

Refereed Conference paper

## 反響特性分析に基づいたBlog記事マイニング

宮田 章裕\* 松岡 寿延\* 岡野 真一\* 石打 智美\* 荒川 則泰\* 加藤 泰久\*

Blog記事は表記揺れが大きく、コメント等により動的に文章やリンクが追加されるという特徴を持っており、従来の検索手法では目的の記事を見つける事が困難である。これに対し、提案手法では記事の内容やリンク構造を分析するのではなく、記事を取り巻く人間の行動情報を利用してBlog記事の性質を分析するアプローチを採っている。記事上で起こっているインタラクションの様子を「広さ」、「強さ」、「速さ」、「長さ」という多次元要素からなる「反響特性」と定義する事で、「幅広い人から関心を集めている」や「少数の人が熱烈に関心を示している」というように記事の性質を詳細かつ多面的に分析する事が可能になる。映画に関するクチコミ記事を探すというシーンを想定して行った評価実験では、「幅広い人から強い関心を長期間集めている」記事を提示したところ、ユーザの満足度が高くなるという結果となり、提案手法の有効性を確認する事ができた。

### Responses of a Blog Post: An Analysis of Interactions between Writers and Readers

Akihiro MIYATA\* Hisanobu MATSUOKA\* Shinichi OKANO\*  
Satomi ISHIUCHI\* Noriyasu ARAKAWA\* Yasuhisa KATO\*

In this paper, we suggest one method for mining blog posts using features of interactions between writers and readers. We analyze features of their interactions and estimate the post's "responses" being composed of width, strength, quickness and length. Using "response" information, users can find posts such as "having gotten a consecutive attention by masses," "having small number of enthusiastic readers," and so on. We ran experiments to evaluate our proposition with a prototype system and in a particular circumstance (searching movie reviews), our method showed better results than an existing way.

## 1 はじめに

2006年3月現在、Blogの国内開設数は800万件を超え[1]、今やBlog記事はメディアとしての地位を確立しつつある。Blog記事の読み手も増加の一途を辿っており、実にインターネットユーザの8割近くがBlog記事を読んだ経験がある[2]。2006年2~3月に我々が実施したアンケート調査(対象:Blog・SNS開設者1,015名)においても、約4割(393名)がBlog記事検索サービスを利用して自ら能動的にBlog記事を読んだ事があると回答した。

読み手の目的も実に多種多様で、前述の被験者にどのようなBlog記事を見つけようとしたか問うと(N=393, MA)、「自分が探している情報・答え(53.69%)」、「クチコミ情報(20.61%)」や「楽しい読み物(30.53%)」、「情報交換できそうな人の記事(12.72%)」という回答が得られた。また、米国で行われた意識調査でも「エンターテインメント」、「ニュース」、「他者との交流」と様々な閲覧目的が確認されている[3]。つまり、国内外問わずBlog記事の閲覧目的は非常に多様であり、目的に応じた記事に読み手をナビゲートする技術が今後より一層求められる事は明らかである。

しかし、Blog記事という従来の静的なWebページ

とは異質なメディアが急激に普及したためか、既存のBlog記事検索技術では更新日時の新しさ、あるいは検索キーワードへの適合度という観点でしか検索が行えない場合がほとんどである。そのため、閲覧目的に適した記事を発見できなかったり、目的に応じて柔軟に検索を行う事が難しかったりするのが現状である。

そこで我々は、Blog記事をより詳細かつ多面的に検索するために、Blog記事上で記事の書き手・読み手のインタラクション(コメントによる議論等)が発生しやすい点、そのインタラクションが記事の内容を反映している場合が多い点に着目し、記事を取り巻く人々の行動情報を利用して記事の性質を分析する手法を考案した。記事上でインタラクションが起こっている状態の記事が「反響を呼んでいる」というイメージで捉え、そのインタラクションの様子を「広さ」、「強さ」、「速さ」、「長さ」という多次元要素からなる「反響特性」と定義する事で、「幅広い人から長きに渡って関心を集めている」や「少数の人が熱烈に関心を示している」といったように記事の性質を詳細かつ多面的に分析するBlog記事マイニング手法を提案する。この情報を利用して読み手を目的に応じたBlog記事にナビゲートする事が本研究の目的である。

