ソーシャルブックマークにおける ブックマークの活性度を考慮したWebページのランキング

髙橋　翼† 北川　博之†,††

† 筑波大学大学院システム情報工学研究科　〒 305-8573 茨城県つくば市天王台 1–1–1
†† 筑波大学計算科学研究センター　〒 305-8573 茨城県つくば市天王台 1–1–1
E-mail: †tsubasa@kde.cs.tsukuba.ac.jp, ††kitagawa@cs.tsukuba.ac.jp

あらまし　近年，インターネット上にブックマーク情報を登録し，共有するサービスであるソーシャルブックマークが注目されている．ソーシャルブックマークからはユーザの Web ページに対する興味を (URL，ユーザ名，ブックマーク日時，タグ集合) というような構造化された形で取得できる．ソーシャルブックマークには，一過性の話題を扱うようなページや，参考文献のように長い期間に渡って人々に参照されるようなページがあり，また過去の情報も削除されずに残っている場合が多い．そのため，ソーシャルブックマークに登録されているページすべてが現在も情報としての価値，鮮度を持っているとは言い難い．そこで，ページに対する過去のブックマークの時系列パターンから，確率推論モデルを用いて，現在におけるブックマークの活性度を評価し，それを基にした Web ページのランキング手法を提案する．

キーワード　ソーシャルブックマーク　時系列情報　活性度　ランキングアルゴリズム

Ranking Web Pages Based on Activation Analysis of Social Bookmarks

Tsubasa TAKAHASHI† and Hiroyuki KITAGAWA†,††

† Graduate School of Systems and Information Engineering, University of Tsukuba　Tennoudai 1-1-1, Tsukuba City, Ibaraki, 305-8573 Japan
†† Center for Computational Sciences, University of Tsukuba　Tennoudai 1-1-1, Tsukuba City, Ibaraki, 305-8573 Japan
E-mail: †tsubasa@kde.cs.tsukuba.ac.jp, ††kitagawa@cs.tsukuba.ac.jp

Abstract　Recently, Social Bookmark, which allows us to register and share our own bookmarks on the web, is attracting attention. We can get structured data such as (URL, Username, Timestamp, Set of tags) from Social Bookmark. These data represent user's interests. There are two aspects of bookmark usages: for reuse and for hot issues. In this paper, by focusing on the timestamps, we propose a evaluation method to measure activation levels of web pages. Moreover, we propose a page evaluation method which improves S-BITS, our previous proposed method, utilizing the evaluation of activation levels. Finally, we show the effectiveness of the proposed method by some experiments.

Key words　social bookmark, time series data, burst of activity, ranking algorithm

1. はじめに

昨今の情報社会では，色々な形で情報が発信され，様々な情報を手に入れられるようになった．しかし，情報は溢れかえり，ユーザは情報の取捨選択を強いられている．このような状況化では，情報に対する信頼度の重要性が強く認識されている．Web 検索エンジンを利用することで，誰もが簡単に情報を探すことができる．しかし，出力されるランキングは，ページ間のリンク構造や単語の出現頻度などによるページの評価に基づいたもので，閲覧した人々がどう感じたか，どんな評価を下したかといった情報が反映されたものではない．

一方，Web 上に個人のブックマーク情報を作成し，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に注目が集まっている．ソーシャルブックマークでは，ユーザは興味や関心を持ったページに独自の観点でタグを用いて注釈を与えることができる．ブックマークされたページはお気に入りのペー

ジであったり，役に立つ，おもしろいなど，ブックマークしたユーザにとって，何らかの価値のある情報を持ったページと考えることができる．また，ユーザがブックマークするという振舞は，ページに対して評価を与える行為と考えることもできる．人それぞれ，興味や嗜好が異なり，分野ごとの専門性には大きな違いがある．現実世界では，特定の分野において，専門性が高く，知識の豊富な人の意見は信頼できる．先行研究 [20] では，ソーシャルブックマークユーザの専門性を考慮したページのランキング手法 S-BITS を提案し，既存の検索エンジンや，単純なブックマーク数のみを考慮した手法よりも高い適合率を示す結果が得られた．

ソーシャルブックマークでは，ユーザがいつブックマークしたかという情報であるブックマーク日時がメタデータとして与えられている．Web ページにはニュース記事のようにある一定期間のみ情報としての価値を持つページや Wikipedia やマニュアルページのように価値が長く持続するページが存在する．例えば，1 年前のニュース記事よりも，この 1 年間継続して読まれてきたページの方が今の時点では価値のある情報と考えることができる．これらのページ間の違いは，ページに与えられたブックマークがどれだけ現在における価値を維持し，活性化しているかの度合と考えることができる．そこで本研究では，現在におけるページのブックマークの活性度を，そのページのブックマークの時系列分布から評価することを考える．また，ページの活性度を考慮することで先行研究で提案した S-BITS を拡張する手法の提案も行う．

著者らの研究 [21] では，ページのブックマークの時系列分布の散らばりの度合から，そのページがどれだけ情報としての価値が持続するのかを測り，それを利用することでページの鮮度の有無を評価する手法を提案している．本稿は，そのページに対する現在のブックマークの活性度がどの程度あるのかを確率推論モデルを用いて，推定し，活性度の度合を利用して，Web ページのランキングの精度を向上するという点で異なる．

