

# Weblog ネットワークにおけるユーザ間の関係と閲覧行動の分析 Analysis of User's Relation and Reading in Weblog

古川 忠延† Tadanobu Furukawa  
松澤 智史†† Tomofumi Matsuzawa  
松尾 豊‡ Yutaka Matsuo

内山 幸樹‡‡ Koki Uchiyama

武田 正之†† Masayuki Takeda

## 1. はじめに

インターネットの普及に伴う情報源の増加によって、Web を利用するユーザが有用な情報に辿り着くのは困難になっており、情報検索や情報推薦などのサービスが必要となっている。なかでも最近、個人による情報発信の形態として Weblog (以下 Blog) が注目を集めている。Blog では、自分の Blog を更新するとともに、他の Blog を訪れて閲覧し、コメントを書いたり、リンクやトラックバックを張ったりするという特徴がある[1]。このようなユーザの活動は、どの Blog とどの Blog の関係が強いのか、またどのユーザとどのユーザの関係が深いのかといった関係性を把握する根拠となる。

本稿では Blog 間の関係性に着目し、それがユーザの閲覧行動にどのくらい影響を与えているかを分析する。第一に、ユーザは関係性の強い Blog をよく訪れている、という相関があるかどうかを検証する。第二に、ユーザが頻繁に読むような Blog を、Blog の関係性から判別できるかどうかを明らかにする。このような判別が可能であれば、Blog の関係性に基づく有効な推薦サービスを構築することができると考えられる。検証には、既存の Blog ホスティングサービスである Doblog<sup>\*1</sup> のデータを用いる。Doblog では、ユーザがログインした上で Blog の書き込みやコメント書き込みを行うため、ユーザの閲覧行動が取得可能である。また「ブックマーク」と呼ばれるお気に入り Blog を登録する機能があり、これも関係性の 1 つとして利用することができる。この Doblog 上の限定的なユーザのデータを利用し、ユーザの行動分析を行っていく。

## 2. 実験手法

本稿では、あるユーザ A がある Blog B を過去に一度でもアクセスしたことがあるとき、それを A→B の訪問と呼び、また複数回 (10 回とする場合と 30 回とする場合について考える) 以上に渡る訪問があるとき、A→B の定期的閲覧と呼ぶこととする。この 2 つのユーザの行動と、Blog 間の関係性とに相関が存在しないかどうかを分析していく。

具体的には、(i) 『どのような関係性にある Blog へは訪問しうるか』をまず求め、さらに (ii) 『一度でも訪問のある Blog の中で』

定期的閲覧関係に発展しているのはどのような関係性にある Blog か』を調べることで、ユーザが普通には訪れる可能性が低いながら、一度知れば興味を持って何度も訪れることを期待できるような Blog を抽出することを目標とする。

Blog 間の関係性としては、ブックマーク・コメント・トラックバックのリンク関係と、定期的閲覧関係の 4 つに注目し、その 2 ホップで繋がる範囲内を対象として調査を行う。2 ホップ内の関係性に限定するのは、ユーザは基本的に自身がブックマークしている Blog など、直接の関係性のある Blog を中心に行動していると考えられ、間接的な関係のある Blog のうちでもっとも影響力が強いと考えられるのが 2 ホップ先の Blog だからである。この関係性について、関係性の“強さ”または“種類”の 2 つの面からアプローチする。

### 2.1. 関係性の強さ

4 つの関係性について、繋がり具合の強弱を示す尺度として、そのノード間を 2 ホップのその関係性で繋ぐ場合の経路数を用いてみた。例として、図 1 (左) において A→B 間の経路数は 3 である (A→1→B, A→2→B, A→4→B)。この A→B 間に直接の繋がりがあるか (A→B の推移律が成り立っているかどうか) の割合と経路数との間に、相関が存在しているかどうかを調べると結果は図 2 のようになる。