以降、2章で既存のBlog記事特徴分析技術について述べ、3章で反響特性分析に基づくBlog記事マイニン

\* NTTサイバーソリューション研究所  
NTT Cyber Solutions Laboratories

グを提案する。4章で提案概念を実装したプロトタイプシステムを紹介し、5章で提案概念の評価実験について報告し、6章にて本稿の結論と今後の方針を述べる。

## 2 Blog記事の特徴分析

### 2.1 アクセス数

大半のBlogサービス事業者では、自社が提供しているBlogのアクセス数をランキング形式で公開している。この情報は、多くの読み手の関心を集めているBlog記事を発見する助けとなりうるが、いくつか問題がある。

まず、アクセスランキング上位はほぼ固定的である。例えば、2006年4～8月の100日間でgooブログ[4]のアクセスランキング上位100位以内に登場した10,000件のBlogのうち、上位10位以内に登場するユニークなBlogはわずか76件である。これは、数百万件と存在しているBlogの中でごく一部のものしか目立っていない事を示唆している。

また、Blogのサイト構造に起因する問題もある。一般的なBlogではトップページに最新10件程度の記事が表示されているケースが多い。そのため、閲覧者がトップページにアクセスした場合、どのBlog記事を読んでいるのかアクセス履歴から識別する事はできない。

### 2.2 ブックマーク数・FOAF数

Blog記事がブックマークされた時、その記事は読み手から関心を持たれたと判断できるだろう。たしかに、有益な記事は多くの読み手からブックマークされる事が各種ソーシャルブックマークサービスにおいても確認できる。しかし、記事単位でブックマークされるような記事は資料的性質が強いものに偏りがちである(プログラミングTips、面接のコツ、冠婚葬祭のマナー等)。

対して、クチコミ情報記事や交流が起きている記事といったBlog記事に目立つコンテンツは長期に渡って保存すべき性質のものではない場合が少なくなく、ブックマークされる機会はさほど多くないと思われる。Blogサイト単位でブックマークされている例は珍しくないが(著名人や友人のBlogである場合が多い)、この場合は読み手の関心を記事単位で評価する事が難しい。最後の点においてはFOAF(人の繋がりをメタデータで表現したもの、友人のBlogサイトリストとしてブックマークのように使われる事が多い)の場合も同様である。

### 2.3 リンク構造

現在、主要なWeb検索サービスではPageRank[5]やHITS[6]等のリンク構造解析に基づくページスコア判定手法が用いられている。この手法は、価値のあるページからリンクを受けているページほど価値が高いという理念に基づいており、Web検索においてはかなりの効果を上げている。しかし、従来のWebコンテンツと比較するとBlog記事が外部からリンクを受ける事は少な

く、他のBlog記事からリンクされているBlog記事の数は約1%程度に過ぎないという報告もある[7]。

この点を考慮して考案された手法がEigenRumor[7]である。この手法では、Blogの編集主体がほとんどの場合一個人である点に着目し、各Blog記事のスコアをBloggerの属性として集約している。このため、リンクが無い・少ないBlog記事であってもBloggerのスコア(過去に執筆した記事への被リンク数等から導出)に基づいて評価する事ができ、検証実験では約9.3%(PageRankの場合は1.15%のみ)のBlog記事に非ゼロのスコアを付与する事ができたという。

### 2.4 コミュニケーション

Blog記事上で行われているコミュニケーションを利用して記事の特徴を分析する手段も検討されている。Mishneらは、記事上で行われるコミュニケーションの1つであるコメントの数と、記事の人气が関係していると指摘している[8]。彼らはアクセス数と被リンク数をBlog記事の人気の指標としており、これらの数値とコメント数には正の相関が見られたという。

また、分析対象はWeb掲示板のスレッドであるが、CMINERというコミュニケーション分析システムも考案されている[9]。このシステムではWeb掲示板のスレッド上における各ユーザの発言数や発言間隔、ユーザ間の関わり合いの深さ等を計測する事ができる。CMINERの分析アプローチは、Blog記事上のコミュニケーション分析を行う際にも参考になる。

