

Weblog におけるユーザの繋がりと閲覧行動の分析

古川 忠延^{†a)} 松澤 智史^{††} 松尾 豊^{†††}
内山 幸樹^{††††} 武田 正之^{††}

Analysis of Users' Relation and Reading in Weblog

Tadanobu FURUKAWA^{†a)}, Tomofumi MATSUZAWA^{††}, Yutaka MATSUO^{†††},
Koki UCHIYAMA^{††††}, and Masayuki TAKEDA^{††}

あらまし 最近注目を集めている Weblog (以下 Blog) では、ブックマーク (ユーザが自分の Blog ページ内からリンクしているお気に入りの Blog) やコメント、トラックバックなどにより Blog 同士の関係性を把握することができる。このような Blog の繋がり全体として Blog ネットワークを構成し、ユーザへの情報推薦に有用な情報源であると考えられる。

本稿では、Blog ネットワークにおいて、ユーザがどのようなページを定期的閲覧しているかを、Blog 間の関係性に着目して分析し、その把握を試みた。機械学習アルゴリズム C4.5 を用いてユーザが閲覧するかどうかの要因を分析した結果、Blog ネットワークにおいてはブックマークを経由しての閲覧が大きな割合を占めていることが分かった。また、ユーザが閲覧している Blog の運営者がコメントやトラックバックといったアクションを起こしている先の Blog も、ユーザにとって興味深い Blog となっている可能性が高いことを裏付けるデータを得ることができた。

キーワード Weblog, Blog, 行動解析, Web コミュニティ, マイニング

1. ま え が き

インターネットの普及に伴い、Web 上において日記やニュースサイト、掲示板など、個人による情報が多く発信されている。情報源の増加により、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提供が必要となっている。

また、それらのサービスの管理者にとっては管理する対象となるサービスの利用者の実態を把握する必要性が生じてくる。

最近 Weblog (以下 Blog) という個人の情報発信の新しい形態が注目を集めている。Blog の定義にはさまざまな議論があるが、その特徴として、情報発信に伴う作業がすべて Web 上で手軽に行うことができユーザが頻繁にコンテンツを更新する、基本的に 1 つの Blog は 1 人のユーザによって更新される、また、コメントやトラックバックといった機能を備え、RSS という形で情報の差分を提供する機能があるなどの点を挙げることができる [1]。特に、トラックバックは Blog に特徴的な機能であり、「リンクを張られた」情報をサーバで受け渡すことでリンクを張られた側の Blog からリンクを張った側の Blog への逆リンクを張るという双方向のコミュニケーションを形成し、議論の続きをみるといったことが可能になる。Blog には MT^(注1)などのサーバ設置型、Cocolog^(注2)や livedoor Blog^(注3)といったホスティング型の形態が存在する。

Blog のユーザは、自分の Blog を更新するととも

[†] 東京理科大学大学院理工学研究科, 野田市
Graduate School of Tokyo University of Science, Yamazaki
2641, Noda-shi, Chiba, 278-0022 Japan

^{††} 東京理科大学理工学部, 野田市
Tokyo University of Science, Yamazaki 2641, Noda-shi,
Chiba, 278-0022 Japan

^{†††} 産業技術総合研究所, 江東区
National Institute of Advanced Industrial Science and Tech-
nology, Aomi 2-41-6, Koto-ku, Tokyo, 135-0064, Japan

^{††††} 株式会社ホットリンク, 品川区
hottolink, Inc., Nishigotanda 2-11-17, Shinagawa-ku, Tokyo,
141-0031, Japan

a) E-mail: j6304640@ed.noda.tus.ac.jp

(注1): Movable type: 米 Six Apart 開発の Blog ツール。CGI として動作する

(注2): <http://www.cocolog-nifty.com/>

(注3): <http://blog.livedoor.com/>

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。このようなユーザの活動は、どの Blog とどの Blog が関係が強いのか（リンクやトラックバックが多いのか）、またどのユーザとどのユーザの関係が深いのか（Blog を頻りに訪れたりコメントを書いたりしているか）といった関係性を把握する根拠となる。実際、ビジネスでもこのような Blog 間の関係性は着目されている。特に、近年ではソーシャルネットワーキングサービスとよばれる人の関係を明示的に登録するサービスと組み合わせる Blog が読まれることも多い。