本稿の以降のセクションの構成は以下の通りである．まず，2 章ではソーシャルブックマークの概要について述べる．3 章では，本研究と関連のある過去の研究について概観する．4 章では，ページのブックマークの活性度とその評価手法について述べる．5 章では，Web ページのランキング手法について述べる．6 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，7 章で本稿のまとめを述べると共に，今後の課題について述べる．

2. ソーシャルブックマーク

ソーシャルブックマーク (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．ユーザは各自のブックマーク情報に独自の注釈をタグによって与え，管理することができる．“万人による注釈” という意味を持つ Folksonomy の概念を実現しており，人手によって注釈を与えられた情報源である．また，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言える．

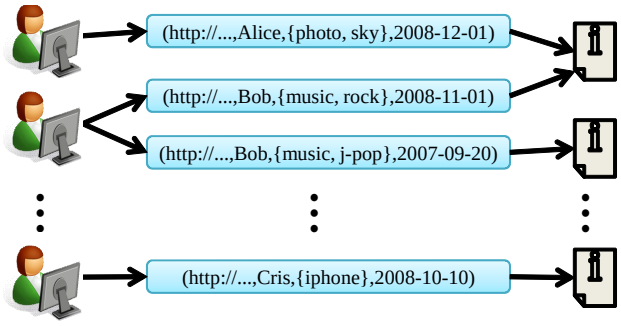

図 1　ソーシャルブックマーク

主要な SBM サービスに，delicious [17] やはてなブックマーク [18] などがある．SBM からはブックマーク情報を (URL，ユーザ名，ブックマーク日時，タグ集合) というような構造化された形で取得できる (図 1)．これはユーザの Web ページに対する興味を表している．ユーザは独自の価値観で任意のタグを用い，ページに注釈を付けることができる．多くのユーザは対象ページを表すキーワードやカテゴリに関する単語やフレーズをタグとして用いている．ブックマーク日時はユーザがブックマーク対象のページにいつ興味を持ったかを表す情報である．

本研究では，ブックマーク情報 b を以下のようにモデル化する．

$$b = (p, u, t, A)$$
$$A = \{a_1, a_2, ..., a_n\}$$

p はページ，u はユーザ，t はブックマーク日時，A は注釈として与えられたタグの集合を表し，a_i は各タグを表す．

SBM は，様々なユーザによって作られる一種のソーシャルネットワークであるため，ユーザによって嗜好も異なり，特定の情報に対する見識にはばらつきがある．なかには特定のトピックに対して非常に詳しいユーザがいたり，少しかじった程度のユーザもいたりする．自分の興味があるトピックについて情報を得たいと思ったとき，そのトピックについて詳しい人からの意見は参考になる可能性が高い．また，多くの人々から評価を得ている情報もまた参考になる可能性が高い．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 1 つの指標と言える．

3. 関 連 研 究

SBM の普及と共に SBM を含む Folksonomy に関する研究はより盛んになってきている．Golder ら [4] は，SBM の構造を詳細に分析し，ユーザの行動やタグの使用頻度などの規則性について報告している．Xian Wu ら [5] は，アノーテーションが与えられた Web リソースに対するセマンティックな検索モデルを SBM を取り上げ，提案している．Hotho ら [6] [7] は SBM のページのランク付けに PageRank を応用した FolkRank を提案し，また，トピックに特有のトレンドを発見する手法を提

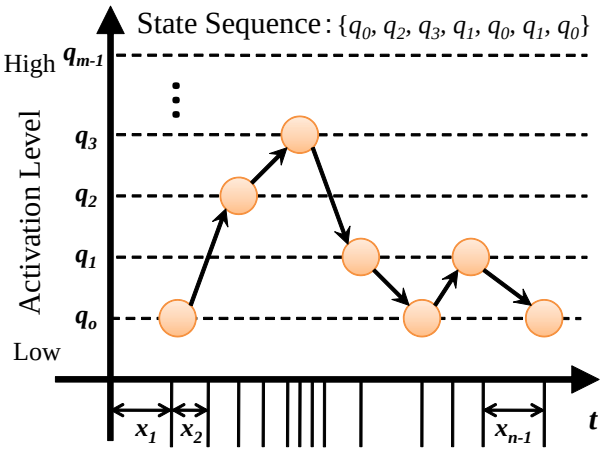

図 2　活性度の状態遷移系列

案している．Heymann ら [8] は，様々な観点から SBM を分析し，その詳細な報告を行っている．山家ら [3] は SBM 上でのページの被ブックマーク数を 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を提案し，SBRank 値の Web 検索の尺度としての有用性を示している．

本稿で着目した時系列情報に関する研究も広く行われている．毛受 [15] らは SBM に新しく投稿されたページの注目度を予測する手法を提案している．本研究は，SBM のあらゆるページが現在どれだけ活性度があるのかを扱っている点で異なる．

4.　ブックマークの活性度の分析

4.1　時系列データの活性度のモデリング

時系列データの活性度に関する研究には，Kleinberg [9] の研究がある．Kleinberg は，文章ストリームにおける文章の到着頻度に着目し，特定のトピックの活性度 (バーストの強度) を分析する手法を提案した．Kleinberg による手法では，隠れマルコフモデル (HMM) [10] を用いて，次の文章ストリームの到着確率が指数関数的に異なる複数の状態を内部状態とし，確率的に状態遷移が行われるモデルを提案している．

