ソーシャルブックマークにおける ユーザ間の類似度を考慮したスパマー検出

渡邊　桂太† 　高橋　翼†† 　北川　博之††,†††

† 筑波大学第三学群情報学類　〒 305–8573 茨城県つくば市天王台 1–1–1
†† 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: †,††{ein_vogel,tsubasa}@kde.cs.tsukuba.ac.jp, †††kitagawa@cs.tsukuba.ac.jp

あらまし　Web の情報量の肥大化に伴い, ユーザの嗜好や興味を反映した Web 検索サービスの必要性が強く認識されている. そこで近年では, ソーシャルブックマークのブックマーク情報が注目を集めており, これを有益な資源とみなした Web 研究が盛んである. しかし, ソーシャルブックマークでは容易にブックマーク情報を生成できることから, 悪意あるユーザ (スパマー) が容易にスパム行為を行えるといった特徴があり, 近年スパマーの増加が問題となっている. そこで本研究では, 学習セットを用いないクラスタリングの手法を基に, ソーシャルブックマークのユーザ間の類似度から, 他のユーザと類似しないスパマーを検出する手法を提案する. 実験では, ソーシャルブックマークの実データを用い, 手法の有効性を検証する.

キーワード　ソーシャルブックマーク, Web2.0, スパム検出, クラスタリング

Detection Spammers using Similarity among Users in Social Bookmarking Systems

Keita WATANABE†, Tsubasa TAKAHASHI††, and Hiroyuki KITAGAWA††,†††

† College of Information Sciences, University of Tsukuba　Tennoudai 1–1–1, Tsukuba City, Ibaraki, 305–8573 Japan
†† 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: †,††{ein_vogel,tsubasa}@kde.cs.tsukuba.ac.jp, †††kitagawa@cs.tsukuba.ac.jp

Abstract　With the information flood, we need the web search services that take into account interests of users. In recent years, Social Bookmark is attracting more attention from Web researchers. Since we can easily develop contents in Social Bookmark. Social Bookmark is a soft target for spammers. In this paper, we propose a method to detect spammers using similarity-based cluster analysis. Through the experiment using the real social bookmark data, we verify usefulness of the proposed approach.

Key words　Social Bookmark, Web2.0, Spam Detection, Cluster Analysis

1. はじめに

近年, Blog や SNS などのサービスの浸透により, 誰もが容易に Web コンテンツを生成できるようになった. これに伴い, Web に情報を発信するユーザ人口は飛躍的な増加を遂げたが, それ以上に Web コンテンツの蓄積量が爆発的な増加を遂げた. 蓄積される情報の中には, ユーザの嗜好や興味を反映した情報も多く含まれている. その中でも特に, ソーシャルブックマークが近年注目を集めている. ソーシャルブックマークは, ブラウザに搭載されているブックマーク機能をオンライン上に実現したもので, Web ページを体系化し共有・管理することができる. ブックマークされた Web ページは, ユーザによって精査され, 評価を受けた Web ページである. よって, ソーシャルブックマークに登録されている Web ページを, ユーザの嗜好を反映

した有益な資源とみなし，これを用いた新しい Web 検索手法や情報抽出手法などの研究が近年盛んである [4], [6], [7].

しかし，ソーシャルブックマークではブックマーク情報の生成が簡単なため，容易に悪質なブックマーク情報を発信することが可能といった問題がある．悪意あるユーザ (スパマー) にとって，ソーシャルブックマークはスパム行為を行う格好の場であり，スパマーの増加が近年問題となっている．ソーシャルブックマークでは，スパム行為と，スパム行為を行う根源であるスパマーの検出が重要な課題である [3].

