

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

中島 伸介[†] 舘村 純一^{††} 原 良憲^{†††} 田中 克己^{††††,†††††} 植村 俊亮[†]

[†] 奈良先端科学技術大学院大学 情報科学研究科 〒 630-0101 奈良県生駒市高山町 8916-5

^{††} NEC Laboratories America. 10090 North Wolfe Road, Suite SW3-350, Cupertino, CA 95014, USA

^{†††} 日本電気株式会社 インターネットシステム研究所 〒 630-0101 奈良県生駒市高山町 8916-47

^{††††} 京都大学大学院 情報学研究科 〒 606-8501 京都市左京区吉田本町

^{†††††} 独立行政法人 情報通信研究機構 〒 619-0289 京都府相楽郡精華町光台 3 - 5

E-mail: †{shin,uemura}@is.naist.jp, ††tatemura@sv.nec-labs.com, †††y-hara@da.jp.nec.com,
††††tanaka@dl.kuis.kyoto-u.ac.jp

あらまし 近年，一般ユーザの情報発信ツールとしてブログが普及しつつあり，今後も利用者は増加するといわれている．これにより，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（ブログマイニング）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．したがって本研究では，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．

キーワード ブログ，トラックバック，Web 情報検索

Analyzing Trackback Usages in Blogspace

Shinsuke NAKAJIMA[†], Junichi TATEMURA^{††}, Yoshinori HARA^{†††}, Katsumi TANAKA^{††††,†††††},
and Shunsuke UEMURA[†]

[†] Graduate School of Information Science, Nara Institute of Science and Technology

^{††} NEC Laboratories America

^{†††} NEC Corporation.

^{††††} Graduate School of Informatics, Kyoto University

^{†††††} National Institute of Information and Communications Technology

E-mail: †{shin,uemura}@is.naist.jp, ††tatemura@sv.nec-labs.com, †††y-hara@da.jp.nec.com,
††††tanaka@dl.kuis.kyoto-u.ac.jp

Abstract Recently, blogs have become widely used by general users as tools for putting out the information quickly and easily. Thus, Web space become 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.

Key words blog, Trackback, Web information retrieval

1. はじめに

近年，一般ユーザの情報発信ツールとしてブログの普及が

急速に進んでいる．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，今後もその増加傾向が続き，2007 年 3 月末には約 782

万人に上るといわれている。このようなブログの普及による効果も手伝って、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変化しつつあり、Web の利用形態が変化してきている。

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 [7] や、商品やイベント等に対する評判情報検索 [9] 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いるが、対象をブログ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 [12] [13] 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 [2] 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 [3]。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。

したがって本研究では、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、トラックバックにより形成される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

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。

以下、本論文の構成を示す。2 節では、ブログおよびトラックバックの概要と関連研究について述べる。3 節では、ブログ空間におけるトラックバック利用状況の調査について述べる。4 節では、まとめと今後の方向性について述べる。

2. ブログおよびトラックバックの概要と関連研究

2.1 ブログおよびトラックバックの概要

ブログは、アメリカにおいては 1999 年以降、急速に発達し標準化が進みつつある Web コンテンツであり、最新ニュースをいち早く取りあげ、独自の視点で解説をするブログや、ある特定の分野に対してコラムを展開するブログ、写真を掲載しそれらにコメントを付けるブログなどその内容は様々である。ブログサイトの構築方法としては、「MovableType [4]」などのブログツールを利用してサーバを立ち上げる方法や、ホスティングサービスを行っているサイトに登録することでブログサイトを立ち上げる方法が挙げられる。

