

ユーザの Web インタラクション履歴の分析に基づく ソーシャルリレーション発見手法の提案

Social relation findings based on analyzing users' Web interaction histories

宮脇 佑介†
Yusuke Miyawaki

中島 伸介†
Shinsuke Nakajima

張 建偉†
Jianwei Zhang

河合 由起子†
Yukiko Kawai

1. はじめに

インターネットの普及に伴い、ユーザが取得可能な情報が溢れるようになった。これに対して、ユーザの嗜好情報に応じて適切な情報を提供する情報推薦手法が注目されている。情報推薦手法を採用することで成功を収めている Web サービスの一つとして、Amazon.com が挙げられる。Amazon.com はサイト内でのユーザの閲覧履歴や購買履歴からユーザの嗜好を読み取り、好みに合った商品を推薦する、情報フィルタリングを行っている。逆に言えば、彼らは Amazon.com 以外のサイトでの行動履歴を考慮することはできないといえる。すなわち、従来の情報推薦手法におけるユーザの嗜好推定手法はまだ限定的である。もしも、普段の Web 検索履歴および閲覧履歴、他のユーザとのチャット履歴、ブログ等への書き込み履歴などを分析することで、そのユーザの嗜好情報を推定することができれば、より精度の高い嗜好推定が可能であり、Amazon.com のような特定の Web システムを初めて利用するユーザに対しても効果的な情報推薦を行える可能性がある。

既に我々は、Web インタラクションの分析に基づいたソーシャルリレーションの発見手法について提案している[6]。しかしながら、概念のみの提案であり、詳細なインタラクション分析に対する検討が不十分であった。そこで我々は、より詳細な Web インタラクションの取得・分析について検討したので、本稿にて報告する。

2. 関連研究

Shiwan ら[1]の研究では、ユーザとアイテムとの関係の可視化を行っている。これはユーザ A が検索キーワード α を検索すると、システムが管理しているブログから、そのキーワード α に関して執筆したユーザ群および、その検索キーワード α に関連したキーワード群を 7つのコミュニティに分類して表示する。

ユーザ A はそのコミュニティから検索キーワード α について頻繁に執筆しているユーザ B を見つかる事が可能となる。そこからユーザ B の詳細を表示することが可能であり、ユーザ B が他にどのような事についてどれほど執筆しているかまで特定することを可能にしている。

しかし、これは検索キーワードについて情報を得たいユーザ A と執筆ユーザ B には検索キーワードだけの繋がりであり、執筆ユーザ B の検索キーワード以外の情報はユーザ A への推薦ではなく、ナビゲーションに近い。

森ら[2]は、人物検索システム「スパイシー」による、エンティティのランキング学習に関する研究を行っている。この研究では、検索クエリに対して様々な Web 上の情報を収集し、それらの情報から得られた属性を元にエンティティのランキング学習を行うものである。これは人物エンティティを対象に、実際のエンティティ検索システムから得られたデータを利用して、Web 情報を利用したエンティティのランキングを行い、エキスパートの発見を正確に行うことが可能とされている。

しかしながら、この研究では人との関係や、その正確性を主題としており、情報推薦を目的としたものではない。

河合ら[3][4][5]の研究は、現在ユーザが閲覧しているページにコミュニケーションツールを加え、ライブで他に閲覧しているユーザとコミュニケーションを可能とし、そのユーザに現在求めている情報の提示を求める手法や、ユーザが興味を示した部分をハイライトさせ、その部分を他の閲覧ユーザと共有することで、閲覧者との話題の共有を実現させる「ソーシャルサーチシステム」が提案されている。

しかしこのシステムを実現させるには、ユーザ自信がコミュニティに参加する必要があり、情報の提供やハイライトなどの操作をユーザ積極的な協力があって成り立つシステムである。

我々の提案手法では、ユーザの Web 検索行動つまりユーザやページの Web インタラクションを記録することで、ユーザの特徴・ページの特徴を捉え、関係の強いユーザやページ同士でお互いの情報を共有し合える情報推薦システム・コミュニティ生成支援システムを提案する。

3. ソーシャルリレーションに基づく情報推薦システム

Web 検索におけるユーザの目的や意図を表現する上で最も重要視されるのは「検索キーワード」そのものである。ただし、検索キーワードのみでは、ユーザの詳細な意図や嗜好を表現することはできない。したがって、Web 検索や閲覧時におけるユーザによる Web インタラクションを詳細に分析して、これを取得することで、より精度の高いユーザの意図や嗜好の抽出が可能となると考えている。我々をこれらのユーザの意図や嗜好の類似度を評価することで、ユーザおよびコンテンツ間の関係性、すなわちソーシャルリレーションを発見し、これを用いた高精度な情報推薦システムの構築を目指している。

