社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2007－DBS－143（43）
2007／7／3

トラックバックコミュニティにおける特徴的なブログ記事集合の抽出について

鎌田 基之† 戸田 智子† 黒田 晋矢† 福田 直樹†† 石川 博††

† 静岡大学大学院情報学研究科 〒432–8011 静岡県浜松市城北3–5–1
†† 静岡大学情報学部情報科学科 〒432–8011 静岡県浜松市城北3–5–1
E-mail: †{gs07016,gs07023,gs07037}@s.inf.shizuoka.ac.jp, ††{fukuta,ishikawa}@inf.shizuoka.ac.jp

あらまし 最近注目されているブログには，関連するブログ記事とリンクするためのトラックバックというメカニズムが存在し，複数の関連するブログがそれによって緩やかなコミュニティを形成している．トラックバックと特徴語によるフィルタリングを用いたフォーカストクローリングを行うことで，着目した話題に限定したブログコミュニティの発見が可能である．そのコミュニティ内では，トピックについての詳細を解説したブログ記事とトピックについての感想を書いたブログ記事の2種類が存在する．本研究では，それらの種類のブログ記事の性質の違いを明らかにすると共に，それらを分類する手法を提案する．

キーワード ブログ，トラックバック，分類

On Extraction of Characteristic Blog Entry Sets in TrackBack-based Communities

Motoyuki KAMADA†, Tomoko TODA†, Shinya KURODA†, Naoki FUKUTA††, and Hiroshi ISHIKAWA††

† Graduate School of Informatics, Shizuoka University 3–5–1 Jouhoku, Hamamatsu-shi, Shizuoka, 432–8011 Japan
†† Department of Computer Science, Faculty of Informatics, Shizuoka University 3–5–1 Jouhoku, Hamamatsu-shi, Shizuoka, 432–8011 Japan
E-mail: †{gs07016,gs07023,gs07037}@s.inf.shizuoka.ac.jp, ††{fukuta,ishikawa}@inf.shizuoka.ac.jp

Abstract Blogs have TrackBack mechanisms which make links relevant entries each other. Some related blog entries form lax blog communities by TrackBacks. With our previous research, we coulld find a collection of Blog comunnities focused on certain topics by crawling entries based on TrackBacks and filtering them through characteristic words. Even within the same blog community, there exist some articles and other articles which express feelings about the former. In this paper, we empirically show that there exist clear distinction between these two types of blog articles and we propose a new method for classification.

Key words Blog, TrackBack, classification

1. はじめに

近年，Web2.0というキーワードとともに個人の情報発信の場が個々のウェブサイトからブログへと変化してきている．平成18年4月の総務省による発表によると，ブログ登録者数が平成18年3月末現在で868万人に達している[1]．この貴重な情報源であるブログからの効果的な情報の収集は重要な課題となっている．

ブログには，関連するブログ記事同士をリンクするための，トラックバックというブログ固有の機能がある（図1）．トラックバックでは，相手のブログ記事に自分の記事から参照リンクを張った際に，自分のブログ記事から相手のブログ記事へトラックバックpingを送信することで，参照リンクを張ったことを相手へ通知し，その結果として相手のブログ記事からのリンクを得ることができる．この得られたリンクはトラックバックリンクと呼ばれている．本論文では，図1に示すように，トラックバックpingの送信元であるブログ記事を**トラックバック元のブログ記事**，トラックバックpingの送信先であるブロ

表 1　解説型ブログ記事と感想型ブログ記事の特徴

分類	特徴
解説型ブログ記事	ブログ記事の持つリンク数が多い 話題に関連があるリンクを持つ 情報源などの文章を引用している
感想型ブログ記事	ブログ記事の持つリンク数は少ない 話題に関連する著者の感想が書かれている

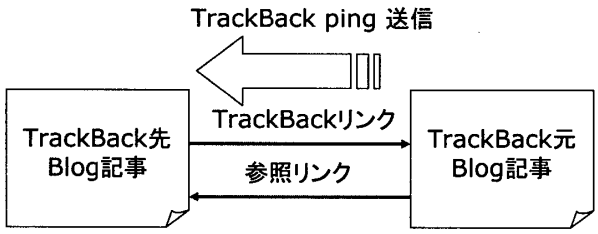

図 1　トラックバック

グ記事を**トラックバック先のブログ記事**と呼ぶこととする．

我々は，文献 [2] で，トラックバックリンクを利用したクローリングをすることで，ある特定の話題に限定したブログ記事の集合を収集するための手法を提案した．さらに，話題の特徴語を用いたフィルタリングを行うことで，収集されるブログ記事集合が着目した話題に限定されることを示した．