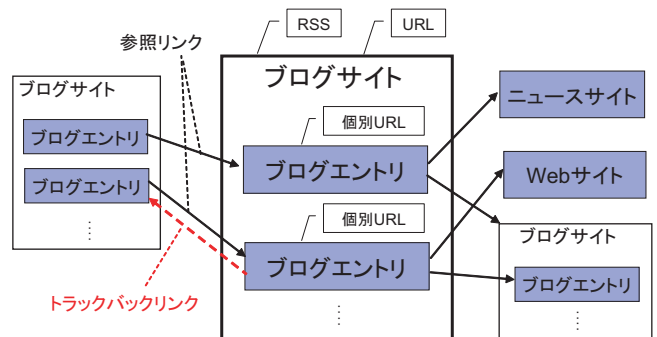


図 1 典型的なブログサイトの例

図 1 に典型的なブログサイトの例を示す。ブログサイトは、そのトップページに最近 (例えば一週間) 書かれた「エントリー」と呼ばれる個別書き込み記事を複数表示している。通常はブログサイトの管理者のみがエントリーを追加することができ、この点が Web 掲示板とは異なる。新しいエントリーが追加されれば、古いエントリーはトップページからは削除されるが、各エントリーが保持している個別 URL を迎れば、トップページから削除された後も閲覧することが可能である。

また、ブログサイトトップページについては、RSS (Rich Site Summary もしくは RDF Site Summary) と呼ばれる XML で記述されたサイトの要約を公開していることが多く、RSS のみを巡回することでブログサイトの更新情報等を取得することが可能となっている。

図 2 にトラックバックの概略図を示す。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

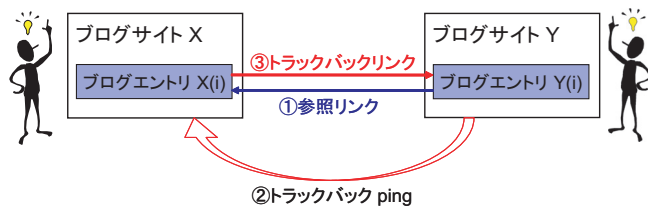


図 2 トラックバックの概略図

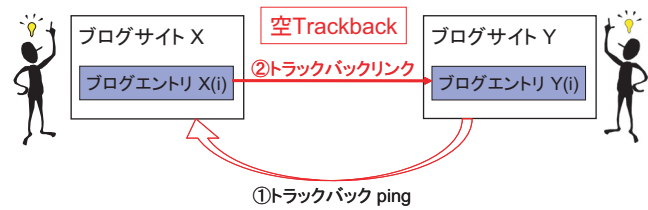


図 3 空トラックバックの概略図

とになる。

次に、図 3 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 ping を送信すれば、図 3 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、対となるべき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”を呼ぶ。

2.2 ブログ関連研究および技術

トラックバックそのものに関する調査および研究はほとんどなされていないが、ブログ関連の研究が盛んに行われているので、それらについて述べる。

• ブログによる情報の広がりに関する研究

Kumar ら [5] および, Gruhl ら [6] は、ブログ空間の進化や広がりに関する調査研究を行っている。

Kumar らは、25,000 のブログサイトとその中の 750,000 本のリンクについて解析している。また、blogspot 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。

Gruhl らは、11,000 以上のブログサイトにおける 400,000 以上のブログエントリーについて解析している。この中で、blogspot におけるマクロな視点によるトピックの伝播の特徴付けと、ミクロな視点による個々のブログ同士のトピックの伝播の特徴付けを試みている。この中で、blogspot において内部的に発生する議論である Chatter と、外的要因により発生する Spikes という尺度を用いて、トピック伝播のモデル化を行っている。

• blog の収集と更新の監視ツール

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 [7] ~ [9]。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。

• blog コミュニティ分析

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 [10]。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 [11]。この手法により、blogger の情報提供能力と情報評価能力の算出を可能にしている。

• リンク構造の時間特性に着目したブログ時系列解析

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 [12] [13]。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。ブログサイトの役割としては、“Topicfinder”、“Agitator”、“Summarizer”などを定義し、ブログスレッドのリンク構造の解析および時系列解析によってこれらの判別方法を提案している。

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。

• トラックバック機能を応用したシステム開発

渡邊らは、公共性の高い情報や評判情報を蓄積し、これをユーザに提供するコミュニケーションシステムに関する研究を行っている [14]。ただし、これはトラックバック機能を応用した新たなシステムの提案であり、現状のトラックバックに関する調査および考察に関するものではない。

上記のように、ブログのコミュニティ分析等、ブログエントリー間のつながりについて注目した研究が盛んに行われているが、ブログ特有のつながりであるトラックバックを利用した研究はほとんど見当たらない。したがって、本研究において明らかにしようとするトラックバックの利用状況は、既存の研究にとつて有効な基礎データとなりうると考えている。