なお、ソーシャルリレーションの発見においては、ユーザ間、ページ間、ユーザ・ページ間の関係性につ

いて分析する必要がある。これらの関係性の概念図を図1に示す。

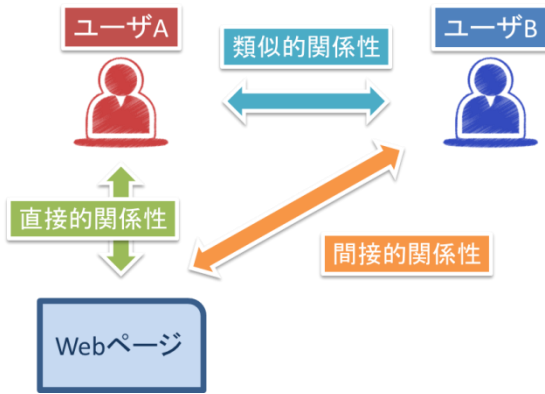


図1 ソーシャルリレーションについて

Web インタラクションに基づく関係性の発見においては、図1に示す通り、直接的関係性、類似的関係性、間接的関係性を分析する必要がある。以下、各関係性について説明すると共に、これら関係性を利用したコミュニティ生成およびこれに基づく情報推薦方式について述べる。

3.1. ユーザとページ間の直接的関係性

ユーザとページ間の直接的関係性は、ユーザの Web ページに対する直接的操作に基づいている。この際、単にどのようなキーワードで検索し、どのページを閲覧したのか、という情報だけでなく、どのページのどの段落に注目したのかという情報を含めて取得した情報に基づいて表現される関係性である。

3.2. ユーザ間・ページ間の類似的関係性

ユーザ間の類似的関係性は、Web 利用時のユーザのインタラクションの類似性に基づいて表現される関係性である。ユーザの Web 検索行動を細かく見ると「いつ・どのような検索キーワードで・どのような経路で・どんなページの・どこに注目したのか」などが挙げられる。これらのユーザの行動を分析することで「検索キーワード」だけではないユーザの特徴を分析することが出来ると考えている。同様にページにも「いつ・どのような検索キーワードで・どのような経路で・どんなページの・どこに注目したのか」などが挙げられる。これらのページに関する情報を類似度判定することで、ページ間の類似的関係性を発見することができる。

3.3. 直接的・類似的関係性に基づく間接的関係性

ユーザとページ間の直接的関係性や、ユーザ間・ページ間の類似的関係性を用いて分析することが可能な間接的関係性の発見方法について述べる。

図1は既にユーザAとユーザBが類似度判定の結果としてユーザ間の関係があり、またユーザAにはあるWeb ページと直接的関係性があることを示す図である。

この時ユーザBにとっては、既に関係のあるユーザAが関係しているWeb ページと何らかの関係があると考えられる。このような、ユーザBが直接的な関係でなくとも、少なからず関係があると考えられる。これをユーザBとWeb ページ間の間接的関係性とする。

3.4. 関係性を利用したコミュニティの生成とこれに基づく情報推薦方式

3.1 節～3.3 節で説明した「類似的関係性・直接的関係性・間接的関係性」を利用したコミュニティの生成や、情報推薦について提案する。

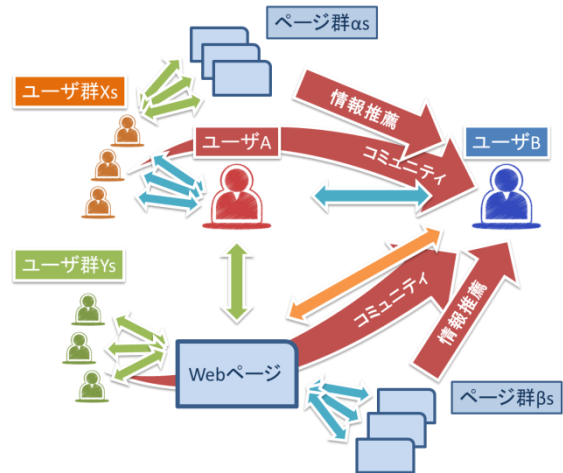


図2 コミュニティ・情報推薦について

図2は図1の「間接的関係性について」の図に関係性の考えられる複数のユーザやページを付け加えたものである。

ユーザー群 Xs はユーザー A と類似的関係性にあり、ユーザー群 Ys はユーザー A に直接的関係性のある Web ページと直接的関係性がある。またページ αs はユーザー群 Xs と直接的関係性があり、同じくページ群 βs は Web ページと類似的関係性があるとすると、この時、ユーザー A と類似的関係性のあるユーザー B はユーザー A と類似的関係性のあるユーザー群 Xs と間接的関係性があると考えられる。同様にユーザー A と直接的関係性のある Web ページに直接的関係性のあるユーザー群 Ys もユーザー B と間接的関係性があると考えられる。そのため、ユーザー B とユーザー群 Xs・ユーザー群 Ys は間接的関係性があると考えられる。

