

# 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

## Analyzing Trackback Usages in Blogspace

中島 伸介<sup>▽</sup> 館村 純一<sup>▽</sup>  
原 良憲<sup>△</sup> 田中 克己<sup>△</sup> 植村 俊亮<sup>▽</sup>

Shinsuke NAKAJIMA Junichi TATEMURA  
Yoshinori HARA Katsumi TANAKA  
Shunsuke UEMURA

近年のブログの普及等により、Web 空間は情報を獲得する場のみならず、情報発信が可能なコミュニケーション空間となりつつあり、ブログの登場により Web の利用形態が変化してきていると考えている。最近では、ブログに特化したブログコミュニティ解析に関する研究や、ブログ解析に基づいたトピック検索や評判情報検索に関する研究も行われている。しかしながら、ブログ記事同士をつなげるリンクにはブログ特有のものとしてトラックバックがあり、ブログ同士の関係を解釈する上では重要な要素でありながら、これらブログマイニングにおいて、トラックバックを積極的に解析している例はほとんどない。本研究では、ブログにおけるトラックバックの利用状況の調査を行い、ブログマイニングにおけるトラックバック解析の意義について考察する。

Recently, blogs have become widely used by general users as tools for putting out the information quickly and easily. Thus, Web space becomes not only a place for getting information but also a place for communication. It can be said that blogs have changed the way people use the Internet. Some researchers do blog analysis such as blog community analysis and reputation retrieval using blog data. However, they do not focus on trackback links in their studies as blog analysis, even though it is a very important element in understanding the relations between blog entries. Therefore, we analyze trackback usages and investigate its importance in understanding blogspace.

### 1. はじめに

2005年5月の総務省の発表[1]によれば、2005年3月末時点で国内のブログ利用者は約335万人に達しており、ブログの普及が急速に進んでいるといえる。ブログ情報は、いわゆる一般ユーザが主なコンテンツ供給者であるので、その情報は

<sup>▽</sup> 正会員 奈良先端科学技術大学院大学 情報科学研究科

{shin, uemura}@is.naist.jp

<sup>△</sup> 非会員 NEC Laboratories America, Inc.

tatemura@sv.nec-labs.com

<sup>\*</sup> 非会員 京都大学経営管理大学院

hara@gsm.kyoto-u.ac.jp

<sup>△</sup> 正会員 京都大学大学院情報学研究科

tanaka@dl.kuis.kyoto-u.ac.jp

世論を反映したものであるという考え方ができる。また、更新が容易であることから、即時性の高い情報源であり、実社会における最近のトピック検出[2]や、商品やイベント等に対する評判情報検索[3]にも応用されており、今後もその重要性が高まると考えている。

したがって、これまでに、Web検索エンジンの構築や、Webマイニングに関する研究が行われており、対象をブログに絞ったブログ検索エンジンの構築や、ブログマイニングに関する研究も行われるようになった。また、他のWebコンテンツに比べ、コンテンツの著者が明確であるために、ある特定分野の有識者を発見しようとする研究[4]も行われている。

ブログ関連研究において、ブログの解析を行う際には、ブログ記事間のリンク構造の解析が不可欠であるが、ブログ記事間のリンクにはブログ特有のものとしてトラックバックがあり、ブログ同士の関係を解釈する上では非常に重要な要素を含んでいるといえる。

しかしながら、これまでに行われているブログ解析において、トラックバックを積極的に解析している例はほとんどない。考えられる理由としては、ブログ空間におけるトラックバックの位置付けが不明確であることが挙げられる。

トラックバックとは、IT用語辞典e-words[5]によると、“ウェブログ(ブログ)の機能の一つで、別のウェブログへリンクを張った際に、リンク先の相手に対してリンクを張ったことを通知する仕組みのこと。”

とあり、通常は相手の意図とは関係なく、他のブログエントリーからのリンクを、自分のブログエントリーに対して貼ることができる機能である。上述の通り元々の定義によれば、トラックバックリンクは、ブログエントリー間の通常のリンクとペアで利用されるはずであるが、通常のリンクを伴わない場合でも、トラックバックping(トラックバックリンクを貼ってもらうための要請)を送信すれば、自動的に先方からトラックバックリンクが貼られてしまうのが現状である。

