

特集論文

Bookmark からの共通話題ネットワークの発見手法の提案とその評価

A Method of Discovery of Shared Topic Networks among People from WWW Bookmarks and Its Evaluations

濱崎 雅弘

Masahiro Hamasaki

奈良先端科学技術大学院大学情報科学研究科

Graduate School of Information Science, Nara Institute Science and Technology
masa-ha@is.aist-nara.ac.jp, <http://www.aist-nara.ac.jp/~mas-ha/>

武田 英明

Hideaki Takeda

国立情報学研究所

National Institute of Informatics
takeda@nii.ac.jp, <http://www-kasm.nii.ac.jp/~takeda/>

松塚 健

Takeshi Matsuzuka

奈良先端科学技術大学院大学情報科学研究科^{†1}Graduate School of Information Science, Nara Institute Science and Technology
takeshi@acm.org, <http://www.toppan.co.jp/>

谷口 雄一郎

Yuichiro Taniguchi

(同 上) ^{†2}youi@praesens.co.jp, <http://www.praesens.co.jp/>

河野 恭之

Yasuyuki Kono

(同 上)

kono@is.aist-nara.ac.jp

木戸出 正継

Masatsugu Kidode

(同 上)

kidode@is.aist-nara.ac.jp

keywords: collaborative filtering, information retrieval, recommender system, world wide web

Summary

In this paper, we propose shared topic networks as a model of human network to organize Internet Information, and developed a system called *kMedia* that can generate shared topic networks by using WWW bookmark files. We also evaluate the system with experiments to know how shared topics network can help users especially to know each other. A shared topic network is formed by linking topics of participants, and used to know interests of others and to exchange information with others. *kMedia* can generate shared topics networks by using structures of WWW bookmarks, i.e., folders of bookmarks are regarded as topics of their owners. Relations among topics of different users are estimated by aggregating similarity among pages in these topics. The experiments were performed to clarify two points; one is whether topics is a better way to exchange information among people and the other is how we can measure human relationship. The first point is examined that topic recommendation is more acceptable than page recommendation. For the second point, we propose category resemblance as measurement of human relationship. Since we compare results between cases with subjects belonging to the same community and cases without communities, we noticed similarity of topics structure is affective. The category resemblance is to estimate this similarity of topic structure and it is proved that it is better than any other parameters with respect to measurement for human relationship.

1. はじめに

近年の WWW に代表される情報伝達技術の進歩により、我々は膨大な量の情報を利用する事が可能になっている。しかし、人間が扱うことのできる情報量には限界があるので、行き交う情報が増えたことによって、結果的に自分の欲しい情報を見失ってしまうような状態になりつつある。本研究ではこの問題の解決策として、人間

間の関係に注目する。

ある話題について情報収集をするとき、その話題と共通性のある話題に興味がある人々と情報交換することができる。出来れば、より効率良く情報を収集することができる。また、共通する話題に興味を持つ人と人との関係を発見することにより、新たに情報交換が活発に行われる集団、つまりコミュニティの形成が期待できる。コミュニティは人の情報編集活動を強力に支援する存在であると言え、その形成を支援することは、情報収集の支援にも繋がる

^{†1} 現在、凸版印刷株式会社

^{†2} 現在、プラエセンス株式会社

と考えられる [西田 00].

本論文では、人が興味を持つ話題の共通性によって人同士のつながりを発見するシステム *kMedia* を提案するとともに、このシステムを用いた実験から、人と人のつながりの発見に有効な指標についての分析を行う。

2. コミュニティの知識

コミュニティは人の繋がりから生まれる。ゆえに人と人との関係は、コミュニティを形成する上で非常に重要な要素の一つである。では、コミュニティを形成するためには、人のどのような関係を見つけたせば良いのか、という事が問題になる。

発見すべき人の関係は、抽象過ぎず、かといって特化しすぎていない、しかるべきレベルで記述されなければならないと考えられる。例えば、ネットニュース [Konstan 97] や Web ページ [Balabanovic 97, Terveen 97], 音楽 [Shardanand 95], といったものを扱った多くの推薦システムでは、情報のオブジェクト自身を共有しているかどうかという事を利用している。しかし、このようなオブジェクトを通じた関係性は、オブジェクトに関する情報を共有するかどうかを決めるのには有効であるが、きわめて個別的なので人と人との関係を理解するのに適しているとはいえない。ReferralWeb [Kautz 97] ではリンクや著書における引用関係から、人の関係を発見しようとしている。これは比較的有効な手法と思われるが、誰にでも適用できる訳ではないという問題がある。

一方、Beehive [Huberman 96] や Babble [Erickson 99], Visual Who [Donath 95] といったコミュニティ支援システムは、より簡単な関係性を用いている。例えば、Beehive では単純に e-mail でのコミュニケーションを観察してアクティブなメンバーかどうかを判断している。Babble では 2 次元空間にユーザの関係を表示する。Visual Who も同様に 2 次元空間を利用するが、より動的である。このような可視化は直感的で理解し易い反面、関係性が抽象的になりすぎるという傾向がある。そこで本研究では、各ユーザが持つ話題に着目し、話題の共通性によってユーザ間の関係を表現することを試みる。話題は直感的で理解が容易であると同時に、どのような関係であるかを明示化することが可能であると思われる。

