Vol.2009-CSEC-47 No.1
2009/12/18

内容の類似性を用いたトラックバックスパム判別法の評価と考察

藤村浩太[†1]　西出隆志[†2]
堀良彰[†2]　櫻井幸一[†2]

以前筆者はトラックバックスパムが送信先のブログ記事の内容とは無関係に送信されることに着目し，記事の内容の類似性を用いたトラックバックスパム判別手法の提案を行った．今回は提案手法について前回行った実験よりも多くのサンプルを用いて実験を行った．まず，提案手法の判別基準が一般性を持つことを示すことすため，複数人のスパム分類結果と提案手法のスパム分類結果を比較した．さらにブログ記事のカテゴリーと提案手法の判別精度の関係についても考察した．

Evaluation and Consideration of Trackback Spam Detecion Scheme Using Similarity Analysis of Contents

Kohta Fujimura,[†1] Takashi Nishide,[†2]
Yoshiaki Hori[†2] and Kouichi Sakurai[†2]

The authors proposed a scheme[1)] for detecting trackback spam using the similarity analysis of contents based on the fact that the contents of trackback spam are irrelevant to the blog posts they refer to. In this paper, we used more samples compared with the previous experiment. In order to show that our criterion for the similarity analysis is general, we compared the classification results of the spams by several people with those by our scheme. Furthermore, we considered the relation between the category of blogs and the accuracy of our detection scheme.

†1 九州大学大学院システム情報科学府
Graduate School of Computer Science and Communication Enginnering, Kyushu University
†2 九州大学大学院システム情報科学研究院
Department of Computer Science and Communication Enginnering, Kyushu University

1. はじめに

近年，インターネットの普及に伴い，スパム行為が増加している．電子メールを使ったスパムは良く知られているが，ブログの機能を利用したスパムも問題になっている．スパマーの多くはボットウイルスに感染したPCや自動送信ツールなど機械化・自動化された手段で多量のスパムの送信を行っている．スパマーが機械化された手段でスパムを行うのは人力で行うと高いコストがかかるためであり，機械化により低いコストで大量の広告をばら撒くことができることがスパムが蔓延している原因である．スパムを減らすためには人間が送信した正当なものと機械的に無差別に送信されたものを区別して機械が送信したスパムを排除し，スパム行為の利益を減少させることが必要である．

機械と人間を区別する方法の例として，
CAPTCHAの一種である画像認識[2)]ではランダムな文字や数字に変形させたりノイズを加えたりして機械では読みにくくしたものが読めるかどうかで機械と人間を区別している．他に，ボットウイルスに感染したPCやツールの動作の規則性を利用して機械と人間を区別する手法などもある．

本稿では，ブログの機能を利用したスパムの一種であるトラックバックスパムの多くがスパム送信先のブログの記事を踏まえずに無差別に送信されていることを利用してトラックバックスパムを判別する手法についての実験を行った．

本稿の構成を以下に示す．第2章ではブログスパムについて概説し，第3章では関連研究について説明する．第4章においてトラックバックの記事の類似性の数値化手法を示し，判別に用いた形態素解析と文書の中の特徴的な単語を抽出するためのアルゴリズムであるtf-idfについて説明する．第5章で，判別精度の実験結果を示す．第6章で考察を行う．最後に，第7章をまとめとする．

2. ブログスパムについて

ブログの普及にともないブログスパムの増加が問題になっている．この章では代表的なブログスパムであるスプログ，コメントスパム，トラックバックスパムについて説明する．

2.1 スプログ

スプログ (スパムブログ) とは広告などで利益を上げることをを目的とした価値の無いブログのことであり，その多くは機械的に作られる．内容は無意味な文や意味のわからない文，反復性のある文を含んだものや，他のブログやウェブサイトの文章をコピーしたり組み

