グラフ構造に基づくソーシャルブックマークにおけるスパマー検出

渡邊桂太[†]　髙橋翼[†]　天笠俊之[†,††]　北川博之[†,††]

[†]筑波大学大学院システム情報工学研究科　　[††]筑波大学計算科学研究センター

1　はじめに

近年，Web 上にブックマーク情報を登録・共有・管理するソーシャルブックマーク（以降，SBM）サービスが，良質な情報源として注目を集めている．しかし，スパムサイトをブックマークしたり，特定サイト内の複数ページをブックマークしたりするような悪質なユーザの増加が問題となっている．そのため，SBM の情報源としての有益性が損なわれつつある．

本研究では，SBM を，ユーザとページをノードとしたブックマークの関係を表す 2 部グラフとみなす．このグラフに対してグラフ構造解析を用い，単一・複数ユーザのスパム度合いを評価し，シングル ID，マルチ ID スパマーを検出する手法を提案する．

2　SBM におけるスパム行為

本稿では，SBM におけるスパム行為をスパムブックマークとスパムタギングに分類する．スパムブックマークとは，スパムコンテンツを含むページ，SEO (Search Engine Optimization) や宣伝を目的としたページをブックマークする行為である．一方，スパムタギングとは，ブックマークしたページに対して，不適切なタグを付与する行為である．具体的には，「通常のページに対して，悪質なタグを付与する行為」と「スパムブックマークの対象となるページに対して，ユーザの閲覧を誘導するタグを付与する行為」が挙げられる [2]．本研究では，前者のスパムブックマークを行うスパマーを，一つの ID を利用してスパムブックマークを行うシングル ID スパマー，複数の ID を利用してスパムブックマークを行うマルチ ID スパマーに分類し，各々のスパム度合いを評価する手法を提案し，それに基づきスパマーの判定を行う．宗方らの研究 [1] では，教師付き機械学習によるシングル ID スパマーの検出手法を示しているが，教師付き学習データの用意に多大な労力が費やされることが考えられる．そこで本研究では，教師付き学習データを必要とせずに，シングル ID スパマーのみならずマルチ ID スパマーも検出対象とした手法を提案する．

3　提案手法

3.1　シングル ID スパマーの評価手法

宗方らの研究 [1] より，シングル ID スパマーがブックマークするページの特徴として，「被ブックマーク数が 1 であることが多い」ことが報告されている．本手法では，シングル ID スパマーの特徴を「被ブックマーク数が極めて少ないページを高い割合でブックマークしている」と考え，シングル ID スパマーのスパム度合いを評価する．

3.1.1　ibf (Inverse Bookmark Frequency)

「被ブックマーク数が極めて少ないページ」を評価する指標として，先行研究では *ibf* 値を提案した [3]．ページ p をブックマークしているユーザの集合を，$users(p)$ としたときの *ibf* 値は，以下のように算出する．

$$ibf(p) = \frac{1}{\log_q(|users(p)| - q + 1)}$$

ただし，q はパラメータであり，4.1 節の実験では，q=2 とする．*ibf* 値は，被ブックマーク数が少ないページには高い値を，多いページには低い値を与える．

3.1.2　ibf スコア (ibf score)

「被ブックマーク数が極めて少ないページ」を *ibf* 値の高いページと考える．そこで，シングル ID スパマーの特徴を「*ibf* 値の高いページを高い割合でブックマークしている」とみなし，ユーザのスパム度合いを評価する指標として *ibf* スコア (*ibf_score*) を提案する．

$pages(u)$ を，ユーザ u がブックマークしているページの集合としたとき，ユーザ u の *ibf* スコアは，以下のように算出される．

$$ibf\_score(u) = \frac{1}{pages(u)} \sum_{p \in pages(u)} ibf(p)$$

ibf スコアは，ユーザ ID が与えられた際，そのユーザがブックマークしているページ全ての *ibf* 値を算出し，平均を取ったものである．そのため，*ibf* スコアの高いユーザは，シングル ID スパマーである可能性が高いことが考えられる．