しかし、ユーザの持つ話題の同定は一般に容易ではない。これまでユーザの関心を発見する方法として主に二つの方法が用いられていた。一つは強化学習やページアンネットワークのような機械学習を用いた方法である。例えば、WebWatcher [Joachims 97] や Letizia [Lieberman 95] などは Web ブラウジングを対象としたシステムである。この手法の長所であり短所である点は持続した関心をくみ取れる点である。問題となるのは、ユーザの関心が変化する点を見つけることができないので、関心を個別に取り出すことが難しい点である。また、学習した

データはユーザにとってわかりやすいものではないため、ユーザがどういった関心を持っているかを明示化するのも困難である。もう一つの方法は、階層分類のような分類技術である。例えば、Scatter/Gather [Cutting 92] や Webmate [Chen 98] などに用いられている。これはユーザの関心をわかりやすく表示する。しかし各分類の意味は明示的ではなく、理解に努力が必要である。

二つの手法には、以上のようにそれぞれ問題点が存在している。さらに共通する問題は、人の持つ話題を同定する方法としてテキスト解析を用いることの限界である。例えば、ユーザが動物のページを収集していた場合、ページには象や猿などの語は多く現れるだろうが、ユーザの関心を示す動物という語はなかなかあらわれない。このように、話題を示すような重要な語は出現しない場合がある。というのも、「象は動物である」といった一般知識となっている事柄は、多くはページには現れないからである。

以上のようなことから、本研究ではユーザの関心を計算により導き出す方法ではなく、ユーザ自身の知識を利用する方法を採ることにした。今回、本研究ではユーザ情報として WWW のブックマークの階層構造を利用している。フォルダ名とフォルダの中にあるページはユーザが持つ内在的な知識・関心を外在化しようと努力した結果である。例えば、フォルダの名前はユーザの関心を示すものであるし、フォルダ中には、ユーザがその話題に属すると考えたページが入っている。ブックマークはツリー構造であり、単一の階層構造に限定されているが、WWW に関するユーザの知識を示すものとしては十分であると考ええる。

WWW のブックマークは、人手によってフィルタリングがかけられた、利用価値の高い情報である。この情報を元に、効果的な情報推薦を行おうとする研究はすでにいくつか行われている。PowerBookmarks [Li 99] は、ユーザのコメントが記入されているブックマークデータをデータベースで管理し、情報の推薦を行う。WebTagger [Keller 97] は、タグ付けを行ったブックマークを用いて、情報共有を効率よく行おうとしている。BookmarkAgent [Mori 99] は、ユーザのブックマーク情報を持つエージェントが、ユーザが現在見ている Web ページに対して情報推薦を行う。Sitereer [Rucker 97] は、ブックマークのフォルダ構造を利用し、フォルダ内の登録された Web ページの比較によって推薦を行う。また、blink.com^{*1}においては、ブックマーク共有サービスが行われている。これらのシステムでは、ブックマークに登録された情報そのものを有効に共有するために、コメントやタグ、テキスト解析などを用いている。つまりユーザが階層構造を用いて整理する対象となる情報のオブジェクトに注目しているが、本システムではその階層構造を含めて所有者を表す知識であると考えている。

*1 <http://www.blink.com/>

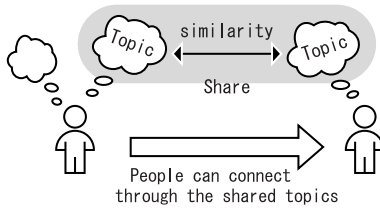


図 1 共通話題による人のつながり

残る問題として、話題間の関係を見つける事が挙げられる。これはフォルダ中のページの類似性を測ることで、それを見つけ出すことにしている。この処理には、通常のテキスト解析の手法を用いている。ただし、このテキスト解析には必ずしも高い精度を期待しているわけではない。というのは、ページ間の類似性がある程度発見するだけで、あとはページの共有状態から話題間の関係を見つけ出せると考えているからである。

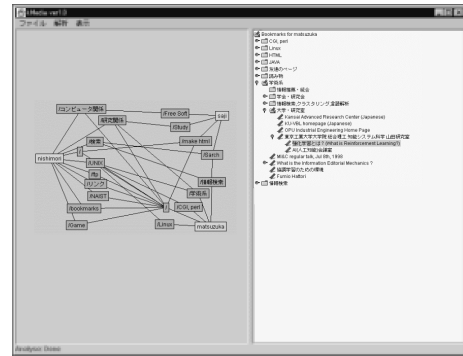
3. 共通話題ネットワーク

本研究で、我々は人のネットワークの実現モデルとして共通話題ネットワークを提案する。ここではユーザの持つ一定のまとまりの興味・関心を話題と呼ぶことにする*2。共通話題ネットワークとは、それぞれのユーザが持つ話題の中で、互いに共通性のあるものの共有したネットワークである。図 1 が示すように、ユーザは共通話題を介して他のユーザとつながることができる。

情報交換の点から見た共通話題ネットワークの利点として、以下のようなものが挙げられる。

- ページやファイルのような個々の情報よりも抽象度の高い話題という情報を共有することで、情報の交換を容易にすることができる。
- 人が作り出した話題という情報の関連性を利用することで、既存のテキスト解析では発見できない関連性を利用できる。
- 自分で作った話題に対して新しい情報が入ってくるという形で情報交換が行われるので、受信した情報の理解が容易である。