HMM はマルコフモデルの各状態に対して，確率的な記号出力を加えたモデルであり，出力記号系列からは，内部状態の状態遷移が不明な，隠れた状態を内部状態として持つオートマトンの一種である．出力記号系列から，HMM の内部状態の状態遷移系列を推定する問題に対しては，ビタビアルゴリズム [11] が有効な手法として知られている．

Kleinberg による手法では，活性度が高い状態では頻繁に文章が到着し，次の文章の到着時間間隔は短くなる可能性が高くなり，活性度が低い状態ではまれにしか文章が到着しないため，次の文章の到着時間間隔は長くなる可能性が高くなるように文章ストリームの活性度をモデル化している．

文章ストリームのトピック分析や注目語の抽出に Kleinberg の手法は有効な手法として知られ，よく用いられている．崔ら [16] は，Kleinberg の手法を拡張し，トピックの関連度を考慮した文章ストリームの活性度分析手法を提案している．

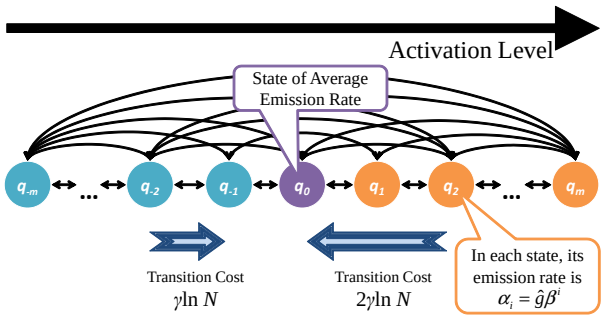

図 3　SBM の活性度モデル

4.2　ブックマークの活性度のモデリング

本研究では，SBM 上の Web ページのブックマークの活性度をモデル化することを考える．Kleinberg の手法と同様に，HMM を用いてモデル化を行う (図 3)．ある Web ページに対するブックマークの活性度が高い状態では頻繁にブックマークがなされ，次のブックマークがなされるまでの時間間隔が短くなる可能性を高くする．活性度が低い状態ではまれにしかブックマークがなされないため，次のブックマークがなされるまでの時間間隔が長くなる可能性を高くする．

一連の n ブックマークの時間間隔 $\mathbf{x} = (x_1, x_2, ..., x_{n-1})$ が与えられたとき，各時間間隔 x_i は付随する HMM の内部状態に応じて確率的に出力される記号とみなす．この HMM を用い，ブックマークの時間間隔 $\mathbf{x} = (x_1, x_2, ..., x_{n-1})$ から最適な状態遷移系列 $\mathbf{s} = (s_1, s_2, ..., s_{n-1})$ (s_j は状態番号) を状態遷移コストが最小になるように，ビタビアルゴリズムを用いて最尤推定する．これによって，あるページの活性度の時間変化の系列を分析することができる．

ページはそれぞれ固有のコンテンツを持つため，どのように注目され，どのような時間間隔でブックマークされるかはページごとに異なる．毎日のようにブックマークされるものもあれば，一か月に何度かしかブックマークされないものもある．これらの違いはブックマークの生起確率とみなすことができる．本研究では，平均的なブックマークの生起確率を基準にページの活性度を規定する．HMM において，この平均的なブックマークの生起確率を持つ状態を q_0 とする．平均的なブックマークの生起確率 $\hat{g}$ は，$\mathbf{x} = (x_1, x_2, ..., x_{n-1})$ をソートしたデータ系列 $\chi = (\chi_1, \chi_2, ..., \chi_{n-1})$ ($\chi_i \in \mathbf{x}$) の四分位範囲の範囲内の値から導出する．

$$\hat{g} = \left(\frac{1}{\frac{1}{2}n} \sum_{i=\frac{1}{4}n}^{\frac{3}{4}n} \chi_i \right)^{-1} \tag{1}$$

このように導出することで，バーストしているときのように過剰に活性している期間や，ブックマークのピークが過ぎてほとんどブックマークされなくなり，極端に活性していない期間からの影響を除いて，平均的なブックマークの生起確率を求めることができる．また，q_0 を基準に，$M = 2m + 1$ 個の異なるブックマークの生起確率を持つ状態から成る内部状態系列 $\mathbf{q} = \{q_{-m}, q_{-m+1}, ..., q_{-1}, q_0, q_1, ..., q_{m-1}, q_m\}$ を規定する．各

状態 q_i は，状態番号 i が増加する程，短い時間間隔で次のブックマークが生起しやすく，状態番号 i が減少する程，次のブックマークが生起するまでの時間間隔が長い可能性が高くなる．状態 q_i は，次のブックマークの出現確率が以下のような式で表わされる指数確率密度関数 f_i に従うような状態である．

$$f_i(x_i) = \alpha_i e^{-\alpha_i x_i} \qquad \alpha_i = \hat{g}\ \beta^i \qquad \beta > 1 \tag{2}$$