## 3 反響特性分析に基づくBlog記事マイニングの提案

### 3.1 研究目標

2章で述べたとおり、Blog記事は従来のWebページとは異なる性質を持つため、既存サービスでは目的の記事を見つける事が難しい。また、既存サービスの多くは大衆から注目を集めているコンテンツを提示する事に重点を置いているが、我々の調査結果では注目を集めている記事以外にも共感できる記事、交流できる記事を発見したいという要望が多く見受けられた(1章参照)。つまり、記事が注目を集めているかどうか測定するだけでは不十分である。

そこで、本研究では従来とは異なる観点からBlog記事を分析し、その分析結果を利用して読み手を閲覧目的に応じたBlog記事にナビゲートする事を目指す。

### 3.2 アプローチ

新たなBlog記事マイニングのアプローチを検討するために、改めてBlog記事というメディアの性質を考えてみる。まず、Blog記事は文中で固有名詞の愛称・略称が多用される等、非常に表記揺れが大きいメディアである。内容も玉石混淆と称される事が多く、個人の備忘

録を目的としたもの、クチコミ情報の発信を目的としたもの、コミュニケーションを目的としたもの、それらが交じったもの等様々なタイプの記事が混在している。しかも、簡単な操作で記事を作成することができるため、このような玉石混濁のコンテンツが日々大量に発生し続けている。さらに、コメントやトラックバックといった機能を利用して、書き手や読み手が動的に文章を書き加えたり、リンクを付与したりしている。このため、従来の静的な Web ページを解析する事に長けていた自然言語処理技術やリンク構造分析技術だけでは、Blog 記事の特徴を的確に分析する事が難しく、結局人間が手間をかけて実際に記事に目を通し、その記事がどのようなものであるか判断しなければならないケースが多々ある。

そこで我々は発想を変え、Blog 記事を実際に書いた、あるいは読んだ人間の行動情報を利用して記事の特徴分析を行う事にした。なぜならば、Blog 記事の書き手・読み手の行動は、記事の内容の影響を反映していると思われるからである。例えば、人気がある記事や洞察力がある記事にはコメント数が多いという調査結果がある [8][10]、物議を醸す様な内容の記事ではコメントによる議論・論争が発生しやすいという傾向も指摘されている [8]。また、資料的価値が高い記事は、他の記事からの参照行為であるトラックバックを多く受ける傾向も見受けられる。読み手からコメントを受けると書き手の執筆意欲が向上するという現象 [11][12] も、書き手と読み手のインタラクションの 1 つと捉えて良いだろう。

記事単位でコメントやトラックバックといったコミュニケーション機能を備えており、記事の編集主体が Blog オーナーという形で見えやすい Blog 記事というメディアは、書き手や読み手のインタラクションが発生しやすいので、このように行動情報を利用するアプローチは特に有効であると思われる。

### 3.3 反響特性

ある Blog 記事を取り巻く人々が互いに、あるいは記事そのものとインタラクションを起こしている状態を、その記事が「反響を呼んでいる」というイメージで捉え、そのインタラクションの様子を「反響特性」と定義する。すなわち、記事を取り巻く書き手・読み手の行動や、彼らの行動の発端である記事を分析する事で導出される指標が「反響特性」であり、「反響特性」には以下の 4 要素が必要だと考えた。

**広さ：** インタラクションしている人の幅の広さ

**強さ：** インタラクションの行為の強さ

**速さ：** インタラクションが発生する速さ

**長さ：** インタラクションが継続する長さ

このように、反響特性を多次元要素の指標としたので Blog 記事の様子を詳細に分析する事が可能になる。例えば、「コメント数」という一次元情報だけでも記事が多くの人々の関心を集めているかどうか推測できるが、少数の常連メンバが活発な議論を行っているような記事はコメント数が多いわりに世間的には人気が無いといった例も報告されている [8]。これに対して、反響特性では「幅広い人がある程度の関心を示している記事 (広さ：大, 強さ：普通)」や「少数の人が熱烈に関心を示している記事 (広さ：小, 強さ：大)」, 「少数の人がチャットのように数十秒間隔でコメントを交し合っている記事 (広さ：小, 強さ：大, 速さ：大)」, 「幅広い人から長きに渡って関心を集め続ける記事 (広さ：大, 長さ：大)」等を区別する事ができる。このように、多種多様な Blog 記事の様子を詳細に把握する事が可能になるため、多様な閲覧目的に応じた Blog 記事をマイニングする事も容易になると思われる。