ここで重要なことは、話題の共通性は単純に話題の名前の比較によって求めることは出来ないという事である。ある情報が与えられた時、その情報がどういった話題に分類されるかは基本的に、人それぞれの判断にゆだねられている。そのため、話題の共通性はその話題に含まれる情報までを調べて初めて判断できる。次章では、この共通話題ネットワークをブックマークを利用して実装した *kMedia* システムについて説明を行う。

図 2 *kMedia* のユーザインタフェース

4. *kMedia* システム

4.1 システム概要

本研究では共通話題ネットワークをブックマークデータを用いて構築するためのシステム、*kMedia* を作成した。*kMedia* はクライアント・サーバ型のシステム構成をとって、クライアント部は Java2、サーバ部は CGI+Perl5.0 で作成されている。

利用方法は、クライアントを起動させて、設定画面からユーザ名と自分のブックマークデータのある場所を設定して処理を実行させる。クライアントはブックマークデータの解析処理を行い、サーバにその結果を送信する。そして、サーバでの処理が終わり、計算結果が返送されると、クライアントのユーザインタフェースに共通話題ネットワークとウェブページの推薦情報がそれぞれ表示される。

なお、ブックマークデータは Netscape Navigator*3 のものを用いているが、Internet Explorer*4 のブックマークでも Netscape 形式にブックマークデータを変換する事で利用可能である。

4.2 インタフェース

kMedia のユーザインタフェース (図 2) は、共通話題ウィンドウとブックマークウィンドウの 2 つのウィンドウと、ステータスバーで構成されている。

共通話題ウィンドウ (図 3) には、共通話題ネットワークが表示される。共通話題ネットワークとは、それぞれのユーザが興味を持つ話題のうち、互いに共通性のあるものを無向グラフで表示したものであり、ユーザを示す黄色のノード (図 3 における A, B, C というラベルが付けられたノード) と話題を示す青色のノード (図 3 における A, B, C 以外のノード) によって、それぞれのユーザの話題の共通性が示されている。

図 3 のグラフは話題を仲介に人がつながっているの、人と人のつながりを表すグラフとしても理解することができる。例えば、図 3 では A さんを中心にみると、UNIX

*2 話題といった場合、「最近の話題」のような時間軸を考慮したものもあるが、ここでは扱わない。

*3 <http://www.netscape.com/>

*4 <http://www.microsoft.com/>

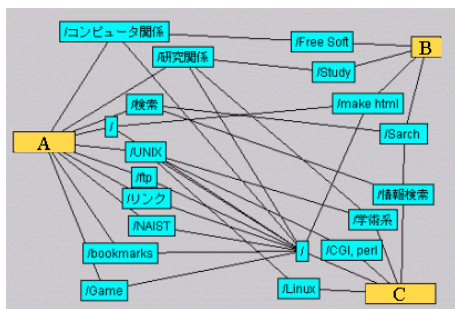


図 3 共通話題ウィンドウ

関係の話題 (UNIX と ftp など) は C さんの話題に繋がっている。このことから、A さんは UNIX で困ったことがあれば C さんに聞けば良いのかな、といった判断を行うことができる。このように共通話題を用いると、人と人の関係をそのつながりの量や内容から理解することができる。

図 2 の右側のウィンドウがブックマークウィンドウである。ここには自分のブックマークデータに、他のユーザのデータで類似したものが推薦情報として付加されて表示される。ユーザは Web ページアイコンをダブルクリックすることで、Web ページをブラウジングすることが可能である。

4.3 共通話題ネットワークの発見

kMedia では、共通ネットワークを発見するために、ブックマークからユーザが興味を持っている話題を調べ、次にユーザ同士の話題の共通性を調べる。

まず、ユーザがどんな話題に興味を持っているかを知るために、*kMedia* はブックマークのフォルダを利用している。フォルダにはユーザが任意の名前を付けることができるが、一般的にフォルダ名には、その中にある Web ページの話題は何であるかが記述されている。そこで、*kMedia* ではブックマークのフォルダにつけられた名前は、その中にある Web ページの話題を表しているとして、また、フォルダ中にある Web ページは、その話題に関する情報であるとしている。

ユーザ同士の話題の共通性は、その話題に含まれる情報がどれだけ類似しているかによって決まる。そこで *kMedia* では、話題の共通性はブックマークフォルダに含まれる情報、すなわちフォルダ中に含まれる Web ページの類似性から求める。つまり、Web ページの類似性が図 4-a のように発見された場合、話題の共通性は図 4-b のように示される。

なお、*kMedia* では、Web ページの類似性を、ページ中の重要な単語 (キーワード) の一致によって決める。Web ページ中の出現頻度の多い単語の中から、上位幾つかをそのページのキーワードとし、互いのページで一致したキーワード数が多いほど、そのページ間には類似性があるとしている。

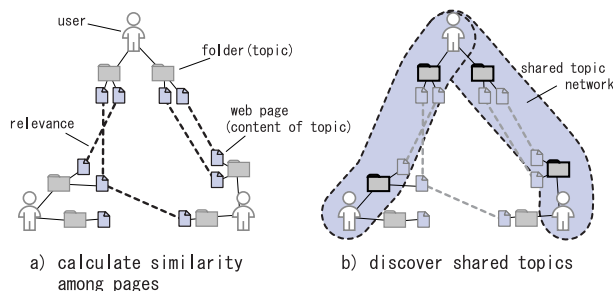


図 4 共通話題の発見

ここで求められた Web ページの類似性を元に、*kMedia* はブックマークデータの推薦を行っている。すなわち、各ユーザが持っているブックマークに登録された Web ページで、設定した閾値以上のキーワードが一致する Web ページを推薦情報として提示する。

