

## ソーシャルブックマークにおけるスパムユーザの分別

## Filtering Spammer in Social Bookmarks

三瓶 智昭†  
Tomoaki Sampei山田 剛一†  
Koichi Yamada絹川 博之†  
Hiroshi Kinukawa

## 1. はじめに

Web 上で情報を共有できるソーシャルブックマーク (Social Bookmark: SBM) サービスがある. SBM には人々の興味や関心を集める有用なコンテンツが数多く登録されている. 一方で SBM を悪用して特定のページへ誘導し, 商品の宣伝を行う・閲覧者に不利益を被らせるといったスパム行為が行われている. そのような行為をするユーザをスパムユーザと呼ぶ[1]. スパムユーザは様々な手法をとるが, それらは本来の SBM の使い方ではないため一般のユーザとはブックマーク行動の特徴に違いが生じる.

本研究では SBM の「はてなブックマーク[2]」を対象とし, ユーザがブックマークしたページの内容や, 記入したコメントなどの特徴に関する機械学習によりフィルタを生成し, それを用いてスパムユーザを分別する.

## 2. スパムユーザとその特徴

本稿では SBM においてスパムブックマークを登録する者をスパムユーザと呼ぶ. スパムブックマークとは, 悪徳商品の宣伝を行う・閲覧者に不利益を被らせるといったスパムコンテンツを含むページをブックマークする行為, または SMO 目的に特定のサイトをブックマークする行為である.

## 2.1 SBM から収集するデータ

はてなブックマークが提供している RSS から新着ブックマークなどの情報を取得し, 各ユーザのブックマーク情報などをデータベースに格納する. RSS から取得するデータは以下のものである.

- ページタイトル・ページ URL
- ブックマークしているユーザ
- 各ユーザがページに付けているコメント
- 各ユーザがページに付けているタグ
- 各ユーザがブックマークした日付・時刻

## 2.2 特徴

**特徴値 1:** スパムユーザがブックマークするページは一般の人の関心が薄いため, ページをブックマークするユーザは少ない傾向にある. そこでユーザのブックマーク内で, 少人数しかブックマークしていないページの割合が大きいほどスパムユーザである可能性が高くなる. ブックマーク者数が1ユーザである場合を**特徴値1-A**, 5ユーザ以下である場合を**特徴値1-B**とする.

$$(1-A) \text{ oneUser} = \frac{\text{ユーザ数が1ユーザのブックマーク数}}{\text{ユーザのブックマーク数}}$$

$$(1-B) \text{ fiveUser} = \frac{\text{ユーザ数が5ユーザ以下のブックマーク数}}{\text{ユーザのブックマーク数}}$$

**特徴値 2:** はてなブックマークでブックマークされているページのドメインのうち, ニュースサイトなど個人が自由に編集できないコンテンツを提供しているサイトは安全と見なせる. そこでそのようなサイトの中でブックマーク数が多い上位 50件のドメインを安全ドメインとしてデータベースに登録する. スパムにはなり得ない安全ドメインのページをブックマークしている割合が高いほどスパムユーザである可能性は低い. 以下の式で表す.

$$\text{safeDomain} = \frac{\text{安全ドメインのブックマーク数}}{\text{ユーザのブックマーク数}}$$

**特徴値 3:** スパムユーザはページタイトルや本文から機械的に単語を集めてタグを付ける傾向がある. 一方で一般ユーザは単語だけでなく“あとで読む”“これはすごい”といった文章もタグとして登録する. そこで文章で作られたタグを“安全タグ”と呼ぶ. 安全タグを利用しているブックマークの割合が多いほど一般ユーザであると考えられる. 安全タグが付いているブックマークの割合を以下の式で表す.

$$\text{safeTag} = \frac{\text{安全タグが使用されているブックマーク数}}{\text{ユーザのブックマーク数}}$$

**特徴値 4:** はてなブックマークで注目されているサイトは Web 検索エンジン上でも数多く検出される傾向にある. そこで国内で多く使われている検索エンジン Yahoo! Japan[4]と Google[5]でページタイトルを完全一致で検索し, ユーザごとにブックマークしている各ページの検索件数の平均を特徴値とする.

**特徴値 5:** 検索に掛かりやすくするためにスパムユーザは数多くブックマークすることが多い傾向にある. そこで各ユーザのブックマーク数を特徴値とする.

**特徴値 6:** はてなブックマークには他の人が集めたブックマークを更新されたか知ることができる「お気に入り機能」というサービスがある. この機能を使用しているユーザは本来の SBM の使い方をしていないと判断できる. ユーザが登録しているお気に入りユーザの数を**特徴値 6-A**, ユーザが他の人にどれだけお気に入りされているかを示す“お気に入られ数”を**特徴値 6-B**とする.

## 3. スパムユーザの分別

本研究ではスパム行為を行うユーザに見られた特徴を基準に, スパムユーザを分別することを目的とする. SBM から収集したデータを基にスパムユーザを分別する分類器を生成する.

