閲覧履歴に基づいた人とページの検索手法の提案

三村　紗織†　　松井　優也†　　河合由起子†　　中島　伸介†

† 京都産業大学　〒 603–8555 京都府北区上賀茂本山

E-mail: †{g738471,i1054056,kawai}@cc.kyoto-su.ac.jp, ††nakajima@cse.kyoto-su.ac.jpp

あらまし　我々はこれまで，閲覧者の量に基づいた検索およびそれら閲覧者とのコミュニケーションを同時に実現するソーシャルサーチを検討してきた．これにより，流行りのページ検索だけでなく，同じページを閲覧している他のユーザとのコミュニケーションによる情報獲得を実現できた．しかながら，検索とコミュニケーション双方において利用者全てが対象であり，ユーザの知識や興味に応じた他閲覧者やページの発見は困難であった．そこで本研究では，ユーザとの興味が類似しており広い/深い知識を持った人およびページを発見することで，効果的にページを通してコミュニケーションすることを目的とする．具体的には，各ユーザのページ閲覧履歴を用いて重心ベクトルと標準偏差を算出し，興味の類似するユーザおよび広く/深く，また多くの知識をもつユーザを発見する．また，それらユーザが閲覧したページのうち協調フィルタリングより興味の類似するページを推薦する．さらに，広い/深い知識をもつユーザが閲覧したページのうち，関連度の高いページをユーザにとって新たな情報を含むページとして推薦する．本稿では，閲覧履歴を用いることで，ユーザとの興味が類似しているユーザとページおよび広い/深い知識をもつユーザとページの発見手法を提案する．

キーワード　コミュニケーション, Web 検索

1. はじめに

Web における知識獲得行動には，ページ検索と，掲示板や twitter などのソーシャルサービスを利用して人に問い合わせる方法がある．ページ検索は速度と網羅性の点で有効であり，ソーシャルサービスは検索サービスよりも時間や手間を要するが，人とのコミュニケーションにより質の高い情報が得られるという利点がある．近年では，ページの情報から専門家を抽出し，検索キーワードに対する専門家を特定し，その専門家の連絡先をユーザへ提供したり，複数人で協調して検索するといった研究開発も行われている [1] [2]．しかしながら，両方の利点を同時に活用したサービスは未だない．

我々はこれまで，検索とソーシャルサービスの双方の利点を同時に得られるサービスとして，ページを閲覧しているユーザのネットワークを構築することで，1）閲覧者の量と質に基づいたページと人の検索ならびに，2）ページを通した他閲覧者とのコミュニケーション，を実現することを目指したソーシャルサーチシステムを検討してきた [15]．検索手法はハイパーリンク構造を用いた Pagerank アルゴリズムに基づいており，各ページ間のリンクに対して閲覧者のアクセス履歴を考慮した重みを付与する．各ページは，この重みが付与されたリンク構造によってランキングされる．これにより，多くのユーザが閲覧している注目度の高いページや，同じ検索キーワードに対して興味を持っているユーザを発見できる．このページと人との同時発見により，検索結果ページを閲覧しているユーザとの即時的コミュニケーションが実現できた．しかしながら，検索とコミュニケーションの双方においてユーザ全てが対象であり，ユーザの知識や興味にあったページを閲覧している閲覧者やページの発見は困難であった．

そこで，本研究ではユーザとの興味が類似しており，また，広い/深い知識を持った人およびページを発見することで，興味のあるページを通して効果的に他閲覧者とコミュニケーションすることを目的とする．提案するユーザ検索方法は，ページの閲覧履歴を用いてユーザの特徴語を抽出し，それらから各ユーザの類似性や知識の獲得状況を推定することで，興味の類似するユーザおよび広く多くの知識をもつユーザを発見する．ページの推薦手法は，興味の類似するユーザが閲覧したページから選別し推薦する．また，広い/深い知識をもつユーザが閲覧したページのうち，類似性が異なるページをユーザにとって新たな情報を含むページとして推薦する．

発見した人の推薦方法は，閲覧しているページを当該ユーザが閲覧しているとは限らないため，当該ユーザが閲覧しているページの URL を推薦する．なお，当該ユーザは自身が現在閲覧していることを通知するか否かの公開・非公開の設定ができる．