本稿では Blog の関係性に着目し、Blog の関係性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。まず第一に、『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。そして関係性の定義としてコメントやトラックバックなど、さまざまな種類を用意することで、どのような関係性が最も有効かを明らかにする。第二に、『ユーザが頻りに読むような 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。

本稿では、既存の Blog ホスティングサービスである Doblog^(注4)のデータベースを用い検証する。このサービスでは、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 を登録することができ、これも関係性のひとつとして利用することができる。本稿では、サービスの運営管理者のために必要となるユーザの利用実態を把握するために、このサービス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う。

以下、2 章で Web における関係性に基づいた関連研究、3 章で本稿における分析のアルゴリズム、4 章で具体的な実験手法、5 章で実験結果に関する分析を行い、最後に本稿のまとめを述べる。

2. 関連研究

Web ページにおいては、そのハイパーリンクに基づくネットワーク的構造から、特定トピックのページの

抽出や Web コミュニティなどといった有用な情報を得ようという研究は盛んに行われてきた。[2] では、クローリングによって収集した約 2 億ページ・1 テラバイトの Web ページのコピーについて分析を行い、hub（優良リンク集）と authority（優良コンテンツ）から成る完全 2 部グラフとなるコミュニティの抽出を行っている。

また、Web ページ（centers）について、そこにリンクを張っているページ（fans）をサーチエンジンから導き、さらにそれらの fans がリンクしているような centers を導くという作業を繰り返すことで、その共起性から hub・authority によるコミュニティを抽出するという方法も提案されている [3]。

本稿では、Web におけるハイパーリンクに近い性質を持つブックマークのほかに、Blog 独自のリンク関係であるコメントやトラックバックなどの関係性も用いてマイニングを行っていく。これらの独自の関係性はそれぞれ異なる性質を持っているであろうことが予想され、様々な切り口からの分析を行っていくことができると期待できる。特に、Blog データベースを用いることにより、ユーザの閲覧行動もその関係性の一つとして用いることができ、そのため、『ユーザにとって有用な情報はどのような情報であるのか』について分析を行っていくことができるのが本稿の特長である。

3. アルゴリズム

本稿では、ある Blog A の運営者（以下、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、A と B の訪問（関係）と呼び、また複数回に渡る訪問（訪問 ×10 とする場合と訪問 ×30 とする場合について考える）があるとき、それを定期的閲覧（関係）と呼ぶこととする。2 つの Blog 間における関係性と、この訪問・定期的閲覧の 2 つのユーザの行動との間に相関が存在する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 に対して訪問しうるか』をまず求め、さらに (ii) 『訪問したことがある 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、一度でも訪問す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 間関係性としては (a) 『ブックマーク・コメント・トラックバックによるリンク関係および定期

(注4): ©NTT DATA CORPORATION, ©hottolink, Inc., <http://www.doblog.com/>. 分析には 2003 年 10 月から 2004 年 7 月までのデータを使用

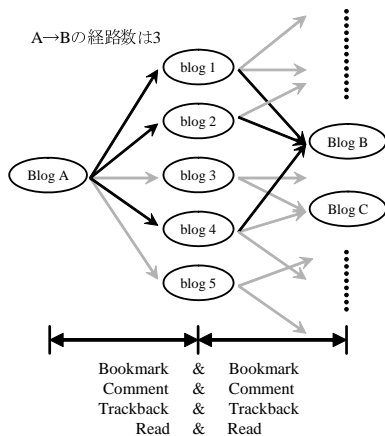


図 1 ある関係性の 2 ホップで繋がる Blog 間の経路数
Fig. 1 Number of routes.