このとき，β は各状態のブックマークの生起確率の分解能を設定するパラメータである．状態番号 i が増加するに従い，活性度が高くなる．また，各状態 q_k から q_l へ状態遷移するとき，以下の式で表わされるような状態遷移コスト $\tau(k,l)$ が発生するものとする．

$$\tau(k,l) = |l-k|\gamma \ln n \qquad \gamma > 0 \tag{3}$$

このとき，γ は状態遷移の起こりやすさを設定するパラメータである．この状態遷移コスト $\tau(k,l)$ により，ブックマークの生起確率が大きく異なる状態への遷移は難しくなり，たまたま頻繁なブックマークの追加が発生したり，ブックマークが発生しなかったりといった，偶然発生するようなノイズに強くすることができる．また，ブックマーク数が大きなページほど，状態遷移にかかるコストが大きくなるため，頻繁な状態遷移はより起こりづらくなる．

本研究では，内部状態数 $M = 15(m = 7)$ であり，q_{-7}〜q_7 を内部状態として持つ HMM を用いる．また，パラメータ $\beta = 2$, $\gamma = 1$ と設定し，活性度の分析を行う．

4.3 現在の活性度の推定

SBM からはあるページが過去にどのようにブックマークされてきたかというブックマークの時間分布を得ることができる．しかし，これはあくまで過去の分布であり，ここから得られる活性度の時間変化は現在の状態を加味したものではない．

現在どのような活性度を持った状態になるのかを推定するために，対象となるページが現在時刻にブックマークされることを仮定する．対象ページ p のブックマーク間隔系列 $\mathbf{x}'$ に，現在時刻と最後のブックマークとの間隔 x_n を追加したものを $\mathbf{x}' = (x_1, x_2, ..., x_{n-1}, x_n)$ とする．$\mathbf{x}'$ を出力記号列として，ビタビアルゴリズムを用いて，活性度の状態遷移系列を推定する．推定された活性度の状態遷移系列 $\mathbf{s} = (s_1, s_2, ..., s_{n-1}, s_n)$ の最後の状態番号 s_n が表す状態 q_{s_n} が，現在における対象ページ p のブックマークの活性度 act_p となる．

推定された活性度の状態遷移系列の例を図 4〜9 に示す．いずれも，2009 年 2 月 14 日時点におけるはてなブックマークの SBM データに基づいている．

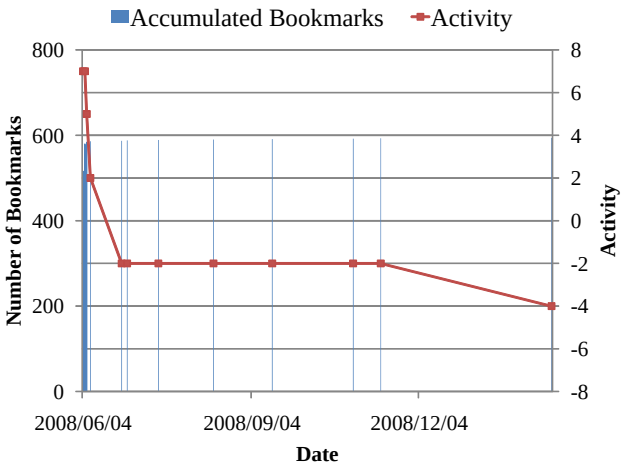

図 4　活性度評価の例 1 (注1)

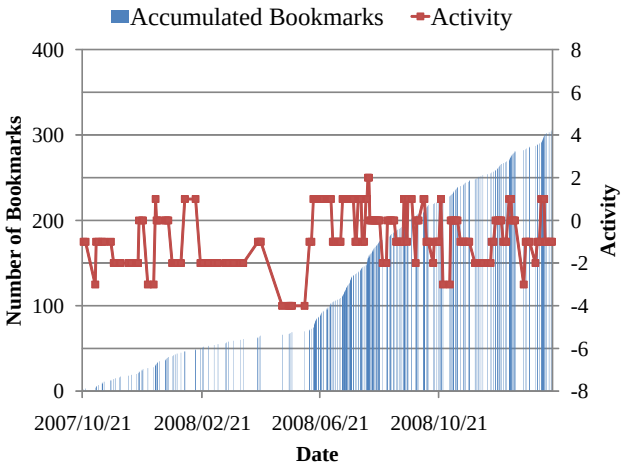

図 5　活性度評価の例 2 (注2)

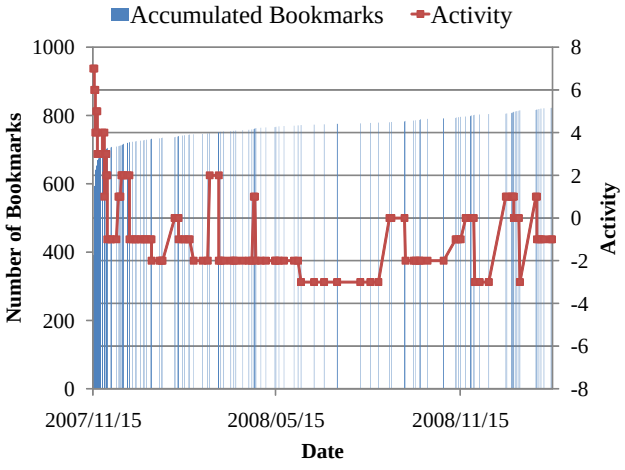

図 6　活性度評価の例 3 (注3)