実際、このような“通常リンクを伴わないトラックバックリンク”は存在し、トラックバック利用のモラルに関する議論もWeb上で展開されている[6]。このように現実問題として、トラックバックリンクの利用状況が明確ではないために、トラックバックリンクにより接続された2つのブログエントリーの関係の解釈が不明であり、この状況がトラックバックを考慮したブログ解析を困難にしているといえる。

本研究では、ブログ空間におけるトラックバックの利用状況の調査を行い、トラックバックに基づくブログ記事間の関係について考察することを目的としている。なお、ここから得られる知見は、ブログ解析においてトラックバックを考慮するための重要な基礎データと成り得ると考えている。

### 2. トラックバックの概要と関連研究

#### 2.1 トラックバックの概要

図1にトラックバックリンクの概略図を示す。トラックバックに関しては、1節にて述べた通り、元々は他人のブログエントリーをリンク参照したことを相手のエントリーに対して通知するための機能である。トラックバックの手順としては、まず他人のブログエントリーをリンク参照した際に、リンク参照したことを知らせるトラックバック ping の送信を行う。するとトラックバック ping を受け取ったブログエントリーにて、自分のエントリーを参照しているブログエントリーへのリンクが自動的に貼られることになる。

次に、図2に空トラックバックの概略図を示す。トラックバックの機能上、リンク参照していなくてもトラックバック

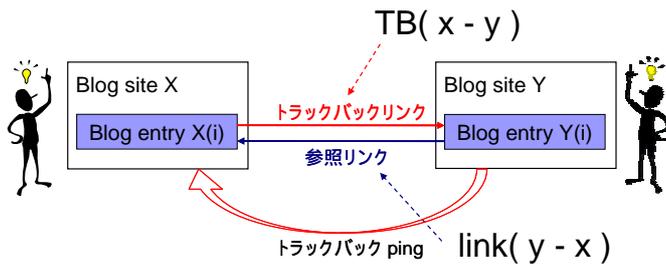


図1 トラックバックリンク  
Fig.1 Trackback link

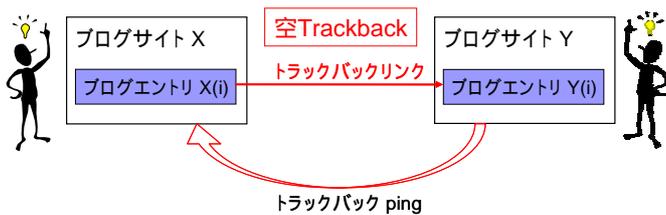


図2 空トラックバックリンク  
Fig.2 blank-trackback

ping を送信すれば、図2のように相手のブログエントリーからトラックバックリンクを貼られることになる。本研究では、このようなトラックバックリンクを、対となる参照リンクが欠如しているという意味で、“空トラックバック”と呼ぶ。

## 2.2 ブログ関連研究

Kumar らは、blogspace と名づけたハイパーリンクによるブログ群のつながりに注目し、このBlogspaceにおけるブログコミュニティの抽出と、このブログコミュニティの進化に関する調査研究を行っている[7]。

奥村らのグループでは、blogの収集と、更新の監視、また収集したデータを用いたテキストマイニングを行うツールである、blogWatcherに関する研究開発を行っている[2][3]。このblogWatcherでは、話題(キーワード)に関する注目度の自動分析や、評判情報の抽出などを可能にしている。

Tseng らは、関連が深いblogサイト同士を近くに配置されるようなクラスタリング手法により、blogコミュニティの視覚化を行っている[8]。

Fujimura らは、EigenRumor と呼ばれるblogコミュニティ分析に基づいたアルゴリズムにより実現される、blogランキングに関する研究を行っている[9]。