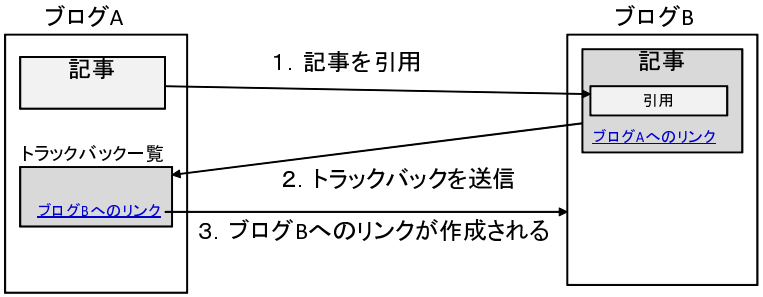

図 1 トラックバックの例

合わせたりしたものである場合が多い．

2.2 コメントスパム

多くのブログにはコメント欄がついており，ブログの投稿に対してコメントを書き込むことができる．

この機能を利用し，著名なブログに記事とは無関係な内容のリンクつきコメントを送信することがコメントスパムであり，別のブログから自分のブログにリンクを張ることで検索エンジンの順位を上げたり，自分のブログに読者を誘導したりといった目的を持っている．

コメントスパムは迷惑メールのブログ版とも言えるもので，迷惑メール対策と同様に「特定のキーワードの禁止」や「特定の IP アドレスからのトラックバックの禁止」などの対策を適用できる．しかし，禁止ワードによってトラックバックスパムを完全に判別することは難しく，IP アドレスによる判別はボットネットを利用したトラックバックに対してはあまり効果的ではない．また，コメント書き込み欄はブログ管理者が用意するため，書き込みの際に CAPCHA などで認証を行うことでも対策を行うことができる．

2.3 トラックバックスパム

トラックバックとはブログの機能の一つであり，他人のブログに別のウェブページへのリンクを作成する機能である．一般的には図 1 のように引用を行った際に引用元のブログへ通知するために使用される．

このトラックバック機能を悪用したスパムがトラックバックスパムである．トラックバックスパムの目的の 1 つはブログ訪問者をトラックバックに含まれるリンクからアフィリエイトサイトやアダルトサイト，ワンクリック詐欺サイトなどの悪質なサイトへ誘導することである．

また，テクノラティ[10] などのブログランキングサイトは Google などの検索エンジンとは異なるランキングの仕組みを持ち，トラックバックのつながりは重要である[3]．そのためトラックバックスパムの 1 つの目的としてブログランキングサイトで順位を上げることが考えられる．

さらに E メールのスパムの 98 %がブロックされるのに対してブロックされるトラックバックスパムは 90 %と少し低いこと[3]，1 件辺り 1 人のユーザーに届く E メールスパムに対してトラックバックスパムは 1 件で 1 人より多くのブログ訪問者に届くことなどから E メールスパムより効率が良いこともトラックバックスパムを行う動機の 1 つである．

既存のトラックバックスパム対策として，総務省の調査によるとブログサイトの多くが以下のような対策をとっている[4]．

- ブログ管理者が承認した後にトラックバックの結果を表示する機能の導入
- URL 等の情報によりあらかじめ用意してあるブラックリストの掲載情報に合致したトラックバックを禁止
- IP アドレス，キーワード，文字コード等の条件によるトラックバックやコメントの制限

しかし，ブログ管理者の承認制はブログ管理者にとっては負担であり，その他の方法でも完全にトラックバックスパムを遮断できるわけではないため，さらなる対策手法が必要である．他には，言及リンクが無いトラックバックを禁止する機能を導入しているブログサイトも存在するが，現在ではトラックバックは様々に用いられており，一部の利用者には不評であった[6]．本研究では，これらの対策手法と組み合わせて用いることができるトラックバックスパム判別基準について実験を行った．

3. 関 連 研 究

最近のトラックバックの研究としては，2009 年の Bursztein らの研究がある．Bursztein らは，サンプルを収集するため実際のブログである「honeyblog」を開設して 1 年以上にわたり 1000 万のサンプルを収集し，それらサンプルの分析やネットワーク分析を行っている．