3. ブログ空間におけるトラックバック利用状況の調査

本節では、ブログ空間におけるトラックバックの利用状況の調査および考察を行うことで、これまでほとんど明確にされていなかった、トラックバックリンクによって接続されたブログエントリー間の関係について考察する。

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 [15] 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 10,737,320 のブログエントリーを収集し、これらを解析データとして利用した。

なお、トラックバックリンクは、ブログエントリーの HTML 文書の中で、通常のリンクとは区別して掲載されているが、共通のタグルールに基づいて記述されている訳ではないので、各サイト毎にトラックバックリンク検出のためのトラックバックリンク判別ルールを設定する必要がある。本論文では、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対して、トラックバック検出用タグ解析ルールを設定した。

nifty.com/ typepad.com/ mo-blog.jp/

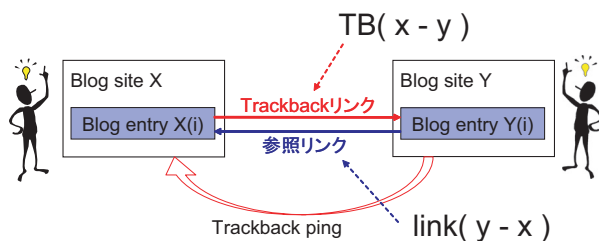


図4 リンクとトラックバックの関係

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/	seesaa.jp/	exblog.jp/
blog.excite.ne.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。収集したブログエントリのうち、このトラックバック解析対象ブログサイトに含まれるものは、約 53.5% に相当する、5,743,913 のブログエントリであった。

3.2 トラックバックを伴うブログエントリ間関係

3.2.1 ブログエントリ間のリンク関係

2.1 節において、空トラックバックの存在について述べたが、本節ではトラックバックを伴うブログエントリ間における、相互のトラックバックリンクおよび通常リンクの関係について調べる。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 (i) から、エントリ (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4に示す通り、TB(i-j) は、エントリ (i) に存在するトラックバックリンクであり、エントリ (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 × にて、その存在の有無を示している。

表1 ブログエントリ間のリンク関係

	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 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

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 [16] 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 ブログエントリ間の内容的関係

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

ここで、“関連”とは、2つのエントリ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 (j) 内において、エントリ (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。また、このうち 4% は、相互トラックバックの関係にあるものであり、内容を確認したところ、やはりお互いにトラックバックスパムを付けているものであった。

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分かった。

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。これらのエントリを実際に閲覧してみたが、内容的には非常に関連の深いものでありながら、

表 3 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の比較

	エントリー数	リンク数	(リンク数/エントリー数)	トラックバック数	(トラックバック数/エントリー数)
2005年10月	3,008,851	1,280,992	0.427	450,297	0.150
2005年11月	2,735,062	1,173,040	0.429	442,426	0.162

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)

	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
通常リンクにより形成されるスレッド数	26	76	539
トラックバックリンクにより形成されるスレッド数	38	82	621

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ようである

3.3 通常リンクとトラックバックリンクの違い

3.1 節および 3.2 節において、トラックバックを伴うブログエントリー間の関係について考察した。本節では、通常リンクとトラックバックの利用状況および性質の違いについて調査および考察する。

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を指す。

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であり、ブログの著者（以下、プログラーと呼ぶ）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、商業サイトへのリンク、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。

表 3 が示すように、2005年10月および11月のエントリー当たりのリンク数は、平均 0.428 (0.427 および 0.429) 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、平均 0.156 (0.150 および 0.162) 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。まずは、ブログエントリーの一時的なコミュニティ形成とみなすことができるブログスレッドについて説明する。