ソーシャルブックマークを利用する一般のユーザは，良質な Web ページ，自身の興味を反映した情報を提供する Web ページをブックマークする傾向がある．良質なものと評価された Web ページは一般的に，多くのユーザから支持を受け，また，ユーザの興味のある Web ページは，類似する興味を持つ他のユーザからも支持を得られる可能性がある．一方，スパマーが高い割合でブックマークする Web ページは，一般のユーザからの支持を得ることは難しく，そのページをブックマークしているユーザが極端に少ないと考えられる．

そこで本稿では，ユーザのブックマーク情報を基にユーザ間の類似度を求め，これを用いて ソーシャルブックマークのユーザ群からスパマーを検出する手法を提案する．本手法では，ユーザ群にクラスタリングを施し，クラスタが形成されていない，もしくは極端に規模の小さいクラスタで形成されているユーザをスパマーとして検出する．

2. 前　　提

2.1 ソーシャルブックマーク

Web が普及し始めていた当初より，Web ブラウザにはブックマークという機能が搭載されていた．ブックマークとは，任意の Web ページの URL を，Web ブラウザに保存・管理する機能である．ブラウザを利用するユーザは，定期的または頻繁に利用する Web ページを保持や それらページへのアクセスを簡易化するためにブックマークを用いる．そのため，ブックマーク情報は，ユーザの一定の評価を受けた有益なページであると言える．しかし，このブックマーク情報は，Web ブラウザ上で管理されるため，他のユーザのブックマーク情報を知ることは困難である．

一方，ソーシャルブックマーク (以下，SBM) という，ブックマーク機能をオンライン上で実現した Web サービスが，近年注目を集めている．2003 年に，Joshua Schachter らによって生み出された SBM サービス，delicious [10](以前は del.icio.us) の運用開始によって，SBM は大きな話題を呼んだ．2005 年には，現在の日本において最大の SBM サービスである，はてなブックマーク [9] がリリースされた．SBM は，サービス開始当初は，認知度は高くなく，IT に関連したトピックに興味のある，一部のコミュニティの人々を中心に利用されていたが，徐々に一般的なトピックに関心のある人々にも浸透し，多様性を有する情報源へと成長し続けている [8].

SBM は，ブラウザ上のブックマーク機能とは異なる大きな特徴をいくつか持つ．1 つに，他のユーザとブックマーク情報を共有できることが，大きな特徴として挙げられる．ブックマーク情報の共有により，他のユーザのブックマーク情報の閲覧を可能になり，また多数のユーザから評価を受けている Web ページを知ることが可能となった．ある Web ページは，複数のユーザによってブックマークされていることがあるため，SBM では，Web ページの評価指標として，そのページをブックマークしているユーザの数が，しばしば用いられることがある．一般的にユーザは，他の多くのユーザに支持を受けている良質な Web ページや，自身の興味を反映した，一部のコミュニティで支持されている Web ページをブックマークする傾向がある．そのため，趣味や興味が類似しているユーザ同士では，ブックマークする Web ページが共通することが考えられる．もう 1 つの特徴として，ブックマーク情報をフォークソノミーの概念に基づいて分類することが可能であることが挙げられる．フォークソノミー (folksonomy) とは，folk と taxonomy を組み合わせた造語であり「万人による注釈」という意味を持つ．SBM におけるフォークソノミーに基づく分類とは，ブックマークした Web ページに対して，タグと呼ばれる自由記述のキーワードを付与し，ブックマーク情報を非階層的に管理することである．このような SBM を図 1 のようなグラフ構造で考えられる．このとき，ユーザがページをブックマークするという振舞はエッジで，ページに付与するタグの集合はエッジに付属する文字列の集合で表現する [7].

また SBM では，ブックマーク情報はユーザ毎に Web ページの形で生成される (図 2)．このブックマーク情報には，ユーザがブックマークしたページへのハイパーリンクと，そのページをブックマークしているユーザの数，またページに付与したタグなどの情報が参照できる．

SBM は，ブックマーク情報の生成が簡単なため，容易に悪質なブックマーク情報を発信することができるという問題がある．そのためソーシャルブックマークは，悪意あるユーザ (スパマー) にとってスパム行為を行う格好の場となってしまい，スパマーの増加が近年問題となっている．ソーシャルブックマークでは，スパム行為の抑制，スパマーの検出が重要な課題であると言える．

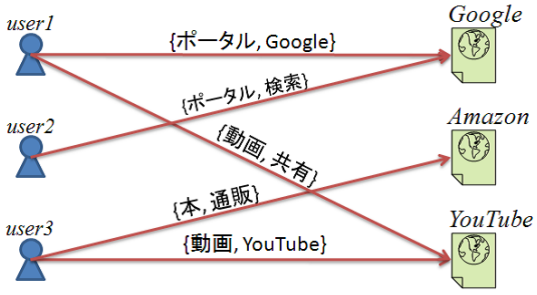

図 1　ソーシャルブックマークの構造

図 2　ブックマーク情報の例

2.2　SBM におけるスパマー

現在 SBM は, Web を構成する重要なコンポーネントの 1 つとなり, 多くのユーザを獲得した. しかしその中には, SBM を悪用しようと考えるユーザも存在する. この悪意あるユーザをスパマーと呼び, 近年の SBM では, このスパマーの増加が問題となっている.

