

# 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

宮脇 佑介<sup>†</sup> 中島 伸介<sup>†</sup> 張 建偉<sup>†</sup> 河合 由起子<sup>†</sup>

<sup>†</sup> 京都産業大学 〒603-8555 京都市北区上賀茂本山

E-mail: <sup>†</sup> {g0847254, nakajima, zjw, kawai}@cc.kyoto-su.ac.jp

## あらまし

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。しかしながら、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、効果的な情報推薦を行える可能性がある。そこで我々は、ユーザの日頃の Web 検索の行動履歴や閲覧履歴から、ユーザとコンテンツとの関係、ユーザとユーザとの関係、コンテンツとコンテンツとの関係を抽出し、これらの社会的な関係（ソーシャルリレーション）を発見することで、効果的な情報推薦システムの構築を目指すことを目標とする。本稿では、まずはソーシャルリレーションの発見手法について検討する。

**キーワード** 情報推薦, インタラクション, ソーシャルリレーション, ユーザプロファイル

## Social relation findings based on an understanding of the interaction context

Yusuke MIYAWAKI<sup>†</sup> Shinsuke NAKAJIMA<sup>†</sup> and Jianwei ZHANG<sup>†</sup> and Yukiko KAWAI<sup>†</sup>

<sup>†</sup> Kyoto Sangyo University Motoyama, Kamigamo, Kitaku, Kyoto, 603-8555, Japan

E-mail: <sup>†</sup> {g0847254, nakajima, zjw, kawai}@cc.kyoto-su.ac.jp

## Outline

Information that users can acquire is overflowing as the Internet spread. The information recommendation technique for providing the relevant information suitable for users' preference has attracted a lot of research interests. However, the technique for estimating users' preference in the existing information recommendation methods is still limited. The preference presumption with higher accuracy and the information recommendation with better effectiveness are possible, if we can estimate users' preference information by analyzing web retrieval and browsing histories, chat histories with other users, and blog posting histories, etc. Our goal is to construct an effective information recommendation system by detecting social relation. Social relation is detected by extracting the relationships between users and contents, between users and users, between contents and contents, from web retrieval and browsing histories conducted in users' daily lives. In this paper, we first describe the method for detecting social relation.

**Keyword** Information recommendation, interaction, social relation, and user profile

## 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。そこで我々は、ユーザの日頃の Web 検索の行動履歴や閲覧履歴から、ユーザとコンテンツとの関係、ユーザとユーザとの関係、コンテンツとコンテンツとの関係を抽出し、これらの社会的な関係（ソーシャルリレーション）を発見することで、効果的な情報推薦システムの構築を目指すことを目標とする。本稿では、まずはソーシャルリレーションの発見手法について検討する。

## 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  $\alpha$  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  $\alpha$  に関して執筆したユーザ群および、その検索キーワード  $\alpha$  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  $\alpha$  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、検索クエリーに対して様々な 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る、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。

一方我々の提案手法では、ユーザの検索行動を逐一記録しプロファイル化することで、ユーザの情報提供の操作を必要とせず、コミュニケーションツールを使わずに、プロファイルの類似度からユーザ間の関係やページ間との関係、ユーザとページ間との関係を特定することが可能である。

## 3. 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

これまでの推薦システムでは、システム内のユーザの行動ログからユーザの嗜好情報を抽出していたが、ユーザの日常の行動からユーザの特徴や、閲覧ページの特徴を取得することで、より詳細なページ間・ユーザ間の嗜好特徴を発見することが可能である。したがって、我々はユーザの嗜好や思考など人間の特徴を考慮した上で、ユーザが求めている情報を特定し推薦することが可能なシステムの開発を目指す。

### 3.1. 本手法の基礎的技術及び定義

本手法では、ページの特徴やユーザの嗜好を利用して、ページ間やユーザ間、ページとユーザ間の関係を発見し、ユーザに適した情報を推薦することを主題としている。以下ではこれらの関係を発見するための、基礎的な技術や定義について説明する。

#### 3.1.1. 対象とする Web 利用時のインタラクション

本研究では、日常の検索行動から取得できるログを利用し、ユーザの特徴や嗜好を抽出することを目指す。図1に本研究で対象とするページとユーザのインタラクションを示す。図1に示すようなユーザが行うインタラクションや、ページがユーザが受けるインタラクションは、それぞれユーザの嗜好やページの特徴を分析する上で重要である。このように、ユーザのページ検索行動から取得される様々ログを、本研究で対象と

