

知識概念に着目したユーザの分類に基づく パーソナライズド Web 検索システムの提案

佐伯 祐太^{†1} 林 周平^{†1}
井下 雄樹^{†1} 藤田 聡^{†1}

Web 上に存在する膨大な情報の中から各ユーザにとって価値のある情報を効果的に推薦するパーソナライズド Web 検索が近年高い注目を集めている。その中でも特に、ソーシャルブックマークやフォークソノミーを利用したパーソナライゼーションは重要な技術のひとつである。しかし一般のフォークソノミーで用いられているタグはフラットな構造をもっており、個人的なメモなど無駄な情報も多く含まれるため、十分な推薦精度が期待できないという問題点があった。これらの点を解決するため本研究では、ユーザの嗜好を示すプロフィールを階層化されたタグによって表現することを提案する。参加ユーザのプロフィールを3パターンに分類することによって、単純なタグ集合の類似性による手法よりも精度よく Web 検索をおこなうことが可能となる。

A Personalized Web Search based on a Classification of Users' Preference using Structured Tag Representation

YUUTA SAIKI,^{†1} SHUHEI HAYASHI,^{†1} YUUKI INOSHITA^{†1}
and SATOSHI FUJITA^{†1}

Recently, personalized Web search has attracted considerable attentions as a way of recommending valuable information to each user on the Web. Such a recommendation is generally realized by using tags with the help of folksonomy. However, we could not expect high precision of recommendation with such a simple tag-based approach because of its flat structure and the inclusion of garbage tags such as personal memo. In order to solve such problem, we introduce a hierarchical structure to tag sets representing the profile of each user, and use those structured tag sets to classify the preference of users. Using such method, we could improve the precision of a personalized Web search.

1. はじめに

Web 検索はインターネットユーザにとって極めて重要な役割を担っている。Google や Yahoo! を始めとするクローラー型の Web 検索システムでは、データベース化された膨大な Web ページ情報の中からユーザが与えたクエリにマッチするページの URL を発見し、それらを PageRank¹⁾ などの適切なアルゴリズムでランキングしたのちユーザに提示するという手法が採られている。その結果、クエリにマッチする大量の URL の中でも特に重要な URL が上位に表示されることになり、ユーザが求めている情報をより早く効率的に発見できるようになった。ただし PageRank などの従来手法では、リンク構造を詳細に解析することでハイパーリンク作成者の意図はランキングに反映させられるものの、検索をおこなっているユーザ自身の情報は一般には検索結果に反映されない。しかし、たとえば一般的に使われている意味とは異なる意味で使われることの多い専門用語 (プログラミング言語の Java とコーヒーの Java など) を使ったキーワード検索などでは、それぞれのユーザのもつ背景知識をランキングにうまく反映させることによって、Web 検索システムの有用性はより高まることが期待される。

以上のような状況を背景として、Web 検索のパーソナライゼーション技術が近年高い注目を集めている²⁾⁻⁴⁾。Web のパーソナライズド検索とは、ユーザの興味や関心、嗜好などを過去の履歴から自動的に取得・蓄積し利用することで、大量の検索結果の中からそれぞれのユーザに適した Web ページを提示する技術である。本研究の目標は、ユーザの嗜好を階層化された概念タグによって明示的に表現し、各ページの特徴をそのページを閲覧したユーザの嗜好群によって表現することで、従来手法よりも精度の高いパーソナライズド Web 検索を実現することである。具体的には、そのページを過去に閲覧したユーザたちとどの程度興味の重なりがあるかを判別し、そこで示されている概念間の上位下位関係をそれぞれのユーザがどのようにとらえているかによって、ユーザの集合をおおまかに3通りに分類する。その上で、分類された3つのクラスのうちのどのクラスのユーザの“視点”に注目するかを検索ユーザ自身に選択させることで、検索結果の絞り込みと再ランキングをおこなう。提案手法の効果は、研究室に所属する22名の学生に対しておこなったアンケート結果により評価される。

^{†1} 広島大学
Hiroshima University

本稿の構成は以下の通りである。まず2節で関連研究を概観する。提案手法の概要を3節で述べた後、キーポイントであるユーザプロファイルの取り扱いについて4節で詳述する。アンケート結果は5節で示される。最後に6節でまとめと今後の課題について述べる。

