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

A Method for Finding Related Pages by Users' Tagging Behavior from Social Bookmarks

杉山 典之♥ 関 洋平♦
青野 雅樹♠

Noriyuki SUGIYAMA Yohei SEKI
Masaki AONO

ある Web ページに興味を持ったとき，その Web ページと関連する Web ページを探したいという要求がある．このような場合，一般的な検索エンジンでは，ユーザは，目的とする Web ページに含まれていると思われる単語やフレーズを推測し，検索キーワードを入力しなければならない．これはユーザにとって負担となるため，Web ページをそのままクエリとして利用し，それに関連したページを検索する手法が求められている．

本稿では，近年注目を集めているソーシャルブックマーク(SBM)を利用し，関連ページを検索する手法を提案する．提案手法では，同じユーザは同じタグを，関連するページ群に付与する傾向があることを利用する．この手法を実装し，評価実験により有効性を検証したので，これを報告する．

There is an increasing demand for finding related Web pages when we encounter a Web page that makes us feel like knowing more about the page. However, with general-purpose search engines, users have to predict query terms and phrases that may be included in the target Web pages. Since this is a formidable task for users, a new-style search method is strongly expected, which makes it possible to retrieve related Web pages from a seed Web page.

We will focus on social bookmark (SBM) services, and propose a method for finding related pages by taking advantage of collective information from SBM members and their tags. Our proposed method relies on the observation that there is a tendency that the same users put the same tags to similar Web pages. We report the evaluation results of our proposed method.

1. はじめに

Web上に存在する情報は日々増大しており，ユーザが所望する情報を探すために，Web検索エンジンは必要不可欠となっている．ユーザが，あるWebページに興味を持ったとき，そのWebページと関連するWebページを探したいという要求がある．このような場合，一般的な検索エンジンでは，ユーザは，目的とするWebページに含まれていると思われる単語やフレーズを推測し，検索キーワードを入力しなければならない．これはユーザにとって負担となる．そこで，Webページをそのままクエリとして利用し，その関連ページを検索する手法が求められている．

近年，ソーシャルブックマーク（以下，SBM）サービスが注目を集めている．SBMに登録しているユーザは，ブックマークするWebページに対し，自由にタグを付与することができる．

本研究では，SBM上にある大量のタグ付けされたWebページ群から，関連ページを検索する手法を提案する．その際，同じユーザは同じタグを，関連するページ群に付与するという傾向を利用する．

以下，2章で関連研究，3章でSBMのユーザのタグ付与行動について述べ，4章でユーザとタグのペアの共起を用いた関連ページ検索手法を提案し，5章で評価と考察を行う．6章でまとめと今後の課題について述べる．

2. 関連研究

2.1 関連ページ検索

一般的なWeb検索エンジンは，ユーザから単語やフレーズをクエリとして受け取り，それらに関連したページを検索する．これに対し，関連ページ検索は，ページを検索質問とし，それに関連したページを返す検索方式である．

関連ページを検索する手法としては，Webページ間のハイパーリンクを用いた手法が数多く提案されている[1][3][5]．これらの手法は，ページを頂点，リンクを辺として有向グラフを作り，リンク構造を解析することで関連ページを発見する．基本的にはCo-citation[4]の考えに基づき，同じページからリンクされているページ群は関連があるとして，関連ページを求める．

また，共通の単語，特有の単語を用いて，兄弟カテゴリーのページを検索する研究も行われている[8]．

本研究もCo-citation分析に基づいているが，ハイパーリンクやページ中の単語を用いずに，SBMから得られるタグとユーザの両方の情報を用いて関連ページを求める点で異なっている．

2.2 SBM を利用した研究

SBMでは，ユーザはブックマークするWebページに対して自由にタグを付与することで，ブックマークを管理できる．

このタグ付け情報を利用した研究として，Webページ推薦システムを構築する研究[7][9]や，検索エンジンの検索結果をリランキングする研究[6]などが行われている．また，タグ付け情報を関連ページ検索に利用するサービスも存在する[11][12]．これらサービスは，ユーザの共起，タグの共起，ユーザとタグのペアの共起を用いていることが多い．本論文では，ユーザとタグのペアの共起を用いた新しい手法を提案する．

3. ユーザのタグ付けの傾向分析

3.1 ユーザ間で異なるタグ付与行動

前章で述べたように，SBMでは，ユーザはWebページに対して自由にタグ付けできる．そのため，ユーザの嗜好によって，タグの付け方が異なる．例として，ある「Webデザイン」

♥ 学生会員 豊橋技術科学大学大学院工学研究科修士課程 情報工学専攻 sugiyama@kde.ics.tut.ac.jp
♦ 非会員 豊橋技術科学大学情報工学系 seki@ics.tut.ac.jp
♠ 正会員 豊橋技術科学大学情報工学系 aono@ics.tut.ac.jp

に関するWebページに対して付与されたタグの上位 10 種類を表 1に示す.