4. トラックバックの記事の類似性の数値化手法について

正当なトラックバックとトラックバックスパムを判別する方法として，下記の 2 点を踏まえて内容の類似性を用いた手法について提案した[1]．

- トラックバックスパムの多くはトラックバック先の記事の内容を踏まえていないことが多い

- 正当なトラックバックではトラックバック元の記事とトラックバック先の記事の趣旨が同じである

しかし，2 つの記事の内容を意味的に比較することは難しいため，提案手法では Lin[5] らがスプログの判別のためにブログの投稿の内容の比較に使っていた式を用い，2 つの記事の中に同じ名詞が含まれていることを 2 つの記事の趣旨が同じであることと見なすことにした．

トラックバック先の記事とトラックバック元の記事の類似性を数値化するために，以下のような手順を踏んだ．

(1) 2 つの記事のタイトルと本文に対して形態素解析を行い品詞に分別する．

(2) 1 で分別した中で 2 つの記事からそれぞれ出現回数が多い名詞 (数，非自立語，接尾辞，代名詞を除く) を 5 つづつ抽出する．この際，2 つの記事の間で選んだ名詞に重複があっても選び直しは行わない．

(3) 2 で抽出した名詞の tf-idf を計算する．

(4) 3 で計算した tf-idf の値に対して Lin らの手法で使われた 2 つのポストの間の類似性の定義

$$S_c(i,j) = \frac{\sum_{k=1}^{10} min(h_i(k), h_j(k))}{\sum_{k=1}^{10} max(h_i(k), h_j(k))}$$

を適用し，記事 i と記事 j の内容の類似度を数値化する．
ここで，$S_c(i,j)$ は，記事 i と記事 j から抽出した 10 個の名詞から計算した 10 組の tf-idf 値 h_i, h_j の大小を比較し，小さいものの合計を大きいものの合計で割った値である．

(5) 上記の実験を正当なトラックバックの記事，トラックバックスパムの記事に対して行う．

以下の節では 1 で用いた形態素解析，3 で用いた tf-idf について説明する．

4.1 形態素解析

形態素解析とは，自然言語処理の基礎技術の一つで，対象言語の文法の知識や辞書を情報源として用い，自然言語で書かれた文を言語で意味を持つ最小単位である形態素の列に分割し，それぞれの品詞を分別する作業のことである[7]．例えば,「パソコンの中に保存する」という文を形態素解析すると表 1 のようになる．

今回の実験では形態素解析ツールには茶筌[8] を使用した．

文字列	読み	原形	品詞の種類	活用の種類	
パソコン	パソコン	パソコン	名詞-一般		
の	ノ	の	助詞-連体化		
中	ナカ	中	名詞-非自立-副詞可能		
に	ニ	に	助詞-格助詞-一般		
保存	ホゾン	保存	名詞-サ変接続		
する	スル	する	動詞-自立ど	サ変･スル	基本形

表 1 形態素解析の例

4.2 tf-idf

tf-idf[9] は，文書の中の特徴的な単語を抽出するためのアルゴリズムである．tf(Term Frequency の略) を単語が文書に出現した回数，N を文書 (記事) の総数，df(Document Frequency の略) を単語が出現する文書数すると，

$$tfidf = tf \times \log(\frac{N}{df})$$

と表される．

今回の実験では tf はブログの投稿のタイトルと本文の中での単語の出現回数とする．また，N をブログ検索サイト (テクノラティジャパン[10]) に登録されているブログの投稿の総数，df をブログ検索サイトで単語を検索したときに検索結果に表示される単語を含むブログの投稿の数とした．

ブログの投稿の総数は現在の値が不明なためテクノラティジャパンが 2006 年 2 月 15 日に発表した 1 億 100 万件とした．

5. 実験と結果・考察

今回は提案手法について前回行った実験よりも多い 1000 組のサンプルを用いて評価実験を行った．また，スパム判定基準は人によって異なると考えられるため，今回は筆者以外の 3 人にも評価実験を行ってもらい，提案手法が適用できるかどうかを調べた．さらにブログ記事のカテゴリ別に分析を行い，提案手法に適したブログ記事のカテゴリがあるかを調べた．

5.1 提案手法の評価実験

最初に，提案手法の評価実験を行った．前回の実験ではサンプルの少なさが問題になったため，今回は 1000 組のサンプルを用いた．1000 組のサンプルを処理するため，スパム判別以外の作業はプログラムを用いて自動化している．実験用サンプルの収集では，まずブログ