(注1)：http://www.softbankmobile.co.jp/ja/news/press/2008/20080604_01/
(注2)：http://ipodtouchlab.com/
(注3)：http://dev.ariel-networks.com/articles/workshop/ruby/
(注4)：http://www.eweb-design.com/
(注5)：http://www.asabanana.net/
(注6)：http://d.hatena.ne.jp/hejihogu/20070227/p9

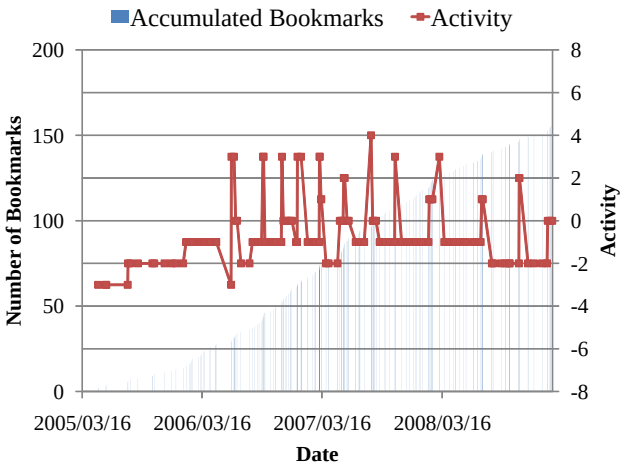

図 7　活性度評価の例 4 (注4)

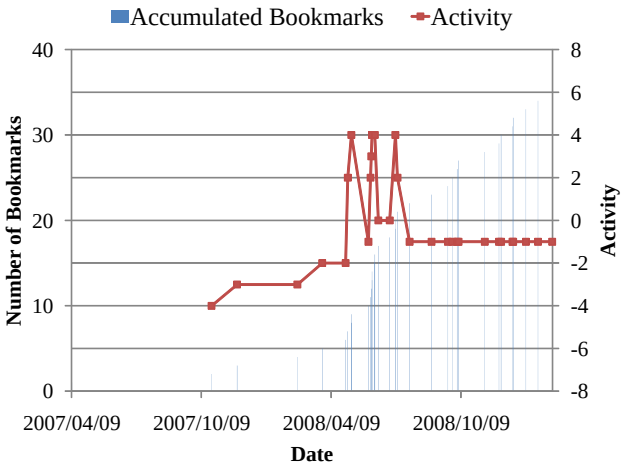

図 8　活性度評価の例 5 (注5)

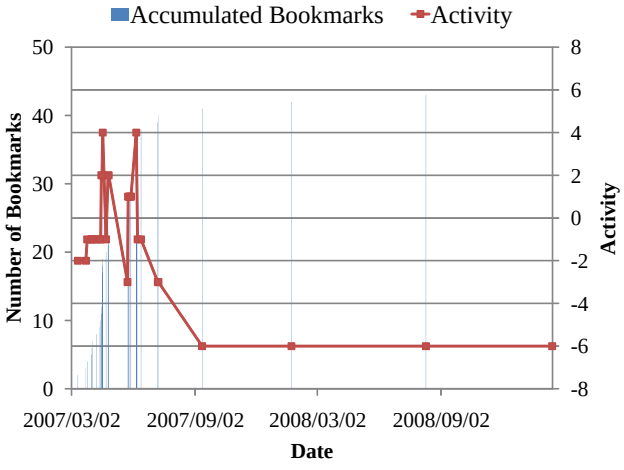

図 9　活性度評価の例 6 (注6)

5.　Web ページのランキング

先行研究 [20] で提案した S-BITS に本稿で提案したページの活性度の評価を取り入れた手法を提案する．単純なブックマーク数やユーザ-ページ間のリンク関係のみを考慮した手法では，現在において，そのページがどれだけ情報としての価値を維持し，参照される可能性があり，活性度を持っているかを考慮することはできない．現在，活性化されている情報かどうかはページの質を測る一つの重要な指標である．上記のような問題点は先行研究で提案した S-BITS [20] にも存在する．そこで，S-BITS にページの活性度の評価を取り入れ，ページの評価精度を向上することを考える．まず，S-BITS の概要を示し，ページの活性度の評価を取り入れた S-BITS の拡張手法を提案する．

5.1　S-BITS

S-BITS(Social-Bookmarking Induced Topic Search) は，SBM におけるページとユーザ間の 2 部グラフを対象にした Web ページのランキング手法である．S-BITS では，SBM ユーザの Hub 度によって表わされる専門性と，ページとユーザとの相互関係から導かれるページの Authority 度によって，ページを評価する．

S-BITS では，SBM から得られる情報について以下の仮定をする．

- ユーザのブックマークをするという振舞はページに対して正の評価を与えることである．
- 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．

上記の仮定の基にページの Authority 度，ユーザの Hub 度を算出する．多くのユーザに支持されており，またそのトピックに詳しく専門性の高いユーザに支持されているページが検索結果の上位に現れる．また，HITS がページ間の in-link， out-link を利用して，対象とするページ集合を拡張しているのに対し，S-BITS では，タグの共起に着目し，共通のタグ集合を利用することで，対象ページ集合の拡張を行う．S-BITS のアルゴリズムの概要は以下の通りである．