本論文では，このトラックバックリンクと特徴語を併用したクローリングによって得られたブログ記事の集合を**トラックバックコミュニティ**と呼ぶこととする．このトラックバックコミュニティ内では，着目した話題に関連した複数のブログ記事が存在しているが，異なる目的に利用可能な，異なった種類のブログ記事が混在している．

ブログ記事の利用目的は，多様である．例えばジャーナリストにとっては，ある話題に関する情報源やブログ記事に書かれていることの根拠となる情報の出所が必要となる場合や，逆に，ある話題に関する意見や感想などの部分を必要としないような場合があると考えられる．マーケティングのために商品の評判を知りたいと考えている人にとっては，商品の仕様・機能解説やその情報の出所などはそれほど必要とせず，個人の使用感や印象などに関連した情報が有益である場合が考えられる．

本論文では，2 種類のブログ記事を仮定する．一つは，ある話題についての情報を多くのリンクなどを用いて解説し，まとめているブログ記事，もう一つは，ある話題についての意見や感想を書いたブログ記事である．本論文では，前者を**解説型ブログ記事**，後者を**感想型ブログ記事**と呼ぶこととする．

解説型ブログ記事と感想型ブログ記事の特徴の概要を表 1 にまとめる．解説型ブログ記事とは，ある話題についての情報を多くのリンクや引用を用いて解説しまとめているブログ記事のことを指す．その特徴は，リンク数が多く，そのリンクがブログ記事の内容の補足となっていること，さらに情報源などの文章を引用していることである．感想型ブログ記事とは，ある話題についての意見や感想を書いたブログ記事のことを指す．特徴は，リンク数が少なく，話題に関連する感想が書かれていることである．

本研究では，文献 [2] の手法を用いて収集したトラックバックコミュニティ内における，解説型ブログ記事と感想型ブログ記事の特徴を明らかにし，ブログ記事をそれらに分類するための手法を提案する．

2. 関連研究

ブログ記事のクローリング手法としては，ブログ検索エンジンでの利用を目的とした手法がいくつか提案されている．

南野らは，ブログ記事を網羅的に収集し，監視するシステムの提案を行っている [3]．南野らの手法では，ブログ記事の発見については，WWW 全体を対象としたクローリングやブログリンク集，ping サーバの更新情報を利用したクローリングによって Web ページを得てから，得られた Web ページをブログであるか個別に判定することでブログ記事の発見を行っている．ブログ記事の収集については，HTML を直接解析することで行っている．

井原らは，画像情報を含むブログ記事の収集とそれらを検索するシステムを構築している [4]．井原らは，ブログ記事の発見を，ブログサービスの提供する Blog 記事更新情報の Web ページを巡回し，ブログトップページの URL を抽出することで行っている．ブログ記事の収集については，初回のみ HTML を直接解析して収集し以降は RSS を利用し，新しいブログ記事のみを取得している．

本研究では，着目した話題に関連するブログ記事のみを対象とする．関連するブログ記事の間で利用されているというトラックバックの特性を生かし，トラックバックリンクを辿ることでブログ記事のクローリングを行う．また，ブログ記事へのトラックバックや本文中のリンクを最大限活用するため，RSS は利用せず HTML を直接解析しながらブログ記事の収集を行う．

谷口らは，ブログ記事本文に書かれたリンクを用いてブログの収集を行い，PageRank と Betweenness Clustering によりブログサイトのコミュニティの抽出と分析を行っている [5]．本研究では，収集においては，本文中のリンクは用いず，トラックバックリンクのみを用いて行う．さらに，ブログサイトではなくブログ記事毎のトラックバックリンクに着目する．

内田らは，トラックバックリンクを辿ることで得られたブログ記事の集合に対してネットワーククラスタリングを行うことで，ある時点の急激に成長する記事クラスタを抽出している [6]．本研究では，ネットワーククラスタリングは行わず，特徴語を用いて限定されたトラックバックコミュニティにおいて実験を行う．