2.2.1　スパム行為

SBM におけるスパム行為は大きく 2 つに分類できる [2], [5]. 本稿では, この 2 つのスパム行為を, スパムタギングとスパムブックマークと呼ぶ.

スパムタギングとは, 通常の Web ページに対して不適切なタグを付与し, ページのコンテンツに悪い印象を与える行為である. 例えば, あるプロ野球選手のプロフィールや戦績を掲載したページに対して「ドーピング」などのタグを付与する. しかし, この選手はドーピングの経験がない場合「ドーピング」というタグは不適切である. このタグを見たユーザは, そのドーピング経験のない選手に対して「ドーピング経験のあるプロ野球選手」と, 誤った認識をしてしまう可能性がある.

スパムブックマークとは, 不当なページをブックマークする行為である. 不当なページは, フィッシングページや過剰に広告を掲載しているアフィリエイトページだけでなく, 生成されるブックマーク情報から得られるハイパーリンク欲しさに SEO(Search Engine Optimization) 目的にブックマークされるページも考えられる.

本研究では, 後者のスパムブックマークを高い割合で行うスパマーを検出対象とする.

2.2.2　スパマーの特徴

本来, ブックマークを受ける Web ページとは, ユーザのお気に入りの Web ページであるため, 良質なものであれば多くのユーザからブックマークされる可能性がある. また, 特定の分野の情報を提供している局所性の高い Web ページにおいても, その情報に高い興味を示す複数のユーザによってブックマークされることが考えられる. 一方, スパマーのブックマークするページは 高い割合で他のユーザからの支持を得られないため, これら不当なページをブックマークしているユーザの数は, 極端に少ないという特徴が挙げられる. SBM において, スパムブックマークを受けるページは 「一般のユーザがほとんど興味を示さないページ」であると考えられる.

一般ユーザは, 多くユーザがブックマークしているページ, 自身の興味を反映したページをブックマークする傾向があることから, 同じ興味を持つユーザとブックマークするページが共通しやすいということが考えられる. スパマーは, 一般のユーザの偽装を目的に低い割合で人気のページをブックマークする者もいるが, 基本的には一般ユーザが参照しない不当なページを高い割合でブックマークする. そのため, 他のユーザとブックマークするページがほとんど共通しないということが考えられる. 以上より, 一般ユーザとスパマーはブックマークするページが高い割合で共通しないという特徴が挙げられ, スパマーは「誰とも類似しないユーザ」であると考えられる. 本研究では, この特徴を用いてスパマーを検出する手法を提案した.

3.　提 案 手 法

本章では, クラスタリングの手法に基づいた SBM におけるスパマーの検出手法を提案する. 特に文書データの分類に利用されるクラスタリングの手法に着目し, この手法を SBM のユーザとユーザのブックマーク情報に対応させたものを提案する.

文書データのクラスタリングは, 文書データ間の語の共起性や出現頻度から類似度を求め, グループ分けをする手法である. 本研究では, 文書データをユーザ, 文書中に含まれる語をユーザのブックマークしている Web ページ集合とみなし, クラスタリングを行い, スパマーを検出する手法を提案する. 本手法のタスクは, ユーザ間の類似度計算と, クラスタ解析によるスパマーの検出の 2 フェーズから成り, 各フェーズの詳細は 3.1, 3.2 で述べる.

3.1　ユーザ間の類似度の算出

文書データ間の類似度の算出に, コサイン類似度がしばしば利用される. コサイン類似度とは, 2 つのベクトルのなす角度であり, 文書をベクトル化することにより, 文書間の類似度を求める手法である (図 3). 本研究では, ユーザがブックマークしている Web ページの集合をベクトルとみなすため, コサイン類似度の算出が可能である. しかし SBM では, ユーザは同じ Web ページを複数ブックマークすることができないため, 文書データのように語の出現頻度は考慮できない.