2. 関連研究

近年、フォークソノミー (folksonomy) を用いて Web 検索の精度を向上させる手法が盛んに研究されている*1。

高橋と北川は、HITS⁵⁾ の応用である S-BITS を用いて Web ページの有用性を評価する手法を提案した⁶⁾。HITS は PageRank と並んで広く認知されているリンク構造解析手法であり、Hub, Authority という二つの重要な概念に基づいている。ここで Hub は多くのページへの out-link をもつページであり、Authority は、特定のトピックにおいて重要なコンテンツをもち、多くのページからの in-link をもつページである。S-BITS では、Hub をユーザ、Authority を Web ページにとらえ、Web ページとユーザの間のブックマーク関係を強化することで、ユーザに信頼されている Web ページを発見する。これにより、PageRank 等で下位に埋もれた Web ページを効果的に発見することが可能となる。

百田と伊東は、類似利用者手法と α ブックマーカー手法という二つの Web 推薦手法を提案した⁷⁾。類似利用者手法では、「類似した嗜好をもつユーザの多くが類似した Web ページに興味をもつ」という仮定のもとで情報推薦がおこなわれる。具体的には、ブックマークデータをユーザのプロファイルとみなし、類似したプロファイルをもつユーザをコサイン類似度を用いて特定した上で、それら類似ユーザのブックマーク群から自身がブックマークしていない Web ページが推薦される。もう一方の α ブックマーカー手法では、類似利用者の代わりに、人気のある情報をいち早く入手できる特別なユーザ (アルファブックマーカー) に着目して Web 推薦がなされる。

一般にソーシャルブックマーク (SBM) を利用した Web 推薦では、ブックマークされた Web ページの共有の度合いをユーザの嗜好類似度とみなすことで情報推薦がおこなわれる。しかし対象となる Web ページ数が膨大になるにつれて、そのような直接的な類似度計算は困難となる。この問題を解決するため丹羽らは、ユーザの嗜好を、各ユーザと SBM 中の各タグとの親和度によって表現することを提案した⁸⁾。ここでユーザとタグの親和度とはユー

ザとタグとの tf-idf 尺度での関連性の強さのことであり、そのタグが希少であるほど親和度は高くなる。またこの方式では、フォークソノミーにおける表記のゆらぎをタグのクラスタリングによって解消し、より広い候補の中から Web ページを推薦することも目指している。

ところで上述の関連研究はいずれも SBM を想定した推薦システムであるため、SBM システムへのユーザ登録が必須であった。そのためプライバシー保護の観点からは課題が残る。また現実の SBM では、“これはすごい” や “後で読む” などといった、他ユーザにとってほとんど価値がないタグも数多く存在するため、タグ情報をそのままの形で情報推薦に使うことは実際には困難である。これらの問題点を解決するため林らは、SBM を想定しないタグベースの Web 推薦方式を新たに提案した⁹⁾。この方式では、各ユーザの嗜好そのものをユーザタグの集合として明示的に保持し、Web ページを閲覧したユーザたちに対応するユーザタグ集合が、閲覧されたページにそのままの形で付加される。様々な嗜好をもったユーザに閲覧されることで各ページに付加されるユーザタグに重なりが生まれることになるが、林らのシステムではそのようなタグの重なりを Web ページの特徴にとらえ、各ユーザの嗜好 (ユーザタグ集合) と Web ページの特徴 (重みつきユーザタグ集合) の近さによって、ユーザに提示される検索結果のランキングを変化させている。

3. 提案システム

3.1 概要

林らの手法では、SBM を利用せず、あらかじめ明示的に入力されたユーザタグのみに基づいて処理がおこなわれるため、覚書のような意味のないタグが一切存在せず、個別の Web ページへのタグづけに伴うユーザの負担が軽減されるという利点がある。しかしこの手法ではユーザタグ間の関係を考慮してないため、多義語を誤って重なりと認識してしまうなど、嗜好の類似性を正しく評価できない場合がある。またこの手法では検索ユーザの嗜好に関連のある Web ページの再ランキングをおこなっているが、どのような関連なのかについては明確に規定されておらず、パーソナライゼーションの目標としてはやや曖昧である。