中島らは，トラックバック利用状況の調査を行うことで，トラックバックリンクでつながったブログ記事の関係について調査している [7]．トラックバックによる緩やかなコミュニティ形成を明らかにしており，ブログのコミュニティ発見に対するトラックバックの重要性を指摘している．本研究では，トラックバックのリンクを辿ることで得られたトラックバックコミュニティにおいて，さらに特徴的なブログ記事の集合の抽出を行う．

3. アプローチの概要

本論文では，関連した話題のブログ記事の間で利用されているトラックバックに基づいたクローリングによって得られたトラックバックコミュニティ内において，解説型ブログ記事と感想型ブログ記事の特徴を明らかにし，ブログ記事をそれらの種類に分類するための手法を提案する．

最初に，トラックバックに基づくブログ記事のクローリングを行い，トラックバックコミュニティの抽出を行う．そのクローリングの過程で，話題の限定とトラックバックスパム対策のためのフィルタリングを行うことで，対象とする話題に限定されたトラックバックコミュニティの収集を行う．次に，抽出したトラックバックコミュニティ内で，ブログ記事が持つリンクを抽出する．抽出したリンク情報の特徴を利用して，解説型ブログ記事と感想型ブログ記事への分類を行う．

3.1 トラックバックコミュニティの抽出

3.1.1 ブログ記事の収集

本研究で用いるトラックバックコミュニティの収集は，トラックバックリンクを辿ることによって行う．SEED ブログ記事（収集の起点となるブログ記事）は，トラックバックを 1 つ以上受けているブログ記事とする．収集はトラックバックリンクを辿って行われ，辿るトラックバックリンクがなくなれば収集を終了する．ただし，起点となるブログ記事を元に特徴語というものを設定し，その特徴語が本文に含まれるブログ記事のみを収集することとする．この方式を用いることによって，トラックバックスパムが収集されることを防ぐと共に，収集するブログ記事を，起点となるブログ記事で扱われている話題に絞り込むことが可能で，トラックバックスパムによるトラックバックコミュニティ内での話題の混濁が起こることを防ぐ．トラックバックスパムとは，起点のブログ記事の内容と無関係な内容が書かれたブログ記事のことである．また，通常トラックバックはブログ記事間で使用されるものであるため，あるブログ記事のトラックバック元であるトラックバックリンク先もまたブログ記事であると仮定できる．本クローリング手法ではクローリング中のウェブページがブログ記事であるかどうかを判定しない．なお，収集される情報はブログ記事の URL，HTML タグを含んだ本文，トラックバック数，トラックバック元のブログ記事 URL である．本アルゴリズムを図 2 に示す．

3.1.2 トラックバックの抽出と本文部分の特定

本文の抽出においては，各ブログサイトが配信している RSS の情報を用いることが考えられる．しかし，ほとんどの RSS では，description に概要として本文の最初の数十文字が書かれているか，本文全文が提供されている場合でも HTML タグを含まない形で本文が記述されている．content として，HTML タグを含んだ本文を提供しているブログサービスもあるが稀である．本研究では，本文中のリンクを必要とするため，ブログ記事の HTML ファイルから直接抽出することを試みた．このために，タグ構造の特徴を利用してデータを抽出するラッパープログラムを，ブログホスティングサービスごとに用意した．

トラックバックの抽出においても，本文の抽出と同様に，ブログホスティングサービス毎に対応する方法で行った．トラックバック数が 0 個の場合は，トラックバック元のブログ記事 URL が存在しないため抽出は行わない．ブログ記事によっては，トラックバックを受け付けない設定にしている場合もある．その場合は，トラックバック数が 0 個の場合として処理する．

```
crawling( Blog b , Topic t ){
  if ( b の本文に t が含まれる ) {
    b の HTML を解析し，すべての TrackBack 元を抽出する;
    解析した情報を格納する;
    foreach tb_origin ( TrackBack 元集合 ) {
      if ( tb_origin が未解析 ) {
        crawling( tb_origin, t );
      }
    }
  }
  else { b を解析済みとする; }
}
```

図 2 クローリングアルゴリズム

3.2 ブログ記事の分類

3.2.1 調査 1：ブログ記事が持つリンク

トラックバックコミュニティ内のブログ記事におけるリンクの調査を行った．ブログ記事の HTML を解析し，$<A>$ タグの属性 href の属性値をリンクとして抽出した．

