2008－CSEC－41 (4)
2008／5／22

内容の類似性を用いたトラックバックスパム判別

藤村 浩太† 堀 良彰‡ 櫻井 幸一‡

†九州大学大学院システム情報科学府 ‡九州大学大学院システム情報科学研究院

あらまし ブログの普及にともないトラックバックスパムの増加が問題になっており、これを正当なものと区別して排除することが必要である．トラックバックスパムの多くは機械的に多数行われているため，人手を介さずに機械的な判別をする手法が必要である．

そこで，スパムで無い正当なトラックバックはトラックバック先の記事と趣旨が同じ事が多いことと，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を利用したトラックバックスパム判別手法について実験を行った．記事の内容を意味的に比較することは難しいため，2つの記事の中に同じ名詞が含まれていることが記事の趣旨が同じであることと見なした．結果，記事の類似性が低いときトラックバックスパム率が高くなることがわかった．

Trackback spam Distinction Using Similarity of Contents

Kohta Fujimura† Yoshiaki Hori‡ Kouichi Sakurai‡

†Graduate School of Information Science and Electrical Engineering, Kyushu University
‡Information Technology and Security Group,
Department of Computer Science and Communication Engineering, Kyushu University

Abstract An increase of trackback spam becomes a problem as blog spreads, and it is necessary to be excluded. Distinction without people is necessary because many of trackback spams are mechanically done.

Then, we experimented on the trackback spam distinction technique using the similarity of the article. However, it is difficult to compare content of the article by the meaning. Then, it was considered that outline of the article was the same that the same noun as two articles was included. As a result, rate of the track back spam rises when the similarity of the article is low.

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[1]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合わせたりしたものである場合が多い [1].

2.2　コメントスパム

ブログにはコメント欄がついているものが多く，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．[3]

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，別のブログの記事に自身のブログへのリンクを作成する機能のことである [4]．一般的にトラックバックは，別のブログの内容を引用・参照したときや，別のブログの記事が自身のブログの記事と関連性のある内容であるときに自身のブログの記事が引用・参照したことや内容に関連性があることを引用元のブログに通知する目的で行われる [5].

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トラックバックを送信することがトラックバックスパムであり，コメントスパムと同様にトラックバックの要約に特定のキーワードが含まれているものを禁止したり特定のIPアドレスを禁止したりすることで対策を行うことができる．

トラックバック独特のスパム判別方法としては「参照元へのリンクがないトラックバックは拒否する」という手法がある．基本的にはトラックバックの際にはトラックバック元の記事にトラックバック先へのリンクを張ることになっていることを利用した手法で，トラックバックスパムの多くはトラックバック先へのリンクを行っていないためにこの方法は無料のブログサービスなどでも広く使われている。しかし、参照元へのリンクを行えばトラックバックスパムを行うことができる点と大量のトラックバックを行

う一部の利用者には使いにくいという点が問題である．

大量のトラックバックを行う利用者の例としてはドラマやアニメの感想を主に掲載するブログの利用者が挙げられる．このようなブログでは内容さえ関連していればトラックバックの条件を満たしていると考えられており [6]，記事の内容によっては 100 を超えるトラックバックを送信することもあるといわれている。その場合に大量の参照リンクを本文中に書くことが必要となるため参照元へのリンクの有無による対策は向いていないのである．そこで，上記のようなトラックバック利用法でも利用できるスパム判別手法について実験を行った．

3 関連研究

関連研究として，Lin らのブログの時間的な変化の自己類似性分析を用いたスプログ検出 [2] がある．

Lin らはスプログの投稿が行われる時刻，投稿間隔，投稿の内容，投稿に含まれるリンクの類似性を tf-idf と呼ばれる手法を用いて数値化し、各項目の変化の規則性や 2 つの項目の間の変化の関係の規則性について分析を行うことでスプログの判別を行っている．

今回の実験では Lin らが投稿の内容の類似性を比較する際に用いていた手法をトラックバック先の記事とトラックバック元の記事の内容の類似性を数値化する際に使用している．また，今回の実験では扱っていないが，トラックバックの時間やトラックバック元の記事のリンクの規則性もトラックバックスパム判別に使える可能性はある．

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実験を行った．

- 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，今回の実験では Lin[2] 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。

表 1: 形態素解析の例

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

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

実験ではインターネット上のブログからトラックバック元とトラックバック先の実験用サンプル記事 100 組を収集した．しかし，収集数が少ないことと，ブログ検索サイトなどからブログへ行き手作業で収集したため，サンプル記事に偏りがある可能性がある点は注意が必要である.

得られたサンプル記事をスパムでないトラックバック 50 組とトラックバックスパム 50 組に筆者が下記の基準で分別し，それらのタイトルと本文を用いて実験を行った.

トラックバックがスパムであるかどうかの判断はトラックバック先の記事とトラックバック元の記事の趣旨が同じかどうかという基準で行った．また，トラックバック元の記事がスプログである場合もトラックバックスパムと見なした.

分別した 100 組の記事のタイトルと本文に対して 4 章で説明した類似性の数値化を行い、その結果から図 1，図 2，図 3，図 4 のグラフを作成した．図 1 では、横軸は 2 つの記事の類似性の値、棒グラフはトラックバックの分布，折れ線グラフはトラックバックスパム率を表している．図 2 では横軸は 2 つの記事の類似性の値、縦軸はトラックバックスパムの分布の関係を表している。図 3 では横軸は 2 つの記事の類似性の値、縦軸はスパムでないトラックバックの分布の関係を表している。図 4 ではトラックバックスパムを判定する閾値を 0 から 1 の間で変化させたとき、横軸はトラックバックスパムでないと判定されるトラックバックスパムの数を、縦軸はトラックバックスパムと判定されるスパムでないトラックバックの数を表している。

6 考察

図 1 が示すように、今回の実験では類似性の値が 0〜0.05 の範囲でのトラックバックスパム率は 93%であり，類似性の値が 0.30 以上ではトラックバックスパムは見られなかった．この結果から，トラックバック元の記事とトラックバック先の記事の類似性の値はトラックバックスパム判別に使うことができると考えられる.