する Web 利用時のインタラクションと定義する。なお、我々はこれらのインタラクションを分析していくことで、単にユーザの嗜好やページの特徴を分析するだけでなく、ページ間の関係やユーザ間の関係、ページとユーザの関係を特定することに取り組む。

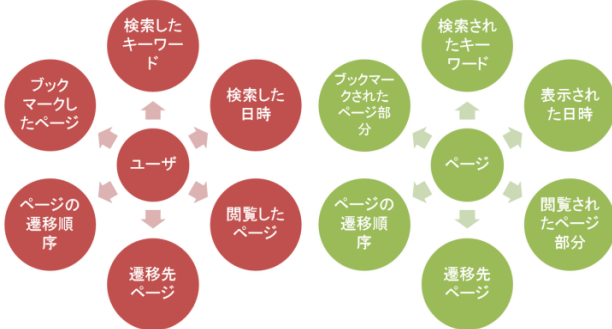


図 1 ページ・ユーザにおけるインタラクション

### 3.1.2. ソーシャルリレーションの定義

ソーシャルリレーションとは、ユーザやページのインタラクションを解析した結果から得られる、複数のユーザ間、ページ間、ユーザ・ページ間の関係の集合体の事を指す。

- ユーザ間のソーシャルリレーション  
ユーザのインタラクションを解析することにより、ユーザの嗜好を特徴ベクトルにて表現したプロフィールからユーザ間の類似性を表現したものである。
- ページ間のソーシャルリレーション  
ページのインタラクションを解析することにより、ページの特徴を特徴ベクトルにて表現したプロフィールからページ間の類似性を表現したものである。
- ユーザ・ページ間のソーシャルリレーション  
ユーザ・ページ間のソーシャルリレーションとは、ユーザがそのページの情報にどれだけ興味を持っているのかという、興味度の強さにて表現したものである。

なお、各プロフィールの取得・表現方法に関しては、3.2 節にて述べる。

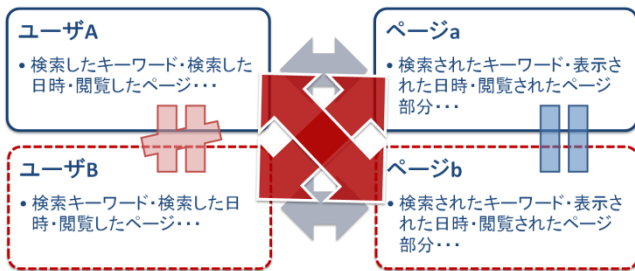


図 2 ページ・ユーザ間のリレーション

図 2 にはページ間の関係とユーザ間の関係から得られる、ページとユーザ間の関係の概略図を示す。

ユーザ A とユーザ B がそれぞれ異なる検索キーワードで検索を行い、ページ a とページ b を閲覧したとする。通常であれば、ユーザ A とユーザ B は検索キーワードの違いにより、関係性を発見することは出来ない。しかしページ a とページ b 間にはプロフィールの比較により、掲載内容や閲覧ユーザなどが類似しており、関係性があるものとする。するとそれらのページを閲覧したユーザ A とユーザ B には検索キーワード以外の新たな関係「閲覧したページ間の関係によるユーザ間の関係」が発見できる。このようにして、直接的な関連がない場合でも、関連性を発見することが可能となる。

### 3.2. ユーザプロフィール

本研究におけるユーザプロフィールとは、あるユーザのキーワード検索において、どのようなキーワードから・どのように遷移して・最終的にどのような特徴を持ったページへ遷移したのか、などを記録したものである。図 3 にユーザの検索行動からプロフィールに記憶される項目、すなわちユーザプロフィールの項目を示す。本節では、図 3 で示した要素を解析することで得られる内容について説明する。