著者らは、Webコンテンツの信頼性評価を目的としたブログ解析手法に関して提案している[4]。この中で、ブログエントリーが形成するブログスレッドを定義し、このブログスレッド内におけるブログサイトの役割の判別方法に関して議論している。

ただし、これらの研究では、エントリー同士の関係を調べる際にはトラックバックリンクによるつながりは考慮できていない。

## 3. ブログ空間におけるトラックバック利用状況の調査

### 3.1 ブログデータの収集とトラックバックデータの抽出方法

本研究では、PING.BLOGGERS.JP[10]が公開している

JPドメインのブログRSSリストに基づいて、2005年10月および11月に投稿されたブログエントリーのうち、5,743,913のブログエントリーを解析データとして利用した。なお、これらのブログエントリーは、URLの一部が以下のものと合致するブログホスティングサイトにて投稿されたエントリーである。

nifty.com/	typepad.com/	mo-blog.jp/
typepad.jp/	weblogs.jp/	no-blog.jp/
blog.ocn.ne.jp/	blogzine.jp/	de-blog.jp/
hatena.ne.jp/	livedoor	rakuten.co.jp/
blog.goo.ne.jp/	blog.excite.ne.jp/	exblog.jp/
seesaa.jp/	blog.dion.ne.jp/	

以下、これらのサイトをトラックバック解析対象ブログサイトと呼ぶ。

## 3.2 トラックバックを伴うブログエントリー間の関係

### 3.2.1 ブログエントリー間のリンク関係

元々のトラックバックの定義によれば、トラックバックリンクに対応する逆向きの通常リンクが存在しているべきであるが、その現状について確認する。

ここで、ブログエントリー(i)から、エントリー(j)へのトラックバックリンク(TB(i-j))が存在する時の、この2つのエントリー間のトラックバックと通常リンクの状況について調査した。図1に示す通り、TB(i-j)は、エントリー(i)に存在するトラックバックリンクであり、エントリー(j)から送信したトラックバック ping に基づいて生成されるものを指し、link(j-i)は、(j)から(i)への通常のリンクを示す。

表1に、トラックバック関係にあるブログエントリー間のリンク関係を示す。すなわち、TB(i-j)が存在する全てのブログエントリーの組み合わせに対して、その時のTB(i-j)、link(i-j)、link(j-i)の有無のパターンについて調べたものである。およびxにて、その存在の有無を示している。

表1 ブログエントリー間のリンク関係  
Table 1 Link-base Relationship

	TB(j-i)		TB(j-i) ×	
	link(j-i)	link(j-i) ×	link(j-i)	link(j-i) ×
link(i-j)	0.03 %	0.30 %	0.01 %	0.78 %
link(i-j) ×	0.30 %	11.36 %	0.62 %	86.59 %

ここで空トラックバックとは、“link(j-i)”が存在しない状態であり、実に合計99.03%となる。すなわち、今回の解析対象サイトにおいては、ほとんど全てに近いトラックバックが“空トラックバック”であるといえる。したがって、トラックバックの利用においては、元々の定義である“通常リンクを貼ったことの通知”とは異なる状況であることが明らかである。著者らによる過去の発表論文[11]において解析を行った2004年12月のデータにおいては、空トラックバックは全体の85.5%であったことを考えると、通常リンクを伴うトラックバックに比べて、空トラックバックの存在が増加傾向にある可能性も考えられる。

また、表1において注目したいのは、“TB(j-i)”が存在する、相互トラックバックの状態である。これは、合計11.99%となる。空トラックバックとは、リンクを相手に対して貼らないのに、トラックバック ping を送信することで自分宛のリンクを生成して、自分のサイトへのアクセス量の増大や、PageRank 値の向上を狙うといった悪意を伴う目的で利用されているとも考えられる。しかしながら、相互にトラックバック