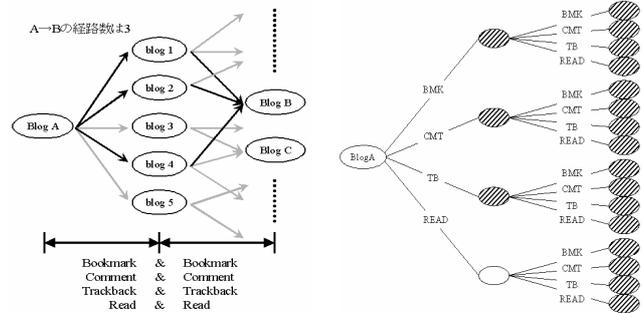


図 1 2 ホップの関係性

左：“強さ”を表す経路数 (“Read”は定期的閲覧)

右：“種類”を分類する 19 種の関係性 (斜線部)

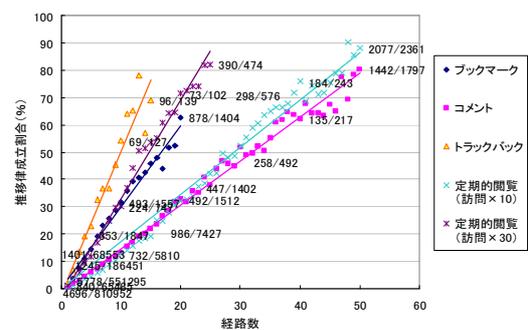


図 2 経路数と推移律成立割合

†東京理科大学大学院 理工学研究科 情報科学専攻,  
Graduate School of Science and Technology,  
Tokyo University of Science  
††東京理科大学 理工学部 情報科学科,  
Dept. of Information Sciences, Tokyo University of Science  
‡産業技術総合研究所,  
National Institute of Advanced Industrial Science and Technology  
‡‡株式会社ホットリンク,  
hottolink, Inc.

\*1) <http://www.doblog.com/>  
解析には 2003 年 11 月～2004 年 7 月のデータを用いた

これより、経路数と推移律の成立割合の間には正の相関があることが分かり、経路数の数をその関係性の強さの指標と捉えることができる。この指標と訪問・定期的閲覧関係の成立割合との間の相関を調査する。

## 2.2. 関係性の種類

ある起点となるBlogから2ホップで辿ることのできる範囲は、図1(右)に示した20通りある。このうち、定期的閲覧の1ホップの関係にあるBlogに対しては訪問・定期的閲覧関係ともに成り立っていて当然であるので、これを除いた19種類の関係性にあるBlogについて、訪問・定期的閲覧関係の成立が成立しているかどうかを調べる。そして、どのような関係性にあるBlogが訪問・定期的閲覧されやすいのかを分析する。

## 3. 結果と分析

### 3.1. 関係性の強さに関する分析

解析は総ユーザー数の10%の1647人分のデータで行った。なおその1647人は、アクセス数や被ブックマーク数などの量や伸び率などに注目して行ったランク付けにおいて、上位の者から抽出している。関係性の2ホップで辿れる位置にあるBlogのうちで訪問または定期的閲覧関係にあるBlogの割合を、経路数ごとに示したものが以下のグラフである。