表 1を見ると，ユーザによって“webデザイン”や“webdesign”といった，同じ意味でも表記の異なるタグを付与していることがわかる．さらには“*webデザイン”や“*webdesign”といった，先頭にアスタリスクのような記号を付けたタグもある．このような先頭に記号や数字を付けたタグは，ユーザ自身が良く使うタグを探しやすくするために，しばしば使われる．

また，言葉の多義性やユーザの嗜好により，同じタグでもユーザによって意味合いが異なることがある[9]. 例えばタグ“library”は，あるユーザにとっては「プログラミングにおけるライブラリ」であるが，別のユーザにとっては「図書館」を意味する．また，タグ“word”は，ユーザによって「名言」を意味したり，「Microsoft Word」のことを指したりする．

3.2 同じユーザの同じタグのブックマーク内容

表 1の「Webデザイン」に関するページには，“web”や“まとめ”といった，広い意味を持つタグが付与されている．これらのタグは，このページ以外にも，他の大勢のユーザが様々なページに付与している．単純にこれらのタグが付与されたページを探すと，「Webデザイン」に関係のないページも数多く見つかると考える．

ここで，表 1の「Webデザイン」に関するページに対しタグ“web”を付与したユーザ 12 人が，他にタグ“web”を付与したページを調べたところ，表 2のような結果になった．表 2に示す通り，12 人中 6 人は，「Webデザイン」に関するページにのみ，タグ“web”を付与していた．

比較のため，タグ“web”をよく使用している別のユーザ 10 人について調べたところ，表 3のような結果になった．「Webデザイン」に関するページを，タグ“web”で管理，分類しているユーザはいなかった．

このように，タグ“web”のような抽象度の高い曖昧なタグであっても，同じユーザが同じタグを付与したページ群を探すことにより，関連ページを見つけられると考える．

4. ユーザとタグのペアの共起を用いた関連ページ検索手法

本章では，前章で述べたユーザのタグ付与行動を踏まえた，関連ページ検索手法について述べる．

提案手法の具体例を図 1に示す．まず，ユーザUにとってのページP_1とP_2の関連度を式(1)で定義する．

$$J(U,P_1,P_2)=\frac{|Tags(U,P_1)\cap Tags(U,P_2)|}{|Tags(U,P_1)\cup Tags(U,P_2)|} \quad (1)$$

ここで，$Tags(U,P)$は，ユーザUがページPに付与したタグの集合を表す．式(1)は，ユーザUがページP_1 およびP_2 に付与したタグの集合間のJaccard係数である．

そして，ページP_1 とP_2 の関連度を表すスコア$R(P_1, P_2)$を式(2)のように与える．

$$R(P_1,P_2)=\frac{\sum_{U_i\in\{TaggedUsers(P_1)\cup TaggedUsers(P_2)\}}J(U_i,P_1,P_2)}{|TaggedUsers(P_1)\cup TaggedUsers(P_2)|} \quad (2)$$

ここで，$TaggedUsers(P)$は，ページPにタグを付与したユーザの集合を表す．式(2)は，0 以上 1 以下の値をとる．この値が高いほど，ページP_1との関連度が高いページと解釈する．図 1の例では，P_1の関連ページは，関連度が高い順にP_2, P_3となる．P_1にタグを付与しているユーザ群が，P_2にも全く同じようにタグを付与している場合，1 となる．

ここで，大勢のユーザがページP_1とP_2をブックマークしている場合を考える．このとき，付与されたタグが大きく異なっていたとすると，ページ間の関連度は低いと考えられる．そこで，ページP_1とP_2の両方にタグを付与したユーザ群のタ

表 1 「Web デザイン」に関するページに付けられたタグ上位 10 種類
Table.1 Top 10 tags for the page about Web design

