

ソーシャルブックマークにおけるスパムとその類似ユーザの抽出 Extraction of Spammer and User Bookmarking Similar to Spammer's Bookmarks 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、商品を宣伝する、閲覧者に不利益を被らせるページをブックマークする行為、または不当に特定のサイトの認知度や評判を高めて、サイトへのアクセスや評価を向上させることを目的にブックマークする等の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と呼ぶ。スパムは様々な使い方をしますが、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。我々は、ユーザがブックマークしたページのドメインや、記入したコメントなどの特徴に基づく機械学習によりフィルタを生成してスパム分別実験し、以下の事を行った[1].

- はてなブックマーク[2]が提供している RSS から新着ブックマークの情報を取得し、各ユーザのブックマーク情報をデータベースに登録した。
- ユーザのブックマーク内で、少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である可能性が高くなる等の事の特徴として 8 種の特徴を見出した。
- 見出した特徴を基に機械学習を行い、スパムの分別を行う分類器を生成して評価を行った。また分類器を用いたスパム判別システムを作成した。

一方で以下の問題があった。

- 分別に有効であった特徴のうち 5 種類はその場で収集したデータから、特徴値を得るための処理に多くの時間をかける必要がある。その為即時にスパムであるか分別できなかった。
- 上記の理由からスパム判別システムに使える特徴値が 2 種しかなく、精度が下がってしまった。

本研究では、新しい特徴値による分別精度の向上を図り、分別したスパムと類似したページをブックマークしているユーザをもとにブラックリストを作成することにより、効率良くスパム検出することを目指す。

本稿では 2 章においてスパムがどのような特徴を持つか論じ、3 章にてスパムを効率良く検出するために類似ユーザの抽出方法を説明し、4 章で提案手法の評価を行い、最後に 5 章でまとめる。

2. スパムとその特徴

これまでに提案した手法の問題点を改善するために今回新しく導入した特徴を以降に記す。

特徴値 1: 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スパムコンテンツでは無いと見なせる。そこでデータ収集期間内の人気エンタリ中でブックマーク人数が 100 人以上となるページ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である可能性は低いと考えられる。この割合を以下の式で表す。

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

特徴値 2: Web ページを宣伝する目的でブックマークを行うスパムはブックマークするページドメインの種類が偏ることが多い。ブックマーク一覧のドメインの割合を特徴とする。この値が小さいほどスパムである可能性が高まる。この割合を以下の式で表す。

$$domainAverage = \frac{\text{ブックマークしたページのドメイン数}}{\text{ユーザのブックマーク数}}$$

3. 類似ユーザの抽出

3.1 スパムの類似性

ソーシャルブックマークを製品の宣伝といった商用目的で悪用するスパムは、複数のアカウントを用いてタグ・コメント等を似た形でブックマークしているなどユーザ間の類似性が見られる。また、ブックマークをした一連の Web ページが似ている傾向が見られた。よってある 1 人のスパムとの類似度が高いユーザを探すことで連鎖的にスパムを抽出でき、リスト化することによってスパム除外のための時間が短縮できると考えられる。

3.2 類似ユーザの抽出

類似ユーザを得る先行研究として畑中ら[3]が新着エンタリからランダムに選んだ 2,000 ユーザからユーザ間類似度を基に類似度の高いユーザグループをリストに登録して関連するブックマークの削減を行った。畑中らはすべてのユーザに対して双方向のブックマーク類似度を計算したが、本研究ではスパムに似たユーザを探すためにユーザ u がスパム v と類似しているかの 1 方向だけを調べる。ユーザ類似度 S を求める式を以下のように提案する。

$$S = \frac{A_{vu}}{BM_u}$$

BM_u はユーザ u のデータ収集期間 T にブックマークをしたページ数、 A_{vu} はスパム v のブックマークと同じページをユーザ u がブックマークした数を表す。 S が閾値 γ を超えた場合、ユーザ u はスパムと類似しているとし、ブラックリストに登録をする。

†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

4. スパマの分別および類似ユーザの抽出の実験

SBM から収集したデータを基にスパマを分別する分類器を生成する。また分類器にてスパマと判定されたユーザがブックマークしているページから類似しているユーザを見つけ出しスパマを効率良く検出していく。全体の流れを図 1 に記す。