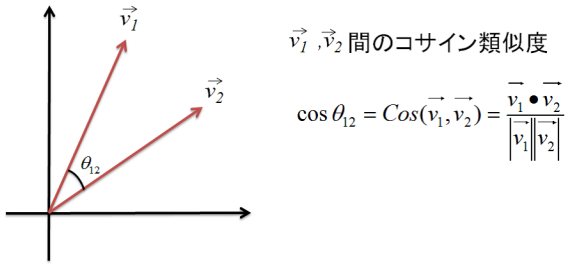

図 3　コサイン類似度

SBM におけるユーザ u_i, u_j 間のコサイン類似度 $Cos(u_i, u_j)$ は, 以下のようになる. このとき, ユーザ u がブックマークしている Web ページ集合を $pages(u)$ とする.

$$Cos(u_i, u_j) = \frac{|pages(u_i) \cap pages(u_j)|}{\sqrt{|pages(u_i)||pages(u_j)|}} \quad (1)$$

上述のようにコサイン類似度では, SBM を利用するユーザ間の類似度を, ブックマークしている Web ページの数と, 共起している Web ページの数で決定する.

しかしこの手法では, ブックマークされている Web ページの重みを考慮していないため, 図 4 のような場合には問題が発生する. このとき, 一般ユーザである *user*3 と *spammer* との間で高い類似度が得られてしまい, *spammer* の特定が困難になってしまう (表 1). この例では, *spammer* が一般のユーザを偽装する目的でブックマークしている一般のページを, *user*3 が完全に重複してブックマークしているため, このような問題が生じてしまう. この原因の 1 つとして, ページが全て同じ重みで評価されていることが考えられる.

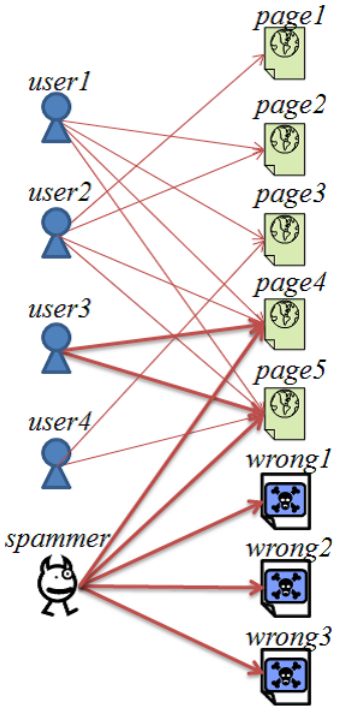

図 4　SBM の例

表 1　図 4 でのユーザ間のコサイン類似度

	*user*1	*user*2	*user*3	*user*4	*spammer*
*user*1	×	0.750	0.707	0.707	0.447
*user*2	0.750	×	0.707	0.354	0.447
*user*3	0.707	0.707	×	0.500	0.632
*user*4	0.707	0.354	0.500	×	0.316
spammer	0.447	0.447	0.632	0.316	×

そこで本研究では, 特徴的な Web ページを共通にブックマークしているユーザ間には, 高い類似度を与える計算法を提案する. 特徴的な Web ページは, 興味や嗜好が特徴的な限られたユーザからブックマークを受けていると考えられる. そこで本研究では, ある Web ページの特徴の度合をそのページをブックマークしているユーザの数から算出し, それを重みとして付加する. これは, TF-IDF 法における大域的重み *idf* (Inverse Document Frequency) の考えに基づくものである. 大域的重み *idf* は, 多くの文書に出現する語は, 一般的に利用される語と考え低い重みを, 特定の文書のみに偏って出現する語は, 特徴的な語と考え, 高い重みを与えるというものである. 本研究では, 多くのユーザからブックマークを受けている Web ページには低い重みを, 特定のユーザのみからブックマークを受けている Web ページには高い重みを付加する大域的重み ibf(Inverse Bookmark Frequency) を提案する. Web ページ p をブックマークしているユーザの集合を, $users(p)$ としたときに, Web ページ p の重み $ibf(p)$ は, 以下のように算出する.

$$ibf(p) = \frac{1}{\log(|users(p)|)} \tag{2}$$

$ibf(p)$ の変化を少なくするために, 分母は, p をブックマークしているユーザ数に対数を施した.