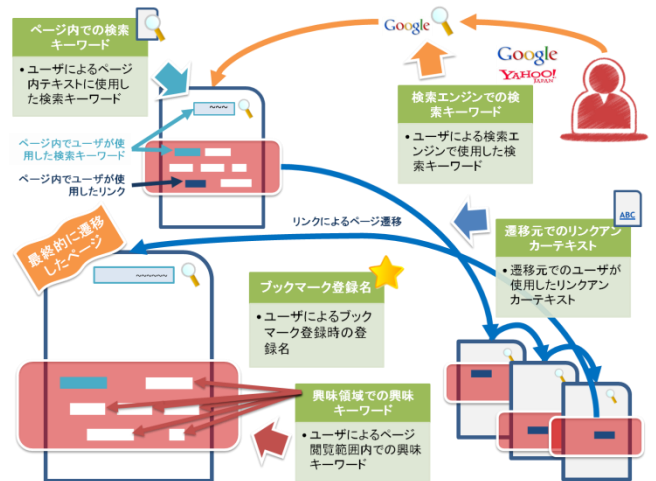


図 3 ユーザの検索行動からプロフィールに記録される項目

#### 検索エンジンでの検索キーワード

ユーザがどのような検索キーワードを検索エンジンで使用したのかを記録する。この要素により、「ユーザがどのようなトピックに関して興味を持っているのか」ということを表すことが可能となる。

#### 遷移元でのリンクアンカーテキスト

遷移元でのリンクアンカーテキストとは、ユーザが閲覧したページ内、また検索サイトで検索した（ユー

ザの使用した検索キーワードに関連する) ページ内でユーザが操作したリンクアンカーテキストを指す。

ユーザが閲覧したページ内でリンクの操作を行えば、そのリンクのキーワードはユーザが閲覧しているページに関連した興味を持ったキーワードを現していると言える。つまり、検索キーワードでは、ユーザが興味を持つ大きなトピックを表すことができるが、このリンクアンカーテキストにより、さらにどのような話題に関して興味を持っているのかを表現することが可能となる。

### ページ内での検索キーワード

閲覧した Web ページはユーザが興味を持っている可能性が高いが、そのページ内に現れる全てのキーワードに興味を持っているとは限らない。閲覧中のページ内で行われたキーワード検索を記録することで、閲覧中のページの中で、どのような話題に関して興味をもったのか、逆に言えば、どのような話題に関してはそれほど興味を持っていないのかということを推定することが可能となる。すなわち、ユーザ間の求めている情報の詳細化・区別・遷移を抽出することができると考えている。

### 興味領域での興味キーワード

ユーザがページ内で興味を示したと思われる場所(マウスの遷移やページスクロール、ページ内でのキーワード検索でアクティブになった領域などから割り出した、ユーザが興味を示した領域)のキーワードである(詳細は3.4で説明する)。これは、上述の“ページ内での検索キーワード”と同様に、閲覧中のページの中で、どのような話題に関して興味をもったのか、また、どのような話題に関してはそれほど興味を持っていないのかということを推定することが可能となる。

閲覧中のページでユーザが興味を示した領域、興味領域のキーワードは、そのユーザにとって興味や必要とされる最も代表的なキーワードであり、ユーザと深い関係があると言える。

### ブックマークタグ名

ユーザがあるページをブックマークした場合のタグ名は、ユーザがページどの部分に視点を置いたのか、ユーザの状況によって異なるユーザ独自に付けられるページの説明文ともいえる。もしあるユーザ同士が同じページに同じタグ名を使用している場合には、ユーザ間には同じ意図で登録した者同士というユーザ間の関係を見ることができる。またあるユーザ同士同じページをブックマーク登録したとしても、登録名が全くことなれば、それらのユーザの関係を区別することができる。

## 3.3. ページプロフィール

ページプロフィールについて説明する。ページプロフィールの取得項目はユーザプロフィールと同様の項目になるが、プロフィールとして記録する対象が異なる。ユーザプロフィールは「ユーザが実際に行動した履歴」つまり「1人のユーザのある検索キーワードでの検索開始から最終的に遷移したページまでの行動」を記録するのに対して、ページプロフィールでは「ユーザによってページが操作された内容」を記録する。