そこで本研究では、林らの手法に新たにユーザタグ間の上位下位関係を導入し、ユーザタグ間の関係に関する各ユーザの知識や嗜好を利用することで、より精度の高い Web 推薦システムを実現することを試みる。具体的には、システムを利用しているユーザの集合を3パターンに分類 (ただし、各ユーザは複数のパターンに属することができるものとする) し、そのいずれかを検索時にユーザに選択させることで、Web 検索のパーソナライゼーションを実現する。以下にユーザの集合を分類する3パターンを示す：

*1 フォークソノミーとは、ユーザ自らが複数のタグを自由に付け加えていくことで Web ページの分類をおこなう手法のことである。

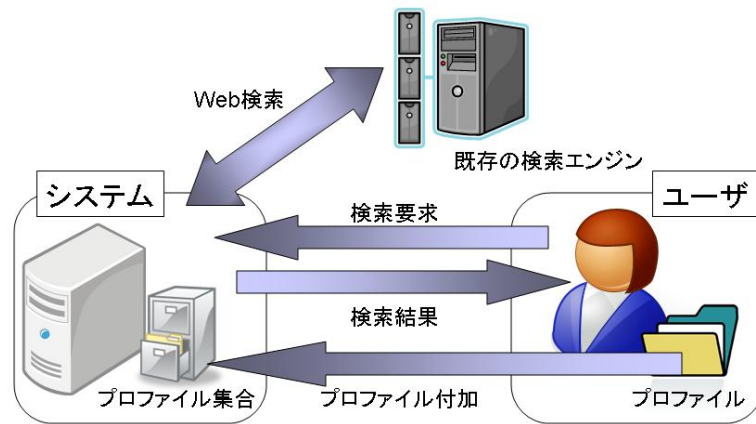


図1 提案システムの概略

- (1) 検索ユーザと同じ視点をもつユーザが多く閲覧した Web ページ
- (2) 検索ユーザと反する視点をもつユーザが多く閲覧した Web ページ
- (3) 検索ユーザより詳しい視点をもつユーザが多く閲覧した Web ページ

混乱を避けるため、ケース (2) の具体例を示しておこう。2名のユーザ X, Y が、それぞれフォークソノミー、ソーシャルブックマーク、タギングという3つのユーザタグをもっているとす。林らの方式では、これらのユーザはいずれも同じ嗜好をもつことになる。“ $a \geq b$ ”によって a が b の上位にあることを示すとしよう。2名のユーザが異なる視点をもつとは、たとえばユーザ X が

タギング \geq フォークソノミー \geq ソーシャルブックマーク、

ユーザ Y が

ソーシャルブックマーク \geq フォークソノミー \geq タギング、

のような上位下位関係を規定している場合のことをいう。この例では、 X は「タグ付け方法」に興味のあるユーザであり、 Y は「ブックマーク管理」に興味のあるユーザであると解釈することができる。視点の詳細については、次節で述べる。

3.2 処理の流れ

前述の通り、提案手法は林らの手法⁹⁾の拡張になっている。以下では林らの手法の概略を述べ、その後で提案手法の処理の流れを示す。

林らの手法では、まずはじめに各ユーザが、自身の嗜好を表すユーザタグをシステムに登録する。ユーザタグとは嗜好に関するキーワードであり、登録時には既存のキーワードから選択してもよいし、新たに自分で記入してもよい。ユーザタグ登録後、各ユーザはシステムを介して通常のキーワード検索をおこなう。検索キーワードはユーザタグに含まれている必要はない。検索結果はシステムを介してユーザに提示される。検索結果をクリックしてユーザが Web ページを閲覧したとき、システムはそのユーザのすべてのユーザタグを嗜好足跡 (preference footprint) として閲覧された Web ページに付加し、記録する。様々な嗜好をもったユーザに閲覧された結果、各 Web ページに付加されたユーザタグ集合はタグの重なりをもつことになる (各ページに付加されたユーザタグの集まりは、集合ではなく多重集合になる)。ここで、通常のSBMシステムで Web ページに付加されるタグがそのページがどのようなページなのかをあらわしているのに対して、林らのシステムで各ページに付加されるタグは、そのページがどのようなユーザに閲覧されたのかをあらわしていることに注意しよう。多くのユーザに閲覧されることで、マニアックで不適切なユーザタグの重みは相対的に低くなり、閲覧したユーザの統計的な情報がその Web ページの特徴として浮かび上がることになる。そのようにして獲得された Web ページの特徴は各ユーザの嗜好と比較され、それ以降の検索では、そのユーザの嗜好とのコサイン類似度が高いページが検索結果の上位に表示されるように検索結果が再ランキングされる。