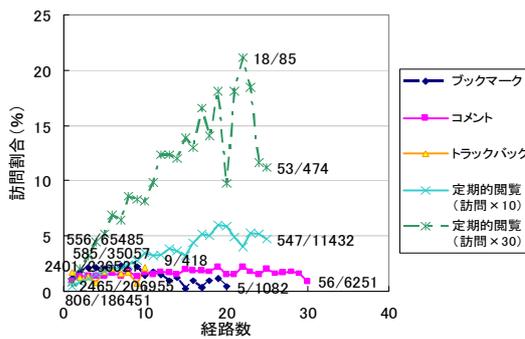


図3 経路数と訪問行動の関係

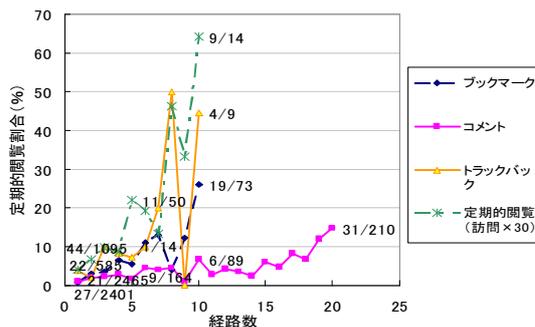


図4 経路数と定期的閲覧(訪問×30とした場合)の関係

図3に示すとおり、経路数と訪問行動の間では、定期的閲覧の2ホップの関係にあるBlogに対して、経路数の影響が強く現れた(すなわちAがBのBlog, BがCのBlogをよく訪れるとき、AはCのBlogに訪問する可能性が高い)。また、図4の結果からは、いずれの関係性においても、その強さと定期的閲覧行動との間に正の相関が現れている。ブックマークやトラックバックの2ホップで繋がる関係にあるBlogでその経路数の多いものは、訪問の可能性は高くないが、一度でも訪問すれば定期的閲覧となる可能性が高いことが分かる。

### 3.2. 関係性の種類に関する分析

解析は処理の都合上500人分のデータで行った。データマイニングツール Weka[2]を用いて構築した、ユーザーの訪問・定期的閲覧(訪問×30とした場合)行動の要因を表した決定木の上位5階層を以下に示す。なお、左への分岐は関係性が存在しない・右が存在する、を表す。B・C・T・Rはそれぞれブックマーク・コメント・トラックバック・定期的閲覧の関係性を表し、BCなどはブックマーク先のBlogがコメントしているという関係性である。図中のnoやyesは訪問・定期的閲覧関係の有無を表し、横の()内の数字はその時点で分類される数を表す。

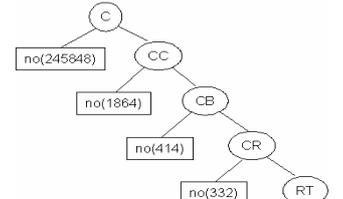


図5 訪問行動の決定木(上位5階層)

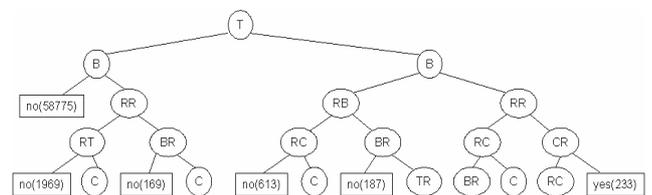


図6 定期的閲覧行動の決定木(上位5階層)

訪問行動の決定木では、500のBlog間による500P<sub>2</sub>=249500通りの組合せのうちで訪問関係が成立しているBlog間では、コメントの影響が大きいという結果となった。すなわちコメントの関係性がないという時点で訪問している可能性が低いことを表し、以降の分岐でも1ホップ目にユーザーのコメントが含まれ、その影響が強く表れている。

また、図6ではユーザーの定期的閲覧行動に最も影響の大きい属性がトラックバックという結果が出ている。その下ではユーザーが閲覧しているBlogを書いているユーザーが、コメントなどの行動を起こしているBlogで分岐している。

以上より、訪問しやすさと、定期的閲覧のしやすさのための属性は大きく異なっていることが分かる。

## 4. まとめ

本稿では、Blogにおけるネットワーク的な関係性に注目し、その強さや種類といった指標がユーザーの行動にどのように影響しているのかを分析した。結果として、ユーザーがBlog上のどのような範囲で行動し、どのようなBlogに興味を持ちやすいのかを把握することができた。ユーザーの行動範囲外にもユーザーが興味を持って閲覧しそうなBlogも存在しており、こういった情報は情報推薦における根拠となりうる情報であると考えられる。

今後はそれぞれの関係性同士の関係(例えばブックマークと定期的閲覧の関係性の類似度など)の考慮や、Blog内のコンテンツにも注目して、より精度の高い研究にしていく。

## 参考文献

- [1] 谷口智哉, 松尾豊, 石塚満, “Blogコミュニティの抽出と分析”, 第6回セマンティックウェブとオントロジー研究会, 2004
- [2] 阿部秀尚, “オープンソースデータマイニングツール WEKA,” 人工知能学会第18回AIシンポジウム(2003)