タグ名	付与ユーザ数
web デザイン	40
css	39
まとめ	37
webdesign	32
デザイン	24
design	19
*web デザイン	17
web	12
リンク集	12
*webdesign	10

表 2 「Web デザイン」に関するページにタグ“web”を付与したユーザのタグ“web”のブックマーク内容
Table.2 Bookmarks tagged with “web” by users that used the tag “web” for a page about Web design

タグ“web”のブックマーク内容	ユーザ数
「Web デザイン」に関するページのみ	6
広義の Web に関するページ	6

表 3 タグ“web”をよく使用するユーザのタグ“web”のブックマーク内容
Table.3 Bookmarks tagged with “web” by users that used the tag “web” frequently

タグ“web”のブックマーク内容	ユーザ数
「Web デザイン」に関するページのみ	0
広義の Web に関するページ	10

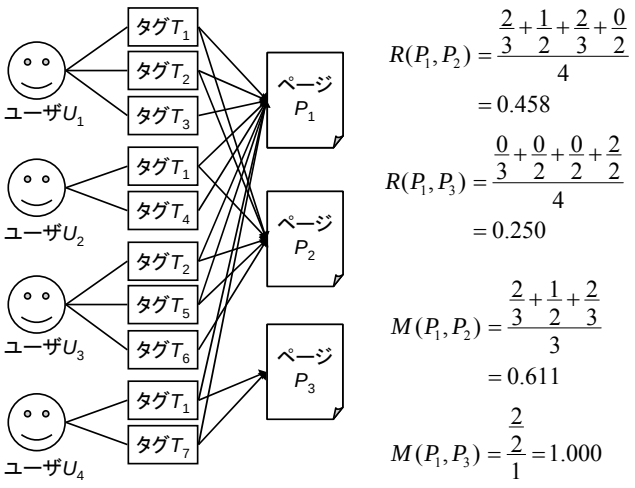

図 1 提案手法の具体例
Fig.1 An example of our method

グの共起率の平均値を式(3)により求め，

$$M(P_1,P_2)=\frac{\sum_{U_i\in\{TaggedUsers(P_1)\cap TaggedUsers(P_2)\}}J(U_i,P_1,P_2)}{\left|TaggedUsers(P_1)\cap TaggedUsers(P_2)\right|} \quad (3)$$

この値が閾値 M_θ より低い場合は，R=0 とし，関連ページ群から除外することとした．ただし，M_θ の値を高く設定すると，関連が強いページも除去される．事前に M_θ を変化させた実験を行ったところ，0.2～0.4 近辺で概ね良い結果が得られた．

5. 評価実験

5.1 データ収集

我々は，はてなブックマーク[10]からデータを収集した．はてなブックマークは 2007 年 12 月時点で，日本で最大規模のSBMサービスであり，10 万人以上のユーザが利用している．我々は，約 5 万人のユーザのデータと，それらユーザによってブックマークされたページ約 150 万ページを収集した．さらに，収集したユーザによってそれらのページに付与されたタグ約 20 万種類を収集した．

5.2 クエリページ

本研究では，クエリページのブックマークユーザの人数が，検索結果に影響を与えると考えられる．そのため，クエリページを選ぶ際に，ブックマークユーザの人数に応じて，以下のクラス分けを行った．

A) ブックマーク人数が 30 人未満
B) ブックマーク人数が 30 人以上 100 人未満
C) ブックマーク人数が 100 人以上 500 人未満
D) ブックマーク人数が 500 人以上

そして，それぞれのクラスからクエリページを 10 ページずつ設定した（表 4～表 7）．表中の人数は，実験を行った時点でのブックマークユーザの人数を表す．

5.3 比較手法

閾値M_θ=1/3 と設定した提案手法と，M_θ=0 と設定した提案手法（すなわち，式(3)の閾値による除去を行わない場合）を用意した．また，比較手法として以下の 2 つの手法を用意し，実験を行った．

タグのベクトルの類似度に基づく手法（タグ手法）

この手法では，ページに付与されたタグを使用し，ページに対してタグの特徴ベクトルを計算し，ページ間の関連度を式(4)のコサイン類似度で求める．

$$R_{Tags}(P_1,P_2)=\frac{V_{Tags}(P_1)\cdot V_{Tags}(P_2)}{\left|V_{Tags}(P_1)\right|\left|V_{Tags}(P_2)\right|} \quad (4)$$