そこで, ibf をコサイン類似度に適用した $IbfSim$ を以下のような算出法で提案する.

$$IbfSim(u_i, u_j) = \frac{\sum_{p \in (pages(u_i) \cap pages(u_j))} ibf(p)}{\sqrt{|pages(u_i)||pages(u_j)|}} \tag{3}$$

$IbfSim$ では, ブックマークしているユーザ数が少ない Web ページを, 共通にブックマークしているユーザ間に高い類似度を与えるといった特徴がある.

$IbfSim$ を図 4 に適用した結果, コサイン類似度で高い類似度を示していた *user*3 と *spammer* との間で, 低い類似度が得られた (表 2). 図 4 では, *spammer* は人気のページをブックマークして一般のユーザを偽装しているため, ページの局所性を考慮した本手法は有効である.

表 2　図 4 でのユーザ間の $ibfsim$

	*user*1	*user*2	*user*3	*user*4	*spammer*
*user*1	×	0.696	0.474	0.730	0.300
*user*2	0.696	×	0.474	0.220	0.300
*user*3	0.474	0.474	×	0.310	0.346
*user*4	0.730	0.220	0.310	×	0.196
spammer	0.300	0.300	0.346	0.196	×

3.2　クラスタ解析とスパマーの検出

前節で定義した提案類似度 $IbfSim$ を用い, 一般のユーザとスパマーを分類するために, ユーザにクラスタ解析を施す. 本節では, 情報の組織化に優れている凝集型の階層的クラスタリングの手法を用いる. 凝集型は, 解析対象のユーザ数 = クラスタ数を初期状態とし, 最も類似するクラスタを逐次併合する手法である. 一般ユーザは, 他の多くの一般ユーザと類似するため, クラスタ併合を重ねる毎に, 大きなクラスタが形成されると考えられる. 本研究では, スパマーを「誰とも類似しないユーザ」としているため, 併合を重ねてもスパマーは, 一般ユーザのクラスタと併合されず, 極端に小規模なクラスタで形成されると考えられる.

以上のような特徴から, スパマーを検出する手法を提案する. 本手法では, クラスタ併合の終了条件として併合類似度の閾値 σ_s を設定し, さらに, スパマーのクラスタを決定するために, 1 クラスタに含まれるユーザ数に閾値 σ_u を設定する. クラスタの併合類似度が σ_s 以下に達した時点で形成されるクラスタから, ユーザ数が σ_u 以下のクラスタに含まれているユーザをスパマーと判別する.

図 5 に本手法のスパマー検出の例を示す. $\sigma_s = \sigma_{s0}$, $\sigma_u = 2$ と設定したとき, *Cluster*3, *Cluster*4, *Cluster*5 がスパマーとして検出される.

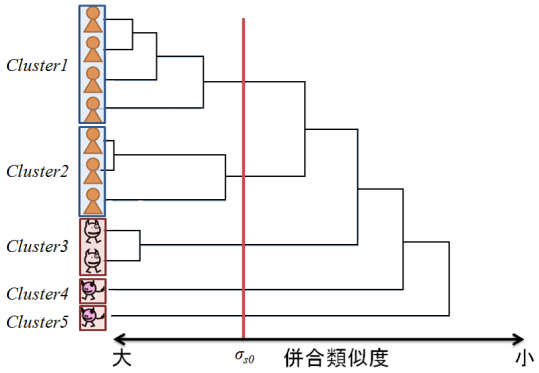

図 5　スパマー検出の例 (閾値 $\sigma_s = \sigma_{s0}, \sigma_u = 2$ に設定)

4. 実　　験

提案手法の有用性を判断するために, 実験を行った. 本実験では, 日本最大の SBM サービス「はてなブックマーク」[9] を対象にしたクローリングシステムを開発し, SBM の実データを収集した. 本実験では, 収集したデータを用いてスパマー検出の正確性を検証する. 実験に使用する SBM のデータ数を, 以下に示す (表 3).