## 4 プロトタイプシステム

### 4.1 利用指標・反響特性導出方法の検討

ここでは、プロトタイプシステムで利用する指標と反響特性導出方法を検討する。3.3 章で述べたとおり、反響特性を導出するためには読み手・書き手の行動、記事に関する情報を収集して分析する必要がある。しかし、アクセス数のように容易に取得できない指標も多々ある。そこで、プロトタイプでは比較的容易に取得可能であり、かつ、Blog 記事を取り巻く人々のインタラクションとして重要な役割を果たしているコメントとトラックバックを中心に利用する事にした。我々が行ったアンケート調査 (N=1,015, MA) でも、関心が持てる Blog 記事を見つけた閲覧者の 70.4% はコメント、21.6% がトラックバックを行う事が分かっている。

プロトタイプで利用する指標と反響特性の関係を以下に示す。なお、トラックバックもコメント同様、送信者のユニーク数や 1 人あたりの送信数等を分析すべきであるが、時間的制約のためプロトタイプではこれらの分析機能を設けていない。しかし、1 つの記事に 1 人が送信するトラックバックは 1 つである場合が多く見受けられるので、大きな問題は生じないと判断した。

**コメント元数** コメントを送信している人のユニーク数。3.3 章で述べたように、単純にコメントの総数だけを考えてしまうと、大勢が少しずつコメントしているのか、少数の人が大量にコメントをしているのか、といった違いが表せなくなってしまうため、このようにコメント元数という指標を利用する。「広さ」に寄与。

**1 人あたりのコメント数** コメント送信者 1 人あたりのコメント送信数。「強さ」に寄与。

**初コメントまでの時間** 記事が投稿されてから初めてコメントが送信されるまでの時間。「速さ」に寄与。

**最新コメントまでの時間** 記事が投稿されてから現時点での最新コメントが送信されるまでの時間。「長さ」に寄与。

**平均コメント時間間隔** 各コメントが送信された時間間隔の平均値。「速さ」に寄与。

**トラックバック数** 記事に送信されたトラックバックの数。「広さ」に寄与。

**コメントリンク率** コメント総数に対する、コメント付与者の URL が書いてあるコメント数の割合。この割合が低いという事は多くの人が身元を明示せずコメントを書いている事を意味する。

**記事属性** 記事の文字数、文中のリンク数・画像数等。

今回の実装では上記の指標を使い、反響特性を導出する事にする。反響特性のスコアは、以下のように各指標の重み付き線形和で表現する事とした。W<sub>1</sub>~W<sub>10</sub> は各スコアの重み定数である。Score(コメント元数)等は各指標の値を0~1点の値で正規化したものである。

$$\begin{aligned}
 \text{Score(反響特性)} &= \text{Score(広さ)} + \text{Score(速さ)} \\
 &\quad + \text{Score(長さ)} + \text{Score(その他)} \\
 \text{Score(広さ)} &= W_1 \times \text{Score(コメント元数)} + W_2 \times \text{Score(トラックバック数)} \\
 \text{Score(速さ)} &= W_3 \times \text{Score(1人あたりのコメント数)} \\
 \text{Score(長さ)} &= W_4 \times \text{Score}\left(\frac{1}{\text{初コメントまでの時間}}\right) + W_5 \times \text{Score}\left(\frac{1}{\text{平均コメント時間間隔}}\right) \\
 \text{Score(その他)} &= W_6 \times \text{Score(最新コメントまでの時間)} \\
 &\quad + W_7 \times \text{Score(コメントリンク率)} + W_8 \times \text{Score(文字数)} \\
 &\quad + W_9 \times \text{Score(リンク数)} + W_{10} \times \text{Score(画像数)}
 \end{aligned}$$