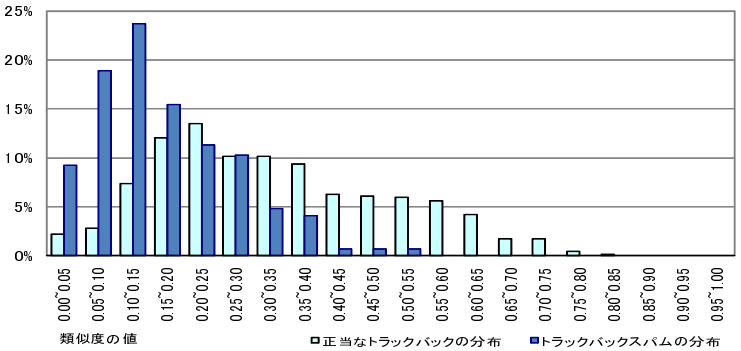

図 2 類似性の値と正当なトラックバック，トラックバックスパムの分布

の更新情報を公開している ping.bloggers.jp[11] からブログの URL を収集した．そして収集したブログの中でトラックバックが送信されている記事と送信元のウェブページを収集した．

収集したサンプルから 1000 組を実験に利用した．まず，筆者が 1000 組のサンプルに対して正当なトラックバックとトラックバックスパムに分別を行った．この際，トラックバックの送信元と送信先が同じブログの場合は明らかに正当なものであるため除外した．その分類の結果をグラフにしたのが図 2 である．図 2 の棒グラフは各類似性の値におけるトラックバックの分布を表しており，折れ線グラフは各類似性の値でのトラックバック中のトラックバックスパムの割合を表している．

図 2 より，サンプル数を増やした場合でもトラックバックスパムは類似度の値が低い傾向があることが言える．トラックバックスパム全体の 67 %の類似度は 0.20 より小さかった．しかし，前回の実験の結果[1] と比べるとトラックバックスパムの類似度の値が全体的に上がっていた．正当なトラックバックはトラックバックスパムに比べると類似度は高い傾向があったが，類似度が 0.20 以下のものも約 24 %あった．

さらに，同様の分別を 20 代男性，20 代女性，50 代男性の 3 人で行い，その場合の提案手法の正当なトラックバック，トラックバックスパムの判定率を調べた．実験は以下の手順で行った．

(1) 20 代男性は 150 組，20 代女性，50 代男性は 200 組のサンプルの分別を行う．実験用のサンプル筆者が分別した 1000 組の一部である．

(2) (1) で分別した結果と筆者のサンプル分別結果を用いて，提案手法に用いる閾値を変化させたときの誤判定数をグラフにし，各人の最適な閾値を求める．

(3) (2) で求めた閾値を用いて提案手法でスパム判別を行ったときの判定率を求める．

(2) のグラフは図 3，図 4，図 5，図 6 のようになった．図 3，図 4，図 5 は同じような結果になったが，図 6 は若干異なっていた．50 代男性は普段インターネットをしないことが影響した可能性がある．