的閲覧関係の 4 つ』と (b) 『両 Blog ユーザの興味分野の類似度』を用いる。

(a) についてはその 2 ホップで繋がる範囲内を対象として調査を行う。2 ホップ内の関係性を利用するのは、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、本稿で抽出を目標とする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 2 ホップ内の関係性について、関係性の“強さ”および“種類”の 2 つの面からアプローチする。

- 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、ノード間をその関係性の 2 ホップで繋いだ場合の経路数を用いた (図 1)。例として、図 1 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)。そして 2 ホップで繋がっているこの A ~ B 間に直接の繋がりがあるかどうか (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。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・定期的閲覧の行動との間の相関を調査する。

- 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 3 に示す 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成り立つ

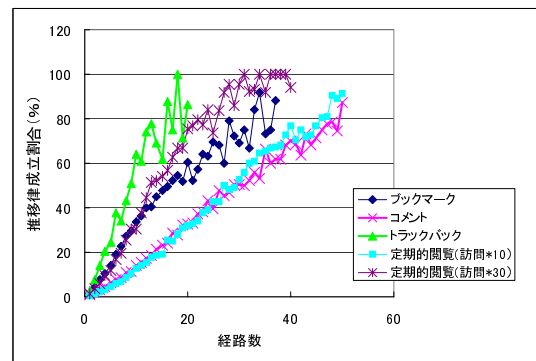


図 2 経路数と推移律成立割合
Fig. 2 Number of routes & transitivity rule.

ていて当然であるので、これを除いた 19 種類の関係性 (図 3 の斜線部) にある Blog について、訪問・定期的閲覧関係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

(b) については Doblog ユーザ登録時に、用意された項目^(注5)の中からユーザが選択した興味分野を利用する。そしてその関係性の強さの尺度として、両ユーザ間の興味分野の一致している個数を用いる。例えば、ユーザ A, B の興味分野がそれぞれ { 国内旅行・野球・バイク・お酒 } { 野球・ガーデニング・テレビゲーム・お酒 } であった場合、一致数は { 野球・お酒 } の 2 となる。この一致数と訪問・定期的閲覧関係との間の相関を調査する。

以上の分析を Doblog 独自の順位付けシステムであるユーザランクの上位 10 % (すなわち 1647 人^(注6)) までの Blog を対象に行っていく。ユーザランクは、一月分のブックマーク・コメント・トラックバック・訪問の 4 つの関係性のデータを用いて算出される。各関係性をエッジ、各 Blog をノードとするネットワークにおいて、活性伝播の原理によってその Blog がどれだけネットワークの構築に貢献しているのか、により得点を与えたものである。よって、ユーザランクの高い、すなわちより他のサイトとの関係性を持っていると思われる Blog を用いることによって、関係性を効果的に抽出しての分析が期待できる。実際、この 1647 人のユーザに限定した関係性のみで、ブックマーク・コメント・トラックバックについてそれぞれ全体の 59

(注5): “スポーツ”, “エンターテインメント”等に関する 103 項目
(注6): 2004 年 7 月現在, 利用者は全体で 16471 人である

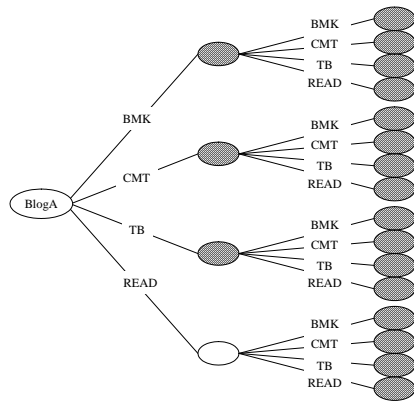


図 3 19 種類の関係性。
 図中の BMK はブックマーク, CMT はコメント,
 TB はトラックバック, READ は定期的閲覧関係を
 表す

Fig. 3 19 kinds of relations.
 BMK is bookmarking, CMT is commenting,
 TB is trackbacking, READ is regular visiting
 relation.