3.2　マルチ ID スパマーの評価手法

マルチ ID スパマーは，複数のユーザ ID を用いてスパムブックマークするため，ユーザ同士で高い割合でブックマーク対象のページが共通することが特徴として挙げられる．SBM の構造を，ユーザとページをノード，ブックマークをエッジとした 2 部グラフで表わしたとき，ブックマークが密に張られている構造が発見されれば，そこに出現するユーザはマルチ ID スパマーである可能性が高いと考えられる．本手法では，密なブックマーク構造を 2 部クリーク，擬似 2 部クリークとして抽出し，それらに含まれるユーザ群のスパム度合いを評価する指標を提案する．

3.2.1　C_ibf(Clieque ibf)

マルチ ID スパマーのスパムブックマーク対象ページの特徴として，「特定の（擬似）2 部クリークに含まれるユーザ群から，極めて高い割合でブックマークを受けているページ」であることが考えられる．そこで，

Spammer Detection based on Graph Structure in Social Bookmarking Systems

Keita WATANABE[†](ein_vogel@kde.cs.tsukuba.ac.jp)
Tsubasa TAKAHASHI[†](tsubasa@kde.cs.tsukuba.ac.jp)
Toshiyuki AMAGASA[†,††](amagasa@cs.tsukuba.ac.jp)
Hiroyuki KITAGAWA[†,††](kitagawa@cs.tsukuba.ac.jp)
[†]Graduate School of Systems and Information Engineering, University of Tsukuba
[††]Center for Computational Sciences, University of Tsukuba

（擬似）2 部クリーク毎に，この特徴を評価する指標として，*C_ibf*(*Clieque ibf*) 値を提案する．（擬似）2 部クリーク Q と Q 内のページ p が与えられたとき，*C_ibf* 値は以下のように算出する．

$$C\_ibf(Q, p \in Q_{pages}) = \frac{1}{log_q(|users(p)| + q - |Q_{users}(p)|)}$$

Q_{pages} は，（擬似）2 部クリーク Q に含まれるユーザがブックマークしているページ集合，$Q_{users}(p)$ は，（擬似）2 部クリーク Q に含まれるユーザのうち，ページ p をブックマークしているユーザの集合とする．また，3.1.1 節と同様に，q はパラメータであり，4.2 節の実験では，q=2 とする．

3.2.2 C_ibf スコア (Clieque_ibf_score)

3.1.2 節と同様，マルチ ID スパマーの特徴を「*C_ibf* 値 の高いページを高い割合でブックマークしている」と考える．そこで，（擬似）2 部クリークに含まれるユーザ群のスパム度合いを評価する指標として，*C_ibf* スコア (*C_ibf_score*) を提案する．Q_{users} を（擬似）2 部クリーク Q に含まれるユーザ集合としたとき、*C_ibf* スコアは以下のように算出される。

$$C\_ibf\_score(Q) = \frac{1}{|Q_{users}|} \sum_{u \in Q_{users}} \frac{1}{|pages(u)|} \sum_{p \in Q_{pages}(u)} C\_ibf(Q, p)$$

ここで，$Q_{pages}(u) = pages(u) \cap Q_{pages}$ とする．*C_ibf* スコアは，（擬似）2 部クリーク内に *C_ibf* 値の高いページを多く含む場合は高い値を，*C_ibf* 値の低いページを多く含む場合は低い値が与えられる．

4 評価実験

提案した手法の検出精度を測るために，評価実験を行った．

4.1 シングル ID スパマーの検出実験

本実験では，我々の研究室で収集しているはてなブックマーク[1]の SBM データから，ランダムに 100 ユーザ選出したデータを用い，シングル ID スパマーの検出精度を評価する．今回用いた実験データの内訳は，一般ユーザ数が 87，スパマー数が 13 であることを目視で確認した．実験の手順は，以下の通りである．

1. 実験データに含まれるユーザの *ibf* スコアを算出．
2. *ibf* スコアの閾値 σ を設定し，$ibf\_score(u) \geq \sigma$ となるユーザ u をシングル ID スパマーとして検出．