次にそれぞれのプロフィールの要素について説明する。

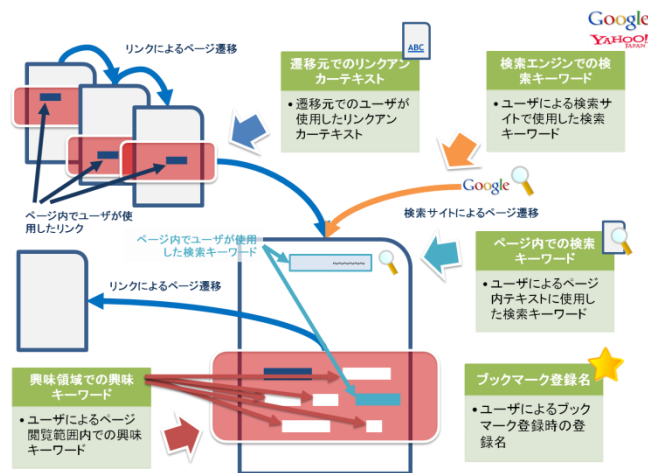


図 4 ページプロフィールに記録される項目

図 4 の取得項目は上記の事から、図 3 の項目と同様であることがわかる。

要素ごとの意味は次のようになる。

### 検索エンジンでの検索キーワード

検索エンジンからページに遷移するのに、どのような検索キーワードを使用したのかを記録する。この要素により、「ページにはどのような情報を求めているのか」という事を表す事が可能となる。また取得した要素が多くなるほど、求めている情報の人気度の度合いや、要素のキーワードの意味的關係からページの該当する大きなトピックを表すことが可能となる。

### 遷移元でのリンクアンカーテキスト

遷移元でのリンクアンカーテキストとは、ユーザが閲覧したページ内、また検索サイトで検索した(ユーザの使用した検索キーワードに関連する)ページ内でユーザが操作したリンクアンカーテキストを指す。

この要素は検索エンジンでの検索キーワードと同様に、「ページにはどのような情報を求めているのか」を表すことができるため、求めている情報の人気度の度合いや、ページの大きいトピックを表すことができる。

### ページ内での検索キーワード

ページ内での検索キーワードとは、ユーザがページ内で特に求めている情報の代表キーワードである。つまりページ内での検索キーワードからページに掲載されている内容から重要なキーワードを表す可能性があり、ページ内にある詳細なトピックを抽出することが可能であると言える。

### 興味領域での興味キーワード

ユーザがページ内で興味を示したと思われる場所（マウスの遷移やページスクロール、ページ内でのキーワード検索でアクティブになった領域などから割り出した、ユーザが興味を示した領域）のキーワードである（詳細は 3.4 で説明する）。

これは、上述の“ページ内での検索キーワード”と同様に、ページ内で重要な詳細トピックの関連キーワードを表している。これにより詳細トピックをさらに詳細化を行い、トピックの細かい分類を表すことが可能となる。

### ブックマークタグ名

ユーザプロフィールの要素と同様に、ユーザが視点を置いた部分や、ユーザの状況に影響したページの説明文となる。これまでの要素では主にページ内に存在するキーワードを元にプロフィールを作成されていたが、この要素はユーザが独自に作成したキーワードを登録させるため、ユーザプロフィールの各要素との類似度を測ることで、ユーザが興味を示すと予想される情報の提供を行うことが可能となる。

## 3.4. インタラクションの分析によるプロフィール情報の取得

ページの特徴を表すページプロフィールおよびユーザの嗜好を表すユーザプロフィールは、ページに対するユーザのインタラクションを解析することにより抽出される。

以下にその解析手法について説明する。

まずユーザがページを閲覧する際に、Web ブラウザが表示している範囲を表示領域とする。さらにこの表示領域の中で、ユーザが積極的に操作した HTML タグの div 要素の範囲を興味領域とし、ユーザが興味をもって閲覧したコンテンツの範囲とみなす。この興味領域の特徴を抽出するために、以下のような語を、興味領域の興味キーワードとして抽出する。

- 興味領域内の名詞（出現頻度をその重みとする）
- マウス操作により選択されたキーワード
- コピーしたキーワード
- クリックしたリンクアンカーテキストに含まれるキーワード

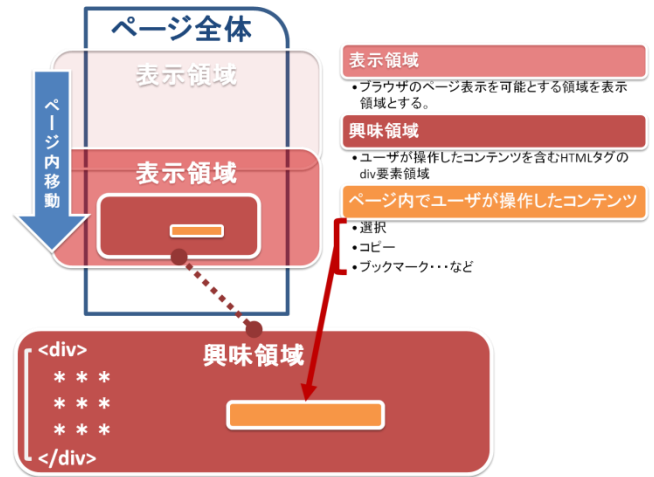


図 5 ユーザのページ操作によるページやユーザの特徴抽出

これらの抽出した興味キーワードをプロフィール化することで、ページ全体を対象とした出現キーワードに基づいた類似度判定以上の、より高度なページ間やユーザ間の関係を判別することが可能となる。