%・64%・64%を扱うことができる。

4. 実験

本実験では、2つの Blog の各組について、以降で定められた条件下において訪問関係または定期的閲覧関係（本章では以降、簡単のため2つを合わせて“閲覧関係”と呼ぶこととする）が存在しているかどうかを調べ、閲覧関係の成立割合を求めていく。但し、成立割合は

$$\text{閲覧関係の成立割合} = \frac{\text{閲覧関係の成立している組数}}{\text{総組合せ数}}$$

で計算する。なお、ユーザの閲覧履歴は Doblog のデータベースより抽出したものを使用する。また、組合せ総数とは、訪問関係の分析においてはその条件を満たしている起点-終点の全ての組合せを数え上げたものであり、定期的閲覧関係の分析ではそのうちで訪問関係が成立している組の数である。これは、本稿が『一度でも訪れれば定期的閲覧に発展する』ような関係性の抽出を目標としているため、訪問したうちのどれだけの割合で定期的閲覧をするようになっているのかを分析したいからである。

4.1 関係性の強さに関する実験

4.1.1 経路数に関する分析

ブックマーク, コメント, トラックバック, 定期的閲覧関係のいずれかの2ホップで繋がっている Blog 同

士について、その経路数毎にどの程度の割合での閲覧関係が成立しているのかを調査する。手順としては

(1) 1ホップ目: 起点となる Blog から上記いずれかのリンク関係で繋がる Blog を全て抽出する。但し、自身へのリンクは除く。

(2) 2ホップ目: (1) で得た Blog 全てから、同様の関係性で繋がる Blog (終点) を抽出。但し起点へ戻るリンクは除く。そして全ての起点-終点の関係と、その経路数を記録しておく。

(3) 各起点-終点の組について起点から終点への閲覧関係が存在しているかどうかを調べ、その有無を経路数毎に集計する。

以上で集計したデータについて、経路数ごとに閲覧関係の成立割合を求める。

4.1.2 興味一致数に関する分析

ユーザ間の興味分野の一致数を関係性として、閲覧関係との相関が存在するかどうかを調査する。すなわち、興味一致数ごとに閲覧関係の有無を集計する。この方法では関係性に方向が存在しないが、例えば Blog A, B 間の関係を考える場合、A → B, B → A それぞれについて閲覧関係が成立しているかどうかを調べる。よって、訪問関係の分析に関しては ${}_{1647}P_2 = 2710962$ 通りの組合せについて調査を行い、興味的一致数ごとに閲覧関係の成立割合を求める。

4.2 関係性の種類に関する実験

全ての2つの Blog 間 (${}_{1647}P_2$ 通り) において、前章で説明した19種類の関係性、すなわち

- 単一のブックマーク・コメント・トラックバックで繋がっている関係性 (3種)

- 1. の各関係性 + 定期的閲覧先のユーザが、ブックマーク・コメント・トラックバック・定期的閲覧している関係性 ((3 + 1) × 4 = 16種)

のうちのどの関係性が該当しているのか (複数該当の場合もあり)、さらに閲覧関係が存在しているのかどうかを調べる。こうして Blog の各組合せごとに、成立している関係性・閲覧関係をまとめて一つの学習データとして機械学習を行い、閲覧関係を決定づけている要素はどの関係性であるのか、分析を行う。但し、各関係性の1ホップで繋がる範囲については訪問関係が成立していて当然であるので、訪問関係の分析には2ホップのいずれかの繋がりのある起点-終点の組合せのみを分析対象とする。なお決定木の構築には、決定木構築アルゴリズムとして確立されている C4.5 [4] を実

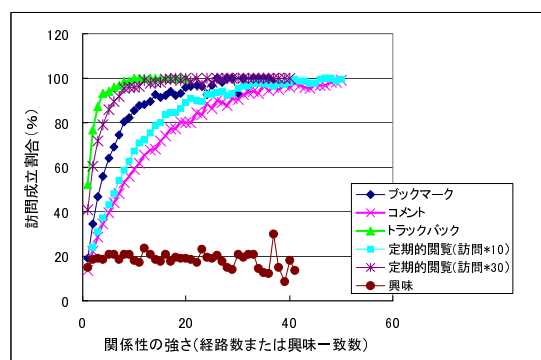


図 4 各経路数・興味一致数と訪問割合の関係
Fig. 4 Number of link routes or similarity of interest & visiting.