一つ目の話題としては，「立てこもり事件で警察官が死傷」を扱った．集まったブログ記事は 47 件であった．また，ホスティングサービス毎の記事数は，表 3 のようになっており，集まったブログ記事には，ホスティングサービスによる偏りはみられない．ブログ記事が持つリンク数は，最も多いもので 53 個，最も少ないもので 0 個だった．リンク数が多い記事（表 2）には，1 日に起こった様々な種類のニュースをまとめた記事や，話題とは関係のないアフィリエイトなどのリンクを多く貼った記事や，話題に関連したニュース記事へのリンクやその引用を数多く用いて詳細についてまとめた記事や，本文の最後に関連ニュースやブログへのリンクを羅列した記事などが見られた．リンクが少ない記事には，話題に関連したニュース記事へのリンクを示し，それについての感想を書いている記事や，リンクは持たず話題についての感想を書いている記事などが見られた．

二つ目の話題としては，「ZARD の坂井泉水さん死去」を扱った．集まったブログ記事は 123 件であった．また，ホスティングサービス毎の記事数は，表 4 のようになっており，livedoor Blog が他のホスティングサービスより多く含まれている．ブログ記事が持つリンク数は，最も多いもので 60 個，最も少ないもので 0 個だった．ここで，リンク数が多い記事（表 2）に着目する．リンク数が多い記事には，先ほどの話題が「立てこもり事件で警察官が死傷」のものと似ており，1 日に起こった様々な種類のニュースをまとめた記事や，話題とは関係のないアフィリエイトなどのリンクを多く貼った記事や，話題に関連したニュース記事へのリンクや引用を数多く用いて詳細についてまとめた記事や，本文の最後に関連ニュースへのリンクや，

表 2 「立てこもり事件で警察官が死傷」「ZARD の坂井泉水さん死去」: リンク数上位 10 記事

「立てこもり事件で警察官が死傷」		「ZARD の坂井泉水さんが死去」	
リンク数	内容	リンク数	内容
53	一日のニュースをまとめたブログ記事	60	一日のニュースをまとめたブログ記事
45	一日のニュースをまとめたブログ記事	56	感想，トラックバック先ブログ記事へのリンクを列挙
40	一日のニュースをまとめたブログ記事	49	感想，ニュース記事へのリンクと引用，アフィリエイト多数
28	関連ニュース・ブログ記事へのリンクを列挙	30	感想，関連ニュース・ブログ記事へのリンクを列挙
19	ニュース記事の引用，アフィリエイト多数	24	感想，公式サイトへのリンク，動画へのリンク
17	ニュース記事の引用，アフィリエイト多数	23	ニュース記事の引用，アフィリエイト多数
17	ニュース記事の引用，アフィリエイト多数	22	感想，ニュース記事へのリンクと引用，動画へのリンク
14	感想，ニュース記事へのリンクと引用	20	感想，ニュース記事へのリンクと引用
13	感想，ニュース記事へのリンクと引用	20	感想，公式サイトへのリンク，アフィリエイト多数
11	感想，ニュース記事へのリンクと引用	17	感想，ニュース記事へのリンク，アフィリエイト多数

表 3 「立てこもり事件で警察官が死傷」: ホスティングサービスの内訳

ホスティングサービス	記事数
アメーバブログ	6
livedoor Blog	7
ココログ	11
goo ブログ	11
Seesaa ブログ	7
ウェブリブログ	2
エキサイトブログ	3

表 4 「ZARD の坂井泉水さん死去」: ホスティングサービスの内訳

ホスティングサービス	記事数
アメーバブログ	15
livedoor Blog	48
ココログ	9
goo ブログ	14
Seesaa ブログ	16
ウェブリブログ	18
LOVELOG	1
AutoPage	1
ヤプログ！	1

関連するブログとしてトラックバック先ブログ記事へのリンクを羅列した記事などが見られた．次に，リンク数が少ない記事に着目する．リンクが 1 つもなかった記事は 27 件あった．この 27 件のブログ記事をみてみると，話題についての感想を書いているブログ記事が多数であったが，その中でリンクは持たないが，どこかのニュース記事を本文中で引用している記事が 4 件見られた．