## 3.5. ソーシャルリレーションの発見手法

これまでで説明したページやユーザの特徴や嗜好を表す各々のプロフィールを使い、類似ページやユーザ、また高度な関係として同じ境遇のユーザや異なる内容でもユーザ間で関係するページの関係など、様々なソーシャルリレーションの発見について説明する。

### 3.5.1. ユーザ間やページ間のソーシャルリレーション

ユーザ間のソーシャルリレーションは、ユーザの嗜好を特徴ベクトルにて表現したプロフィールからユーザ間の類似性を表現したものであり、ページ間のソーシャルリレーションは、ページの特徴を特徴ベクトルにて表現したプロフィールからページ間の類似性を表現したものである。したがって、それぞれのソーシャルリレーションは、特徴ベクトルにて表現されたプロフィール同士の類似度計算に基づいて、その関係の強さを計算することが可能である。本研究では、コサイン相関値により類似度計算を行うことを検討している。

### 3.5.2. ユーザとページ間のソーシャルリレーション

ユーザとページ間のソーシャルリレーションとは、ユーザがそのページの情報にどれだけ興味を持っているのかという、興味度の強さにて表現したものである。このリレーションは、ユーザ自身が実際に閲覧したことにより得られるリレーションだけでなく、自分と嗜

好みが似ているユーザを介したリレーションや、興味度が高い閲覧済みのページを介したリレーションを発見することも可能である。

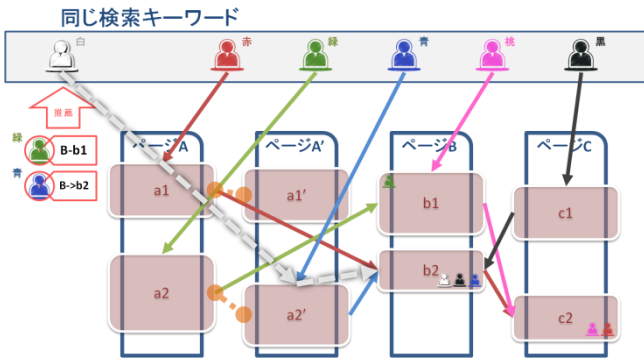


図 6 ユーザやページのインタラクションとソーシャルリレーションによる推薦の一例

図 6 に、ユーザやページのインタラクションとソーシャルリレーションによる推薦の一例を示す。図中のページ A と A' は類似ページである。ページ A の a1 や a2, ページ B の b1 や b2 は、ユーザがそのページ内で興味をもった興味領域である。白ユーザの検索キーワードを既に検索しているユーザが複数存在したとする (図中の赤, 緑, 青, 桃, 黒のユーザ)。ここで, a1 と a1', a2 と a2' は、意味的に類似していることが分かっているものとする。

ここで白ユーザの検索の行動から、初めに遷移したページがページ A' の a2' となった場合、この白ユーザには青ユーザと同じ遷移のため、探索履歴の類似度が高いと判定することができ、青ユーザが最終的に興味を示したページ B の b2 を推薦することができる。またページ A' の a2' と関係のあるページ A の a2 へ遷移した緑ユーザが、最終的に閲覧した興味ページ B の b1 を推薦することができる。