このことから、ユーザー B へ「ページ群 αs・ページ群 βs を推薦」することや、「ユーザー群 Xs・ユーザー群 Ys のコミュニティに属させる」ことが可能となる。

4. Web インタラクション分析に基づくソーシャルリレーションの発見

関係性を発見するためのユーザやページの特徴を取得するために、ユーザの Web インタラクション履歴の分析を行う。

本論文で分析に使用する代表的なユーザの Web インタラクションとして以下のものがある。

- **keyword_**
ブラウザで何らかのイベントが発生し、現在の URL のホスト名が google や yahoo などの際の検索クエリ
- **text_**
ユーザのページ閲覧中にマウスイベントが発生した HTML 要素のテキスト
- **event_**
ユーザがページ内で操作した HTML 要素に対するマウスイベントの名前
- **eventcount**
ユーザがページ内で操作した HTML 要素に対するマウスイベントの回数
- **url_**

ブラウザやページ内で何らかのイベントが発生した際のページの URL

- **time_**
ブラウザやページ内で何らかのイベントが発生した際の時刻
- **link_**
ページ内のリンクを操作した場合の遷移先ページの URL

これらの Web インタクション履歴を分析することで、ユーザ間やページ間、またユーザ・ページ間などの様々な関係性、つまりソーシャルリレーションを発見する。

4.1. 取得される Web インタクションの詳細や記録のタイミングについて

Web インタクションの記録のタイミングはブラウザ側でのイベント発生時や、ユーザによるページ操作が行われた際のマウスイベント発生時となる。



図3 ブラウザ側のイベントによる Web インタクションの記録

図3はユーザが一般的に行う検索行動の一部である。「検索開始」では、ブラウザや google・yahoo などの検索サイトに検索キーワードを入力し、検索を開始する。この時に google のページ内でユーザによるクリックイベントが発生しているが、Web インタクションの記録については次項で説明する。

「検索結果」では、入力した検索キーワードに対する結果が google から送信された場面を示す。この時の google からの検索結果にはユーザによって、新しいタブでの表示が行われているものとする。このような場合、ブラウザ側のイベントとして「タブの新規作成」が行われており、Web インタクションとして「keyword_、url_、time_」が記録される。

ブラウザ側のイベントとしては「タブの新規作成」の他に「タブの再読み込み・タブの移動・タブを閉じる」なども Web インタクションの記録のトリガとなる。

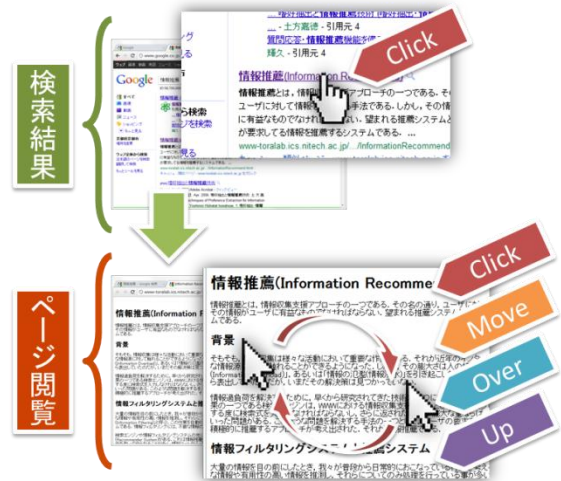


図4 ページ側のイベントによる Web インタクションの記録

図4はユーザの Web 検索行動を示した図3の続きとなる。

「検索結果」ではユーザが google から返って来た検索結果の一覧から一つを選んだことを示している。ここでは図3の「検索開始」で省略したユーザによるマウスイベントが発生しており、Web インタクションとして「text_、event_、eventcount_、url_、time_、(link_)」が記録されている。この時のマウスイベントは「mouseClick」の他に、ユーザがマウスでページ全体を眺めている時にも発生する「mouseMove、mouseOver、MouseUp」なども含まれる。