Web ページ間の類似

kMedia は指定されたブックマークファイルを解析し、ブックマークに登録された Web ページを読み込み、その Web ページの文章から単語を切り出す。単語の切り出しには、茶筌 [松本 99] などによる形態素解析ではなく、字種に注目した単語切り出し [片岡 97] を用いた。

字種に注目した単語切り出しとは、解析する文章を先頭から一字ずつ読み取り、連続する漢字、カタカナ、全角英数文字を単語として切り出す方法である。この単語切り出し方法は形態素解析に比べて厳密には欠けるが、特にカタカナの新語、略語に対して柔軟な切り出しができるため、WWW の情報のように、雑多でやわらかい文章の解析に適している。

切り出された単語のうち、頻度順に、頻度が同じ場合は文字長の長い順に最大 10 単語をそのページのキーワードとしている。ただし、「ホームページ」や「カウント」など、あまりその文書の特徴付けられないような頻出単語はキーワードとして除外した。なお、単語と見なす最短文字長は調整可能であるが、今回は 3 文字とした。

このようにして選ばれた各 Web ページのキーワードの一致から、Web ページ間の類似を調べる。今回は、キーワードが 2 語以上一致し、一致したキーワードが両 Web ページで合計 10 回以上現れた場合に、Web ページ間に類似性があったと判断し、Web ページの推薦を行う。なお、このキーワードの出現回数のことを、以後、ページ関連度と呼ぶ。

話題の共通性

話題の共通性は、話題を示すブックマークフォルダ中の Web ページの類似によって決める。フォルダ間で互いに類似性がある Web ページの数をフォルダ関連度と呼ぶ。以下の実験では、フォルダ関連度が 3 以上、つまりフォルダ中の Web ページが 3 つ以上、互いに類似性が

ある場合に、そのフォルダ（話題）間に共通性があるとし、フォルダの推薦を行う。

なお、以下の実験では話題の共通性を調べる際、フォルダの階層は 2 階層以上は考慮していない。つまり、2 階層以上深いフォルダ内の Web ページは、全て親フォルダ内の Web ページと見なした。kMedia では、2 階層以上を扱うことも可能であるが、表示や分析の煩雑さを考慮し、このようにした。

5. 実験概要

このような共通話題ネットワーク発見手法の有効性について、実験を行って検証した。まず本章では実験の概要について述べる。

実験の目的は、人と人を行き交う情報として話題を用いることの有効性を検証すること、人と人との関係を測るための指標を見つけることの 2 点である。

被験者として、4 つの異なるコミュニティから 3 人ずつを集めた。ここで言うコミュニティとは、実生活上一つのグループに属して活動していることを指す（今回の例では大学・大学院の研究室）。同じコミュニティに所属する人同士と、異なるコミュニティに所属する人同士とで、ユーザの評価がどの程度変わるかを調べるために、互いに異なるコミュニティに属する者だけのグループを、同じ 12 人の被験者を用いて 4 グループ作った。

この実験では、kMedia を通常の利用法とは異なる方法を用いた。まず、被験者から集めたブックマークを実験者が kMedia に与える。そして、その結果を HTML 形式に加工し、フォームを用いて評価してもらった。各被験者には、まず同じコミュニティに所属する人同士で kMedia を使った場合の推薦結果に対する評価、次に異なるコミュニティに所属する人同士での結果に対する評価をしてもらう。なお、被験者には推薦相手の名前は伏せており、推薦された情報が同じコミュニティに所属している人からのものか、そうでないかはわからない。

評価は、以下に示すような推薦されたページに対する評価と推薦されたフォルダに対する評価、そしてページやフォルダを推薦した推薦者に対する評価の合計 3 種類 5 項目、それぞれ 5 段階（5:大変そう思う, 4:そう思う, 3:どちらとも言えない, 2:そう思わない, 1:全く思わない）で評価を行ってもらった。

- あるページに対してのページ推薦が妥当であるか
- 推薦されたフォルダは似ているか（フォルダ評価 1）
- 推薦されたフォルダは役立つか（フォルダ評価 2）
- 相手と連絡を取りたいか（推薦者評価 1）
- 相手と会いたい（推薦者評価 2）

まず被験者には推薦されたページに対して、自分が持つ被推薦ページと推薦ページの組み合わせが適切であるかどうか、それぞれ 5 段階で評価を行ってもらった。次に、推薦されたフォルダに対して、「似ているか」「役立つ

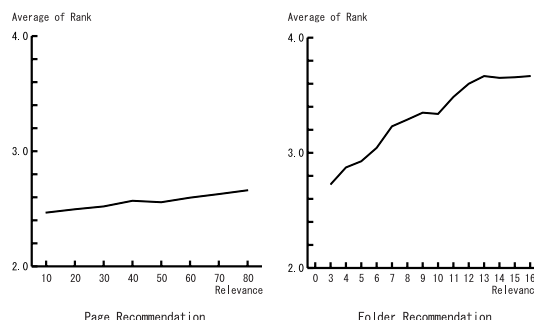


図 5 フォルダ推薦の有効性

つか」という 2 つの観点から、5 段階で評価を行ってもらった。なお、評価フォームの下には相手のブックマークが表示されており、被験者はそのフォルダ内の Web ページを見ながら、推薦されたフォルダを評価する。