本論文では，閲覧履歴を用いることで，ユーザとの興味が類似しているユーザとページおよび広い/深い知識をもつユーザとページの発見手法を提案する．

本稿の構成は以下の通りである．2 章ではシステム設計，3 章で実験について述べる．4 章で今後の課題について述べた後，5 章でまとめとする．

2. システム概要

図 1 にこれまで開発してきたソーシャルサーチシステムの概要を示す．本システムはユーザから検索キーワードを取得すると，リアルタイムに閲覧しているユーザおよび過去のアクセスユーザに基づいて順位付けをした検索結果を提示する．検索結果として表示される各リンクには今現在，そのリンク先にいる

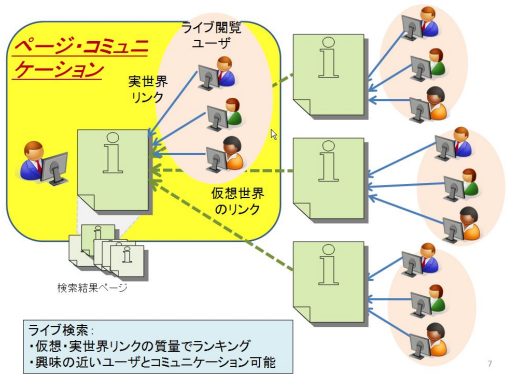

図 1　システムの概要

閲覧者の数が表示される．ユーザは検索結果リストのリンクをクリックするとページの閲覧とともに，そのページを閲覧している他ユーザとその場でチャットによるコミュニケーションができる．また，各ページ内の全リンクに関しても閲覧ユーザの数が提示されており，リンク先の閲覧ユーザの存在を確認できるため，より多くのユーザとコミュニケーションを行う機会を得られる．さらに，ページ内にアノテーション（ハイライト）を簡単に付与できることで，注目個所や他ユーザと共通の話題にしたい個所として共有でき，ページ内の情報発見やページ内情報を用いたコミュニケーションを活性化できる．

本研究では，ユーザの興味や知識に応じたユーザとのコミュニケーションを第一の目的とし，また，興味にあったページや新たな知識を得られるページを発見し推薦することを目的とする．以下では，これらの特徴を実現する各機能の特徴を述べる．

2.1　興味と知識の抽出と人の発見

各ユーザの興味と知識を抽出し，そのユーザと興味の類似するユーザおよび異なる知識をもつユーザの発見で，効果的な検索とコミュニケーションを促進することを目指す．

各ユーザの興味と知識の抽出手順を示す．まず，下記の手順にて各ユーザの興味と知識を抽出する．

（1）　閲覧した各ページの tf・idf 値の高い単語を特徴語として抽出し，ページベクトルを生成

（2）　ユーザごとに閲覧したページのうち，閲覧間隔が一定時間内のページのみ抽出

（3）　抽出したページのページベクトルから重心ベクトルと標準偏差を算出

算出した重心ベクトルを各ユーザの興味とし，標準偏差を知識獲得状況とする．ユーザが任意のページを閲覧した際に，上記の手法で重心ベクトルを算出し，各重心ベクトルとのユークリッド距離の近いものを興味の類似するユーザと判定する．

次に，閲覧したユーザの重心ベクトルと標準偏差が他のユーザの標準偏差内に含まれており，重心ベクトルとの距離が近く標準偏差が閾値以上の場合，そのユーザの知識範囲をカバーしつつ広い/深い知識をもつユーザとする．ただし，ページ閲覧数が一定数以上とする．

