

blog におけるトラックバックへの影響因子の解析

Analysis of Influence Factor of TrackBack in blog

武田 善行*¹ 唐澤 鵬翔*¹ 梶川 裕矢*¹ 松島 克守*¹
 Yoshiyuki Takeda Hosyo Karasawa Yuya Kajikawa Katsumori Matsushima

*¹ 東京大学大学院工学研究科
 School of Engineering, The University of Tokyo

This paper reports on the findings of a property that the categories make change an influence factor of TrackBack in the BlogSphere. TrackBack determine communication area and density. We define some factor that are characterizing each blogs and community formulation and analyze correlation between the factors in each category. As a result, we acquired some useful factors explaining categories' own custom.

1. はじめに

近年 blog によるコミュニケーションが普及しつつある。総務省の報告によれば 2005 年度末時点の国内 blog 利用者数は約 340 万人、閲覧経験者数は 3500 万人にも及ぶ。blog の普及に伴い、そこでおこなわれているコミュニケーションを定量的に把握することの重要性がますます高まっている。blog は大衆向けのメディアであると同時に多岐にわたる話題が扱われるため、世論の把握に有効なメディアであろうと考えられている。

blog に限らず、Web におけるコミュニティ研究は CMC 研究の分野で長いあいだ関心の的であった。CMC 研究においては、現実のコミュニケーションとの対比、主に社会的制約からの解放や脱個人化に代表される“手がかりの欠如”によって生じる様々な現象の解明から [Spears 1993]、情報ないしメッセージを媒介とする人間関係や社会関係の構築にその興味がおかれていた [Chesbro 1989]。大規模なコミュニケーションがなされており、かつ様々な話題を含んでいるという点で blog と類似のコミュニケーションツールに 2 ちゃんねるがある。松村らは、2 ちゃんねるにおいて、定型的固有表現や匿名性が重要な役割を果たしており、議論の発散や深化に影響を与えると報告している [松村 2004]。

本研究では、blog コミュニティの形成要因を明らかにするため、トラックバックにより計量されるコミュニケーション範囲やコミュニケーション密度について分析をおこなった。blog の構成要素により定まる観測変数を設定し、これらの尺度との相関を測ることでコミュニティ生成に関わる要因を分析した。結果として、カテゴリ毎にコミュニティを形成するための要因は異なるという性質を明らかにした。

2. 研究方法

本研究では、トラックバックにより構成されるネットワークにおいて、2 回以上トラックバックを受けた blog 数 (T_2) と、隣接 blog から受けた平均トラックバック数 (T_a) とを非説明変数として設定した。それぞれ、 T_2 はコミュニケーション範囲、 T_a はコミュニケーション密度と解釈できる。ここで、コミュニケーション範囲に関して 2 回以上トラックバックを受けたリンク先とした理由は、偶発的なトラックバックを除去するためである。一般に、コミュニティの成立には継続的なコミュニケー

表 1: 観測指標

指標	説明
T_2	2 回以上トラックバックを受けた blog 数
T_a	隣接 blog 毎の平均トラックバック数
S	エントリ毎の平均バイト数
C	エントリ毎の平均コメント数
P	エントリ毎の平均画像数
L	エントリ毎の平均アウトリンク数
K	エントリ毎の平均流行語出現数
Y	継続年数
A	平均更新日数

表 2: 流行語

クールビズ	嫌韓流	フォー	偽装
ハードゲイ	タミフル	口ハス	刺客
インスパイヤ	ツンデレ	想定外	下流
ポピーマジック	小泉劇場	想定内	格差
小泉チルドレン	テラワロス	裕福層	萌え
ちよいもてオヤジ			

ションが必要と考えられる。閾値をどのように設定するかに関しては議論の余地があるが、本研究では、偶発的なトラックバックの代表であるトラックバックスパムの除去を主目的としてこの値を設定した。

トラックバックに影響を与えるであろう指標として表 1 を選択した。これらは blog の構成要素であり、計算機により自動的に測定できる指標である。また、 K を算出するにあたり用いた流行語を表 2 に示す。

2005 年 12 月 27 日から 2006 年 1 月 25 日の間に更新された blog のうち 2258 件を抽出した。また、そのなかから情報を抽出することのできた 1270 の blog を選択し、表 1 に示す変数間の相関分析をおこなった。

3. 結果と議論

分析の結果を表 3 に示す。結果として全ての指標間に有意な相関はみられなかった。もっとも相関が高い例は T_2 と K 間の 0.275 であった。任意指標間に有意な相関が観測されなかった理由として、分析した blog 集合中のエントリが非均質な性質を持っているためではないかと考えた。扱う話題の異なるエントリにおいて、トラックバックを集めるために必要な条件や指

連絡先: 武田 善行, 東京大学大学院工学研究科, 〒113-8656 東京都文京区弥生 2-11-16 総合研究機構, takeda@sogo.t.u-tokyo.ac.jp

表 3: 全体の分析結果

	<i>S</i>	<i>C</i>	<i>P</i>	<i>L</i>	<i>K</i>	<i>Y</i>	<i>A</i>
T_2	0.114	0.136	-0.110	0.83	0.21	0.132	0.172
T_a	0.010	-0.007	0.096	0.018	0.01	0.033	0.011