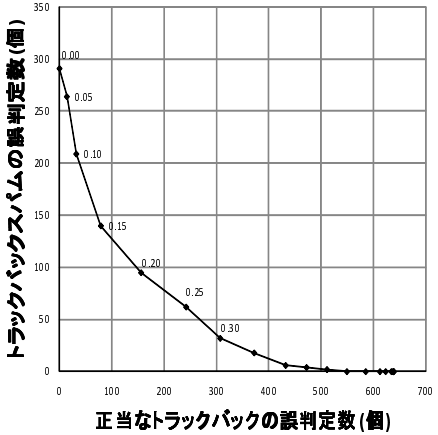

図 3 閾値を変化させたときの誤判定数（筆者：20 代男性）

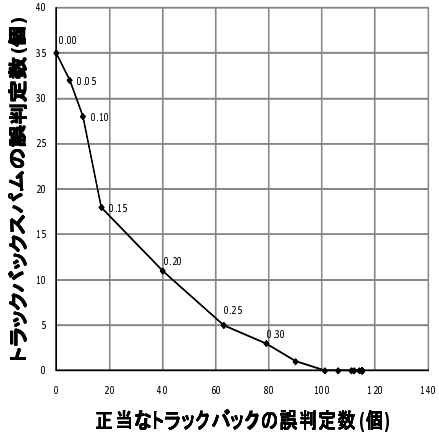

図 4 閾値を変化させたときの誤判定数（20 代男性）

図 3，図 4，図 5，図 6 から，閾値 0.15，0.20 のとき正当なトラックバックとトラックバックスパムを正しく判定する割合を求めた結果，表 2 のようになった． 表 2 から，閾値 0.15 の場合には提案手法は全員に対して約 90 %の正当なトラックバックを判別したが，筆者以外の 3 人の場合にスパムを正しく判定する割合は 50 %を下回った．また，閾値 0.20 の場合には提案手法は全員に対して 50〜70 %のトラックバックスパムを正しく判別したが，正当なトラックバックを正しく判定する割合は 80 %を下回った．今回の結果から類似度を用いたトラックバックスパム判別が筆者以外の人物の判定基準にも適用できることがわかったが，判定結果の数値そのものは良い結果ではなかった．実際にトラックバックスパム判別に用いるには他の手法との併用が必要だと考えられる．

判定結果の数値が低下した要因として 2 つ考えられる．1 つは 1 年前回の実験から 1 年

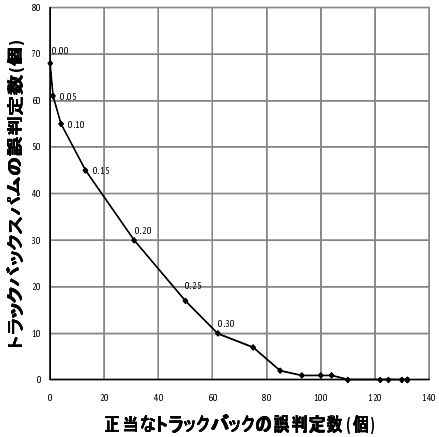

図 5 閾値を変化させたときの誤判定数（20 代女性）

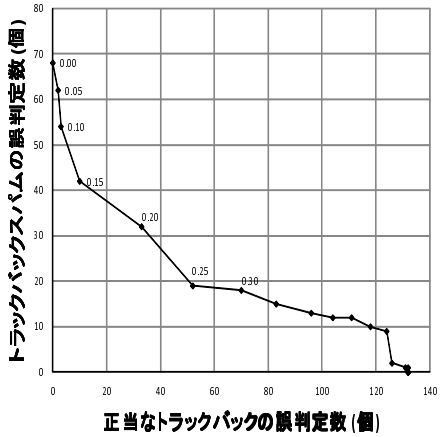

図 6 閾値を変化させたときの誤判定数（50 代男性）

閾値	サンプル分別者	正当トラックバック判定率	スパム判定率
0.15	筆者(20代男性)	0.88	0.52
	20代男性	0.85	0.46
	20代女性	0.9	0.34
	50代男性	0.92	0.38
0.2	筆者(20代男性)	0.76	0.67
	20代男性	0.65	0.69
	20代女性	0.77	0.56
	50代男性	0.75	0.53

表 2 サンプル判別者を変えた場合の判定精度

以上経過しているためトラックバックの傾向が変わった可能性である．もう 1 つは実験を自動化した際に判別に用いる名詞をブログ記事の本文からではなく本文を含むページ全体から得ていることである．これはプログラムの問題であるため，もしこの点が影響している場合はプログラムの修正が必要である．

5.2 ブログ記事のカテゴリーと判別精度

さらに，ブログ記事のカテゴリーと判別精度の関係についても検証した．ブログ記事のカテゴリ分けは記事にブログの著者が付けた分類のタグや本文の内容を基に筆者が行った．

サンプル数が 10 個以上あった日記，アニメ，料理，経済，写真，映画，ペット，本，健康，スポーツ，旅，ドラマのカテゴリのブログ記事の閾値 0.20 のときの正当なトラックバックの判定率，トラックバックスパムの判定率は表 3 のようになった．図 2 の筆者の閾値 0.20 の場合と比べると，アニメ，スポーツ，ドラマが特に高い判定率を示した．スポーツ，ドラマはサンプル数が少ないため断定はできないが，アニメカテゴリは特に提案手法に適していると考えられる．逆に，日記カテゴリは全体と比べると少し判定率が低かった．