装し、描画機能も搭載している^(注7)データマイニングツール Weka^(注8)[5]を用いる。

5. 結果と分析

前章で述べた実験について、その結果と考察を示す。

5.1 関係性の強さ

5.1.1 関係性の強さと訪問関係に関して

Blog 間の各関係性による経路数および興味一致数と訪問との関係についての結果を図 4 に示す。

この図は

- 各関係性の経路数が増えるに従って訪問割合は上がる。
- 興味の一一致数は、訪問の起こりやすさとは無関係である。

の 2 点を表している。

前者に関しては、ブックマークやコメントといった何らかの関係性が存在する場合 Blog 上にその Blog へのリンクが作成され、道筋が多くなるために物理的にユーザにとって訪問しやすくなるという理由が考えられる。また、定期的閲覧関係による関係性においても正の相関が表れている。ここで定期的閲覧関係の成立している 2 つの Blog 間におけるブックマーク・コメント・トラックバックを調べると表 1 のようにいずれかのリンクが高い確率で張られており、やはり物理的理由から相関が現れているのではないかと考えられる。

また、後者の興味一致数に関しては、

- (1) ユーザは必ずしも興味ある内容について Blog

(注7): Weka では J48 として実装されている

(注8): Machine Learning Project (<http://www.cs.waikato.ac.nz/ml/>)

表 1 定期的閲覧が成立している関係における各関係性の成立割合 (1 行目は定期的閲覧成立の条件)

Table 1 Rate of each links on relation of regular reading.

関係性	訪問 ×10 以上	訪問 ×30 以上
	成立割合 (%)	成立割合 (%)
ブックマーク	45.9	70.1
コメント	60.2	84.4
トラックバック	14.9	29.1

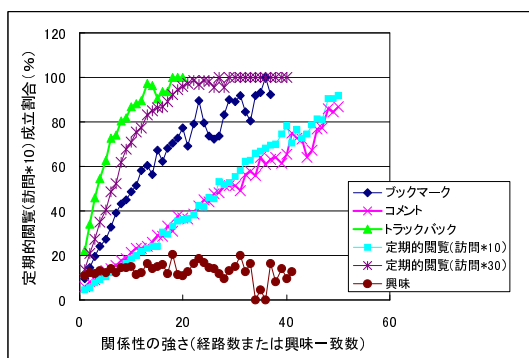


図 5 各経路数・興味一致数と定期的閲覧 (=訪問 ×10 以上) の関係

Fig. 5 Number of link routes or similarity of interest & regular reading (over 10 visitings).

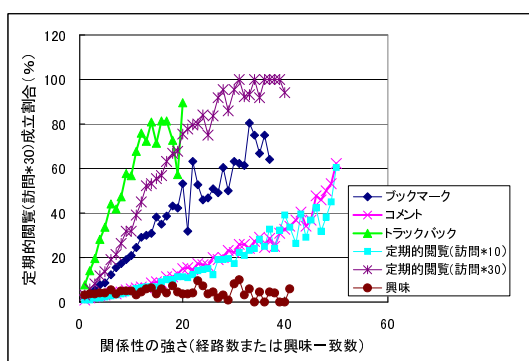


図 6 各経路数・興味一致数と定期的閲覧 (=訪問 ×30 以上) の関係

Fig. 6 Number of link routes or similarity of interest & regular reading (over 30 visitings).

を書いていない。または必ずしも興味のある内容の Blog ばかりを訪問していない。

- (2) (1) でない場合、Blog ネットワークにおいてユーザはあまり興味のある Blog に辿り着けていない。この 2 点が推測できる。このことについては次節の定期的閲覧関係の分析で更に考察する。

5.1.2 関係性の強さと定期的閲覧関係に関して

図5, 図6に, Blog間の各関係性による経路数および興味一致数と定期的閲覧との関係に関する結果を, 定期的閲覧関係の条件(訪問回数10回以上または30回以上)別にそれぞれまとめたものを示す.

この図からは次の2点が推測できる.

- 各関係性の経路数が増えるに従って定期的閲覧割合は上がる.
- 興味の一一致数は, 定期的閲覧の可能性に無関係である.

すなわち複数の経路で辿り着くことのできるBlogほど, 一度でも訪問するとその後も定期的閲覧に訪れる可能性が高いということを表している.