ただし，

$$V_{Tags}(P)=\{rel(P,T_1),rel(P,T_2),\cdots,rel(P,T_n)\}$$

$$rel(P,T)=TF(P,T)\times IDF(T)$$

$$TF(P,T)=\frac{w(P,T)}{\sum_{T_i\in TAGS}w(P,T_i)}$$

表 4 クエリページ クラス A（30 人未満）
Table.4 Query Pages of Class A (less than 30 users)

	ページタイトルと URL	人数
A1	アスクル http://www.askul.co.jp/	28
A2	講談社 http://www.kodansha.co.jp/	26
A3	明治製菓 http://www.meiji.co.jp/	24
A4	セブン-イレブン・ジャパン http://www.sej.co.jp/index.html	22
A5	東京大学 http://www.u-tokyo.ac.jp/index_j.html	20
A6	リクナビ http://www.rikunabi.com/	17
A7	JR 東海 http://www.jr-central.co.jp/	15
A8	ニッセン http://www.nissen.co.jp/	13
A9	BANDAI http://www.bandai.co.jp/	9
A10	すき家 http://www.zensho.com/	7

表 5 クエリページ クラス B（30 人以上 100 人未満）
Table.5 Query Pages of Class B (30 to 100 users)

	ページタイトルと URL	人数
B1	三菱東京ＵＦＪ銀行 http://www.bk.mufg.jp/	78
B2	JAXA http://www.jaxa.jp/	70
B3	ココログ http://www.cocolog-nifty.com/	66
B4	三井住友 VISA カード http://www.smbc-card.com/	59
B5	フジテレビ http://www.fujitv.co.jp/index.html	53
B6	Sony Japan http://www.sony.co.jp/	50
B7	紀伊國屋書店 http://www.kinokuniya.co.jp/	38
B8	コクヨ http://www.kokuyo.co.jp/	38
B9	toyota.jp http://toyota.jp/	31
B10	総務省 http://www.soumu.go.jp/	30

表 6 クエリページ クラス C（100 人以上 500 人未満）
Table.6 Query Pages of Class C (100 to 500 users)

	ページタイトルと URL	人数
C1	Hatebu Friends http://www.kde.ics.tut.ac.jp/~sugiyama/sbm/hatebufriends.html	436
C2	アップル http://www.apple.com/jp/	362
C3	しょこたんぶろぐ http://yaplog.jp/strawberry2/	316
C4	世界中のインパクトあふれる映画みたいな景色の写真いろいろ-GIGAZINE http://gigazine.net/index.php?/news/comments/20061120_impact_scenery/	182
C5	nikkansports.com http://www.nikkansports.com/	182
C6	css Zen Garden：CSS デザインの美 http://www.csszengarden.com/tr/japanese/	172
C7	tenki.jp http://tenki.jp/	145
C8	Bloglines http://www.bloglines.com/?Lang=japanese	141
C9	NTT ドコモ http://www.nttdocomo.co.jp/	128
C10	任天堂 http://www.nintendo.co.jp/	102

表 7 クエリページ クラス D（500 人以上）
Table.7 Query Pages of Class D (500 or more users)

	ページタイトルと URL	人数
D1	東大で学んだ卒論の書き方★論文の書き方 http://staff.aist.go.jp/toru-nakata/sotsuron.html	2619
D2	Weblio｜辞書 http://www.weblio.jp/	2144
D3	Yahoo! JAPAN http://www.yahoo.co.jp/	1949
D4	ニコニコ動画 http://www.nicovideo.jp/	1473
D5	kizasi.jp hjttp://kizasi.jp/	1014
D6	CNET Japan http://japan.cnet.com/	808
D7	はてなブックマーク http://b.hatena.ne.jp/	794
D8	NIKKEI NET http://www.nikkei.co.jp/	679
D9	価格.com http://kakaku.com/	603
D10	Vector http://www.vector.co.jp/	519

$$IDF(T)=\log\frac{\sum_{P_j\in PAGES}\sum_{T_i\in TAGS}w(P_j,T_i)}{\sum_{P_j\in PAGES}w(P_j,T)}$$

である．ここで，$w(P,T)$はページPに付与されたタグTの数，$TAGS$は全てのタグの集合，$PAGES$は全てのページの集合を表す．$rel(P,T)$の式は，文献[7]を参考にした．