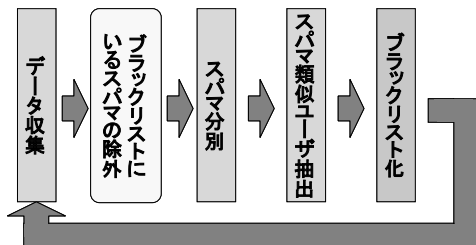


図 1. スパマ分別・除去の流れ

4.1 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 2012 年 2 月 11 日から 3 ヶ月間、1 時間毎に新着ブックマークから 30 件、合計 64,800 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 9,894 人に対して学習ツール Weka[4] 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	68,177
ユーザ数	8,794
スパマ数	1,100
タグ数	144,044
ブックマーク数	1,859,752

4.2 分別結果

10 分割交差検定にて評価をした先行研究と今回の実験結果を以下に示す。精度は先行研究より数値が悪くなっているが、再現率については飛躍的に向上して結果、調和平均である F 値が前回の値を超え性能は向上したと言える結果となった。

表 2. スパマ/一般ユーザの分別結果

	ユーザ	精度	再現率	F 値
前発表結果[1]	スパマ	0.942	0.769	0.846
	一般ユーザ	0.984	0.997	0.993
本研究の結果	スパマ	0.885	0.925	0.905
	一般ユーザ	0.992	0.987	0.989

4.3 類似ユーザ抽出

生成された分類器にてスパマと判定された 1,287 ユーザがブックマークしたページの中で各々最新 40 ページ分を収集し、それらのページをブックマークしている 17,510 ユーザに対してブックマーク類似度 S を計算する。閾値 γ は 17,510 ユーザに対して類似度 S を属性にして機械学習にて 10 分割交差検定を行った結果から、0.547 とする。

4.4 抽出結果

スパマとブックマークが類似していると判定されたユーザは 624 ユーザであり、4.2 で “スパマであり、スパマと判定された” ユーザと類似していた。また 624 ユーザす

べてを人手で調べた結果、621 ユーザがスパマの特徴に一致するユーザであることが確認できたが、3 ユーザは人手でもスパマか一般ユーザか判断のつかないユーザであった。

4.5 考察

- (1) 生成された分類器は F 値が既発表結果から 5.9% 向上した 90.5% であった。導入した *SafeDomain* に関してはデータ収集期間の人気エン트리ページから安全ドメイン一覧作成したので、この先スパマの傾向が変わってきたとしてもその時の人気エン트리からまた安全ドメイン一覧を得ることで対応できると考えられる。精度が 5.7% 低下したのは特徴値 *domainAverage* が SBM を Blog の更新通知に使っているユーザや 1 つの話題を扱うニュースページだけをブックマークするユーザをスパマと判断したためである。人気のあるページであってもそのブックマークリストはドメインが 1 種類であるので今回誤認識してしまった。
- (2) 類似ユーザ抽出で得られたユーザのうち 621 ユーザはスパマであり目的通りの結果が得られた。残りの 3 ユーザは “スパマ判定されたが一般ユーザ” であるユーザと類似し、人手でもスパマか一般ユーザか判断のつかないユーザであった。しかしブックマークしているページがそれぞれ 1 ~ 3 ページだけと少なく、ブックマーク内容もネット通販の商品ページやゲーム攻略を行なっている Wiki の 1 ページなど大多数の人にとって有益な情報とは言い難いページなのでスパマとして除外しても問題は無いと考えられる。

5. おわりに

SBM で有用な情報を得るために、スパマの特徴を分析してユーザ分別のための機械学習を行い、同時にスパマ類似ユーザを抽出・ブラックリストに登録することで次回以降のスパム除外を効率良くできるようにした。新しく追加した特徴値はいままで見逃していたスパマを検出でき、スパマ分別の再現率を上げる効果があることが分かった。またスパマと類似したユーザを正確に抽出することができた。

本研究のスパマ抽出機能をシステム化したところ、データ収集やスパマの判別に多くの時間がかかってしまう。これを実用的段階まで処理速度を早めていくことが今後の課題である。

参考文献

- [1] 三瓶, 山田, 絹川 "D-017 ソーシャルブックマークにおけるスパムユーザの分別(D 分野: データベース)," *情報科学技術フォーラム講演論文集*, vol. 10(2), pp. 159-160, 2011.
- [2] はてなブックマーク : <http://b.hatena.ne.jp/>
- [3] 畑中, 久松 "ブックマーク類似度を用いたソーシャルブックマーク汚染対策手法の提案と評価," *情報処理学会研究報告. データベース・システム研究会報告*, vol.2010-DBS-151, No.19, pp. 1-7, 2010.
- [4] Weka : <http://www.cs.waikato.ac.nz/ml/weka/>