## 4.2 実データの調査

プロトタイプシステムの効率的な開発のためには、4.1章で挙げた指標の性質を把握しておく必要があると考え、各指標を実際の Blog 記事から収集して調査を行った。

調査に利用した Blog 記事は、goo ブログ [4] の新着記事リストに掲載されてから少なくとも3ヶ月以上経過した記事102,104個である。最新の記事を利用しなかった理由は、投稿された直後の新しい記事を対象にしてしまうと、今後付与されるかもしれないコメントやトラックバックを数え損ねてしまう恐れがあるからである。調査結果を表1に示す。なお、表中の※はコメントが1個以上付与されている記事32,808個に関するデータである。また、◇に関してはコメント送信日時が記事投稿日時よりも過去になっている記事(まれに未来の日時で記事を投稿するケースがあるため)は除外してある。

表1を見ると、第3四分点で考えれば Blog 記事が投稿されてから初めてコメントが付くまでの経過時間が22.63時間、最新コメントまでの経過時間が72.00時間である事が分かる。ここで、今回の分析対象が3ヶ月以

表 1: 実データ調査結果

	平均値	第3四分点
コメント元数※	2.89	3
1人あたりのコメント数※	1.33	1.5
初コメントまでの時間※◇	31.56 時間	22.63 時間
最新コメントまでの時間※◇	84.50 時間	72.00 時間
平均コメント時間間隔※◇	29.75 時間	23.77 時間
トラックバック数	0.15	0
コメントリンク率※	0.19	0.33
記事の文字数	449.8	570
記事中のリンク数	0.51	0
記事中の画像数	0.42	0

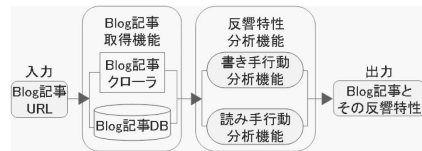


図 1: プロトタイプシステム構成図

上前に投稿された Blog 記事である事を考慮すれば、最新コメントは最終コメントとほぼ同義であると考えて問題無いだろう。つまり、Blog 記事が投稿されてから約1日経った後で分析して3日間程監視を続ければ、75%の Blog 記事に関しては付与され得るコメントはほぼ網羅できると言える。もちろん、記事を監視する期間を長くすればより正確な情報を取得できる事が期待されるが、これは時間コスト等とトレードオフの関係にある。

## 4.3 システム構成

プロトタイプのシステム構成を図1に示す。反響特性分析機能が提案概念の中核に相当し、この機能は書き手行動分析機能、読み手行動分析機能から成っている。入力として Blog 記事の URL を与えると、Blog 記事クローラ、あるいは DB から Blog 記事を取得して分析を行い、出力として Blog 記事と反響特性を併せて提示する事が可能である。なお、各指標の収集方法は、4.2章にて得られたデータに基づいて最適化が行われている。



表 2: 各パラメータとユーザ評価の相関係数

	$\rho$
文書適合度	0.03
コメント元数	0.62
1人あたりのコメント数	0.49
初コメントまでの時間※	0.40
最新コメントまでの時間※	0.69
平均コメント時間間隔※	0.61
トラックバック数	0.85
コメントリンク率※	0.73
記事の文字数	0.47
記事中のリンク数	0.10
記事中の画像数	0.16

## 5 評価実験

### 5.1 実験目的

この実験の目的は、提案手法がどの程度的確に、読み手を閲覧目的に応じた Blog 記事にナビゲートできるか評価する事である。比較対象（以降、「既存手法」）、および実験用 Blog 記事収集手段として、goo ブログ [4] で提供されている「goo ブログ検索」を利用した。