ユーザの共起を用いた手法（ユーザ手法）

この手法では，ページをブックマークしたユーザを用いて，ページ間の関連度を式(5)のJaccard係数で計算する．

$$R_{Users}(P_1,P_2)=\frac{|Users(P_1)\cap Users(P_2)|}{|Users(P_1)\cup Users(P_2)|} \quad (5)$$

ここで，$Users(P)$は，ページ P をブックマークしたユーザの集合を表す．

5.4 評価尺度

情報検索における評価尺度として，精度と再現率がよく用いられる．このうち再現率は，検索対象のデータ中に適合ページがいくつあるのかが分かっていなければならない．本研究における検索対象は，SBMから収集した約 150 万ページであるため，再現率を求めるためには，150 万ページ中に適合ページがいくつあるかを確認する必要があり，これは困難である．また通常，ユーザは検索結果の上位から見ていくため，適合ページがより上位に検索されることが望まれる．そのため評価尺度としては，上位ページほど重みが大きくなる DCG (Discounted Cumulative Gain)[2]を用いた．

$$DCG_i=\begin{cases}G_1 & if\ i=1\\ DCG_{i-1}+\dfrac{G_i}{\log_2 i} & otherwise\end{cases}$$

ここで G の値は，各順位の検索結果の適合度の高さによって，多値を取ることができる．

今回は，検索結果の Web ページ上位 20 件に対し，適合，不適合の判定を，3 名の評価者に行ってもらった．判定基準は以下の通りである．

・ 適合　…クエリページと関連がある
・ 不適合…クエリページと関連が見られない
　または，ページにアクセスできない

そして，適合と判定した評価者の人数を G の値とした．

5.5 実験結果

5.3 節で述べた各手法の検索結果のDCGを，図 2～図 5に示す．各グラフは，横軸は検索結果の順位，縦軸は 10 個のクエリに対する各検索結果の順位における，DCGの平均値を示す．「理論的上限」は，検索結果上位 20 件を 3 名の評価者が 3 名とも適合と判定した場合の，理論的上限値を示す．