6. ページとフォルダに関する共通性の分析

6.1 フォルダ推薦の有効性

図 5 は、ページ推薦とフォルダ推薦の評価結果を示したものである。なお、フォルダ推薦については 2 種類の評価をしてもらったが、両方とも同じ傾向であったため、「似ているか」の評価結果を示す。

それぞれ縦軸が平均評価値、横軸が推薦を行うための関連度の閾値である。今回の実験で用いた閾値（ページ関連度 10、フォルダ関連度 3）でも t 検定（有意水準 5%）をしたところ、有意にフォルダ推薦の方が評価が高かった。また、フォルダ推薦に対する評価値はどのような閾値においてもフォルダ推薦の方が高く、また、グラフの傾きは顕著にフォルダ推薦の方が大きい。これはフォルダ関連度の方がユーザの評価により関連している事を示している。このことから、フォルダ（話題）が概念的に Web ページよりも抽象的であるため、ユーザにとって受け入れ易かったのではないかと考えられる。

6.2 コミュニティの効果

コミュニティの有無で情報推薦の効果がどのように変わるか検証する。同じコミュニティに所属する人同士で行ったページ推薦と、異なるコミュニティに所属する人同士で行ったページ推薦の評価を比較した。図 6 は、ページ推薦を行うためのページ関連度の閾値を変化させた場合の、推薦されたページに対する評価の平均値を示したものである。

グラフから、コミュニティの有無に関係なく、ページ関連度によって設定された閾値が高くなれば、評価も高くなる傾向にあることがわかる。ただし、同じコミュニティ間の方が評価は常に高い。今回用いた閾値（ページ関連度 10）でも t 検定（有意水準 5%）を行ったところ、有意であった。情報推薦は、一般に似たような関心をもつ人同士で行った方が有効であるといわれている。kMedia

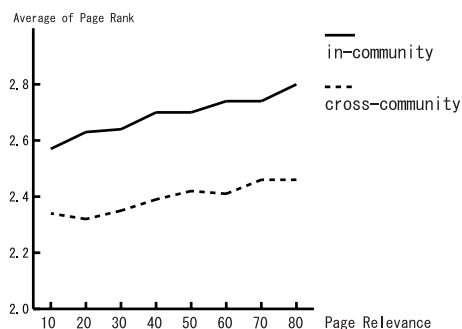


図 6 ページ関連度と推薦ページ評価

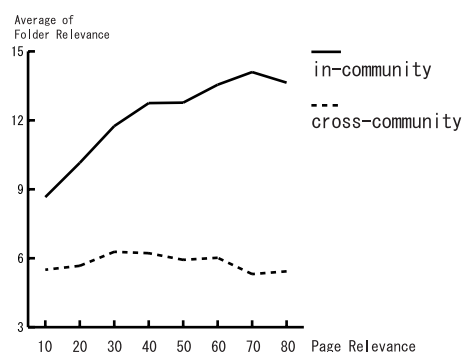


図 8 ページ関連度とフォルダ関連度

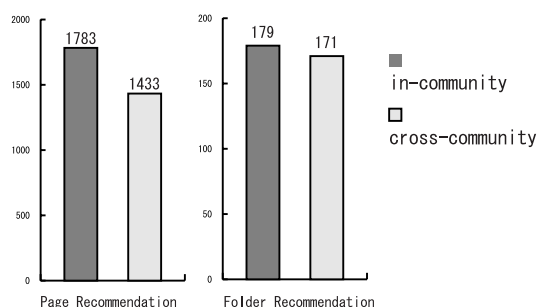


図 7 ページ・フォルダ推薦数

によるページ推薦の効果も、同じコミュニティに所属する人同士で行った方が高いという結果が得られた。また、フォルダ推薦においてもほぼ同様の傾向が現れた。

大小の差はあるが、結果としてページ推薦、フォルダ推薦ともにコミュニティの有無による影響が現れた。実生活においても、親しい友人から推薦された情報は他の人からのものよりも有効である可能性が高い。この結果はそれと同様であるといえる。しかし今回の実験においては推薦者が誰であるかは伏せられているので、その所属で判定しているわけではない。ここで行った関係の提示の中に、“コミュニティらしさ”が含まれていたと考えられる。すなわち、人が相手を同じコミュニティだと判断する要素、つまりは人のつながりを示す要素が含まれている筈である。

そこで、人のつながりを示す要素を見つけ出すために、推薦の評価結果以外にどのような特徴がコミュニティの有無で現れているかを調べる。

§ 1 ページ推薦数とフォルダ推薦数

図 7 は、同じコミュニティ間でのページ推薦と異なるコミュニティ間でのページ推薦数、およびフォルダ推薦数を比較したものである。コミュニティの有無で推薦ページ数の差は 300 ほどあるのに対し、推薦フォルダ数の差はわずか 8 であった。

フォルダ推薦はフォルダ中のページの推薦数によって決まる。よって、ページ推薦数が多いのにフォルダ推薦数が小さいと言うことは、推薦されたフォルダが多くの推薦ページを持っていることを示している。つまり、こ

の結果から、同一コミュニティはそうでない場合と比べて、推薦フォルダがより多くの推薦ページを持つ傾向にあることがわかる。

§ 2 ページ関連度とフォルダ関連度

図 8 は、ページ関連度による閾値を変化させた場合、推薦されるページのフォルダ関連度がどのように変化するかを示したものである。推薦ページのフォルダ関連度とは、類似性のあるページ組の、それぞれが登録されているフォルダ間のフォルダ関連度を示す。

