

## 情報伝播プロセス解析によるべき乗則の発見 Power-law Property in Information Spreading Process

米田 貴雄<sup>†</sup>  
Takao YONETA

吉井 伸一郎<sup>†</sup>  
Shinichiro YOSHII

The purpose of this research is to analyze the process of information spreading on the Blog system. To collect real Blog data from WWW, We developed "Blog Crawling System" which consists of a database and a spider program. Using experimental data acquired, this paper discusses some hidden property that is a power-law found in information spreading process in the Blog service.

### 1. はじめに

本研究は、近年、急速に普及しつつある新しいメディアであるブログを研究対象とし、情報伝播ダイナミクスによる伝播様を解析することを目的とする。そこで、解析に必要なデータを収集するシステムを開発、実装し、情報伝播の実データを収集する。そして、得られたデータを用い情報伝播プロセスを解析するとともに、膨大なユーザの局所的な行為から形成される大局的な性質について議論する。

### 2. 解析手法

本研究は現実のブログサイトを対象に情報伝播プロセスを解析する。ブログとはエントリと呼ばれる記事を日記のように連ねていくタイプの Web サイトである。ほとんどのブログは RSS を用いてエントリの投稿された時刻などのメタ情報を管理しているため、従来の情報伝播研究では問題となっていた過去に遡っての事例調査や、情報伝播プロセスの厳密な定義・追跡が可能である。

#### 2.1 ブログ上での情報伝播

本研究では、エントリ間で行われたトラックバックを情報の伝播と定義する。トラックバックとはエントリ間の参照関係を明示するものである (Fig.1 参照)。

#### 2.2 Blog Crawling System の開発

Blog Crawling System は、WWW にアクセスし、ブログサイトの情報を取得するシステム [1] である。システム構成図を Fig.2 に示す。本システムは、以下の4つのタスクを繰り返す行うことでブログデータを収集する。

- (1) エントリの選択
- (2) エントリ URL へアクセスし、HTML ソースコードをダウンロード
- (3) HTML ソースコードを解析し、投稿された日付、トラックバックしているエントリ URL を抽出
- (4) 抽出した URL をデータベースへ格納

なお、本実験ではトラックバックによる繋がりを重視しているため、エントリの選択は抽出されたエントリから優先的に選択する。また、クロウリングの初期 URL には、PING.BLOGGERS.JP<sup>\*1</sup>が配布している OPML<sup>\*2</sup>ファイル [2] のブログリストを用いた。

<sup>†</sup> 北海道大学大学院情報科学研究科

<sup>\*1</sup> <http://ping.bloggers.jp/>

<sup>\*2</sup> Outline Processor Markup Language の略。文書のアウトライン構造の情報を異なる OS や異なる環境で交換できることを可能とするための XML 規格

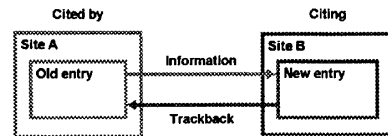


Fig. 1 Information flow

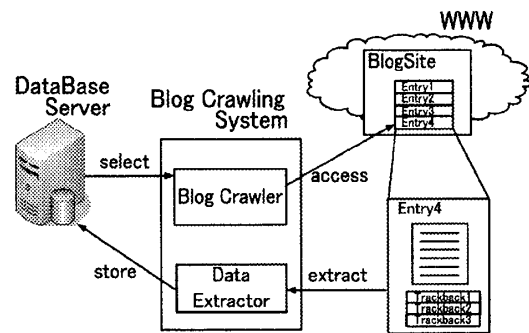


Fig. 2 Blog Crawling System

### 3. 実験

#### 3.1 実験条件

解析対象は、Livedoor がホスティングするブログサービス、Livedoor Blog<sup>\*3</sup>とした。

Table.1 に本システムが収集したデータを示す。収集したデータはサービスが開始された 2003 年 12 月 19 日から 2005 年 1 月 18 日の間で投稿された約 1 年間分のエントリとトラックバックである。

情報伝播プロセスは、ある一つのエントリを根とし、トラックバックしているエントリを枝、または葉としたツリー構造で表す。ただし、互いにトラックバックしあっている場合には、時間の古い方のトラックバックのみを用いることにする。

情報伝播規模を一つの情報伝播プロセスに含まれるエントリの総数とする。ただし、情報伝播プロセスの根となるエントリはどのエントリにもトラックバックをしていないものとする。

#### 3.2 情報伝播プロセス

Fig.3 は 2004 年 7 月 5 日に投稿されたエントリから発生した伝播プロセスの挙動を示している。横軸は日付、縦軸はその日行われたトラックバック数である。

この図では 4 つの特徴的なピーク (7 月 5 日 ~ 7 月 19 日, 9

<sup>\*3</sup> <http://blog.livedoor.com/>

Table. 1 Experimental data

取得データ	データ数
ブログサイト	11,086
エントリ	358,677
トラックバック	140,528

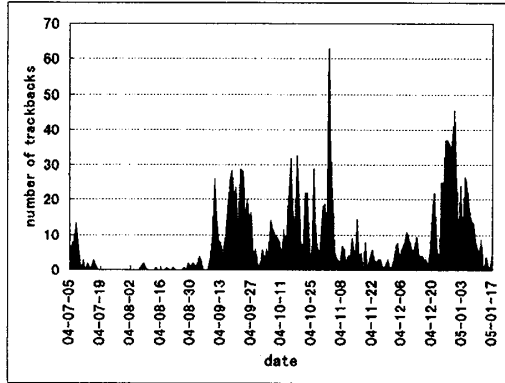


Fig. 3 Information spreading process

Table. 2 Event data

日付	実際に起きた出来事
7月7日	プロ野球球団合併構想の会議
9月10日	プロ野球選手会、ストライキ回避を決定
11月2日	新規参入球団が楽天に決定
12月31日	紅白歌合戦