D1 をクエリページとした場合の，閾値を設定しない提案手法（M_θ=0）の検索結果例を表 8に示す．また，A7 および

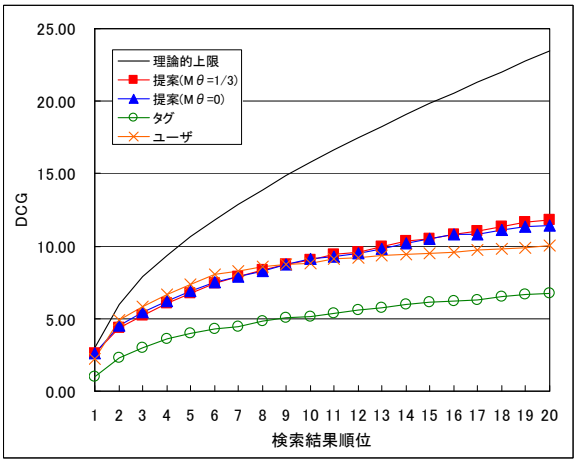

図 2 クラス A の DCG の平均
Fig.2　Mean DCG for Class A

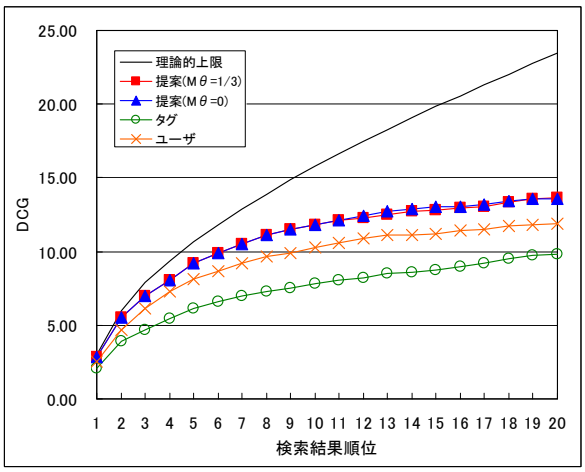

図 3 クラス B の DCG の平均
Fig.3　Mean DCG for Class B

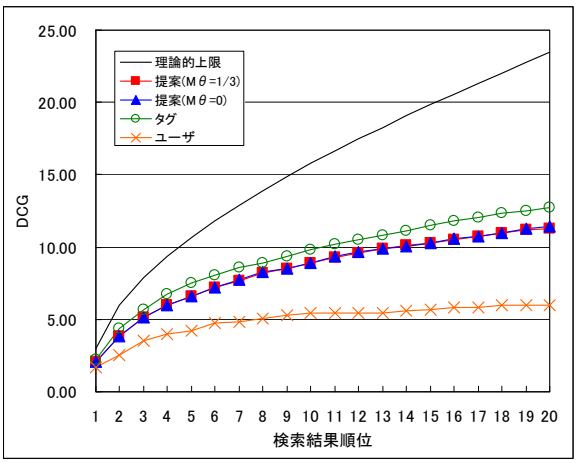

図 4 クラス C の DCG の平均
Fig.4　Mean DCG for Class C

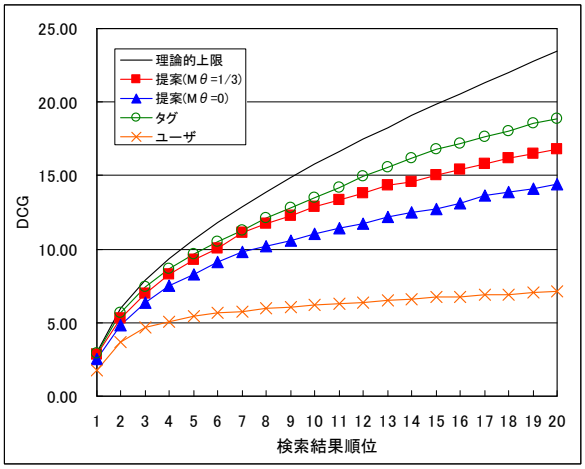

図 5 クラス D の DCG の平均
Fig.5　Mean DCG for Class D

D1 をクエリページとした場合の，各手法の検索結果上位 3 件を表 9，表 10に示す.

5.6 考察

(a) 提案手法の特性

図 2～図 5の結果から，提案手法は，タグ手法，ユーザ手法と比べ, 比較的, ブックマークユーザの人数に影響を受けず, 安定して関連ページを検索することができていることがわかる.

図 2，図 3より，クエリページのブックマーク人数が少ない場合では，提案手法はタグ手法と比べ，良い結果となった.

また図 4，図 5より，クエリページのブックマーク人数が多い場合では，提案手法はユーザ手法と比べ，良い結果となった．タグ手法にはやや劣っているが，Mの閾値M_0=1/3 と設定することにより，結果を改善できていることがわかる．表 8に示すように，ブックマーク人数が多いページをクエリにした場合，閾値を設定しない提案手法（M_0=0）は，複数の評価者から関連していないと判定されたページが上位に来ている．しかし，M_0=1/3 と設定することにより，関連していないページを除去することができた.

図 2より，ブックマーク人数が少ないページをクエリにした場合は，提案手法はユーザ手法と比べ，検索結果上位の精度が低くなっている．これは，関連のあるページ群をブックマークしているユーザが，それらのページ群にタグを付与していないことがあったため，提案手法では，関連が弱いと解釈されたことが原因である.

図 4，図 5より，ブックマーク人数が多いページをクエリにした場合は, 提案手法はタグ手法に比べ, 精度が劣っている. これは，クエリページのブックマーク人数が多い場合には，それぞれのユーザが複数の関連ページに対し，同じようにタグを付与していることが少ないためと考える.

SBMでは，同じ意味で表記の異なるタグが複数ユーザ間で付与されることがある[9]．しかし提案手法では，[9]と同様に，同じユーザの同じタグが付与されたページの集合を利用するため，ユーザ間のタグ名の揺らぎは問題にならない.

(b) 比較手法の問題点

タグ手法は, ブックマーク人数が少ないページをクエリにした場合には，検索精度を大きく低下させている（図 2，図 3）．この原因として，ページに付与されたタグの数や種類が少ないために，ページの特徴を，タグのベクトルで上手く表現できなかったことが考えられる．実際，クエリページA9 では，主にタグ“企業”が付与されていたため，単にタグ“企業”が 1 つ付与されたページとの関連度が高くなり，検索精度が低下していた．またA7 には，主にタグ“旅”，“交通”や“鉄道”が付与されていたが，それらのタグが単に 1 つずつ付与されたページが上位に検索されてしまっていた.