ページ関連度による閾値を高くすると、同じコミュニティに所属する人同士で推薦されるページは、フォルダ関連度の高いものが多くなっている。対して、異なるコミュニティでは、そのような傾向は現れない。

7. 人の間の距離に関する分析

7.1 カテゴリズ近似度

推薦数および関連度についての分析から、実社会において同じコミュニティに属している（つまり人のつながりがある）ことは、そうでない場合と比較して以下のような二つの傾向があるといえる。

- ページ推薦数に対してフォルダ推薦数は少ない
- ページ関連度とフォルダ関連度の相関が高い

前者は、一つの共通話題に対して多くの類似ページがあることを示し、後者は、よく似たページはよく似た話題に分類されていることを示している。この二つの傾向から、人のつながりの尺度として、分類方法の類似性が使えるのではないかと考えられる。例えば図 9 において、点線は関連性があるページ組を示しているが、左右両方とも同じ 4 つである。だが、その関連性のあるページの分類の仕方がそれぞれ異なっている。この場合、同じような分類をしている左の方が、右の方と比較して人のつながりが強いと判断できる。

ここで分類方法が類似性を示すものとして、カテゴリズ近似度という指標を新たに作る。これは、共有している情報の量でユーザ間を評価するのではなく、一致する率で評価するものである。ユーザ a , b 間のカテゴリズ近似度 $CR(a,b)$ は、以下の式で求めることができる。

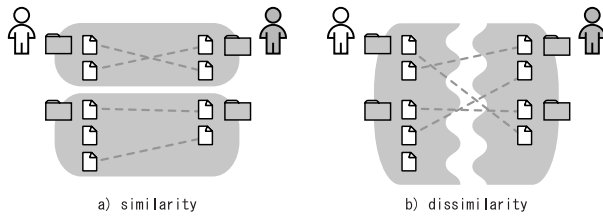


図 9 分類方法の類似

表 1 推薦者の評価との相関 (1)

	連絡したい	会いたい
推薦ページ数	0.42	0.30
平均ページ関連度	-0.13	-0.19
推薦フォルダ数	0.45	0.35
平均フォルダ関連度	0.38	0.30

$$CR(a,b) = \frac{Nf_{ab} \times Rf_{ab}}{Np_{ab}}$$

Nf_{ab} : 推薦フォルダ数

Rf_{ab} : 推薦フォルダの平均フォルダ関連度

Np_{ab} : 推薦ページ数

7.2 推薦者評価とカテゴライズ近似度

今回の実験では、推薦されたページやフォルダと同様に、それらのページやフォルダを推薦した推薦者自身の評価も行った。この、推薦者への評価という値は、人のつながりを示す値であると言える。そこで、それらの値とカテゴライズ近似度の相関を調べることで、提案するパラメータが人のつながりを測る指標となりうるのかを調べた。

まず、比較対象としてシステムが計算した四つの基本的なパラメータと被験者による推薦者への評価との相関係数を表 1 に示す。最も相関係数が高いパラメータは推薦フォルダ数であった。また、推薦者に対する評価は「連絡したいか」「会いたい」という 2 つの項目を用意したが、全てのパラメータにおいて前者の方が相関係数は高いという結果になった。

次に示す表 2 は、推薦されたページやフォルダに対する被験者の評価と、その推薦者に対する評価との相関を示したものである。被験者による評価係数は、「会いたい」との相関においてシステムが求めたパラメータよりも高い値を出しているが、「連絡したい」との相関においては逆に低い結果となった。

「連絡したい」と「会いたい」という二つの評価項目を比較すると、直感的には後者の方が人のつながりを強くあらわしていると思われる。ならば、人のつながりを測る指標は

(「連絡したい」との相関) < (「会いたい」との相関)

表 2 推薦者の評価との相関 (2)

	連絡したい	会いたい
平均ページ評価	0.29	0.40
平均フォルダ評価 1	0.28	0.32
平均フォルダ評価 2	0.09	0.20

表 3 推薦者の評価との相関 (3)

	連絡したい	会いたい
カテゴライズ近似度	0.49	0.55

となるべきである。しかし、システムが計算した値では逆の傾向が現れてしまう。このことから、システムが計算した値はなにかしら不十分であると言える。では人が評価した値はどうであるかという、「連絡したい」との相関係数と「会いたい」との相関係数の関係は正しい。しかし、不十分であると判断したシステムが計算した値よりも、「連絡したい」との相関係数において低い結果があらわれた。このことから、人が評価した値も、まだ不十分であると言える。

最後に、今回新たに提案したカテゴライズ近似度と人の評価との相関を表 3 に示す。

カテゴライズ近似度の相関係数は両項目とも最高値であり、かつ大小関係も妥当なもの（「会いたい」との相関の方が高い）であった。また、このカテゴライズ近似度は相関の有意性検定において、有意水準 5 % で有意であった。

7.3 ページ推薦とカテゴライズ近似度

さらにカテゴライズ近似度がページ推薦に与える影響を調べた。今回の実験では一人当たり四人の推薦者からの推薦情報を評価してもらったが、その推薦者四人のうちカテゴライズ近似度が高かった者上位二人を、被推薦者にとって同じコミュニティに属する人だと判断して新たに仮想的にコミュニティを作成した。