以上の手順より，発見された人は，本システムのコミュニケーション画面に提示される．しかしながら，閲覧しているページ

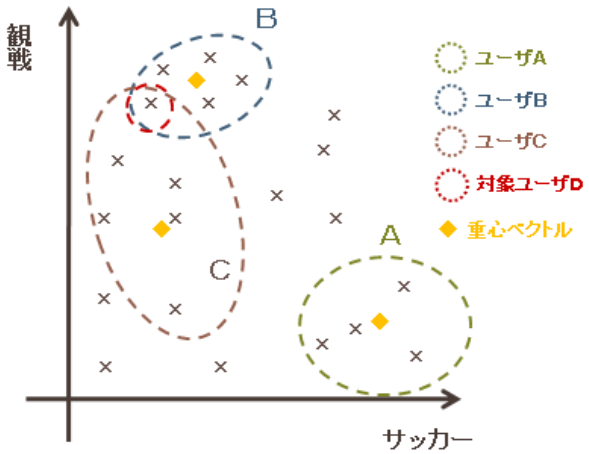

図 2　興味ベクトル生成

を当該ユーザが閲覧しているとは限らない．そこで，発見された人が他のページを閲覧している場合は，そのページへのリンクを生成提示する．これにより，知識人となるユーザに問い合わせしたい場合は，リンク先のページに移動することで，直接問い合わせることができる．ただし，ユーザがコミュニケーション許可をしている場合のみリンクは生成され提示される．

2.2　ページの推薦

ユーザの興味や知識に応じたユーザを発見した後，ページを選別する．まず，興味の類似するユーザの閲覧履歴より，協調フィルタリングにより推薦ページを決定する．次に，広い/深い知識をもつユーザの閲覧履歴より，同様に，協調フィルタリングより推薦ページを決定する．

以上より推薦されたページは，本システムの検索結果として提示されるだけでなく，各ページを閲覧する度にコミュニケーション画面と同様に提示される．

3.　システム設計

本システムはサーバ・クライアント型システムとして設計した．以下ではシステムの詳細な設計について述べる．

3.1　ユーザの興味ベクトル生成

本システムでは，サーバはユーザの閲覧履歴 (図 1) を管理し，その履歴情報を用いてユーザの興味や知識を抽出する．表 1 に閲覧履歴のスキーマを示す．ユーザの興味抽出は，

（1）　各ページのページベクトルを生成

（2）　ユーザの閲覧履歴から興味ベクトルを生成

(1) では，まず，各ページから単語の tf 値，閲覧ページ全体から idf 値を算出し，各ページごとに $tf \cdot idf$ 値を算出する，算出した tf・idf 値の大きい順に単語 n 個を抽出しページベクトルの要素とする．(2) では，ページ間の閲覧時間間隔 T 以内ごとにページを分類し，分類したページ集合ごとにページベクトルを用いて，重心ベクトルを算出する．この重心ベクトルをユーザの興味ベクトルとする．つまり，ユーザの興味ベクトルは一定時間内の閲覧履歴情報ごとに生成される．図 4 に興味ベクトル生成の例を示す．まず，例としてサッカーと観戦に興味を持つ人がいるとする．図の×しるしはサッカーと観戦という

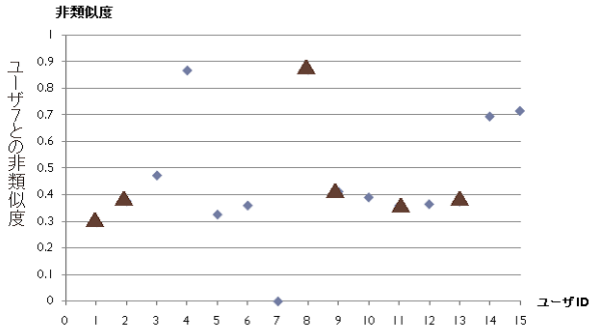

図 3　実 験 結 果

キーワードを特徴語に持つページである．ユーザA，ユーザB，ユーザCそれぞれがページを閲覧した．さらに対象ユーザDが任意のページを閲覧した場合．ユーザDからの興味空間が近く，ユーザDよりも空間が広い人を推薦する．

3.2　類似ユーザの選出

ユーザがページを閲覧すると，まず，前節の (1) と同様に閲覧しているページのページベクトルを算出した．次に，このページベクトルと閲覧履歴の全ページとのユークリッド距離を算出し，類似ページを抽出する．類似ページを閲覧したユーザを対象として，前節で抽出した各ユーザの興味ベクトル（重心ベクトル）とのユークリッド距離を算出し，興味の類似するユーザとして選出する．なお，初めてのユーザ出ない場合は閲覧中のページと類似するページをすでに閲覧していたページを用いて興味ベクトルを生成する場合，この興味ベクトルと他ユーザとの興味ベクトルのユークリッド距離を算出し，興味類似ユーザとして選出する．図 4 の例では興味ベクトルが対象ユーザDと近い位置にある，ユーザBとユーザCが興味類似ユーザとして選出される．