ユーザ手法は, ブックマーク人数が多いページをクエリにした場合に, 関連していないページが多く検索された（図 4，図 5）．これは，ユーザの共起のみを用いたために，SBMでよくブックマークされる人気のページや定番ページが上位に検索されたためである．実際，クエリページD1 は，D2 や D4 と関連がないと全ての評価者から判定されたが，お互いに検索結果中で上位にランクされていた.

(c) 計算量の比較

クエリページに対して関連ページを求める計算量について考察する．ここで，ページ総数を P，ユーザ総数を U，タグ総種類数を T とする.

提案手法では, クエリページにタグを付与したユーザ群を調べ, それらユーザ群がタグを付与したページ群を調べることで関連ページを求めるので，計算量は $O(UP)$となる．実際には，ほとんどのユーザは多くのページにタグを付与していないため，$O(U)$に近くなる．ユーザ手法も同様である.

タグ手法では，クエリページに付与されたタグ群を調べ, それらのタグが付与されたページ群を調べることになるので，計算量は $O(TP)$となる．実際には，タグ-ページ行列はスパースなため，$O(T)$に近くなる.

5.1 節より $T > U$ であることから，提案手法はタグ手法よりも少ない時間で検索を行うことができると言える.

表 8　閾値を設定しない提案手法（M_0=0）の検索結果例
クエリページ D1（東大で学んだ卒論の書き方★論文の書き方）に対する検索結果上位 10 件
Table.8　Top 10 results of our method for the query page D1

順位	R	ページタイトルと URL	人数	M	判定
1	0.024	論文の書き方　http://www.csg.is.titech.ac.jp/~chiba/writing/	212	0.44	◎
2	0.017	プレゼンハック ～プレゼン改善のための 10 個の小技～	IDEA*IDEA http://www.ideaxidea.com/archives/2005/09/_10.html	869	0.15	△
3	0.016	参照文献の書き方（Description of Bibliographic References） http://www.lib.hit-u.ac.jp/service/guide-j/sanshobunken.html	182	0.34	◎
4	0.014	正しい技術文章作成のためのヒント http://www.ispl.jp/~oosaki/research/tips-japanese/	204	0.35	○
5	0.013	議論のしかた　http://iwatam-server.dyndns.org/software/giron/giron/	662	0.16	×
6	0.012	論文の読み方、探し方 - Horiguchi-Abe lab. FAQ: Reading papers (In Japanese) http://mitsuko.jaist.ac.jp/hori-abe-lab/etc/semi/index-j.html	66	0.59	△
7	0.011	中学生レベルの英語力の奴が 4 ヶ月で TOEIC「B クラス」を出す方法 (b) – log http://d.hatena.ne.jp/bambix/20070312/1173628642	1164	0.08	×
8	0.011	会社で使える文例集　http://www.aimcom.co.jp/newstool/IT21/	469	0.18	×※
9	0.010	Eテキスト「レポートの書き方」 http://www2.dokkyo.ac.jp/~msemi008/index2/e_text/	130	0.40	×※
10	0.010	小説（文章）を書くのに役に立つかもしれない豆知識 http://homepage1.nifty.com/hiroyuki/doc/sentences.html	274	0.23	×

判定: ◎…3 人，○…2 人，△…1 人，×…0 人の評価者が適合と判定　※ページにアクセスできなかった（404 Not Found）
（M の値が下線付きの検索結果は，閾値を設定した提案手法（M_0=1/3）では除去される）

表 9 クエリページ A7（JR 東海，ブックマーク人数 15 人）に対する検索結果上位 3 件
Table.9 Top 3 results of our method and comparative methods for the query page A7

順位	提案手法（M_θ=1/3）		タグ手法		ユーザ手法	
	ページタイトル	判定	ページタイトル	判定	ページタイトル	判定
1	JR 東海	◎	きっぷあれこれ –どこなびドットコム	△	JR 西日本	◎
2	小田急	◎	Hiro's Ticket Web -きっぷ総合サイト-	△	JR 東日本	◎
3	JR 北海道	◎	ＪＲ九州 SL58654 号の復活について	×	JR 北海道	◎

表 10 クエリページ D1（東大で学んだ卒論の書き方★論文の書き方，ブックマーク人数 2619 人）に対する検索結果上位 3 件
Table.10 Top 3 results of our method and comparative methods for the query page D1