実験結果を，宗方ら [1] の検出結果と比較した形で図 1 に示す．結果から，宗方らの手法よりも高い性能でシングル ID スパマーの検出が可能であることを確認した．

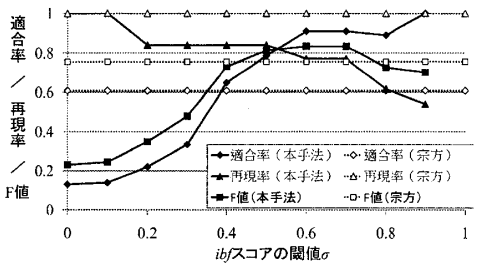

図 1: シングル ID スパマーの検出結果

4.2 マルチ ID スパマーの検出実験

4.1 節で示した SBM データは，マルチ ID スパマーを解析する上で十分な規模でなかったため，本実験では，livedoor クリップ[2]が提供している研究用データセットを用いて，マルチ ID スパマーの検出精度を評価した．実験データは，2009 年 12 月に公開されたものを用い，詳細を表 1 に示す．

表 1: マルチ ID スパマー検出の実験データ

ユーザ数	45,031
ページ数（被ブックマーク数 ≥ 3 ユーザ）	330,498
ブックマーク（エッジ）数	2, 467,438

今回，ユーザと各ユーザがブックマークしているページから構成される隣接行列を利用し，クラスタリングによって近似的に（擬似）2 部クリークを抽出する手法を用いた．隣接行列を *ibf* 値で重み付けしたデータから，ユーザのクラスタリングを bisecting k-means で行う．このとき，抽出したクラスタに含まれるユーザと，各ユーザのブックマーク，ブックマーク対象のページを近似 2 部クリークとして扱う．この近似 2 部クリークの抽出では，2 部クリークと擬似 2 部クリークを合わせて抽出するため，以下のような性質を持つ．

近似 2 部クリーク = 2 部クリーク ∪ 擬似 2 部クリーク

マルチ ID スパマー検出実験の手順を以下に示す．

1. 生成するクラスタ数 k を設定し，k 個のクラスタを近似 2 部クリークとして抽出．
2. 1. で抽出した近似 2 部クリークの *C_ibf* スコアを算出．
3. *C_ibf* スコアの閾値 σ を設定し，$C\_ibf\_score(Q) \geq \sigma$ となる近似 2 部クリーク Q に含まれるユーザ群をマルチ ID スパマーとして検出．

表 2 に，$k = 1000$, $\sigma = 1$ と設定したときの検出結果を示す．近似 2 部クリークを 2 部クリークと擬似 2 部クリークに仕分けたところ，2 部クリークに注目した際に，最も高い検出精度が得られた．

表 2: マルチ ID スパマーの検出結果 ($k = 1000, \sigma = 1.00$)

	近似 2 部クリーク	2 部クリーク	擬似 2 部クリーク
総検出数	143	103	40
スパム検出数	134	101	33
適合率	0.937	0.980	0.825

5 まとめと今後の課題

本研究では，SBM におけるスパマーのスパム度合いを評価する手法を提案し，高いスコアを持つユーザ，ユーザ群をスパマーとして検出する実験を行った．実験結果では，シングル，マルチ ID スパマーいずれも高い精度での検出が可能であることを確認した．今後は，シングル ID スパマーの再現率減少の原因について調査し，より高い性能で検出できるような手法の検討を行う．また，マルチ ID スパマーの検出実験を様々なパラメータで行い，より詳細に検出性能を評価する．

謝辞 本研究の一部は科学研究費補助金特定領域研究 (#21013004) による．

参考文献

[1] 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におけるスパム検出のための特徴とその評価. FIT2009, 2009.

[2] 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におけるタグ付与行動に基づくスパマー検出. DBSJ Journal Vol.7 No.4, 2009.

[3] 渡邊桂太, 髙橋翼, 北川博之, 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, DEIM 2009, 2009.

[1] http://b.hatena.ne.jp/

[2] http://clip.livedoor.com/