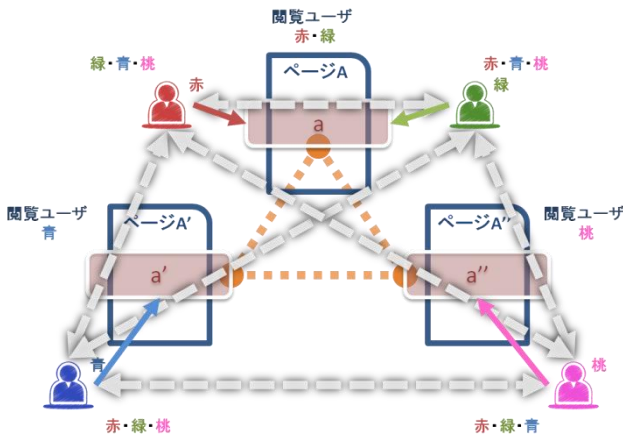


図 7 ページ間のリレーションを考慮した高度なユーザ間のリレーション

図 7 にページ間のリレーションを考慮した高度な

ユーザ間のリレーションを示す。図 7 のようにページ間のリレーションを考慮することにより、緑ユーザの遷移したページ A の a2 は、ページ A' の a2' と類似していることがわかる。したがって、緑ユーザはページ A' の a2' にも関心を持つ可能性が高い。このように「あるユーザがあるページを閲覧した」という単純なページとユーザの閲覧関係ではなく、赤ユーザは同じページ A を閲覧した緑ユーザ以外に、ページ A と特徴などが類似したページ A' やページ A'' などを閲覧した緑青桃ユーザと関係がある。このような「関係のあるページを閲覧したユーザ同士」といったリレーションの発見も可能となる。

#### 4. 想定されるアプリケーション例

図 8 に、類似したユーザの検索結果や行動を利用したアプリケーション例を示す。

ここで、赤ユーザが検索キーワード「クラウド」で Web 検索を行ったとする。この時、従来のように赤ユーザの検索キーワードによるページ検索を行うのではなく、赤ユーザがどのような人で何を推薦すべきなのか、またどのような情報が赤ユーザに適しているのかを考慮した上で推薦を行う必要がある。

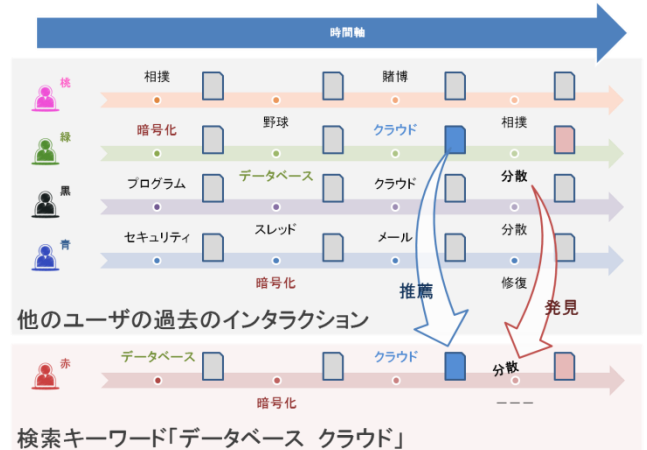


図 8 類似したユーザの検索結果や行動を利用したアプリケーション例

赤ユーザのこれまでのインタラクションから嗜好を抽出し、赤ユーザに類似するユーザから同じ検索キーワードで得た情報を提示する。

赤ユーザに類似するユーザとは、過去に「データベース」や「暗号化」についていくつか検索を行っており、ユーザプロフィールには「データベース」や、それと関係のあるページの特徴が見られるユーザ。つまりユーザは過去に検索キーワード「データベース」や「暗号化」を検索しているか、その内容を含んだページに興味を示しているユーザを指す。もしこのようなユーザを発見でき、そのユーザが、赤ユーザの「クラウド」について検索を行って、その結果を得ていたと

すれば、同じ趣向を持ったユーザ同士、現在「クラウド」で検索を行っている赤ユーザにも、その情報を提示することができるわけである。

つまり図 8 で示すように、対象ユーザに対して、閲覧履歴やその目的が類似したユーザを発見し、このユーザが関心を持った情報を高精度で推薦することが可能なアプリケーションの開発に役立つものと考えている。

## 5. まとめ

推薦精度の向上を目指し、ユーザとページのインタラクションを解析することにより、社会的関係すなわちソーシャルリレーションを発見するための基本方針について検討した。

今後の課題としては、提案したソーシャルリレーションの発見手法に基づいて、高度な情報推薦システムの開発に取り組むつもりである。

## 謝 辞

この研究の一部は、文部科学省科学研究費補助金若手研究 (B) (課題番号: 20700089) の一環としてなされたものである。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 4<sup>th</sup> ACM Conference on Recommender Systems, pp.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 24<sup>th</sup> 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 WI2 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 WI2 研究会, 2010.