3.2.2 調査 2：トラックバックコミュニティにおけるリンクの共起

先ほど用いた「立てこもり事件で警察官が死傷」の話題を持つ 47 件のブログ記事に対して，同じリンクがどれだけ現れているかの調査を行った．その結果を表 5 に示す．goo ニュース記事へのリンクが 47 記事中 5 記事に出現しており最も多かった．その他，4 記事に現れたリンクは Yahoo ニュース記事へのリンクが 2 種類，goo ニュース記事へのリンクが 1 種類，ブログランキングへのリンクが 2 種類であった．ここで，goo ニュース記事へのリンクを持っていたブログ記事は，すべてブログホスティングサービスが goo のものであった．goo ニュース記事では，「この記事についてブログを書く」という，goo のブログ向けに簡単にブログ記事にハイパーリンクを作成するサービスを提供している．このことが理由となって，goo ニュース記事へのリンクが多く共起したと考えられる．goo 以外のホスティングサービスでも同様のサービスを提供しているところもあるが，今回のトラックバックコミュニティ内では goo ニュース記事以外には見られなかった．Yahoo ニュース記事へのリンクが現れた 1 種類目の 4 記事は，livedoor が 1 記事，ココログが 2 記事，goo が 1 記事であり，2 種類目の 4 記事は，ウェブリブログが 1 記事，Seesaa が 1 記事，goo が 1 記事，ココログが 1 記事であった．Yahoo ニュース記事については，ホスティングサービスによる偏りはみられなかった．

続いて，「ZARD の坂井泉水さん死去」の話題を持つ 123 件のブログ記事に対しても同様の調査を行った．その結果を表 5 に示す．タグ「ZARD」へのリンクが 124 記事中 19 記事に出現しており最も多かった．livedoor Blog では，ジャンル分けのためにブログ記事毎に自由に付加できる「タグ」が存在する．表 4 に示したように，livedoor Blog のブログ記事が多いため，livedoor Blog 固有のタグによるリンクが多く出現したと考えられる．しかし，それを除けば，13 記事に現れた Yahoo ニュース記事へのリンク，7 記事に現れた ZARD オフィシャルサイトへのリンク，6 記事に現れたサンスポニュース速報へのリンクと，ブログ記事の情報源となると思われるウェブページへのリンクの共起が多くみられた．

3.3 分類方法の提案

調査 1 の結果より，トラックバックコミュニティにおいて，リンク数が多いブログ記事については解説型ブログ記事の傾向が強く，リンク数が少ないブログ記事については感想型ブログ記事の傾向が強いことが確認された．しかし，リンク数が多いブログ記事の全てについては，必ずしも解説型ブログ記事の傾向が強いとはいえなかった．その原因として挙げられるのは，アフィリエイトやブログランキングへのリンク，ブログサービス固有のタグによるリンクがリンク数に含まれていることである．話題に関連の少ない無駄なリンクを計測してしまっているため，結果に影響を与えていることが考えられる．

調査 2 の結果より，トラックバックコミュニティ内では多く

表 5 「立てこもり事件で警察官が死傷」「ZARD の坂井泉水さんが死去」: 共起リンク上位 5 つ

「立てこもり事件で警察官が死傷」		「ZARD の坂井泉水さんが死去」	
現れた記事数	リンク先	現れた記事数	リンク先
5	goo ニュース記事	19	タグ「ZARD」 (livedoor Blog)
4	Yahoo ニュース記事	13	Yahoo ニュース記事
4	goo ニュース記事	12	タグ「坂井泉水」 (livedoor Blog)
4	ブログランキング「人気 blog ランキング」	7	ZARD オフィシャルサイト
4	ブログランキング「にほんブログ村」	6	サンスポニュース速報

表 6 分類のルール

分類	ルール
解説型ブログ記事	一定の閾値数以上のブログ記事で共起したリンク数がある閾値以上のブログ記事
感想型ブログ記事	一定の閾値数以上のブログ記事で共起したリンク数がある閾値未満のブログ記事

のリンクが共起し，特にニュース記事へのリンクが多くのブログ記事に現れることがわかった．ホスティングサービスの影響を受けて共起したリンクもいくつか見られたが，アフィリエイトやブログランキングへのリンク，ブログサービス固有のタグによるリンクを除けば，話題の情報源となりうるウェブページへのリンクが多く見られた．