### 5.2 実験準備

今回は検証の第一歩として、比較的ニーズが高いと思われる「Blog 記事からの映画に関するクチコミ情報探し」というシーンを想定した。

実験準備として、読み手からの評価が高い記事の性質を調査するため映画に関する記事の収集を行った。収集方法としては、映画タイトル（「プレイブストーリー」、「ダ・ヴィンチ・コード」、「トリック劇場版 2」、「間宮兄弟」）でそれぞれ goo ブログ検索を行い、得られた記事を無作為に計 100 個抽出した。そして、1 つずつ著者が実際に閲覧して映画のクチコミ情報として有用かどうか採点した（以降、この採点結果を「ユーザ評価」とする）。具体的には「映画のストーリーや感想が第三者に分かりやすく述べられているか」という観点で 4 段階採点（0~3 点）した。また、検索時に利用した映画タイトルと記事本文の文書適合度も求めた。表 2 に示すのは、各記事のパラメータとユーザ評価との間のスピアマンの相関係数  $\rho$  である。なお、表中の※はコメントが 1 個以上付いている記事 64 個に関するデータである。

表 2 を見ると、文書適合度とユーザ評価にはほとんど相関が無い事が分かる。これに対し、人間が実際に閲覧しコメントやトラックバックという行動を起こした記事はユーザ評価も高い傾向にある事が分かる。例えば、コメント元数、トラックバック数とユーザ評価の相関係

数は 0.62, 0.85 と高い値になっている。確かに、ユーザ評価が高い記事は映画の詳細な情報が丁寧にまとめられている場合が多く、記事への感想コメントや参照行為としてのトラックバックを受けているケースが頻繁に見受けられた。また、最新コメントまでの時間とユーザ評価との相関係数も 0.69 と高い。これは、有用な記事ほど長期間に渡って注目を集め続ける事を示唆している。さらに、コメントリンク率とユーザ評価も 0.73 と高い相関を示した。この理由は一概には判断できないが、読み手が有用と感じた記事にコメントを送る際は身元を明示する（=コメントにリンクを付ける）傾向があるのかもしれない。あるいは、「万人にとっては有用でない記事」にコメントを送る場合（友人の日記記事におしゃべり感覚でコメントを書き込む、等）には手間をかけてコメントにリンクを付ける読み手が少ないのかもしれない。いずれにせよ仮説の域を出ておらず、今後詳しく調査を行う必要がある。

### 5.3 実験内容

まず、映画タイトル（「子ぎつねへレン」、「嫌われ松子の一生」の 2 パターン）をキーワードにして 5.2 章と同様に Blog 記事検索を行い、検索結果として得られる Blog 記事それぞれ 500 個の全てを収集した。次に、収集した記事全てをプロトタイプシステムで分析し、4.1 章で述べた *Score* (反響特性) をプロトタイプシステムによるスコアとした。各指標の重み  $W_1 \sim W_{10}$  は表 2 で対応する  $\rho$  値を用いた。つまり、「広さ」に寄与しているコメント元数、「広さ」と「強さ」に寄与しているトラックバック数、「長さ」に寄与している最新コメントまでの時間がユーザ評価と相関が高いので、「広さ：大、強さ：大、長さ：大」という反響特性を持つ記事ほどスコアが高くなる事になる。

そして、20~50 代の会社員 15 名に対し、既存手法による検索結果上位 5 件（文書適合度順）の Blog 記事と、プロトタイプシステムによるスコア上位 5 件の Blog 記事を比較し、どちらが満足度の高い映画クチコミ情報を提供しているか採点（0~3 点の 4 段階評価、採点基準は 5.2 章の「ユーザ評価」と同様）するタスクを課した。この際、順序効果の影響を相殺するため、被験者を 2 組に分けリストを閲覧する順番が異なるようにした。

### 5.4 実験結果と考察

表 3 に実験結果を示す。表中の数値は、各手法が提示した Blog 記事に対する満足度を全被験者で平均したものであり、どちらの映画においても提案手法が既存手法よりも高い満足度を被験者に与えていた事が分かる。ここから、映画に関するクチコミ情報源として満足度の高い Blog 記事を提示するシーンにおいては、提案手法に一定の優位性があると言える。

表 3: 全被験者の満足度 (N=15, 0~3 点の 4 段階評価)

