Archive ” アイカメラから得られた 23 の教訓	35	82	-0.23
ITmedia +D Games: 「PSPはカーナビになりました」——GPSレシーバー+カメラで広がるPSPの世界 (1/2)	14	31	-1.17
【楽天市場】雑貨 HIGHTIDE DAILY DEPOT:雑貨や文具の通販. 手帳やトイカメラ, エド・エンバリーまで販売	46	104	-2.23

表 3 被ブックマーク数が近い Page Rank, 被ブックマーク数, 総タグ数, SKRank 値

Web ページ名	Page Rank	被ブックマーク数	総タグ数	SKRank 値
ブログの画像からカメラ携帯の機種を調べてみる: blog.bulknews.net	4	123	400	20.39
ITmedia Biz. ID:Flickrと携帯カメラで記憶のスナップショット	3	105	367	11.81
ライブカメラコミュニティー Stickam	5	154	484	11.34
Googleでウェブカメラを根こそぎ探し出す方法 - GIGAZINE	3	130	392	7.68
スタジオグラフィックス プロが教える デジタルカメラの撮影&レタッチテクニック	4	150	374	5.94
ビックカメラ.com	6	136	300	-0.055

4. 2. SKRank の特性

表 1, 2 を見ると, SKRank 値に基づいて並び替えた場合, 商品販売系の Web ページはランキングの下位に来る傾向となる。これは, ユーザがタグ付けをする際に, 有名な Web ページはタグ付けせずとも記憶できるからだと推測できる。また, 上位に来る Web ページはノウハウ系の Web ページが多い。これは, ユーザが後に Web ページに書いてある内容を試したりするためにタグ付けをするなど, できる限りその記事の内容やおもしろさを記憶するために, タグの付与が多くなり, 結果として上位に来ていると推測できる。

次に, ソーシャルブックマークにおけるユーザのタグ付け行為を分析する。まず, その Web ページをブックマークするということはその Web ページに対して何らかの興味を持ったということである。さらに, その Web ページに対してタグ付けを行うということはその Web ページを忘れないように注釈として, タグ付けを行った可能性が高い。このことから, SKRank は淘汰された Web ページの中でさらに興味がある人によって評価されたランキングであると言える。以上の特性の正しさをみるために, SKRank 毎の付与タグ数の割合と SKRank と Page Rank 及び平均タグ数との相関をとった。

図 4 は取得した SKRank 値の中から, 最も少ない値 (-2.23) と 2 番目に少ない値 (-1.55) 及び最も大きな値 (20.4) と 2 番目に大きな値 (11.81) に付けられるタグ数の割合を基に, SKRank 値の振れ幅に対して, 個々のグラフの振られ方の分布を取ったものである。SKRank 値が -2.23, -1.55 の場合は, タグ数の少ない方の割合が高く, SKRank 値が 11.81, 20.4 の場合にはタグ数の多い方に割合が多くなる。このことから, SKRank が付与タグ数と相関があることが言える。11.81, 20.4 の場合のように, 似たグラフの分布をしていても, SKRank に差が生じる。これは, ブックマークを行うユーザの User Rank 値に影響を受けているためと考えられる。

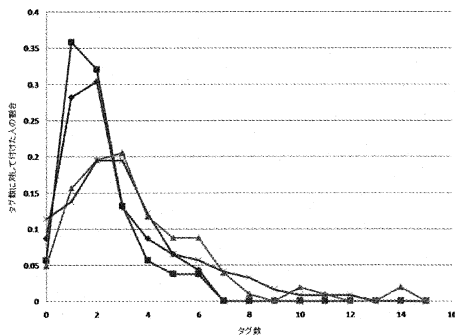


図 4 SKRank 値-2.23,-1.55,11.81,20.4 における人が付けたタグ数の割合

図 5 は Page Rank と SKRank の相関を取ったものである。傾向として SKRank が高いと Page Rank が低くなることはないが, SKRank が低くても Page Rank が高くなっている。これは, SKRank によって, 下限のふりがかけられた状態になっていることを示しており, 信頼性のない Web ページが評価されていないことがわかる。