以上をふまえ，ブログ記事の分類のルールを表 6 のように定める．トラックバックコミュニティ内で，リンク数が多いブログ記事を解説型ブログ記事として分類し，リンク数が少ないブログ記事を感想型ブログ記事として分類する．ただし，調査 2 より，ブログランキングなどのリンクを除けば，トラックバックコミュニティ内で共起するリンクは，話題の情報源となりうるウェブページへのリンクであることが多いため，ある一定の閾値数以上のブログ記事で共起したリンクに着目し，そのリンク数が多いブログ記事を解説型ブログ記事として分類する．例えば，閾値を 3 とした場合は，トラックバックコミュニティ内で 3 記事以上に現れたリンクのみを計測の対象として，再びブログ記事の持つリンク数を計測することになる．また，調査 1 より，アフィリエイトなどのリンクが大量に存在することで，情報源へのリンクが全く存在しなくても解説型ブログ記事として分類されてしまうことが考えられるため，ブログ記事がもつリンクの抽出については，著名なアフィリエイト，ブログランキング，ブログサービス固有のタグに含まれる URL をチェックすることで，それらを無視することとした．

4. 評価実験

ここでは，3.3 で述べた分類方法の評価を行う．共起したリンクの閾値については，2 とした．トラックバックコミュニティ内で 2 記事以上で共起したリンクを，共起したリンクとして分類に用いる．トラックバックコミュニティ内で 1 記事にしか現れなかったリンクは，分類には用いない．次に，分類の基準となる，ブログ記事が持つ共起したリンク数の閾値は，1 とした．共起したリンクを 1 個以上持つブログ記事を解説型ブログ記事，1 個も持たないブログ記事を感想型ブログ記事とする．

4.1 話題：フジ『発掘！あるある大事典 』の納豆特集で捏造

話題は「フジ『発掘！あるある大事典 』の納豆特集で捏造」を用いた．集まったブログ記事は 156 件であった．その中で，解説型ブログ記事として分類されたブログ記事の中から，特に，共起したリンク数が多い上位 10 記事のリンク数と内容を表 7 に示す．表 7 中の，「共起したリンク数」とは，トラックバックコミュニティ内で 2 記事以上に現れたリンクを計測したもの，「内容」とは，ブログ記事の本文の内容を示したものである．

表 7 より，すべてのブログ記事が，話題に関連したニュース記事へのリンクを複数持っており，そのリンク先の内容を引用しているものが 10 記事中 8 記事存在している．本文中には著者による感想が含まれているブログ記事も存在するが，「内容」から，それ以上に解説型ブログ記事としての傾向が強いと考えられる．したがって，トラックバックコミュニティ内において共起したリンク数が多い記事を解説型ブログ記事として分類することは有効である．

次に，感想型ブログ記事として分類されたブログ記事である，共起したリンク数が 0 個であったブログ記事は 75 件あり，その内容を調査したものを表 8 に示す．表 8 中の，「引用あり」とは，ニュース記事などをブログ記事中で引用し，さらに引用元を示していたブログ記事，「リンクあり」とは，ニュース記事などへのリンクをブログ記事中で示していたブログ記事，「引用・リンクあり」とは，上記の 2 つの条件を共に満たしたブログ記事，「感想のみ」とは，引用やリンクが存在しない，感想だけが書かれていたブログ記事である．

表 8 より，「感想のみ」が書かれていたブログ記事が 40 記事存在し，感想型ブログ記事としての傾向が強いと考えられる．「引用あり」「リンクあり」「引用・リンクあり」についても，その引用数やリンク数は 1 個から 3 個程度であったため，解説型ブログ記事として分類されるべきブログ記事とは言えないと考えられる．そのため，「引用あり」「リンクあり」「引用・リンクあり」が，感想型ブログ記事として分類されることは妥当だと考えられる．

4.2 考 察

解説型ブログ記事として分類された，共起したリンク数が多いブログ記事には，解説型ブログ記事としての傾向が強く見られたが，そのブログ記事では，話題に関連した感想も少なからず存在していた．一方，全く感想を持たないブログ記事は，表 7 では，共起したリンク数 28 個のブログ記事 1 件だけであった．しかし，ブログというものは個人の意見を表現するもので