月10日～9月28日, 10月25日～11月3日, 12月20日～1月5日)が見られる。これらのピーク時に投稿されたエントリの内容を調べてみると、まず一つ目のピークでは「大阪近鉄バッファローズの球団合併問題」、二つ目では「プロ野球ストライキ」、三つ目では「楽天進出決定」、四つ目では「紅白歌合戦」や「正月」に関する話題について記述していた。このように、一つの伝播プロセスの中で話題に変化していることがわかる。次に、ピークが発生している時期について現実世界で実際に起こった出来事を調査したところ、これらのピークと実際の出来事が発生した時期がほぼ一致することを確認した (Table.2 参照)。Fig.3 の図で最も興味深いのは11月2日に楽天が進出を決めた日の前後の挙動である。楽天が決まる前日までは人々の興味が小さなピークとして現れていたのに対し、楽天の進出が決めた後では、トラックバックの数が激減していた。これは、世間の話題となっていた新球団の新規参入問題が決着してしまったことによって、人々の興味が薄れてしまったのだと考えることができる。このように、ブログには、現実世界とリンクし情報を伝播する性質があることがわかった。

### 3.3 伝播規模と発生頻度の関係

3.2 節と同様な解析を他の情報伝播プロセスについて行ったところ、それらの伝播プロセスは多種多様であり、共通する普遍的性質は無いように思われた。そこで見方を変え、伝播プロセスをマクロな視点から観測することにした。Fig.4 は横軸に情報伝播規模を、縦軸にはその発生頻度をとり、両対数でプロットしたものである。その結果、地震や雪崩などの自己組織化臨界現象に共通するベキ乗則が現れることが発見された [3]。ベキ乗則は情報伝播規模がおよそ 300 以下の伝播プロセスで現

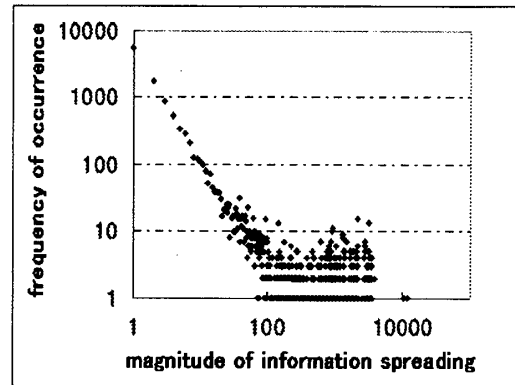


Fig. 4 Power-law in information spreading process

れており、全体の約 90 % の伝播プロセスを含んでいた。このときの指数は  $\gamma = 1.31$  であった。

## 4. 考察

### 4.1 情報伝播プロセス

複数の話題が一つの伝播プロセスに存在していることが明らかになった。この点において、情報がミーム [4] のように進化していると考えられる。しかし、ブログ上での情報だけで進化しているのではなく、TV や新聞などのマスメディアの影響を大きく受けていることがわかる。事実、実際に起きた出来事とブログ上での話題が起きた時期はほぼ一致していた。このように一つのエントリから始まる情報伝播プロセスを追跡することにより、話題が変化していく過程を観測可能であることを確認した。

### 4.2 伝播規模と発生頻度の関係

これまで、地震などの自己組織化現象や生態系の絶滅現象など、自然界でのベキ乗則は多く発見されてきた [3, 5]。しかし、情報伝播という人間の行動による現象がベキ乗則となることは知られていない。だが、おそらく他の現象同様、情報伝播現象にもなんらかのベキ乗則を生成する普遍的な原理が働いているのではないかと考えている。

## 5. まとめ

本研究では、現実のブログサイトを対象に情報伝播プロセスの解析を行った。その結果、情報伝播プロセスを追跡することにより、情報の変化が観測可能であることを示し、また、情報がミームのように進化するものであることを示した。さらに、情報伝播現象をマクロな視点から観測することにより、これまで知られていなかった情報伝播ダイナミクスにベキ乗則が現れるということを発見した。

## 参考文献

- [1] T. Minamino, Y. Suzuki, T. Fujiki, M. Okumura. Automatically Collecting and Monitoring Japanese Weblogs (2004).
- [2] Scripting News, Inc. OPML 1.0 Specification. *Outline Processor Markup Language*, <http://www.opml.org/spec> (2000).
- [3] P. Bak. *How Nature Works: The Science of Self-Organized Criticality*. Springer-Verlag Telos, Reprint edition (1999).
- [4] R. Dawkins. *The selfish gene*. Oxford University Press (1991).
- [5] スチュアート・カウフマン. 『自己組織化と進化の理論』. 日本経済新聞社, 米沢富美子監訳 (1999).