（1） 検索クエリ q を与え，q を用いて検索エンジンから上位 n 件のページを収集する (初期ページ集合 P_0)．SBM サービスを利用し，初期ページ集合 P_0 の各ページ p_i のブックマーク情報 $b_{ji}(=(p_i, u_j, t_{ji}, A_{ji}))$ を収集する (ブックマーク集合 B)．収集した全ユーザからユーザ集合 U，全タグ集合からルートタグ集合 V を生成する．

（2） タグ集合を利用して関連性のあるページを収集する．V 中において，頻出なタグ集合 F を抽出する (V')．頻出タグ集合 F は相関ルールマイニングの極大頻出アイテム集合抽出によって得る [13] [14]．頻出タグ集合 $F \in V'$ を包含するタグ集合 A でブックマークされているページを収集し，P_0 とマージする (ページ集合 P)．

（3） ページ集合 P，ユーザ集合 U，ブックマーク集合 B からなるグラフ G を対象に，ページの Authority スコア (p_score)，

ユーザの Hub スコア (u_score) を計算し，Authority スコアを基に，ページのランキングを行う (図 10).

S-BITS
$p\_score^0 = \{1, 1, 1, ..., 1\}$
$u\_score^0 = \{1, 1, 1, ..., 1\}$
$k = 1$
Repeat
foreach $p_i \in P$
$p\_score_i^k = \sum_{b_{ji} \in B} u\_score_j^{k-1}$
foreach $u_i \in U$
$u\_score_i^k = \sum_{b_{ij} \in B} p\_score_j^{k-1}$
normalize(p_score^k)
normalize(u_score^k)
until $|p\_score^k - p\_score^{k-1}|_1 < \epsilon_p$
and $|u\_score^k - u\_score^{k-1}|_1 < \epsilon_u$
return p_score^k and u_score^k
end

図 10　S-BITS の評価値算出アルゴリズム

5.2　活性度を考慮したランキング

前節で提案した活性度の評価を Web ページのランキングに反映させ，ランキングの適合率を向上させることを考える．本研究では，Web ページに対して評価を行った活性度の値を，ページの重みとしてユーザ-ページ間のブックマークに与え，S-BITS のページスコア，ユーザスコアの計算の際に，この重みを反映させる．提案した活性度の評価手法では，活性度を整数で表わしている．検索者の時間軸に関する評価の多様性に対応するために，活性度がページの評価に対して与える影響力を変更可能にすることを考える．これを実現するために，ニューラルネットワークのバックプロパゲーションで良く用いられるシグモイド関数を用い，活性度を標本化する．シグモイド関数は以下のような式で表わされる．

$$\varsigma_\lambda(z) = \frac{1}{1 + \exp(-\lambda z)} \tag{4}$$

シグモイド関数では，パラメータ λ の値を 0 に近づけることで，$\varsigma_\lambda(z) = 0.5$ に漸近し，∞ に近づけることで，大きさ 1 のステップ関数に漸近する．シグモイド関数を用いてページの活性度を標本化することで，活性度の高低をより顕著にしたり，鈍化させたりと言ったように，0 から 1 の範囲で活性度が与える影響を調整することができる．シグモイド関数を用いて標本化した活性度をページとユーザの評価に適用する．基本的な評価の考え方は以下の通りである．

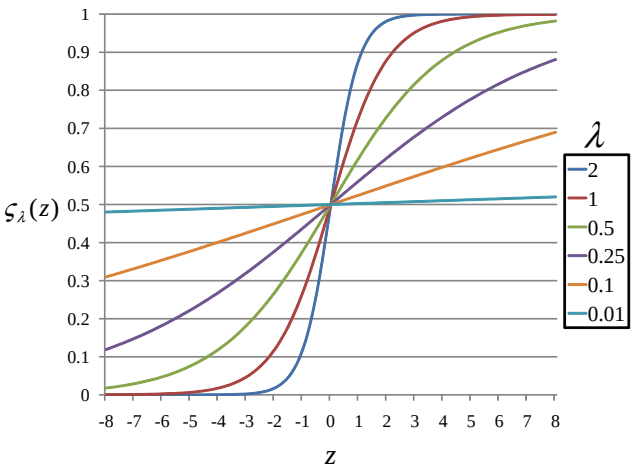

図 11　シグモイド関数

- 良いユーザにブックマークされ，活性度が高く保たれているページは良いページである．
- 活性度が高く保たれ，かつ良いページをブックマークしているユーザは良いユーザである．

上記を基に，先行研究で提案した S-BITS のページ-ユーザの評価式を以下のように定義する．

$$p\_score_i^k = \sum_{b_{ji} \in B} \varsigma_\lambda(act_{p_i})\ u\_score_j^{k-1} \tag{5}$$

$$u\_score_i^k = \sum_{b_{ij} \in B} \varsigma_\lambda(act_{p_j})\ p\_score_j^{k-1} \tag{6}$$