次に, 平均タグ数と SKRank の相関を取ったものを図 6 に示す。平均タグ数と SKRank には正の相関があることがわかる。ソーシャルブックマークに付けられるタグは, 興味ある人がその Web ページをブックマークした中で, さらに反響の多かった Web ページであり, 平均タグ数が多いということはそれだけ, 玄人好みのページと言える。実際に表 1, 2 を見てみると, ビックカメラや楽天市場のような販売系のサイトでは, コアなユーザよりも広く浅いユーザがいる Web ページでは SKRank 値はあまり高くなっていない。一方, 狭く深い層がいるようなノウハウ系の Web ページだと SKRank が高くなる傾向になっている。

以上のことから, SKRank というのは, 信頼性のある Web ページの中でさらに, 玄人好みの Web ページを抽出できる Web ページの評価手法と言える。

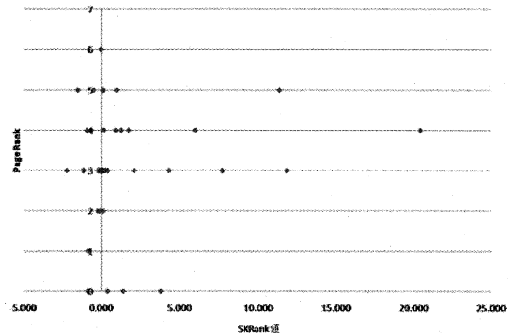


図 5 Page Rank と SKRank

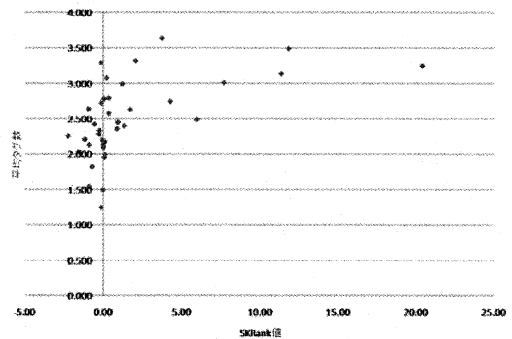


図 6 平均タグ数と SKRank

5. おわりに

本稿では、ユーザ、Web ページとそれらの関連情報の三者間の構造を利用したモデルである SKN モデルの評価モデルを用いて、Web ページを評価する手法である SKRank モデルを提案した。また、SKRank モデルをソーシャルブックマークにおけるユーザ、Web ページ及び Web ページに付与されているタグ情報等に適用して Web ページの定量的評価を試みた。実際に本評価手法に従い、ソーシャルブックマークサイトである Delicious を使って、SKRank 値を求めた。さらに、人間による主観的評価と平均タグ数や Page Rank との相関による客観的評価から、SKRank の基本的コンセプトには妥当性があり、信頼性のある友人好みの Web ページの評価値が高くなることがわかった。今後課題としては、客観的な評価や正確なモデル化が必要である。

なお、現在はソーシャルブックマークのサイトを閲覧し、人間がブックマーク及びタグ情報を取得しているため、被ブックマーク数の多い Web ページなどのサンプル数も少ない。そこで、様々なソーシャルブックマークサイトのタグ情報などを自動で取得できるようなクローラーの開発予定である。

参考文献

[1]Yusuke Yanbe, Adam Jatowt, Satoshi Nakamura, and Katsumi Tanaka, "Towards Improving Web Search by Utilizing Social Bookmarks," ICWE 2007, LNCS 4607, pp. 343-357 (2007).

[2]高橋翼, 北川博之, "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評価," DEWS2008 B8-6 (2008).

[3]丹羽智史, 土肥拓生, 本位田真一, "Folksonomy マイニングに基づく Web ページ推薦システム," 情報処理学会論文誌, Vol. 47, No. 5, pp. 1382-1392 (2006).

[4]杉山 典之, 関 洋平, 青野 雅樹, "ユーザのタグ付けの傾向を利用したソーシャルブックマーク内の関連ページ検索手法," DEWS2008 B9-2 (2008).

[5]百田信, 伊東 栄典, "ソーシャルブックマークに基づく情報発見," DEWS2008 I1-15 (2008).

[6]Page, L., Brin, S., Motwani, R. and Winograd, T.: The pagerank citation ranking: Bringing order to the Web. Technical report, Stanford Digital Library Technologies Project (1998)

[7]Delicious, <http://delicious.com/>

[8]はてなブックマーク, <http://b.hatena.ne.jp/>