バックをかけるということは、お互いに相手の存在を認めていると考えられ、“空トラックバックは悪意を伴う”という認識が必ずしも当てはまらないことになる。

そこで、次節にて、空トラックバックを伴う2つのブログエントリー同士は、内容的にどのような関係になっているのかということについて調べる。

3.2.2 ブログエントリー間の内容的関係

ここでは特に、空トラックバックを伴う2つのブログエントリー同士の内容的関係について調べる。対象データは、2005年10月から2005年11月までのエントリーで空トラックバックを伴うものから無作為に選択した50ケースに関して、人手により内容的状況の確認を行った。表2に、トラックバックの内容的状況を示す。なお、表中の全てのケースにおいて、TB(i-j)は存在している。

表2 ブログエントリー間の内容的関係  
Table 2 Content-base Relationship

	TB(j-i)		TB(j-i) ×	
	言及	言及 ×	言及	言及 ×
関連あり	0%	8%	2%	56%
関連なし	-	4%	-	30%

ここで、“関連”とは、2つのエントリー同士の内容的関連の有無であり、話題に関連性があるかどうかということを示している。“言及”とは、トラックバック ping を送信したエントリー(j)内において、エントリー(i)に対する言及の有無を表している。

関連のないエントリーに対するトラックバックは、合計34%存在した。内容を確認したところ、全てがアダルト系のサイトであり、ほぼスパムと断定できるものであった。このようにトラックバックスパムの存在を確認できたが、空トラックバック全体に占める割合としては、半分以下であり、やはり“空トラックバックは迷惑なスパムである”とは必ずしも言えないことも判明した。

また、トラックバック ping を送信していながら、送信先のエントリーに関する言及がないながらも、内容的には関連しているものが56%も存在している。これらのエントリーを実際に関連してみたが、内容的には非常に関連の深いものでありながら、他のエントリーの内容に関しては全く触れていない。例えば、同じ映画や、テレビドラマ等のレビューを書いているもの同士が、お互いの記事には言及することは無しに、空トラックバックによってつながりを持つというケースを数多く確認した。

さらにこのような関係は1つのエントリーに対してだけでなく、複数のエントリーと同時に相互トラックバックの関係を保っているケースも少なくない。そうすることで、これらのエントリーは明示的かつ強固なつながりではなく、緩やかなつながりで関係を保ったコミュニティのようなものを形成しているものと思われる。

3.3 通常リンクとトラックバックリンクの違い

表3に、ブログエントリー数に対するリンク数およびトラックバック数を示す。表中のエントリー数とは、トラックバック解析対象ブログサイトにて2005年10月および11月に投稿されたエントリー数であり、5,743,913エントリーである。リンク数とは、この解析対象ブログサイトのエントリーにおいて記述されているリンク数であり、ブログの著者(以下、プログラマーと呼ぶ)が、投稿の際に記述したものに限定している。つまり、過去のアーカイブへのリンクや、トラックバックリン

クのように、自動的に生成されるものは含まない。トラックバック数は、そのエントリーへ ping が送信されたことで、そこから ping 送信元へ貼られているトラックバックリンクの数である。また、括弧内は、各々のエントリー数当たりの数である。

表3 リンク数とトラックバック数の比較

Table 3 Comparison between number of links and number of trackbacks

リンク数	2,454,032 ( 0.427 )
トラックバック数	892,723 ( 0.155 )

表3が示すように、この期間のエントリー当たりのリンク数が、0.427であるのに対し、エントリー当たりのトラックバック数は、0.155と、およそ1/3程度であった。他のエントリーとのつながりを持つ手段としては、通常のリンクが主流であるといえるが、トラックバックに関してもかなりの数が存在しているということが分かる。

次に、このリンクにより形成されるブログコミュニティと、トラックバックにより形成されるブログコミュニティの違いについて考察する。ブログスレッドとは、著者らが提案したリンクを介したエントリーの集まりのことであり、似通ったブログエントリー同士が形成するコミュニティのようなものである。

表4 リンクに基づくスレッド数とトラックバックに基づくスレッド数の比較(2005年10月)

Table 4 Comparison of number of threads between hyperlinks and trackbacks (Oct. 2005)

	エントリー数 50以上	エントリー数 30以上	エントリー数 10以上
通常リンクに基づくスレッド数	26	76	539
トラックバックリンクに基づくスレッド数	38	82	621