この評価式を S-BITS のアルゴリズム (図 10) に当てはめ，ページスコアベクトルとユーザスコアベクトルが平衡を迎えるまで計算を繰り返し，定常状態となったページスコアを基に，Web ページのリランキングを行う．これにより，現在においても，情報が価値を維持しているかどうかを考慮した上で，ユーザの検索対象ページから成るコミュニティにおける専門性を考慮した Web ページのランキングが可能となる．本手法を AS-BITS (Activity-aware S-BITS) と呼ぶ．

6.　評 価 実 験

提案手法の有用性を測るために評価実験を行う．既存の検索エンジン (Yahoo!) のオリジナルランキング，S-BITS のランキング，提案手法 AS-BITS のランキングの妥当性を比較・検討する．提案手法 AS-BITS はスコア計算に用いるシグモイド関数のパラメータ λ の値が 1, 0.5, 0.25 の 3 種類のランキングを作成し，ページの活性度の重みづけの方法の違いによる，適合率の変化を観測する．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 Web Search API [19]
- 対象 SBM サービス：はてなブックマーク [18]
- 検索語：iphone, ruby, web design, 論文 書き方

検索対象ページに関するソーシャルブックマーク情報は，はてなブックマークからクローリングしたデータを用いる．クローリングしたデータに対して，あらかじめページの活性度を測定しておく．このとき，音声認識や言語モデルの認識によく用いられる HMM Tool Kit(HTK) [12] を利用することで HMM およびビタビアルゴリズムを扱う．HTK では，音声や言語を含む一般の時系列データに対して HMM を用いた認識・推論を行うことができる．

Yahoo!のオリジナルのランキングは，Yahoo! Web Search API を利用して取得する．また，S-BITS，提案手法 AS-BITS のランキングは Yahoo! Web Search API から上位 200 件を取

得し，それぞれについてはてなブックマークに登録されているブックマーク情報を抽出し，抽出した情報を基に，それぞれの評価手法でスコアを算出し，ランキングを作成する．

各ランキングに対して手作業による評価を行うために，被験者を募った．被験者に対して，評価対象ページのコンテンツが有用性のある情報かどうか，及び，現在も情報としての価値が持続している（鮮度がある）情報かどうかの 2 つの観点から評価を依頼した．このとき，どの評価手法であるかという心理的影響を避けるために，評価手法に関する情報を全く与えずに同一フォーマットで出力される検索結果に対して評価を依頼した．人手による判定には作業量的限界があるため，上位 20 件のみを評価対象とした．

ページの鮮度に関する適合率 (図 12) では，AS-BITS($\lambda = 1, 0.5$) が高い適合率を示している．ただし，上位 1, 2 件では，Yahoo!のランキングが最も高い．S-BITS は上位において，非常に適合率が低い．SBM には，極めて短い期間に注目を集め，多くの人々にブックマークされるページが存在する．たとえ既に廃れてしまった情報を持つページでも，S-BITS はブックマーク数の多いページを上位に位置づける傾向にあるため，このような結果になったと考えられる．検索語 “iphone”において，このような現象が顕著に見られた．また，$\lambda = 0.25$ の AS-BITS は他の手法に比べて，良い結果であるとは言い難い．$\lambda = 0.25$ のシグモイド関数による活性度の標本化では，S-BITS と組み合わせたときに，活性度を生かし切れていないのではないかと考えられる．

ページの有用性の適合率 (図 13) でも，AS-BITS($\lambda = 1, 0.5$) が高い適合率を示している．S-BITS は上位 5 件において，適合率が低い．活性度が低く，鮮度が保たれていないページは有用でないものが多い．ページの活性度に基づく，鮮度評価のモジュールが S-BITS には組み込まれていないことに起因すると考えられる．ただし，$\lambda = 0.25$ の AS-BITS は上位においては S-BITS を上回る適合率を示してはいるが，それ以降は 10% 程度劣った結果を示している．このことからも，S-BITS に単純に活性度による重みづけをすれば，適合率の向上が望めるわけではないことがわかる．

AS-BITS が高い適合率を示していることから，活性度を評価し，Web ページのランキングに適用させることが一定の有益性を有していると考えることができる．ただし，活性度をどのように標本化し，ページの評価に適用させるかは，検討の余地がある．

7. ま と め

SBM の各ブックマークから得られるメタデータであるブックマーク日時に着目し，ブックマーク日時の時系列情報から，HMM を用いて，ページがどの程度活性しているかを推論することによる，現在のページの活性度を評価する手法を提案した．ページの活性度の概念を基に，ブックマークの時間軸を考慮した S-BITS の改良手法 AS-BITS を提案した．評価実験を通して，AS-BITS が高い適合率を示し，有効な手法であることを示した．今後の課題として，提案モデルの各種パラメータの最適