また, 定期的閲覧割合においてもユーザ間の興味の一一致数とは相関を得ることができず, これはユーザが興味の一一致するユーザのBlogに辿り着いたとしてもそのBlogを以降も定期的閲覧し続けてはいないことを表す. これは前節における推測(2)と相反することであり(1)の推測を支持している. 但し, ユーザが自身の興味に沿ってBlogを記述しているかどうかの調査はテキストマイニングに関する技術であり, 本稿では触れない.

5.1.3 経路数により訪問・定期的閲覧を決定付ける決定木

以上までに考察したデータについて, データマイニングツール Weka を用いて訪問・定期的閲覧のための決定木の構築を行った. 各Blog間におけるブックマーク・コメント・トラックバック・定期的閲覧行動による経路数と, そのときの訪問または定期的閲覧関係の有無をまとめたものを1つの学習データとし, 訪問の分析においては全てのBlogの組合せ(1647P₂)のうちで経路数が一つでも1以上となっている約560000組, 定期的閲覧の分析においては訪問関係が成立している154549組のデータを用いて学習を行った.

結果は, 訪問関係の分析については定期的閲覧の条件である訪問回数を10回以上とした場合が図7, 30回以上とした場合が図8であり, 定期的閲覧関係の分析についても10回以上とした場合が図9, 30回以上とした場合が図10である. なお, 相関がないと考えられる興味の一一致数に関しては学習データの属性に含めないものとする.

決定木はもっとも影響の大きい属性から上から並んでおり, 丸で囲んだ各ノードが関係性の条件, B・C・T・Rはそれぞれブックマーク・コメント・トラック

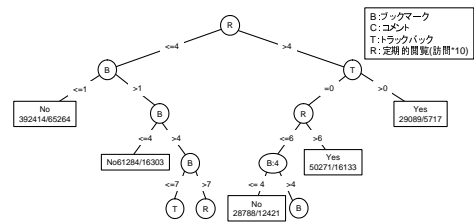


図7 経路数と訪問行動に関する決定木(定期的閲覧 = 訪問10回以上; 学習データ数 568056)

Fig. 7 Decision tree of visiting by number of routes. (regular reading = over 10 visitings; 568056 pairs)

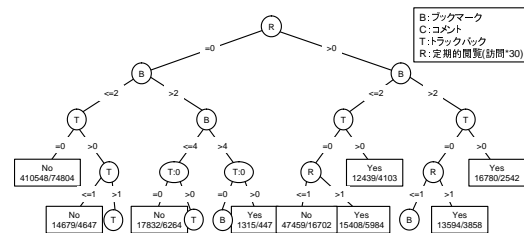


図8 経路数と訪問行動に関する決定木(定期的閲覧 = 訪問30回以上; 学習データ数 568043)

Fig. 8 Decision tree of visiting by number of routes. (regular reading = over 30 visitings; 568043 pairs)

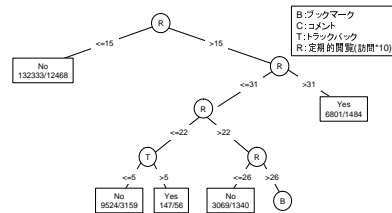


図9 経路数と定期的閲覧行動に関する決定木(定期的閲覧 = 訪問10回以上; 学習データ数 154549)

Fig. 9 Decision tree of regular reading by number of routes. (regular reading = over 10 visitings; 154549 pairs)

バック・定期的閲覧の各2ホップの関係を表す. ノードから左下へ向かうエッジはそのラベルの数字以下の経路数しかない場合の分岐で, 右下へ向かうエッジは数字より多い経路数がある場合の分岐である. 四角で囲まれたノードは直前までの条件での訪問または定期的閲覧関係の成立(Yes)・不成立(No)を表し, その下の数値は, “/”の左側がその時点で判別されるBlogの組合せ数, 右側が例外数である. 例えば, 図7を左方向に辿ることで, 「定期的閲覧の経路数が4以下で, かつ, ブックマークの経路数が1以下であるような関係にあるBlogの組は392414あり, 65264組を除いて