「ページ閲覧」では図4の「検索結果」と同様に、ページ内で発生したマウスイベント毎に Web インタクション「text_、event_、eventcount_、url_、time_、(link_)」を記録する。

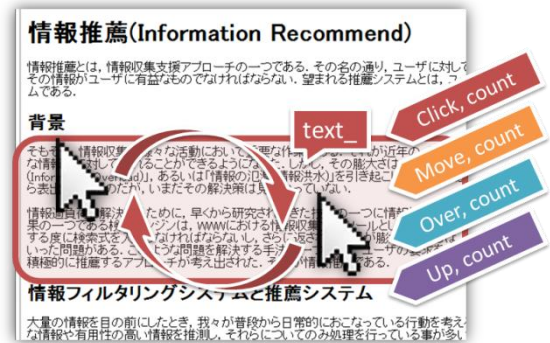


図5 マウスイベント発生時に取得される text_ と event_、eventcount_ について

Web インタクションの「text_」はユーザがページ内でマウスイベントを発生させた HTML 要素のテキストとなる。そのため図5のように本文のマウスイベントが発生した要素のテキスト「そもそも～」の部分が「text_」として記録される。

「event_、eventcount_」は HTML 要素で発生したマウスイベントの名前と累計数が記録される。図5のように、ある部分に集中してマウスイベントの発生があれば、ユーザがページのどのような部分に最も興味を示したのかを測ることができる。

「event_、eventcount_」から判定できる、ユーザのHTML要素への評点の数式を以下に示す。

$$\text{HTML要素評点} = \sum_{i=1}^4 X_i Y_i$$

mouseClick	重み: X_1	イベント回数: Y_1
mouseUp	重み: X_2	イベント回数: Y_2
mouseOver	重み: X_3	イベント回数: Y_3
mouseMove	重み: X_4	イベント回数: Y_4

$$X_1 > X_2 > X_3 > X_4$$

これにより、ユーザがページのどの部分に興味を示したかを数値化する。

この数値化された興味度に基づいて、興味度の高いテキスト分析を行うことで、ユーザが興味をしたキーワードやトピックを分析することが可能となる。

4.2. Webインタラクションの履歴から得られるユーザやページの特徴

ユーザのWeb検索行動を一連の動作を、4.1のWebインタラクションとして記録・分析することでユーザやページの特徴を抽出する。

ユーザの特徴が最も色濃く現れるWebインタラクションは、ユーザが実際に入力・操作したテキストや注目度を記録した「keyword_、text_、event_、eventcount_」となる。これらのデータからユーザの求めている情報が「どのような検索キーワードで・どんなページの・どこに注目したのか」を分析することができる。これにより「ユーザは何に興味を示すのか・ページには何が記載されているのか」という、ユーザやページの特徴を取得することができる。

またユーザによるページ内の部分的な注目度の測定や、その時の時間、遷移先のページURLを「time_、url_」で取得しているため、「いつ・どのような経路で」といった動作までを把握することが可能となり、ユーザやページの特徴として抽出されたキーワードの類似だけではなく、より細かな行動や時系列を含めた特徴の抽出が可能となる。

5. まとめ

これまでに提案したWebインタラクションの分析に基づいたソーシャルリレーションの発見手法に対し、より詳細なWebインタラクションの取得・分析について検討した。

今後は、提案したソーシャルリレーションの発見手法に基づいたプロトタイプシステムの実装と評価実験を行う予定である。

謝辞

本研究の一部は、文部科学省科学研究費助成事業（学術研究助成基金助成金）基盤研究（C）（課題番号：#23500140）による。ここに記して謝意を表します。

参考文献

- [1] Shiwan Zhao, Xiatian Zhang, Michelle X. Zhou, Wentao Zheng, Quan Yuan, Rongyao Fu, Who is Talking about What: Social Map – based Recommendation for Content – Centric Social Websites, The 4th ACM Conference on Recommender Systems, 00.143-150, 2010.
- [2] 森 純一郎, 松尾 豊, Web 情報を用いたエンティティのランキング学習に関する研究, 東京大学大学院工学系

研究科, The 24th Annual Conference of the Japanese Society for Artificial Intelligence, 2010.

- [3] 松井 優也, 河合 由起子: リアルタイム閲覧者ネットワークによる検索システムの提案, 電子情報通信学会 W12 研究会, 2009.
- [4] 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
- [5] 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 W12 研究会, 2010.
- [6] 宮脇 佑介, 中島 伸介, 張 建偉, 河合 由起子: インタラクションの文脈理解に基づくソーシャルリレーションの発見, DEIM, 2011.