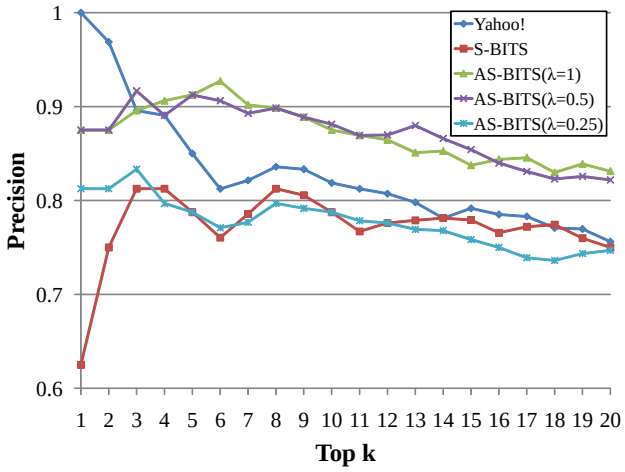

図 12　鮮度の適合率

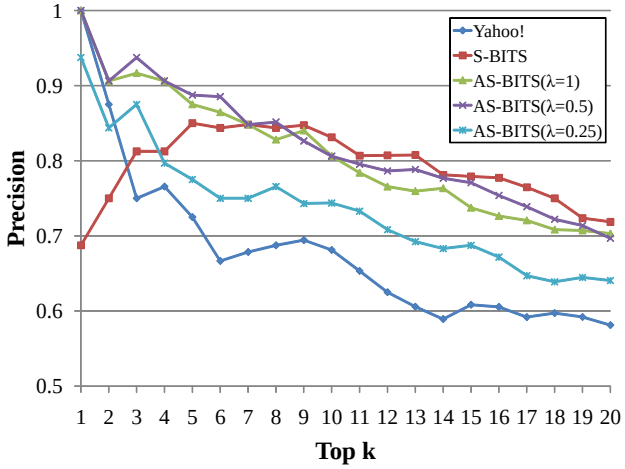

図 13　有用性の適合率

な値を検討することが挙げられる．また，ページのブックマーク日時の時系列情報だけでなく，ユーザがブックマークをした日時の時系列情報や，タグの共起，局所性など，SBM から得られる様々なメタデータを分析し，より多角的な検索モデルを構築したいと考えており，今後の課題である．

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．

文　献

[1] J. M. Kleinberg : “Authoritative Sources in a Hyperlinked Environment”, Journal of the ACM, Vol. 46, No. 5, pp. 604-632, 1999.

[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] Y. Yanbe, A. Jatowt, S. Nakamura and K. Tanaka : “Towards Improving Web Search by Utilizing Social Bookmarks”, In Proc. of 7th International Conference on Web Engineering, pp.343-357, 2007.

[4] S. A. Golder and B. A. Huberman : “The structure of collaborative tagging systems”, http://www.hpl.hp.com/research/idl/papers/tags/, 2005.
[5] X. Wu, L. Zhang and Y. Yu : “Exploring Social Annotations for the Semantic Web”, In Proc. of the 15th World Wide Web Conference, pp.417-426, 2006.
[6] A. Hotho, R. Jäschke, C. Schmitz and G. Stumme : “Information Retrieval in Folksonomies: Search and Ranking”, In Proc. of 3rd European Semantic Web Conference (ESWC 2006), pp.411-426, 2006.
[7] A. Hotho, R. Jäschke, C. Schmitz and G. Stumme : “Trend Detection in Folksonomies”, In Proc. of First International Conference on Semantic and Digital Media Technologies (SAMT 2006), pp. 56-70, 2006.
[8] P. Heymann, G. Koutrika, and H. Garcia-Molina : “Can Social Bookmarking Improve Web Search?”, In Proc. of the First International Conference on Web Search and Data Mining, pp.195-206, 2008.
[9] J. Kleinberg : “Bursty and Hierarchical Structure in Streams”, In Proc. of the 8th ACM SGKDD International Conference on Knowledge Discovery and Data Mining, 2002.
[10] Lawrence R. Rabiner : ”A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition”, Proceedings of the IEEE, 77 (2), pp. 257-286, February 1989.
[11] G. D. Forney : “The Viterbi algorithm”, Proceedings of the IEEE 61(3):268-278, March 1973.
[12] HMM Tool Kit http://htk.eng.cam.ac.uk/
[13] R. Agrawal and R. Srikant: Fast Algorithms for mining Association Rules. In Proc. of the 20th International Conference on Very Large Data Bases, pp.487-499, 1994.
[14] D. Burdick, M. Calimlim and J. Gehrke: MAFIA: A Maximal Frequent Itemset Algorithm for Transactional Databases. In Proc. of the 17th International Conference on Data Engineering (ICDE’01), p.443-452, 2001.
[15] 毛受崇, 吉川正俊 : “ブックマークの時系列情報を利用したソーシャルブックマークにおける注目度予測”, 電子情報通信学会 第 19 回データ工学ワークショップ (DEWS2008), 2008.
[16] 崔春花, 北川博之: “到着頻度と関連性を考慮した時系列文書のトピック分析”, 日本データベース学会 Letters Vol. 3, No. 2, pp. 37-40, 2004.
[17] “delicious”. http://delicious.com/
[18] “はてなブックマーク”. http://b.hatena.ne.jp/
[19] Yahoo! Search Web Services. Yahoo! DEVELOPER NETWORK. http://developer.yahoo.co.jp/search/web/V1/webSearch.html
[20] T. Takahashi and H. Kitagawa : “S-BITS: Social-Bookmarking Induced Topic Search”, In Proc. of the 9th International Conference on Web-Age Infromation Management (WAIM2008), pp.25-30, 2008.
[21] 髙橋翼, 北川博之 : “ソーシャルブックマークによる情報の鮮度を考慮した Web ページ評価手法”, Web とデータベースに関するフォーラム (WebDB Forum 2008), 2008.