本稿で新たに提案するシステムでは、林らのシステムで用いられているユーザタグの集合に新たに階層関係を導入する。階層関係が導入されたユーザタグ集合を以下では**プロフィール**と呼ぶ (プロフィールの詳細については後述する)。システムの基本構成を図1に示す。システムの基本的な流れは以下の通りである：

ステップ1： ユーザは、自身の嗜好をあらわすユーザタグをシステムに登録するとともに、登録したユーザタグ間の階層関係 (プロフィール) をグラフ形式で入力する。プロフィールの入力は適切な GUI を用いておこなわれる。

ステップ2： ユーザはシステムを介してキーワード検索をおこなう。試作システムでは検索エンジンとして Yahoo!検索を用いており、ランキングのトップ k が取得される (ただしこのタイミングでは表示はしない)。

ステップ3： 取得した k 個のページに付加されているプロフィールと検索ユーザのプロフィールを比較して**類似度**を計算する (類似度の計算方法は後述)。さらに類似度が閾値を満たすプロフィールと検索ユーザのプロフィールの比較をおこない閲覧ユーザを前述の3つの各クラスに分類する。 k 個の Web ページはユーザの半数以上が属するクラ

スに分類され、割合の多い順に再ランキングして表示される (表示されるクラスの切り替えは、検索ユーザがおこなう。ユーザの各クラスへの分類方法については後述)。

ステップ4: ユーザが URL をクリックしてある Web ページを閲覧すると、そのページにユーザのプロファイルが付加される。現在の実装では、各ページにはそのページを閲覧したすべてのユーザのプロファイルがリストの形で付加されている (情報の圧縮方法については今後の検討課題である)。

前述のように、林らの手法との違いは、以下の2点である: 1) Web ページの特徴がユーザタグの多重集合ではなくプロファイルの集合で与えられていること、2) ユーザプロファイルと Web ページに付加されたプロファイルリストの類似度が tf-idf とコサイン類似度ではなく、後述の集合演算によって定義されること。

3.3 プロファイル

提案手法においてプロファイルとは、概念の有限集合とそれら概念間の階層構造を規定したものである。各概念は林らの手法におけるユーザタグに対応し、概念間の関係としては、いわゆる多重継承は許すが巡回的な関係は許さない。システム中で用いられるユーザタグ全体の集合を S としよう。 S 上にはシステム管理者によって適切な半順序関係 $R (\subseteq S \times S)$ が与えられているものとする^{*1}。各ユーザは、 S の任意の部分集合 S' に対して半順序関係 $R' (\subseteq S' \times S')$ を設定することができる。ただし設定される関係 R' は、 R に矛盾してはならないものとする。ここで R' が R に矛盾していないとは、以下の条件を満たすことをいう:

$(a, b) \in R$ と $(b, a) \in R'$ を同時に満足するような $a, b \in S$ は存在しない。

あるユーザのプロファイルとは、そのユーザが選択した S の部分集合 S' と、 R に矛盾しないある半順序関係 R' の組 $\langle S', R' \rangle$ のことである。集合 S と半順序関係 R の一部が図2で規定されている場合のプロファイルの具体例を図3に示す。図3は R に矛盾しない半順序関係 R' の例を2つ示している。煩雑さを避けるため、この図では推移律を適用して得られるリンクは省略してある。

上記のモデルでプロファイルを作成したとき、プロファイル X と Y における要素 a, b 間の関係には、以下の3通りの場合があり得る:

合意している (a, b) が X と Y の両方に入っている、もしくはそのどちらにも入っていない

^{*1} 半順序関係とは、反射律、推移律、反対称律の三つを満たす関係であり、提案方式の文脈では、反射律は各タグに自己ループが存在することを意味し、反対称律と推移律は、生成されるプロファイルが非巡回グラフであることを意味する。空集合も半順序関係であることに注意しよう。

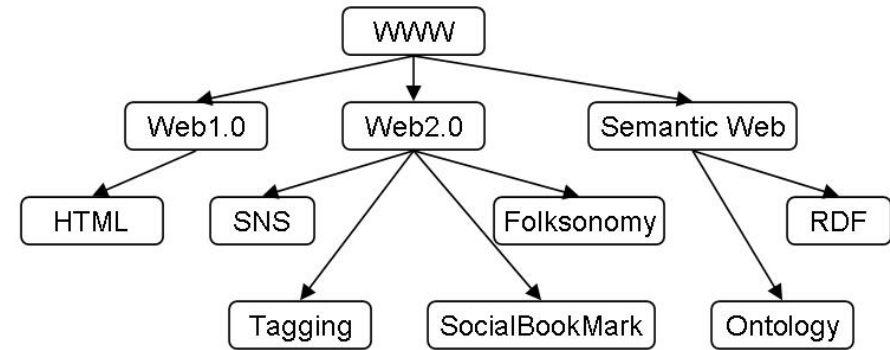


図2 集合 S とその上で規定された半順序関係 R の例。

競合している (a, b) が X に属し、 (b, a) が Y に属している (もしくはその逆)

詳細化している (a, b) がいずれか一方に入っている

前述の3つのユーザクラスは、直感的には上記の3つの場合とそれぞれ対応している。

4. プロファイルの比較

4.1 プロファイルの類似度

プロファイル X に含まれている要素数を $|X|$ 、プロファイル Y に含まれている要素数を $|Y|$ 、プロファイル X, Y の両方に含まれている要素数を $|X \cap Y|$ とあらわすことにする。このとき、プロファイル X, Y の類似度 $sim(X, Y)$ は、以下のように定義される:

$$sim(X, Y) \stackrel{\text{def}}{=} \frac{|X \cap Y|^2}{|X| \times |Y|} \quad (1)$$

$X = Y$ のとき $sim(X, Y) = 1$ であり、それ以外の場合は $0 \leq sim(X, Y) < 1$ となることに注意しよう。図3のプロファイル X と Y の類似度を計算すると、 $|X|=12$ 、 $|Y|=12$ 、 $|X \cap Y|=7$ より $sim(X, Y)=49/144$ となる。以下では、閾値以上の類似度 $sim(X, Y)$ をもつ閲覧ユーザのみを3つのパターンに分類する。

4.2 ユーザの分類

これまでそのページを閲覧したユーザは、検索ユーザと比べてどのような視点をもって見るかによって分類される。以下でその判別手法を述べる。

4.2.1 同じ視点か反する視点か

プロフィール X と Y における合意している要素ペア数を $agre(X, Y)$, 競合しているペア数を $conf(X, Y)$ とする. ただし, ユーザは半順序関係 R に矛盾しない R' のプロフィールをもっているため, R に含まれるペアは合意しているペアに含まないものとする. あらかじめ決められた閾値 θ_1 に対して

$$\frac{agre(X, Y)}{agre(X, Y) + conf(X, Y)} \geq \theta_1$$

を満たすとき, それらのプロフィールは**同じ視点**であるという. 一方, あらかじめ決められた別の閾値 θ_2 に対して

$$\frac{agre(X, Y)}{agre(X, Y) + conf(X, Y)} \leq \theta_2$$

を満たすとき, それらのプロフィールは**反する視点**であるという.

具体例として, 図3のプロフィール X と Y を比較してみよう. 半順序関係 R によってあらかじめ指定されているペアを除外すると, 合意しているペアは

(Web Search, Personalization), (Web2.0, Personalization), (Collaborative Filtering, Personalization), (Folksonomy, Personalization)

の4つとなる. 一方, 競合しているペアは

(Tag, Folksonomy), (Tagging, Folksonomy), (Tag, Tagging)

の3つである. したがって, $\theta_1 = 0.70, \theta_2 = 0.60$ とすると, $4/7 \leq \theta_2$ よりこれら2つのプロフィールは反する視点であると判断されることになる.