表 7 「納豆特集で捏造」: リンク数上位 10 記事

共起したリンク数	内容
28	ニュース記事へのリンクが多数，ニュース記事の引用，公式サイトなど関連ページへのリンク
18	ニュース記事へのリンクと引用，公式サイトなど関連ページへのリンク，感想
18	ニュース記事へのリンクと引用，公式サイトなど関連ページへのリンク，感想
12	公式サイトへのリンクと引用，ニュース記事へのリンクを列挙，感想
8	ニュース記事へのリンクと引用，感想
7	ニュース記事へのリンクを列挙，感想
6	ニュース記事へのリンクと引用，感想
6	ニュース記事へのリンクを日付毎に整理，感想
4	ニュース記事へのリンク，公式サイトへのリンク，感想
4	ニュース記事へのリンクと引用，感想，動画

表 8 「納豆特集で捏造」: リンク数 0 個のブログ記事の内容

内容	感想のみ	引用あり	リンクあり	引用・リンクあり
記事数	40	19	8	8

もあるため，解説型ブログ記事としての傾向が強ければ感想が書かれていても，解説型ブログ記事として分類することは問題はないと考える．

表 7 に示されなかったブログ記事についても内容を調査した．共起したリンク数が少ないブログ記事には，情報源へのリンクを持つものの，感想型ブログ記事としての傾向が強いものも見られた．共起したリンク数が 0 個であったブログ記事には，共起したリンク数は 0 個であっても，情報源となりうるリンクを持つものが見られた．それらは解説型ブログ記事とは呼べないものの，その性質を少しは含んでいると考えられる．

今回は，ブログコミュニティ内で共起したリンク数を用いることでの，解説型ブログ記事と感想型ブログ記事の分類の方針を示したが，リンク数の明確な基準を決定するまでには至らなかった．

5. おわりに

トラックバックに基づいたクローリングによって得られたトラックバックコミュニティ内において，解説型ブログ記事と感想型ブログ記事の特徴を明らかにした．トラックバックコミュニティ内においてブログ記事が持つ，複数のブログ記事において共起したリンク数に基づいた分類手法の提案を行い，小規模な実験によりその効果を確かめた．

解説型ブログ記事と感想型ブログ記事の分類に利用する特徴の妥当性については，今後も検討が必要である．共起したリンク数の他に，本来のリンク数やトラックバック数を特徴として用いることを検討していく．解説型ブログ記事と感想型ブログ記事に加え，両方の性質を併せ持つ準解説型ブログ記事と呼べるような新たな分類を検討していく．今回 1 つの話題に関するトラックバックコミュニティにおいて実験を行ったが，さらに事例を増やし，共起したリンク数がいくつ以上になれば解説型ブログ記事として分類するかなどを検討していく．

また，文献 [2] で行っているブログ記事の推薦の提示結果に，今回の分類を適用することを考えている．

謝辞 本研究の一部は科学研究費補助金基盤研究（B）（課題番号 19300026）の助成による．

参考文献

[1] ブログ及び SNS の登録者数（平成 18 年 3 月末現在），総務省報道資料（平成 18 年 4 月 13 日），http://www.soumu.go.jp/s-news/2006/060413_2.html，accessed 2007.6.8.

[2] 鎌田基之，福田直樹，石川博，“TrackBack と特徴語に基づく Blog クローリングと Blog 記事の推薦”，DEWS2007 A8-4，2007.

[3] 南野朋之， 鈴木泰裕，藤木稔明，奥村学，“ブログ の自動収集と監視”，人工知能学会論文誌，Vol.19，No.6，pp.511-520，2004.

[4] 伊原 伸介，林 貴宏，尾内 理紀夫，“もぶろげっと: 画像情報を含むブログ記事検索システム”，インタラクティブシステムとソフトウェアに関するワークショップ (WISS2005) 論文集，pp. 69-74，2005.

[5] 谷口智哉，松尾豊，石塚満，“Blog コミュニティの抽出と分析”，第 6 回セマンティックウェブとオントロジー研究会 人工知能学会資料，SIG-SWO-A401-08，2004.

[6] 内田 誠，柴田 尚樹，“ブログ記事ネットワークからの emering topic の抽出と可視化”，人工知能学会第 20 回全国大会，2006.

[7] 中島伸介，舘村純一，原良憲，田中克己，植村俊亮，“ブログ空間におけるトラックバック利用状況の調査および考察”，DEWS2006 1B-i6，2006.