図 10 は、全体 (all)、実際のコミュニティがある場合 (in real community) とない場合 (cross real community)、新しく作った仮想コミュニティがある場合 (in new group) の合計四種類のグループにおけるページ推薦の評価値の平均値を示している。実際のコミュニティ同様、カテゴライズ近似度で作った新しいコミュニティにおいても、その有無によってページ推薦の評価値に差が現れた。この差は t 検定 (有意水準 5 %) で有意であった。また、コミュニティを全く考慮しない場合 (図 10-all) との差も、同じく t 検定 (有意水準 5 %) で有意であった。

8. 結 論

本稿では、情報交換を容易にするために人のつながりはどうのように説明されるべきか、という点について議論し、共通話題を提示するシステム、*kMedia* を提案した。

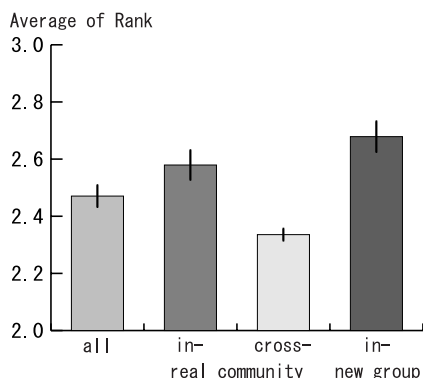


図 10 ページ推薦におけるカテゴライズ近似度の有効性

今回、共通話題ネットワークの生成においてはブックマークというユーザの知識とページ関連度の自動発見という組み合わせで行ったが、この人の知識と機械的解析による共通話題の発見手法は实际的であったといえる。ページ間の関係発見ではなく、話題間の関係を発見することが有効であることが結果の分析でも確かめられた。また、人の関係においては、ただ類似性の発見されたページやフォルダの数を測るだけでなく、フォルダによるページの分類の仕方、つまり分類観点の類似性を見ることで発見することが出来た。これは、人がどのように話題を構成しているかを見比べていることであり、ここからも共通話題という概念の有効性がわかる。

9. おわりに

WWWの拡大に伴い、ユーザの興味を反映したフィルタリングを可能とするWWW検索システムの必要性が高まっている。このとき、実世界と同じように、同じ関心を持つ人間同士で情報を交換・共有することが可能ならば、効率的に情報収集することが可能である。

より良い情報を得るために、情報共有を行うユーザ数は多い方がよい。だが、ユーザが多くなると、共有された情報から選択するのが困難になる。特にWebページのように分類が困難である場合、それは顕著に表れる。ユーザ間での情報共有を効率よく行うためには、まず情報源をある程度分別する必要がある。

本研究では、ブックマークという、ユーザによって作られた階層構造を持った知識を利用する事によって、自動的に話題のつながりおよび人のつながりを見出す手法を提案した。膨大な数のユーザによる情報共有が行われた場合、この手法を用いてることによって、適切な情報推薦を支援することができる。

本手法は情報推薦を支援するが、推薦された情報はユーザにとって未知の情報であり、その情報がどのようなつながりによって自分に推薦されたのかが適切に示されなければ、ユーザはその情報を理解することが出来ない。これでは良い推薦情報を発見したとしても、情報の推薦が

適切に行われたことにはならない。そこで今後は、発見された推薦情報をどのように提示すればユーザが負担無く受け入れられるか、という点についても考察する必要がある。

謝 辞

実験に協力してくださった被験者の方々、および研究のアドバイスを与えてくださった奈良先端科学技術大学院大学 知能情報処理学講座の皆様に感謝します。

◇ 参 考 文 献 ◇

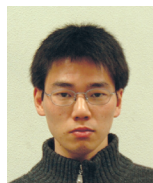
- [Balabanovic 97] Balabanovic, M. and Shoham, Y.: *Fab: Content-based, collaborative recommendation.*, *Communications of the ACM*, Vol. 40, No. 3, pp. 66 – 72 (1997).
- [Chen 98] Chen, L. and Sycara, K.: *WebMate: A Personal Agent for Browsing and Searching*, in *Proceedings of the 2nd International Conference on Autonomous Agents and Multi Agent Systems, AGENTS '98*, pp. 132 – 139 (1998).
- [Cutting 92] Cutting, D. R., Karger, D. R., Pederson, J. O., and Tukey, J. W.: *Scatter/Gather: A cluster-based approach to browsing large document collections*, in *15th Annual International ACM/SIGIR Conference*, pp. 318–329 (1992).
- [Donath 95] Donath, J. S.: *Visual Who: Animating the affinities and activities of an electronic community*, in *ACM Multimedia 95* (1995).
- [Erickson 99] Erickson, T., Smith, D. N., Kellogg, W. A., Laff, M., Richards, J. T., and Bradner, E.: *Socially Translucent Systems: Social Proxies, Persistent Conversation, and the Design of "Babble"*, in *CHI*, pp. 72–79 (1999).
- [Huberman 96] Huberman, B. A. and Kaminisky, M.: *Beehive: A system for cooperative filtering and sharing of information* (1996), <ftp://parcftp.xerox.com/pub/dynamics/beehive.html>.
- [Joachims 97] Joachims, T., Freitag, D., and Mitchell, T.: *WebWatcher: A Tour Guide for the World Wide Web*, in *IJCAI-97* (1997).
- [Kautz 97] Kautz, H., Selman, B., and Shah, M.: *Referral Web: Combining Social Networks and Collaborative Filtering*, *Communications of the ACM*, Vol. 40, No. 3, pp. 63–65 (1997).
- [Keller 97] Keller, R. M., Wolfe, S., Chen, J. R., Rabinowitz, J. L., and Mathe, N.: *A Bookmarking Service for Organizing and Sharing URLs*, in *Proceedings of The 6th International World Wide Web Conference (WWW-6)* (1997).
- [Konstan 97] Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gorden, L. R., and Riedl, J.: *GroupLens: Applying collaborative filtering to usenet news.*, *Communications of the ACM*, Vol. 40, No. 3, pp. 76–87 (1997).
- [Li 99] Li, W.-S., Vu, Q., Agrawal, D., Hara, Y., and Takano, H.: *PowerBookmarks: A System for Personalizable Web Information Organization, Sharing, and Management*, in *Proceedings of The 8th International World Wide Web Conference (WWW-8)* (1999).
- [Lieberman 95] Lieberman, H.: *Letizia: An agent that assist web browsing.*, in *Proceedings of IJCAI-95*, pp. 924–929 (1995).
- [片岡 97] 片岡, 今中, 水谷, 若見: *テキスト情報を対象としたキーワード抽出と関連情報収集システム*, *日本ファジイ学会誌*, Vol. 9, No. 5, pp. 710–717 (1997).
- [Mori 99] Mori, M. and Yamada, S.: *Bookmark-Agent: Information Sharing of URLs*, in *Poster Proceedings of The 8th International World Wide Web Conference (WWW-8)* (1999).