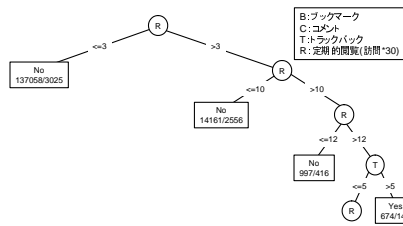


図 10 経路数と定期的閲覧行動に関する決定木 (定期的閲覧 = 訪問 30 回以上; 学習データ数 154549)
 Fig. 10 Decision tree of regular reading by number of routes. (regular reading = over 30 visitings; 154549 pairs)

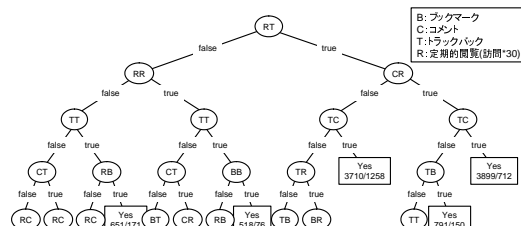


図 12 19 種類の関係性と訪問行動に関する決定木 (定期的閲覧 = 訪問 30 回以上; 学習データ数 323861)
 Fig. 12 Decision tree of visiting by 19 kinds of relations. (regular reading = over 30 visitings; 323861 pairs)

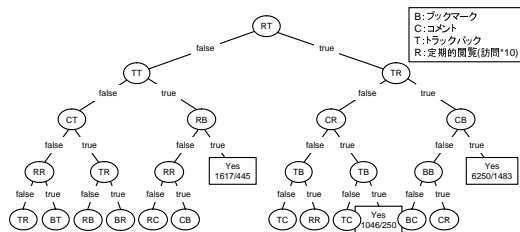


図 11 19 種類の関係性と訪問行動に関する決定木 (定期的閲覧 = 訪問 10 回以上; 学習データ数 382234)
 Fig. 11 Decision tree of visiting by 19 kinds of relations. (regular reading = over 10 visitings; 382234 pairs)

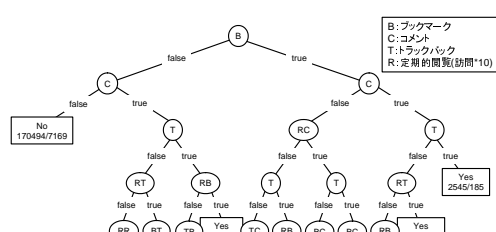


図 13 19 種類の関係性と定期的閲覧行動に関する決定木 (定期的閲覧 = 訪問 10 回以上; 学習データ数 206804)
 Fig. 13 Decision tree of regular reading by 19 kinds of relations. (regular reading = over 10 visitings; 206804 pairs)

訪問関係が成立していない」ことが分かる。なおこの決定木は紙面の都合上、上位 5 階層までのみを示した。

グラフからは各属性の優位性が分からなかったが、決定木から、訪問関係にもっとも影響しているのは定期的閲覧による経路数であることが分かる。ついでブックマーク・トラックバックが続き、特にトラックバックはその経路数が小さい場合でも大きな影響を与えているのが分かる。

また、定期的閲覧関係についても影響の大きい属性は定期的閲覧による経路数となり、大半がその値で分類されるという結果となった。訪問関係においては定期的閲覧の経路数の多い関係はあまり影響しておらず (必ずしも訪問できない)、定期的閲覧の経路数の多い関係にある Blog は推薦の価値のあるものではないかと推測できる。

5.2 関係性の種類

5.2.1 関係性の種類と訪問関係に関して

定期的閲覧の条件を 10 回以上とした場合の結果を図 11 に、30 回以上とした場合の結果を図 12 に示す。なお、学習には 2 ホップの関係性が 1 つ以上成立しているデータのみを用いている。

なお、決定木の見方は前節のものと同様であるが、RT・CB など 2 文字の関係は 2 ホップの繋がりを示しており、それぞれ定期的閲覧している先のユーザがトラックバックしている関係・コメントしている先のユーザがブックマークしている関係を表す。そして各ノードはその関係性の有無 (左が無し: false、右が有り: true) で分岐している。

最も影響力のあるのは「自分が定期的閲覧している先のユーザがトラックバックしている関係」であり、それ以外では「自分がトラックバックしている先のユーザが何らかの行動をしている関係」が多く影響しているのが分かる。