順位	提案手法（M_θ=1/3）		タグ手法		ユーザ手法	
	ページタイトル	判定	ページタイトル	判定	ページタイトル	判定
1	論文の書き方	◎	論文の書き方	◎	Windows XP の動作を軽快にしたい - mtblue.org	×
2	参照文献の書き方	◎	科学技術論文の書き方	◎	KANOU.JP: 良質な教科書系ウェブサイト集	△
3	正しい技術文章作成のためのヒント	○	研究レビューの書き方「超」序論	◎	ウノウラボ Unoh Labs: 海外経験のない典型的理系人間が日常会話レベルの英語を話せるようになるまでの道のり	×

6. おわりに

SBM のユーザとタグの共起情報を利用し，関連ページを検索する手法を提案した．実験の結果，提案手法はタグのベクトルやユーザの共起を用いた手法と比べ，検索精度へのブックマークユーザの人数の影響が少ないことがわかり，有効性が確認できた．

今後の課題としては，タグを付与したブックマークユーザの人数が少ないクエリページにおける検索精度の向上，複数のクエリページの入力への拡張や，複数の側面を持つクエリページに対し，ユーザにとって有益な関連ページをインタラクティブに提示する方法の検討が挙げられる．

【文献】

[1] J. Dean and M. R. Henzinger, “Finding Related Pages in the World Wide Web,” Proc. of the 8th International World Wide Web Conference (WWW8), Toronto, Canada, May 1999.

[2] K. Järvelin and J. Kekäläinen, “IR Evaluation Methods for Retrieving Highly Relevant Documents,” Proc. of the 23rd Annual International ACM SIGIR Conference (SIGIR 2000), pp.41-48, July 2000.

[3] G. Jeh and J. Widom, “SimRank: A Method of Structural-Context Similarity,” Proc. of the 8th ACM SIGKDD International Conference on Adaptive Hypermedia and Adaptive Web-based Systems, pp.538-543, Edmonton, Canada, July 2002.

[4] H. Small, “Co-citation in the Scientific Literature:A New Measure of the Relationship Between Two Documents,” Journal of the American Society for Information Science, vol.24, pp.265-269, 1973.

[5] M. Toyoda and M. Kitsuregawa, “Creating a Web Community Chart for Navigating Related Communities,” Proc. of the 12th ACM Conference on Hypertext and Hypermedia, pp.103-112, Denmark, August 2001.

[6] Y. Yanbe, A. Jatowt, S. Nakamura and K. Tanaka, “Can Social Bookmarking Enhance Search in the Web?,” Proc. of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2007), Vancouver, Canada, June 2007.

[7] 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.

[8] 大島裕明，小山聡，田中克己，“文書群を問合せとした兄弟カテゴリー文書の検索,” 電子情報通信学会論文誌 D, Vol.J90-D, No.2, pp.196-208, 2007.

[9] 佐々木祥，宮田高道，稲積泰宏，小林亜樹，酒井善則，“Social Bookmark におけるコンテンツクラスタ間の類似度を用いた web コンテンツ推薦システム,” 情報処理学会論文誌:データベース，Vol.48, No.SIG20(TOD36), pp.14-27, 2007.

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

[11] HatenaTail, http://hatenatail.com/

[12] similicio.us, http://www.similicio.us/

杉山 典之 Noriyuki SUGIYAMA

2008 年豊橋技術科学大学大学院工学研究科修士課程情報工学専攻修了．Web マイニングの研究に従事．日本データベース学会学生会員．

関 洋平 Yohei SEKI

豊橋技術科学大学情報工学系助教. 2005年総合研究大学院大学複合科学研究科博士課程修了．博士（情報学）．自然言語処理, 文書要約, 多言語意見分析の研究に従事. ACM, ACL, 情報処理学会，言語処理学会，電子情報通信学会各会員.

青野 雅樹 Masaki AONO

豊橋技術科学大学情報工学系教授．1984 年東京大学理学系大学院情報科学専攻修士課程修了．Ph.D（レンセラー工科大学）．データマイニング，情報検索，マルチメディアデータ処理，情報視覚化などの研究に従事．ACM，IEEE，情報処理学会，日本データベース学会，言語処理学会，人工知能学会，電子情報通信学会各会員.