- [Rucker 97] Rucker, J. and Polanco, M. J.: Siteeer: Personalized navigation for the web, *Communications of the ACM*, Vol. 40, No. 3, pp. 73-75 (1997).
- [Shardanand 95] Shardanand, U. and Maes, P.: Social Information Filtering: Algorithms for Automating "Word of Mouth", in *CHI*, pp. 210-217 (1995).
- [Terveen 97] Terveen, L., Hill, W., Amento, B., McDonald, D., and Creter, J.: PHOAKS: A system for sharing recommendations., *Communications of the ACM*, Vol. 40, No. 3, pp. 59-62 (1997).
- [西田 00] 西田豊明: コミュニティの知識創造を支援するインタラクティブなメディアを目指して, 情報処理, Vol. 41, No. 5, pp. 542-546 (2000).
- [松本 99] 松本, 北内, 山下, 平野: 日本語形態素解析システム『茶筌』version2.2.0 使用説明書, NAIST Technical Report NAIST-IS-TR99012, 奈良先端科学技術大学院大学情報科学研究科 (1999).

〔担当委員: 大澤幸生〕

2001 年 9 月 16 日 受理

著者紹介



濱崎 雅弘(学生会員)

2000 年同志社大学工学部知識工学科卒業。同年、奈良先端科学技術大学院大学情報科学研究科博士前期課程入学、現在に至る。学部時代は遺伝的アルゴリズム、現在は情報共有やオンラインコミュニティの研究に従事。



武田 英明(正会員)

1986 年東京大学工学部卒業。1988 年同大学大学院修士課程修了。1991 年同大学大学院博士課程修了。1991 年財団法人日本システム開発研究所嘱託研究員。1992 年ノルウェー工科大学 doctoral fellow。1993 年奈良先端科学技術大学院大学情報科学研究科助手、1995 年同助教授、2000 年国立情報学研究所助教授、現在に至る。東京大学工学博士。知能の共有と再利用、設計学、知的 CAD などの研究に従事。1995 年人工知能学会全国大会優秀論文賞受賞。人工知能学会, AAAI 各会員。



松塚 健

1998 年大阪府立大学工学部経営工学科卒業。2000 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了。修士(工学)。在学中は、情報共有とネットワークコミュニティに関する研究に従事。



谷口 雄一郎

1999 年 3 月福井大学工学部卒業。2001 年 3 月奈良先端科学技術大学院大学情報科学学科博士前期課程修了。修士(工学)。プラエセンス(株)を設立しデスクトップエージェントプラットフォーム「ペルソナウェア」を商品化。現在取締役。在学中は、音声対話システムに関する研究に従事。



河野 恭之(正会員)

1989 年大阪大学基礎工学部情報工学科卒業。1994 年同大学大学院基礎工学研究科博士後期課程修了。博士(工学)。同年(株)東芝入社。同社関西研究所研究主務などを経て、2000 年 4 月奈良先端科学技術大学院大学情報科学研究科助教授。知的 CAI, マルチモーダル理解, 音声対話 HI, 知的インタフェース, ウェアラブルインタフェースの研究に従事。電子情報通信学会, 情報処理学会, ヒューマンインタフェース学会, ACM 各会員。



木戸出 正継

1970 年京都大学大学院工学研究科修士課程修了。同年、東京芝浦電気(現、東芝)総合研究所入社。1975-1977 年米国パーデュー大学客員研究員。1997-1999 年東芝アメリカ社副社長。2000 年奈良先端科学技術大学院大学情報科学研究科教授、現在に至る。知能情報処理、特に画像処理とヒューマンインタフェースの研究に従事。1985 年オーム技術賞、1994 年電子情報通信学会業績賞。著書「画像データベース(オーム社)」, 「コンピュータ画像処理入門(総研出版、共著)」など。京都大学工学博士。電子情報通信学会, 情報処理学会, IEEE など会員。国際パターン認識協会(IAPR)フェロー。