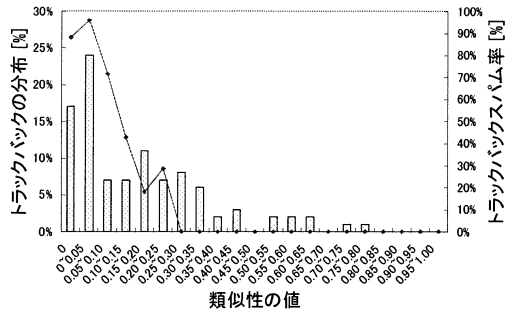

図 1: 2つの記事の類似性の値とトラックバックの分布，トラックバックスパム率

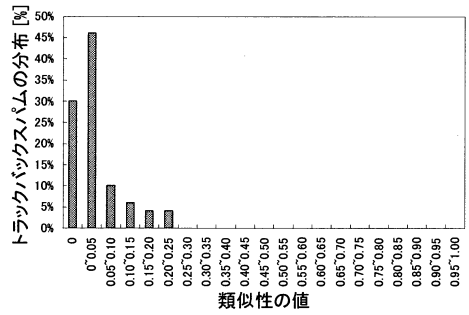

図 2: 2つの記事の類似性の値とトラックバックスパムの分布

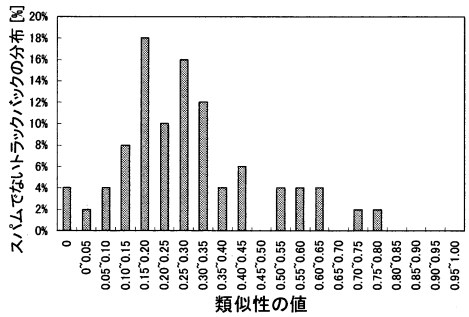

図 3: 2つの記事の類似性の値とスパムでないトラックバックの分布

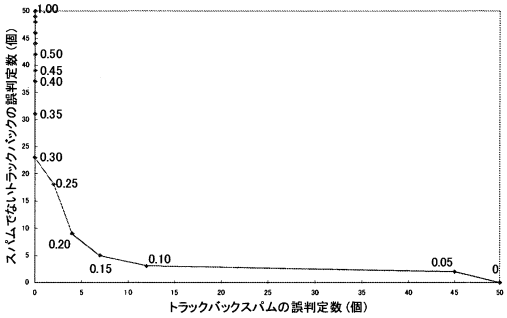

図 4: スパムでないと判定する閾値を変化させたときの誤判定数

6.1 類似性の値とトラックバックの分布の関係

今回の実験でのトラックバックスパムの分布は図2が示すように、全体的に見ても類似性の値は低い傾向にあり、特に類似性の値が 0.05 以下の場合はトラックバックスパム全体の 76%(38個) を占めていた。

スパムでないトラックバックの分布は図3が示すように、類似性の値は 0 から 0.80 とトラックバックスパムの分布と比べるとばらつきがあった。類似性の値 0.15 から 0.35 の間は比較的分布が高い密度で分布しており、スパムで無いトラックバック全体の 56%(28 個) であった。しかし類似性の値が 0.05 以下のものが 6%(3 個) あり，低いながらもこの方法のみでは誤判別の可能性がある.

6.2 トラックバックスパム判定のための閾値設定

スパム判定の閾値に関しては、図4が示すように、トラックバックスパムの誤判定数が小さくなるとスパムでないトラックバックの誤判定数が大きくなることがわかる。閾値を 0.15 に設定するとトラックバックスパムを 86%スパムと判定でき、かつスパムでないトラックバックをスパムと誤判定する割合を 5%にすることができ、最も適当であると考えられる。

6.3 考えられる問題点

この手法に対して考えられる問題点としては含まれる単語の数のみの比較であることと単語が含まれているかどうかしか見ていないことから単語の詰め込みに弱いと考えられる．それに加えて検索エンジンの検索結果を利用して送られるトラックバックはある程度の類似性を持っていると思われるためにこのようなトラックバックスパムにも弱いと考えられる．

7 おわりに

本論文では，機械的に大量に送信されるトラックバックスパムと人間が送信した正当なトラックバックをより正確に判別するために、記事の内容の類似性が利用できることを示すことを目的として実験を行った．実験の結果によりトラックバック元の記事とトラックバック先の記事の類似性の値とトラックバックスパム率に関係があることを示した．

今後の課題としては，判別精度を上げるために比較する単語の数を5個づつから増やしたり，単語の並び方を考慮することなどが挙げられる．また，今回の実験では固有名詞と思われるものが未知語として扱われている場合があったため，形態素解析の際に使う辞書を分析を行うブログに合わせたものを使うことでより精度を上げることができると考えられる．

また検知手法だけでなく根本的な解決のためスパム製作者が無意味なウェブサイトで利益を得ることができないような方法についても考える必要があると思われる．

参考文献

[1] 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.

[2] 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)

[3] IT 用語辞典 e-Words, コメントスパム (2008/4/16), http://e-words.jp/w/

[4] Wikipedia, トラックバック (2008/4/16), http://ja.wikipedia.org/wiki/

[5] IT 用語辞典 e-Words, トラックバック (2008/4/16), http://e-words.jp/w/

[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 (2008/4/16), 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 (2008/4/16), 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 (2008/4/16), http://www.technorati.jp/