ここで、ブログエントリー間のリンクに基づいて形成されるブログスレッドと、トラックバックに基づいて形成されるブログスレッドの比較を行う。データとしては、2005年10月に投稿されたブログエントリーと、その中に存在するリンクおよびトラックバックリンクである。表4に、リンクに基づくスレッド数と、トラックバックに基づくスレッド数を示す。各スレッドに参加しているエントリーの数が、50以上の場合、30以上の場合、10以上の場合のスレッド数を示している。表4によると、各条件とも、リンクによるスレッドよりも、トラックバックによるスレッド数の方が多いことが分かる。表3より、この期間のリンク数がこの期間のトラックバック数の3倍弱であることを考えれば、リンク数当たりのスレッド形成能力は、通常のリンクよりもトラックバックの方が単純に3倍程度高いということがいえる。したがって、ブログコミュニティ解析により、エントリー同士の関係について考察する上では、トラックバックの存在は非常に重要であるといえる。

3.4 考察

3.4.1 トラックバックの元々の定義と利用状況との相違

実験結果から、対となるべき通常のリンクを伴わない“空トラックバック”が99%以上存在することが分かった。また、空トラックバックは必ずしもスパムではなく、緩やかなコミ

コミュニティを形成しているケースも数多く存在することが判明した。元々の定義とは異なる使い方はあるが、これほど大多数になると、むしろトラックバックの定義を改める必要があると思われる。実際の利用状況を踏まえて、トラックバックに対するより適切な解釈を下記のように提案する。

「他のブログサイトのエントリのうち、自分のエントリと関連する内容のものに対して、“私も関連する内容のエントリを書いています”という意味で、トラックバックpingを送信し、自分のエントリ宛のリンクを獲得する行為のこと。」

トラックバックpingを受信したブロガーが許可するかどうかという問題は別にして、現状では上記のような使われ方が大多数である。したがって、トラックバックを考慮したブログ解析を行う際には、上記のような解釈をするべきである。

### 3.4.2 トラックバックスパムの存在

解析したトラックバックのうち34%程度がトラックバックスパムと思われるものであった。ブログ解析の際にトラックバックを考慮すると、このトラックバックスパムが大きなノイズになる恐れもある。しかしながら、これはトラックバックだけでなく、Web空間全体が共有する問題である。表3が示すように、エントリ同士をつなげる手段としては、通常のリンクと共に数的にもトラックバックの存在は無視できない。したがって、ブログエントリ同士の関係を解析する上では、トラックバックを無視するのではなく、他のスパムと同様に、積極的なスパム対策を講じることで、トラックバックを考慮したブログ解析を目指すべきである。

### 3.4.3 トラックバックに基づくブログコミュニティ形成

トラックバックによるブログエントリ同士のつながりにより、通常のリンク以上の効率で、ブログのスレッドを形成していることが分かった。ブログスレッドは、短期的なブログコミュニティともみなすことができる。実際にトラックバックに基づいて形成されているブログスレッドを確認したが、通常のリンクに基づいたブログスレッドとは少々異なる緩やかなコミュニティを形成しているものが多かった。したがって、ブログ空間の詳細な理解を目指した、ブログのコミュニティ発見やコミュニティ解析を行う上では、トラックバックは必要不可欠な解析データといえる。

## 4. まとめ

本論文では、ブログ空間におけるトラックバックの利用状況の調査を行い、トラックバックにより形成されるブログ記事間の関係について考察した。

- ・ 解析したトラックバックのうち全体の99%のトラックバックが空トラックバック(対応する通常リンクを伴わないトラックバック)であることを明らかにした。
- ・ トラックバックを伴うブログエントリ間の内容的関係に関する調査実験では、トラックバックpingを送信しながらも、相手のブログエントリに関する言及がないブログエントリの存在や、相互の空トラックバックに基づく緩やかなコミュニティ形成が行われていることを明らかにした。
- ・ 通常のリンクに基づくブログスレッド形成と、トラックバックに基づくブログスレッド形成の比較を行い、ブログコミュニティ解析におけるトラックバックの重要性を確認した。