表 3　実験データ

ユーザ数	ブックマーク対象の Web ページ数
1496	597574

本実験では, クラスタ併合回数ごとに, 1 ユーザで構成されるクラスタをスパマーとして検出する. 検出されるユーザが, 実際にスパマーであるかの判別は手作業で行う. 手作業による判定には, 作業量的限界があるため, 総検出数が 30 以下の場合のみとする. 今回, クラスタの併合に単連結法と完全連結法を用いたが, 完全連結法では判定が可能な結果が得られなかった. 完全連結法は, 2 つのクラスタ間で各々のクラスタから 1 つずつ選んだデータ間において, 最も小さい類似度をクラスタ間の類似度として扱う手法である. 今回の実験でクラスタ解析を行うデータは, SBM におけるユーザ間のブックマーク情報の共通性から算出される類似度である. SBM では, 一般ユーザ間であってもブックマーク情報が全く共通しない場合が多々ある. そのため, 完全連結法では, クラスタ併合を重ねてクラスタの規模が大きくなるほど, 2 つのクラスタ間に各々含まれるユーザの間で, 全く類似しないユーザが含まれる可能性が高くなる. これにより, 少ない併合回数で全クラスタ数が 1 つに収束してしまうため, 本実験では, 完全連結法によるクラスタ解析は, 不適切であることが判明した. 本章では, 単連結法によるクラスタ解析で得られたスパマーの検出結果を示す (表 4).

単連結法の結果から, 適合率と再現率 (図 6) を算出し, スパマー検出の正確性と網羅性を測定した. 再現率は, 今回の実験で検出したスパマーの最大数 12 をスパマーの総数として考え, 算出した.

本実験での適合率と再現率の定義は以下の通りである.

$$\text{適合率 } (Precision) = \frac{\text{検出スパマー数}}{\text{総検出数}} \tag{4}$$

$$\text{再現率 } (Recall) = \frac{\text{検出スパマー数}}{\text{総スパマー数}} \tag{5}$$

表 4　検 出 結 果

併合回数	総検出数	検出スパマー数
1057	30	12
1060	25	12
1070	18	10
1080	14	10
1090	12	10
1100	10	10
1110	7	7
1120	7	7

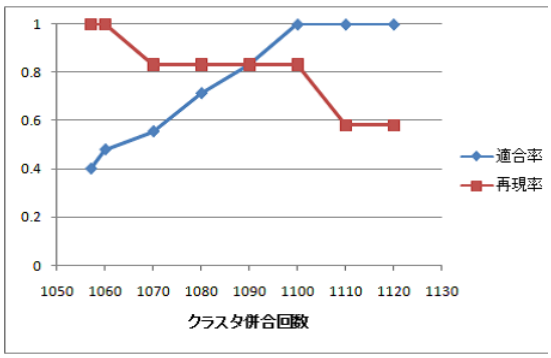

図 6　適合率と再現率

結果より, クラスタ併合を重ねるごとに, 高い適合率が得られることが確認できた. 特に, クラスタ併合回数が 1100 回以上の場合に, 100%の適合率が得られたことから, 本手法は, 併合回数を多くすることで一般ユーザが, スパマーとして誤検出されることを回避できると考えられる. しかし, 併合回数が多くし高い適合率を得られる一方で, 検出されるスパマーの網羅性が失われていることが確認できた. より多くのスパマーを確実に検出するためには, クラスタ併合回数がより少ない段階で, 高い適合率を得る必要がある.

本研究では, ブックマークの共起の局所性を考慮したユーザ間の提案類似度 $IbfSim$ を基に, クラスタ解析を用いてスパマー

を検出する提案手法が, 十分にスパマーの検出が可能であることを示した. 関連研究 [3] では, スパマーの学習セットを必要とするが, 実際にスパマーのデータを用意するのは難しい. 本手法は, 学習セットを利用しないため, 実サービスに本手法を導入することは比較的容易であると考えられる. さらに本実験では, はてなブックマーク [9] の実データを実験データとし, スパマーの検出が可能であることを示したため, 本手法を現実の SBM サービスに適用することは有効であると考えられる.

5. 関 連 研 究

SBM のブックマーク情報が, 有益な情報源として注目を集めるとともに, これを用いた Web 研究 [4]〜[7] が, 近年盛んである. 百田ら [4] は, SBM で, 類似するユーザおよび興味の近い Web ページを発見する手法と, 将来的に人気が上がるであろう Web ページを推薦する手法を提案している. 宗片ら [5] は, はてなブックマークを対象にしたブックマーク情報を収集するシステムの開発と, スパムブックマークの特徴を報告している. 山家ら [6] は, SBM のブックマーク情報から, 特定の時期に人気となる Web ページを分析し, 検索が行われた時期に応じた検索のランキング手法を提案している. 高橋ら [7] は, ソーシャルブックマーク上での Web ページを Authority，ユーザを Hub とみなし，HITS の概念を利用した Web ページのランキングアルゴリズムを提案し, これに加え, Web ページの鮮度を考慮したランキング手法を提案した.