3.3　広い／深い知識を持つユーザの分類

興味の類似するユーザの閲覧履歴を用いて，そのユーザの知識を推定する．前節で選出されたユーザの興味ベクトルは，閲覧しているページと類似したページを含んだページ集合から生成されている．それらのページ集合から標準偏差を算出し，標準偏差の閾値により，広い知識を持つユーザと深い知識をもつユーザに分類する．図 4 の例では，ユーザBが深い知識を持つユーザ，ユーザCが広い知識を持つユーザとして対象ユーザDに推薦される．

3.4　ページ推薦

前節では，興味が類似しており，広い／深い知識をもつユーザを発見した．それらのユーザが閲覧したページのうち，各ユーザごとに類似する興味ベクトルを含むページ集合を用いて，協調フィルタリングを行い，ページを選別する．ただし，選別されたページのうち既に閲覧しているページは除く．

3.5　ユーザ推薦

本実験では任意のユーザが，任意のページを閲覧した場合に推薦されるユーザについて検証する．

ユーザ ID が 7 のユーザが履歴にはないページを閲覧した際の，他のユーザとの非類似度を図 3 に示す．図 3 の横軸はユーザ ID，縦軸は類似度となっている．図 3 の▲と点は対象となるユーザ ID7 と各ユーザの類似度を示している．類似度が 0 に近いほどユーザ同士の類似性が高くなる．実験より閲覧したページと類似したページを閲覧したユーザとしてユーザ ID 1, 2, 3, 4, 5 が選別された．図 3 では選別されたユーザは▲で表されている．さらにその中から，類似度が低い 5 ユーザが今回対象となるユーザに推薦された．推薦されたユーザが閲覧したページ内容を見てみると，対象となるユーザとページ内容が似ている．しかし，まったく違う内容を見ているユーザも推薦されている．さらにユーザ ID4 は類似度が離れているのに似ているページを閲覧したユーザとして推薦された．

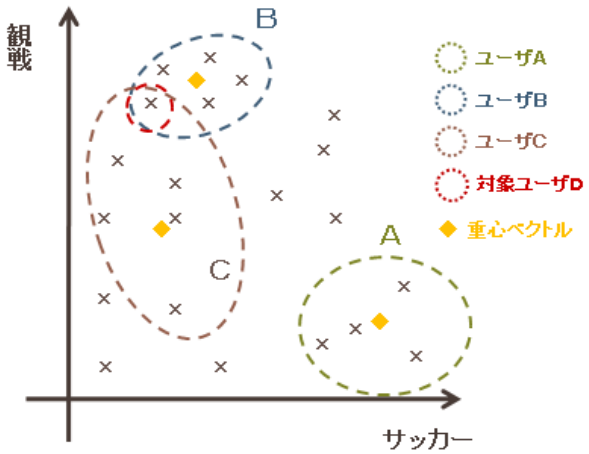

図 4　興味ベクトル生成

3.5.1　考　　察

実験結果より，内容が類似していページを多く閲覧しているユーザが選出されていることが示せた．また，それらのユーザの標準偏差より，閲覧しているページと類似しているページを多く閲覧しているユーザを深い知識を持つユーザ，類似しているページ以外のページも閲覧しているユーザを広い知識を持つユーザとして分類した．

しかしながら，類似度が離れているのに似ているユーザとして選別されたり，類似度が低いのにもかかわらず選別されないユーザもあった．これは，類似度を計算する際に使用するページの特徴語抽出において，ページに含まれる文章の量を考慮していなかったため，tf 値の大きな偏りが発生したためと考えられる．そのため，それぞれのページに適した特徴語が抽出されず，興味ベクトル間の正しい距離が計算されなかったと思われる．