ブログ空間の理解を目指したブログ解析における、トラックバックの重要性と、トラックバックの解釈に関する注意点

について述べた。

## [謝辞]

本研究の一部は、科学研究費補助金 基盤(A)(2) (課題番号: 15200010) および 若手研究(B) (課題番号: 17700132) による。ここに記して謝意を表します。

## [文献]

- [1] ブログ・SNS (ソーシャルネットワーキングサイト) の現状分析及び将来予測, 総務省報道資料 (平成 17 年 5 月 17 日), [http://www.soumu.go.jp/s-news/2005/050517\\_3.html](http://www.soumu.go.jp/s-news/2005/050517_3.html)
- [2] 奥村 学, 南野朋之, 藤木稔明, 鈴木泰裕: blog ページの自動収集と監視に基づくテキストマイニング, 人工知能学会, セマンティックウェブとオントロジー研究会, SIG-SWO-A401-01, 2004 年.
- [3] 鈴木泰裕, 高村大也, 奥村 学: "Weblog を対象とした評価表現抽出", 人工知能学会, セマンティックウェブとオントロジー研究会, SIG-SWO-A401-02, 2004 年.
- [4] Shinsuke Nakajima, Junichi Tatemura, Yoichiro Hino, Yoshinori Hara, and Katsumi Tanaka: "Discovering Important Bloggers Based on Analyzing Blog Threads", WWW2005 Workshop on the Weblogging Ecosystem, May 2005. <http://www-idl.hpl.hp.com/blogworkshop2005/nakajima.pdf>
- [5] IT 用語辞典 e-words, <http://e-words.jp/>
- [6] トラックバックの有効な使い方を考える, <http://kotonoha.main.jp/2003/12/09trackback.html>
- [7] Ravi Kumar, et al: "On the Bursty Evolution of Blogspace", The Twelfth International World Wide Web Conference (2003). <http://www2003.org/cdrom/papers/refereed/p477/p477-kumar/p477-kumar.htm>
- [8] Belle L. Tseng, Junichi Tatemura, and Yi Wu: "Tomographic Clustering To Visualize Blog Communities as Mountain Views", WWW2005 Workshop on the Weblogging Ecosystem, 2005. <http://www-idl.hpl.hp.com/blogworkshop2005>
- [9] Ko Fujimura, Takafumi Inoue, and Masayuki Sugisaki: "The EigenRumor Algorithm for Ranking Blogs", WWW2005 Workshop on the Weblogging Ecosystem, 2005. <http://www-idl.hpl.hp.com/blogworkshop2005>
- [10] PING.BLOGGERS.JP, <http://ping.bloggers.jp/>
- [11] 中島伸介, 館村純一, 原 良憲, 田中克己: Blog スレッドの解析に基づく重要な blogger の発見, DBWS2005 情報処理学会研究報告, 2005-DBS-137(I), pp.31-38, 2005 年 07 月.

## 中島 伸介 Shinsuke NAKAJIMA

奈良先端科学技術大学院大学情報科学研究科助手・博士(情報学)・日本データベース学会, 情報処理学会, IEEE CS 各会員。

## 館村 純一 Junichi TATEMURA

NEC Laboratories America 勤務・工学博士・情報処理学会, ACM, IEEE Computer Society 各会員。

## 原 良憲 Yoshinori HARA

京都大学経営管理大学院教授・博士(情報学)・主にハイパーメディアシステム関連の研究開発に従事・情報処理学会, ACM 各会員。

## 田中 克己 Katsumi TANAKA

京都大学大学院情報科学研究科教授・工学博士・主にデータベースの研究に従事・情報処理学会, 日本データベース学会, 人工知能学会, ACM, IEEE CS 等各会員。

## 植村 俊亮 Shunsuke UEMURA

奈良先端科学技術大学院大学情報科学研究科教授・データベースシステムの研究に従事・工学博士・情報処理学会, 電子情報通信学会, 各フェロー, IEEE Fellow。