表 4: 頻出カテゴリ名とその同義語

順位	カテゴリ名	同義語	含有率
1	映画	movie	28.5
2	日記		26.1
3	音楽	music	21.1
4	その他		20.3
5	テレビ	TV, CM	18.5
6	ゲーム	game	16.6
7	アニメ	anime	15.2
8	本	book	14.8
9	パソコン	PC, ネット, Web	13.6
10	日常	日々	13.5

標に現れる傾向は異なるのではないかと考えた。

そこで次に, blog 集合をカテゴリ毎に分割し, それぞれの部分集合中での相関を分析した. カテゴリの分割には主観的な立場をとった. 具体的にはエントリー毎に付与されるカテゴリ表記を列挙し, 頻出する形態素を主要カテゴリ名として抽出した. また, それらカテゴリをよく表すであろうキーワードを人手で抽出し, 同義語として抽出した. 最終的に, カテゴリ表記がこれら語のいずれを含むかにより, カテゴリ分割をおこなった.

カテゴリ表記中に頻出するカテゴリ名と, 人手によって抽出された同義語を表 4 に示す. ここで, “その他” カテゴリはトラックバックを集めるために必要な固有の条件を持ち得ないと考え, 分析対象から除外した.

カテゴリ毎に分析した結果, 全体での分析に比べ T_2 , T_a と高い相関を持つ指標がみられた. 相関係数が 0.3 以上の指標を表 5 に示す. また, カテゴリ毎のコミュニティの性質を表 6 に示す.

カテゴリ毎に議論の対象が類似しているにもかかわらず, コミュニティの形成に影響を及ぼす指標が異なる例として, 映画カテゴリとテレビカテゴリがあげられる. それぞれ, 現在流行している映画やテレビドラマに関する論評が主なコンテンツである. コミュニケーション範囲を広げる際に, 映画カテゴリにおいては, コメント数を増やし, 映画に関して掘り下げた議論を展開する場を提供することが必要とされる. 一方, テレビカテゴリにおいては, 定期的に放映されるテレビドラマを広く漏れなく記事にすることが必要とされる. また, コミュニケーション密度を高める際には, 映画カテゴリにおいては, 映画の公式サイトや劇場の情報, 関連する映画論評へのリンクといった補足的な情報を十分に盛り込むことが必要とされる.

カテゴリ名からはほぼ同一の印象を受けるにもかかわらず, コミュニケーション範囲を広げる際に影響を及ぼす指標が異なる例として, 日記カテゴリと日常カテゴリがあげられる. 日記カテゴリでは, T_2 と K の相関が高いのに対して, 日常カテゴリではそのような傾向がみられない. 個別の blog を観察した結果から, 日記が他分野の語を共に含む場合が多いという傾向がみられた. 例えば, 映画日記やゲーム日記, 旅行日記などがあげられる. このことから日記カテゴリは複数のカテゴリを含む複合的なカテゴリといえる.

逆にカテゴリ間で類似している例としては, ゲームカテゴリ

表 5: コミュニティ形成に影響を及ぼす指標

カテゴリ	T_2	T_a
映画	<i>C, K</i>	<i>L</i>
日記	<i>K</i>	<i>P</i>
音楽	<i>K</i>	<i>P, A</i>
テレビ	<i>K, A</i>	<i>A</i>
ゲーム	<i>K, A</i>	<i>P</i>
アニメ	<i>K, A</i>	<i>P</i>
本	<i>C</i>	<i>K</i>
パソコン	<i>C</i>	<i>P, A</i>
日常		<i>P</i>

表 6: カテゴリ毎のコミュニティの性質

カテゴリ	平均 T_2	平均 T_a
映画	13.45	6.64
日記	8.58	18.15
音楽	5.29	7.82
テレビ	16.74	11.95
ゲーム	17.76	9.39
アニメ	28.57	8.49
本	4.10	10.33
パソコン	5.85	8.84
日常	8.19	13.87
全体	9.49	11.04

とアニメカテゴリがあげられる. 流行指向や活発な更新, 視覚要素重視といった特徴はアニメやゲームといった媒体やユーザの年齢層 (blog に多くの時間を割くことが可能な若年層) を考えると自然な結果といえる. カテゴリ間の類似は, メディアミックス戦略が主な原因であった. 調査対象 blog 中で, 多メディア展開されている作品が多く言及されていることを確認した.

4. まとめ

本研究では, 幅広く密度の高いコミュニケーションをおこなうコミュニティを形成する方法を発見することを目的とし, blog の構成要素によって定まる各指標を分析した. コミュニケーション範囲やコミュニケーション密度を表す尺度 T_2 , T_a を定義し, カテゴリ単位で見た場合に, それらと相関の高い指標がどのように影響を与えているかについて分析をおこなった. 結果として話題のカテゴリがコミュニティ形成の性質を左右する要因であることを明らかにし, 各カテゴリにおける主要な指標の抽出に成功した.

参考文献

- [Spears 1993] Spears, R. and Lea, M.: Social influence and the influence of the “social” in computer-mediated communication, *Computer-Mediated Communication*, pp. 30–65, 1993.
- [Chesbro 1989] Chesbro, J. and Donald, B.: Computing as Rhetoric, *CMC: Human Relationships in a Computerised World*, pp. 213–237, 1989.
- [松村 2004] 松村真宏, 三浦麻子, 柴内康文, 大澤幸生, 石塚満: 2ちゃんねるが盛り上がるダイナミズム, *情報処理学会* 45 巻 3 号, pp.1053–1061, 2004.