	子ぎつねヘレン 嫌われ松子の一生	
既存手法	1.09	0.98
提案手法	1.71	2.20

実際に既存手法が提示した記事を見ると、「今度この映画を観よう」程度の内容しか無くクチコミ情報としては有用でないものや、文章が乱雑で第三者には難解なものが多々見受けられた。これに対して、提案手法が提示した記事は大半が映画のストーリーや感想、関連情報を丁寧に説明していた。このような記事では、多くの人がコメントを寄せる場合が多く、最大で 37 人が記事への共感や感想（「内容が役立った」、「私も面白いと思った」等）を語り合っているものがあつた。また、長期間に渡ってトラックバックを受けている場合も多く、3ヶ月に渡って 84 件のトラックバック（スパムは無い）を集め続け、他の記事から参照・引用されているものもあつた。なお、この記事の作者は過去に 300 件以上の映画クチコミ記事を執筆しており、手馴れた文章で映画の内容や感想を上手に紹介しているという印象を受けた。

このように、提案手法では実際に多くの人が目を通し、コメントやトラックバックという関心を示す行為を行った記事を優先的に提示する。そのため、文書適合度が高いと機械的に判定されただけの記事が表示される既存手法よりも被験者が読んだ時の満足度が高く、表 3 の結果に至ったと思われる。

## 6 結論

本研究では、目的に応じた Blog 記事に読み手をナビゲートするために、記事を取り巻く人々のインタラクションの様子「反響特性」を利用した Blog 記事マイニングを提案し、評価実験では一定の優位性を確認できた。特に、記事の内容を分析するのではなく、記事に対する人間の行動情報を利用して Blog 記事の性質を詳細かつ多面的に分析している点が本研究の新規性であると考えている。

今回は対象を映画に関する記事に限定したが、今後は他のジャンルの記事に関しても反響特性分析が有用であるか検証する予定である。例えば、時事ニュースに関する内容では「速さ：大」となる記事が読み手に望まれる等、映画の場合とは異なる現象が起こる可能性がある。また、我々が現在開発している Blog コミュニティプロファイリング技術「XappaLinks」[13]では頻繁に交流しあっている Blogger 群を Blog コミュニティとして抽出しているが、その交流場所である記事自体の性質を分析する際にも本提案の技術を適用する方針である。

## 参考文献

- [1] 総務省: ブログ及び SNS の登録者数 (2006). <http://www.soumu.go.jp/s-news/2006/060413.2.html> (2006/8/10 現在) .
- [2] goo リサーチ: 第 26 回 Blog に関する調査 (2006). <http://japan.internet.com/research/20060721/1.html> (2006/8/10 現在) .
- [3] AOL, Inc.: Blog Trends Survey (2005). AOL Press Releases(2005/09/16).
- [4] goo ブログ: <http://blog.goo.ne.jp> (2006/8/10 現在) .
- [5] Brin and Page: The Anatomy of a Large-Scale Hypertextual Web Search Engine, *Proceedings of 7th International World Wide Web Conference* (1998).
- [6] Kleigberg: Authoritative Sources in a Hyperlinked Environment, *Journal of the ACM*, Vol. 46, No. 5 (1999).
- [7] Fujimura, Inoue and Sugisaki: The EigenRumor Algorithm for Ranking Blogs, *WWW 2005 2nd Annual Workshop on the Weblogging Ecosystem* (2005).
- [8] Mishne and Glance: Leave a Reply: An Analysis of Weblog Comments, *WWW 2006 Workshop on the 3rd Annual Workshop on Weblogging Ecosystem: Aggregation, Analysis and Dynamics* (2006).
- [9] 井上, 小林, 池田, 重野, 岡田: ウェブ掲示板を対象としたネットワークコミュニティ分析システム: CMINER, *情報処理学会論文誌*, Vol. 45, No. 1, pp. 131-141 (2004).
- [10] Krishnamurthy: The multidimensionality of blog conversations: The virtual enactment of september 11, *Internet Research 3.0* (2002).
- [11] Gumbrecht: Blogs as protected space, *WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics* (2004).
- [12] Trevino: Blogger motivations: Power, pull, and positive feedback, *Internet Research 6.0* (2005).
- [13] 宮田, 松岡, 岡野, 石打, 荒川, 加藤: XappaLinks: Blog コミュニティ参加支援システム, *電子情報通信学会 2006 総合大会* (2006).