4.2.2 詳しい視点かどうか

R で順序関係が指定されていない要素ペアのうち, X に含まれていて Y には含まれていないペア (X が Y を詳細化しているペア) の個数を $refine(X, Y)$ とする.

$$refine(X, Y) > refine(Y, X)$$

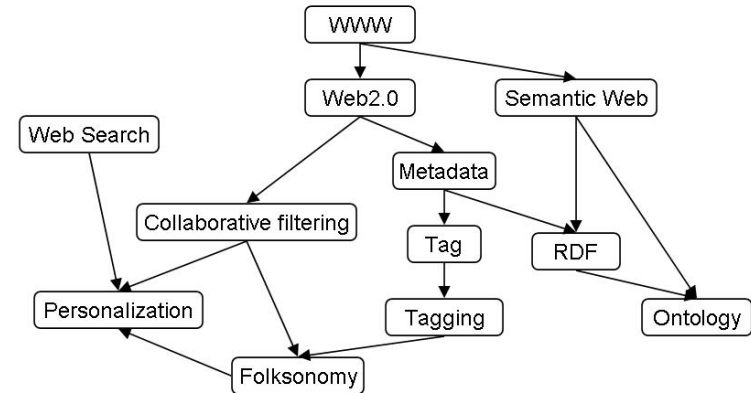
を満たすとき, X は Y よりも**詳しい視点**をもつという. 図3のプロフィール X と Y を比較すると, プロフィール X が Y を詳細化しているペアは

(WWW, Tag), (Web2.0, Metadata), (Collaborative Filtering, Folksonomy)

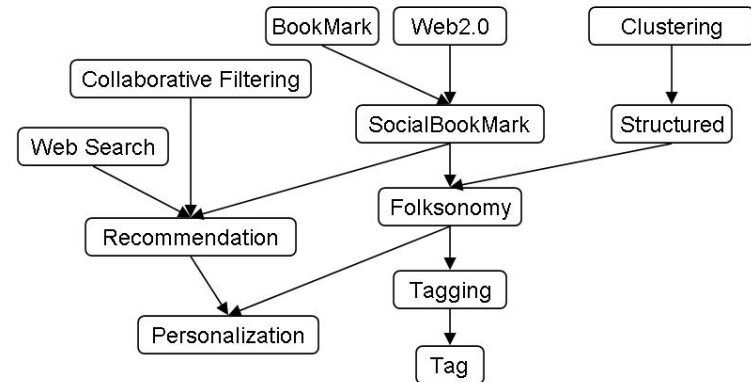
など計17個存在する. 一方, プロフィール Y が X を詳細化しているペアは

(Web2.0, Recommendation), (BookMark, Personalization), (Structured, Tag)

など計19個存在する. よってこの例では, プロフィール Y はプロフィール X よりも詳しい視点をもつと判断される.



(a) プロフィール X .



(b) プロフィール Y .

図3 R に矛盾しない半順序関係 R' の例.

5. 評価

提案手法の妥当性を検証するため実験による評価をおこなった. 実験では, 著者らと同じ研究室に所属する22名の学生を対象にアンケートを実施し, 自身の研究テーマに関連するプロフィールをそれぞれ作成してもらった. 具体的には, 各被験者に対して“Web マイニ

表 1 被験者の研究分野の主観的
分類

研究分野	人数
Web マイニング	6
P2P 検索	13
どちらにも含まれる	2
どちらにも含まれない	1
合計	22

表 2 各分野におけるユーザのタグ数

研究分野	人数	タグ最大数	タグ最小数	タグ平均数
Web マイニング	8(2)	46	12	19.36
P2P 検索	15(2)	24	11	16
その他	1	-	-	16
合計	22	46	11	17.41

ング”と“P2P 検索”に関する 64 個のタグとそれらの間の関係を例示したグラフを提示した後で、被験者自身の研究に則したグラフ(ユーザプロフィール)を作成してもらう。なお作成にあたっては、提示した 64 個以外のタグの使用も許すこととした。このため、被験者によってタグ数は異なっている。