閾値	サンプル分別者	正当トラックバック判定率	スパム判定率
0.15	筆者(20代男性)	0.88	0.52
	20代男性	0.85	0.46
	20代女性	0.9	0.34
	50代男性	0.92	0.38
0.2	筆者(20代男性)	0.76	0.67
	20代男性	0.65	0.69
	20代女性	0.77	0.56
	50代男性	0.75	0.53

表 3 利用者を変えた場合の判定精度

閾値	カテゴリ	ラックバック	パム判定率	サンプル数
0.2	日記	0.62	0.65	292
	アニメ	0.9	0.73	286
	料理	0.47	0.82	62
	経済	0.79	0.79	38
	写真	0.81	0.4	26
	映画	0.75	0.67	23
	ペット	0.79	0.67	20
	本	1	0.5	19
	健康	0.62	0.33	16
	スポーツ	0.86	0.89	16
	旅	0.64	1	14
	ドラマ	1	0.75	10
	その他	0.54	0.66	107

表 4 サンプルのカテゴリ分類結果

6. おわりに

本論文では，まず以前提案したトラックバックスパムをブログ記事の内容の類似性を用いて判別する手法に対して，評価実験を前回の問題点を改善して行った．その結果，内容の類

似性がトラックバックスパムの分布に関係があることを再び示せた．しかし，今回の結果は前回の結果よりも悪く，トラックバックスパム判別を提案手法単独で判別を行うには足りないものであった．

今後の課題としては，まず実験の結果が悪くなった原因を考える必要がある．手法の実験の自動化に際して手法を変更した部分に問題がなかったかを見直さなければならない．また，参照リンクの有無などの既存の手法との組み合わせで判別精度を上げることを考えている．さらに，今回の実験で提案手法に適していることが分かったアニメ，スポーツ，ドラマカテゴリの判別精度をより高める改良も行うつもりである．

参 考 文 献

1) Kohta Fujimura, Yoshiaki Hori, Kouichi Sakurai, “Trackback Spam Distinction Using Similarity of Contents” Joint Workshop on Information Security 2008
2) L.V. Ahn, M. Blum, and J. Langford, “Telling Humans and computers apart automatically,” Communications of the ACM, Vol. 47, No. 2, pp. 57-60, February 2004.
3) “TrackBack Spam: Abuse and Prevention，” Proceedings of the 2009 ACM workshop on Cloud computing security 2009, Chicago, Illinois, USA November 13 - 13, 2009,Pages 3-10
4) ブログの実態に関する調査研究,
http://www.soumu.go.jp/iicp/chousakenkyu/data/research/survey/telecom/2009/2009-02.pdf
5) Yu-Ru Lin, Hari Sundaram, Yun Chi, Junichi Tatemura, and Belle L. Tseng, “Splog Detection Using Self-similarity Analysis on Blog Temporal Dynamics,” AIRWeb 2007, pp.1-8,(2007)
6) 絵文録ことのは, トラックバックをめぐる４つの文化圏の文化衝突――「言及なしトラックバック」はなぜ問題になるのか (2009/11/24),
http://www.kotono8.com/2006/01/ 06trackback.html
7) 鈴木 肇, “形態素解析と自動要約の可能性,” 産業経済研究所紀要, 第 17 号, pp.59-64, 2007 年 3 月
8) ChaSen – 形態素解析器 (2009/11/24),
http://chasen-legacy.sourceforge.jp/
9) 佐藤 翔輔, 林 春男, 牧 紀男, 井ノ口 宗成, “TFIDF/TF 指標を用いた危機管理分野における言語資料体からのキーワード自動検出手法の開発－ 2004 年新潟県中越地震災害を取り上げたウェブニュースへの適用事例－,” 地域安全学会論文集 No.8, pp.367-376, 2006.11
10) テクノラティジャパン (2009/11/24),
http://www.technorati.jp/
11) PING.BLOGGERS.JP,
http://ping.bloggers.jp/