5.2.2 関係性の種類と定期的閲覧関係に関して

定期的閲覧の条件を 10 回以上とした場合の結果を図 13 に、30 回以上とした場合の結果を図 14 に示す。なお、学習データの数には訪問関係の成立している Blog の組合せ数である。

ブックマークやコメント、トラックバックといった、自分から 1 ホップの直接の関係性が、ユーザの定期的閲覧行動に最も影響しているのが分かる。図 13、図

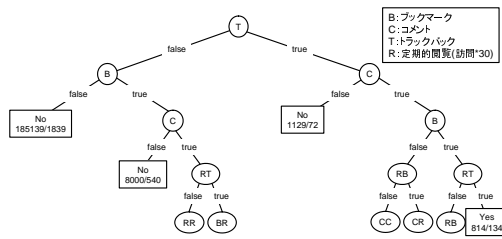


図 14 19 種類の関係性と定期的閲覧行動に関する決定木 (定期的閲覧 = 訪問 30 回以上; 学習データ数 206804)

Fig. 14 Decision tree of regular reading by 19 kinds of relations. (regular reading = over 30 visitings; 206804 pairs)

13 いずれにおいても、それぞれ「ブックマークもコメントもない」場合、「トラックバックもブックマークもない」場合は全体の約 90 %が定期的閲覧をしておらず、ブックマークの影響の強さが分かる。

また、2 ホップの関係に着目すると、「自分が定期的閲覧している先のユーザが何らかの行動をしている関係」の影響力が大きい。この関係にある Blog というのは、前述の通り自分から 1 ホップ先の関係にある Blog は定期的閲覧となりやすい Blog であるから、すなわち 1 ホップ先のユーザが定期的閲覧をしている可能性の高い Blog である。このことから、自分が興味を持っている先のユーザが興味を持っている Blog というのは、やはり自分にとっても興味深いものなりやすい、ということが考えられる。このことは前節での経路数と定期的閲覧の関係においての分析と合致するものである。

6. ま と め

本稿では、Blog ネットワークについてその独自の関係性に注目した分析を行った。Blog から 2 ホップで繋がっている範囲について、その経路数による強さや、種類といった指標からユーザの閲覧行動の把握を試みた。結果として、ブックマークによる影響が強く、ユーザがブックマークを中心に Blog ネットワーク内を巡回していることを裏付けることができた。

また、ユーザは自身が興味を持って繰り返し閲覧している先の Blog ユーザが、アクションを起こしているような関係にある Blog に対しても繰り返し閲覧しているという傾向が分かり、このような傾向は情報推薦の際に推薦の根拠となりうる情報であると考えられる。

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] R.Kumar, P.Raghavan, S.Rajagopalan and A.Tomkins, “ Trawling the Web for Emerging Cyber-Communities ”, The 8th International World Wide Web Conference (1999)
- [3] 村田剛志, “ 参照の共起性に基づく Web コミュニティの発見 ”, 人工知能学会誌, Vol16, No.3, pp316-323(2001)
- [4] 古川康一, “ Data Mining: Concepts and Techniques 第 8 回 ”, (URL) <http://web.sfc.keio.ac.jp/~soh/dm03/dm03-8.ppt>
- [5] 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)
(平成 xx 年 xx 月 xx 日受付)

古川 忠延

2004 年東京理科大学理工学部情報科学科卒業。現在同大学院修士課程在学中。

松澤 智史

1998 年東京理科大学理工学部情報科学科卒業。2004 年東京大学大学院工学系研究科電子工学専攻博士課程修了。同年より東京理科大学理工学部情報科学科助手。博士(工学)。

松尾 豊

1997 年東京大学工学部電子情報工学科卒業。2002 年同大学院博士課程修了。同年より、産業技術総合研究所サイバーアシスト研究センター勤務。博士(工学)。

内山 幸樹

1994 年東京大学工学部卒業。1996 年同大学院修士課程修了。2000 年株式会社ホットリンク設立、代表取締役社長に就任。

武田 正之

1977 年東京理科大学理工学部電気工学科卒業。1982 年東京工業大学大学院博士課程修了。同年東京理科大学理工学部情報科学科助手、同助教授を経て現在同大学教授。工学博士。