3.5.2　今後の課題

提案ユーザ推薦手法で対象となるユーザに適したユーザが推薦されるかどうかを検証を行った．検証結果から，ユーザの閲覧履歴を用いて類似性や知識の獲得状況を推定することで，興味が似ているユーザを発見が実現されたと考えられる．しかし，ページの特徴語抽出においてページ内のデータ量に依存するため，特徴語の値が偏ってしまった．これを解消するために，適した特徴語が抽出されるように改善を行う予定である．また，今回はユーザ推薦の検証しか行っていないため，今後はページ

ID	URL	アクセス日時	delay	keyword1	tf・idf1	keyword2	tf・idf2
1	http://e-words.jp/w/PHP.html	02/15 21:20	0	PHP	0.023445	Perl	0.085658
2	http://www.amazon.co.jp/	02/15 21:20	0	amazon	0.1568	日用品	0.2256
1	http://www.php.gr.jp/	02/15 21:21	60	PHP	0.07836	言語	0.067855
3	http://www.tbs.co.jp/sebare/	02/15 21:23	0	スポーツ	0$\dot{0}$4567	試合	0.036574

表 1　ページ閲覧履歴

ID	ページタイトル	ID	ページタイトル	ID	ページタイトル
1	初心者用 PHP 入門 PHP マニュアル PHP 研究所 PHP VS Perl	6	テイルズオブイノセンス テイルズオブイノセンス (wiki) テイルズオブイノセンス攻略 テイルズシリーズ (wiki)	11	Yahoo!スポーツ スポーツナビ スポニチ スポーツ (wiki)
2	PHP vs Perl Perl vs PHP PHP の方が早いって本当? Perl(wiki)	7	Perl(wiki) とほほの Perl 入門 Perl 基礎入門 Perl 講座	12	スポーツナビサッカー Yahoo!スポーツサッカー サッカー (wiki) スポニチサッカー
3	PHP 入門 PHP マニュアル PHP　MySQL メモ PHP スクリプト	8	テイルズチャンネル テイルズオブシリーズ (wiki) バナフェスタウン! テイルズオブイノセンス	13	TBS 世界バレー 日本バレーボール協会 バレーボール (wiki) バレーボール V リーグ
4	テイルズオブハーツ テイルズオブハーツ (wiki) テイルズオブハーツ攻略 wiki テイルズオブハーツガイド	9	勝手 Perl リファレンス Perl PerlLesson Perl / 正規表現	14	サッカー選手一覧 (wiki) サッカー名選手館 海外サッカー選手の本音 サッカー選手メーカー
5	Amazon マライヤの赤ちゃん 豆まき プログラミング (wiki)	10	悪霊シリーズ (wiki) ニコニコ動画 リクナビ 2011 マイナビ 2011	15	サッカー日本代表 サッカーフリークジャパン COERVER COACHING JAPAN Adidas Japan

表 2　各ユーザのページ閲覧履歴

推薦の検証を行う予定である．

4. 関 連 研 究

ユーザのインタラクションによりランキングする研究は多く行われている [4] [5] [6]．これらは，ユーザの閲覧履歴から最も人気のあるページをランキングする手法 [4] など，多くのユーザの長期の検索結果に対する振る舞いを用いている．ユーザのアクセス履歴からページの重要度を決定する点は本手法と類似している．しかしながら，リンク先に同時刻にアクセスしているユーザを把握することはできず，ページの検索はできるがユーザ発見にまでは至っていない．

ユーザコミュニティに基づく情報推薦サービスに関する研究も多く行われている [2] [7] [9] [8]．多くのユーザの振る舞いから情報を推薦する協調フィルタリング手法を基本としており，既に書籍販売等のサービスで用いられている．また，数人の小規模なユーザの協調作業により，情報検索の効率向上を目指した研究もある [2] [7]．ユーザ間支援により検索効率を向上する点では本手法と類似しているが，不特定多数のユーザ同士での直接的なコミュニケーションはできない点が異なる．また，ユーザの特性抽出に関する研究も多く行われている [1] [10] [11]．これにより，専門家を抽出し，検索キーワードに対する専門家を特定し，その専門家の連絡先をユーザへ提供することで，不明なことを直接専門家に問い合わせることができる．しかしながら，即時性がなく手間を要するため，ページを閲覧しているその瞬間に問合せができる提案手法は効率性が高いといえる．