SBM のような, コンテンツ共有サービスのスパムの問題を扱った研究が, 近年盛んになってきている [1]〜[3]. Heymann ら [1] は, コンテンツ共有サービスにおけるスパムの問題を概要的に述べた. これらのスパムの対策方法として, スパムを検出する (detection), コンテンツのランク付けによってスパムの効果を減衰させる (demotion), タグ付与回数に上限を設けるなど, ユーザの行動を一部制限することによって, スパム行為を予防 (prevention) する, という 3 つのアプローチを述べた. Koutrika ら [2] は, コンテンツにタグの付与ができるタギングシステムでのスパムの問題を扱っている. 「仮想的なタギングシステム下で有効なスパム対策手法は, 現実世界のタギングシステム下でも有効である」という仮定から, ユーザのタグ付与行動をモデル化し, 仮想タギングシステムを設計した. また, この仮想タギングシステム内で, 有効なスパム対策の手法を提案した. 中でも本研究と密接に関連している研究として, 数原らの研究が挙げられる. 数原ら [3] は, スパマーと非スパマーがラベル付けされている SBM のデータセットを用い, ユーザのタグ付与行為の時系列情報に基づく特徴を抽出し, 教師つき機械学習でスパマーを検出する手法を提案している. 彼らの研究と本研究は, SBM のスパマーを検出するという同じ目的を持つものの, 本研究は教師つきのデータを必要としないという点でスパマー検出のアプローチが異なる.

6. ま と め

本研究では, ソーシャルブックマークにおけるスパマーを「誰とも類似しないユーザ」と考え, クラスタリングの手法を基にスパマーを検出する手法を提案した. 本手法で, ユーザ間のブックマークの共起の局所性を考慮した, ユーザ間の類似度 IbfSim を提案した. 本稿では教師データを必要とせずに, スパマーの検出が可能であることを示した. しかし, 高い適合率を得る一方, スパマーの検出の網羅性が失われているといった問題が確認できた. 今後の課題として, スパマー検出に適切なクラスタの併合回数を決定すること, スパムタギングを行うスパマーの検出が可能な手法へと拡張することが挙げられる.

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] Paul Heymann, Georgia Koutrika, and Hector Garcia-Molina. Fighting spam on social web sites: A survey of approaches and future challenges. *IEEE Internet Computing*, 11(6):36–45, 2007.

[2] Georgia Koutrika, Frans Adjie Effendi, Zoltán Gyöngyi, Paul Heymann, and Hector Garcia-Molina. Combating spam in tagging systems. In *Proceedings of the 3rd international workshop on Adversarial information rtrieval on the web*, pages 57–64, New York, NY, USA, 2007. ACM.

[3] 数原良彦, 植松幸生, 井上孝史, 片岡良治. ソーシャルブックマークユーザのタグ付与行動に基づくスパマー判別手法. Web とデータベースに関するフォーラム (*WebDB Forum2008*), 2008.

[4] 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. データ工学ワークショップ (*DEWS2008*), 2008.

[5] 宗片健太朗, 福原知宏, 山田剛一, 絹川博之, 中川裕志. ソーシャルブックマークを用いた情報収集・分析ツールの開発. 人工知能学会第 22 回全国大会 (*JSAI 2008*), 2008.

[6] 山家雄介, 中村聡史, アダム ヤトフト, 田中克己. ソーシャルブックマークの周期性発見に基づく時期連動型検索ランキング手法. Web とデータベースに関するフォーラム (*WebDB Forum2008*), 2008.

[7] 高橋翼, 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 Forum2008*), 2008.

[8] 横田真俊. ソーシャルメディアとマーケティング. 会誌「情報処理」Vol.49 No.12, pp.1411-1414 (Dec. 2008).

[9] はてなブックマーク. http://b.hatena.ne.jp/

[10] delicious. http://delicious.com/