## 3.1 実験データ

機械学習に用いたデータを表 1 に示す. これらのデータは 2011 年 3 月 1 日から 2 ヶ月間, 1 時間毎に新着ブッ

†東京電機大学大学院 未来科学研究科  
Graduate School of Science and Technology for Future Life,  
Tokyo Denki University

クマークから 30 件, 合計 43,200 件の収集した情報から一部をランダムに選び調査したものである. ユーザ 2,230 人に対して 2.2 で示した 8 種の特徴値を求めた. 学習ツール Weka[3] を用い, 学習アルゴリズムには決定木 J48 を選び, 機械学習を行う.

表 1 機械学習に用いたデータの量

URL 数	24,177
ユーザ数	2,232
スパムユーザ数	147
タグ数	8,961
ブックマーク数	150,752

3.2 評価方法

精度と再現率および F 値の定義を以下に示す.

$$\text{精度} = \frac{N_{tp}}{N_{tp} + N_{fn}} \quad \text{再現率} = \frac{N_{tp}}{N_{tp} + N_{fp}} \quad \text{F 値} = \frac{2 \times \text{精度} \times \text{再現率}}{\text{精度} + \text{再現率}}$$

$N_{tp}$ : Spam/Normal ユーザを Spam/Normal ユーザと判定した数  
 $N_{fp}$ : Normal/Spam ユーザを Spam/Normal ユーザと判定した数  
 $N_{fn}$ : Spam/Normal ユーザを Normal/Spam ユーザと判定した数

データから決定木(図 1)で分類器を生成し, 10 分割交差検定で評価した.

3.3 結果

実験結果を表 2 に示す. 一般ユーザに関しては精度・再現率ともに高い数値を得られ, スпамユーザに関しては精度が 94.2%, 再現率は 76.9%となっている. スпамユーザの分別に有効な特徴は 1-A, 1-B, 2, 4, 5 の 5 種である.

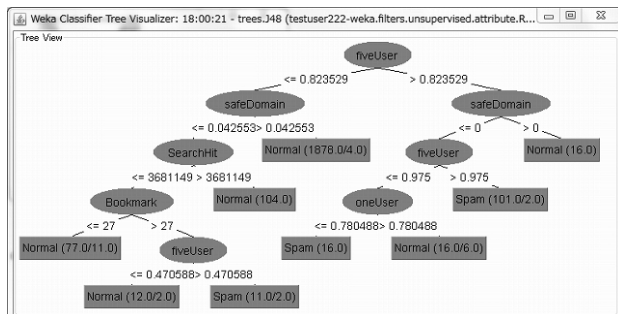


図 1. 機械学習で生成された決定木

表 2. スпамユーザ / 一般ユーザの分別結果

	精度	再現率	F 値
スパムユーザ	0.942	0.769	0.846
一般ユーザ	0.984	0.997	0.993

3.4 考察

生成された分類器は精度 94.2%に対して再現率が 76.9%であった. 検出されなかったスパムユーザの特徴として, スパムブックマークを行う際, 複数のユーザがタグ・コメント等をほぼ同じ形でブックマークしている点が挙げられる. 複数人が組織的にブックマークを行うため, ページのブックマーク人数が 5 ユーザ以上になってしまい特徴 1-B で検出することが出来なかった. そのため一般ユーザとして誤検出したと考えられる.

決定木に現れなかった特徴が出た理由として特徴値 3, 6-A, 6-B のいずれもその機能を使用していなかったユー

ザが多い点が挙げられる. また, 機能を使用しているユーザはそもそも特徴値 1-B の fiveUser で一般ユーザであると判別されているので決定木には現れなかったと考えられる.

4. スпам分別システム

スパムユーザ分別に有効な特徴を用いて, はてなブックマークの新着エントリに掲載されているユーザがスパムであるか即座に判定するシステム(図 2)を作成した. 本来ならばすべての特徴を組み込んだものが望ましいが, 特徴値 2~4 は RSS から取得した情報をデータベースに取り込み, 値を得る処理を行うために即座にスパムを検出するためには使用できない. そのため特徴は 1-A, 1-B の 2 種を使用するとし, 精度・再現率ともに低い値になっている(表 3). この結果は逆に特徴値 2~4 が有効であり, これによってスパム分別精度が 10.7%上昇していることを表している.



図 2. リアルタイムスパム判定システム

表 3. システムでのスパムユーザ / 一般ユーザの分別結果

	精度	再現率	F 値
スパムユーザ	0.835	0.789	0.811
一般ユーザ	0.985	0.989	0.987

5. おわりに

SBM で有用な情報を得るために, スпамユーザの特徴を分析してユーザ分別のための機械学習を行った. 特徴値として挙げた 8 種のうち 5 種が決定木に現れ, スпамユーザ分別に効果があることが分かった.

更なるスパムユーザ分別の精度と再現率向上のために, 考察で挙げた組織的にブックマークを行う類似したスパムユーザの特徴を見出すことが今後の課題である.

参考文献

- [1] 宗片健太郎, 福原宏宏, 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の検出, 情報科学技術フォーラム講演論文集 8(2), pp151-152, (2009)
- [2] はてなブックマーク: <http://b.hatena.ne.jp/>
- [3] Weka: <http://www.cs.waikato.ac.nz/ml/weka/>
- [4] Yahoo Japan: <http://www.yahoo.co.jp/>
- [5] Google: <http://www.google.co.jp/>