ページを同時に閲覧しているユーザとのコミュニケーションを支援する研究 [12] や，グリッドブラウジングにより他者の Web ページの遷移情報を集計したものを表示し，ページを推薦すると共にユーザ同士のコミュニケーションを促す研究もある [13]．さらに，今見ているページに知人が訪れたという履歴を残すことにより，あまり親密ではない知人とのコミュニケーションのきっかけとするサービス [14] もある．これらは Web 上にいる他者を活用したコミュニケーションやブラウジングという点では本研究とも共通しているが，これらの研究の目的はあくまでコミュニケーションである．本研究は Web 上の他者を検索に活用しすることにより，ユーザの興味のある項目ごとにリアルタイムな注目ページを参照することができる点で異なる．

5. ま と め

本論文では，これまで開発してきた閲覧者の量に基づいた検索およびそれら閲覧者とのコミュニケーションを同時に実現するソーシャルサーチ上で，興味が類似しており広い/深い知識を持った人およびページを発見する手法を提案した．また，それら発見した人とのコミュニケーション手法についても検討した．実験の結果，ユーザの閲覧履歴を用いて類似性や知識の獲得状況を推定することで，興味が似ているユーザを発見が実現

された．今後は，提案した検索手法の検証を行う予定である．

謝　　辞

本研究は，戦略的情報通信研究開発推進制度 (SCOPE) 若手 ICT 研究者育成型研究開発 (課題番号:102107001) の助成により実施したものである．ここに記して謝意を表す.

文　　献

[1] E. Y. Chang. Confucius and“ Its ”Intelligent Disciples. In *Proc. CIKM2009*, 2009.
[2] J. Pickens, G. Golovchinsky, C. Shah, P. Qvarfordt, and M. Back. Algorithmic Mediation for Collaborative Exploratory Search. In *Proc. SIGIR2008*, pp.315-322, 2008.
[3] Wired: http://klab.kyoto-su.ac.jp/˜mito/
[4] R. W. White, M. Bilenko and S. Cucerzan. Studying the Use of Popular Destinations to Enhance Web Search interaction. In *Proc. SIGIR2007*, pp.159-166, 2007.
[5] F. Radlinski and T. Joachims. Query Chains: Learning to rank from implicit feedback. In *Proc. KDD2005*, pp.239-248, 2005.
[6] E. Agichtein, E. Brill and S. Dumais. Improving web search ranking by incorporating user behavior information. In *Proc. SIGIR2006*, pp.19-26, 2006.
[7] M. R. Morris and E. Horvitz. SearchTogether. An interface for collaborative web search. In *Proc. UIST2007*, pp.3-12, 2007.
[8] M. B. Twidale, D. M. Nichols and C. D. Paice. Browsing is a collaborative process. Information Processing and management, 33(6):761-783, 1997.
[9] B. Smyth, E. Balfe, O. Boydell, K. Bradley, P. Briggs, M. Coyle and J. Freyne. A live-user evaluation of collaborative web search. In *Proc. IJCAI2005*, pp.1419-1424, 2005.
[10] A. Goel and K. Munagala. Hybrid Keyword Search Auctions. In *Proc. WWW2009*, pp.221-230, 2009.
[11] M. Richardson, E. Dominowska, and R. Ragno. Predicting clicks: estimating the click-through rate for new ads. In *Proc. WWW2007*, pp.521-530, 2007.
[12] 佐藤 俊輔: web ページにおける閲覧者間の繋がりを用いたインフォーマルコミュニケーション支援, 筑波大学第三学群情報学類卒業研究論文, 2006,
[13] 井上 恭輔: antwave-超次元コラボレーションブラウザ, 第 16 回全国高等専門学校プログラミングコンテスト自由部門, 2005.
[14] 赤塚 大典: 弱い紐帯に注目したコミュニケーションメディア「わくらわ」, 2006.
[15] 松井 優也, 河合 由起子: 閲覧者ネットワークによる情報収集支援サービスの提案, WebDB Forum, (2009).
[16] 松井 優也, 河合 由起子，望月崇由: 閲覧者ネットワークによる検索システムの検討, 電子情報通信学会 WI2 研究会, (2010).