図 5 に、ブログスレッドの例を示す。ブログスレッドとは、

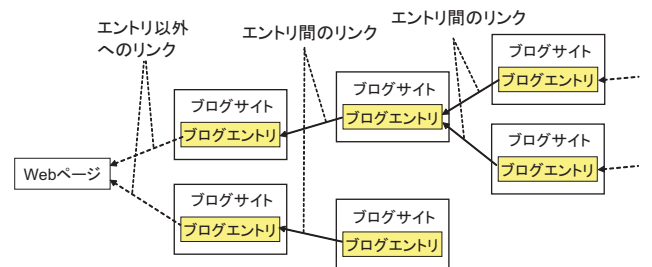


図 5 ブログスレッド

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 [12] [13]、以下のようにブログエントリーをノード、リンクをエッジとした、有向連結グラフとして定義できる ..

$$thread := (V, E)$$

$$V = V_W \cup V_E, \quad E = L_s \cup L_r$$

$$L_s \subseteq \{(e, e') | e \in V_E, e' \in V_W\}$$

$$L_r \subseteq \{(e, e') | e \in V_E, e' \in V_E\}$$

ただし、 V_W は Web ページ集合であり、 V_E はブログエントリー集合である。また、 L_s は、“ブログエントリー以外のページへのリンク”の集合を示し、 L_r は、“ブログエントリーへのリンク”の集合を示す。

すなわち、各プログラーが関連するブログエントリーに対して、自分のエントリーからリンクを貼っていくことで、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。

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くことがわかる。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、単位個数当たりのスレッド形成能力は、通常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

本節では、ここまでに行った、ブログ空間におけるトラックバックの利用状況に関する調査実験結果を踏まえた考察を行う。

- 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、本来は通常のリンクの通知として定義されているトラックバックリンクに対して、実際には、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、緩やかなコミュニティを形成しているケースも数多く存在することを明らかにした。元々の定義とは異なる使い方ではあるが、ここまで大多数になると、むしろトラックバックの定義そのものの解釈の方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트리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。

- トラックバックスパムの存在

実験結果から、確認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트리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트리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。

- トラックバックに基づくブログコミュニティ形成

実験結果から、トラックバックによるブログエン트리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. おわりに

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。以下に、本論文のまとめを示す。

- ブログエン트리からのトラックバックリンクの抽出のためのルールベースを作成し、JPドメインのブログサイトにおけるトラックバックリンクに関する調査実験を行った。

- トラックバックを伴うブログエン트리間に関する調査実験では、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트리間の内容の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。

- トラックバック解析対象ブログサイトにおける、リンク数とトラックバック数の違いを調べた。また、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。

- 上記の調査実験に基づいて考察し、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。

今後は、ブログ解析におけるトラックバックの位置づけに関する調査および考察をさらに進めると共に、日本のドメインだけではなく、海外のブログサイトのトラックバックについても解析範囲を広げていくつもりである。

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クワーキングサイト) 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
- [2] IT用語辞典 e-words, <http://e-words.jp/>
- [3] 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>
- [4] movabletype.org, <http://www.movabletype.org/>
- [5] 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>
- [6] D. Gruhl, et al: "Information Diffusion Through Blogspace", *The Thirteenth International World Wide Web Conference (2004)*.
<http://www2004.org/proceedings/docs/1p491.pdf>
- [7] 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.
- [8] 南野 朋之, 鈴木 泰裕, 藤木 稔明, 奥村 学: blog 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 年.
- [9] 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.
- [10] Belle L. Tseng, Junichi Tatemura, and Yi Wu: "To-

- mographic Clustering To Visualize Blog Communities as Mountain Views”, WWW2005 Workshop on the Weblogging Ecosystem, 2005.
<http://www-idl.hpl.hp.com/blogworkshop2005>
- [11] 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>
- [12] 中島伸介, 館村純一, 日野洋一郎, 原 良憲, 田中克己: リンク構造の時間特性に着目した Weblog 解析に基づくコンテンツの信頼性評価の検討, DBSJ Letters, Vol.3, No.1, pp.109-112, 2004 年 6 月.
- [13] 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>
- [14] 渡邊正弘, 大倉典子: blog の機能を利用した位置情報コミュニケーションシステム, データベースと Web 情報システムに関するシンポジウム (DBWeb2004) 情報処理学会シンポジウムシリーズ Vol.2004, No.14, pp.73-80, 2004 年 11 月.
- [15] PING.BLOGGERS.JP, <http://ping.bloggers.jp/>
- [16] 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.