次に作成されたプロフィールを前述の方法でクラス分けし、その分類が各被験者が自己申告した研究テーマの分類とどのような関係にあるのかを調べた。自己申告された研究分野を表 1 に示す。アンケートでは、各被験者に自身の研究と関連の強い他の被験者の名前を記入してもらい、その結果を利用して 2 つの研究分野内でさらに細かくグループ分けをした。グループ分けの結果、“Web マイニング”は 5 名、2 名、1 名の 3 つのグループに分かれ、“P2P 検索”は 7 名、3 名、3 名、2 名の 4 つのグループに分かれた。なお今回のアンケートでは、タグ集合 S と適切な半順序関係 R を提示しなかったため、アンケート結果から得られたグループ分けを利用して以下の方法で S と R を抽出した：1) 過半数が利用したタグを集合 S とする。2) S 中の 2 つのタグの間の関係に最多数グループ(上述の 5 名のグループと 7 名のグループ)の半数以上が同意しており、かつどの被験者もその関係と競合していない関係を R とする。

プロフィール作成は手書きによっておこなわれたため、タグや関係性の記述忘れや、検索の利用に適さない過剰なタグ数のグラフ作成などの問題が生じた。そのため以下の評価では、3 つのクラスのうち、同じ視点と反する視点のユーザの特定のみをおこなうこととする。

5.1 評価結果と考察

類似度

各グループに属するメンバーの間で提案手法による類似度を算出した。Web マイニングの 5 名のグループに関する結果を表 3 に示す。表からわかるように、被験者 A を除くとメンバー間の類似度はいずれも 0.163 以上となった。対照的に被験者 A に対する類似度は、一致タグ数が他の被験者同士の一致タグ数とほぼ等しいにもかかわらず、最大でも 0.093 と小

表 3 Web マイニングの中の 5 人グループの類似度 (一致タグ数)

被験者名 (タグ数)	A(46)	B(15)	C(18)	D(15)	E(20)
A(46)	-	0.093(8)	0.077(8)	0.093(8)	0.088(9)
B(15)	-	-	0.181(7)	0.284(8)	0.270(9)
C(18)	-	-	-	0.237(8)	0.336(11)
D(15)	-	-	-	-	0.163(7)
E(20)	-	-	-	-	-

表 4 表 3 に該当するユーザの合意ペア基準の比率
(合意しているペア数：競合しているペア数)

被験者名 (タグ数)	A(46)	B(15)	C(18)	D(15)	E(20)
A(46)	-	1.00(3:0)	0.40(2:3)	0.67(4:2)	0.67(4:2)
B(15)	-	-	0.00(0:2)	0.75(3:1)	0.70(7:3)
C(18)	-	-	-	0.50(3:3)	0.96(24:1)
D(15)	-	-	-	-	0.50(3:3)
E(20)	-	-	-	-	-

さい値である。これは、被験者 A のタグ数が他の被験者に比べて非常に大きいためである。よって、閾値以上の類似度を得るためには検索したい情報を得るための必要最低限のタグ数のプロフィールを構築することが重要であり、この問題は、巨大なプロフィールを複数に分けることによって解消が期待できる。また、この 5 名とその他の 3 名の類似度の平均は 0.043 であり、アンケートの結果と同じように類似度によってグループ分けできることがわかった。

一方の P2P 検索では、アンケート結果とは若干異なる類似度が算出された。その原因としては、Web マイニングのグループと比べて被験者同士がお互いに近い目的や近い手法での研究を行っておらず、 R で定義されるタグ以外のタグが一致していないことが挙げられる。よって、嗜好分野が同じでも似通った考え方を持たなければ類似度が高くないことが確かめられた。

ユーザのクラスの分類

次に、メンバー間の類似度が高かった上述の 5 名のグループに着目し、それらのメンバーが合意しているペア数と競合しているペア数の比率を調べた。結果を表 4 に示す。今回のアンケートでは、被験者同士が想定していたほどの密な嗜好(研究目標や背景)をもっていないため、合意、競合しているペア数が少なく細かな数値が得られなかった。しかし表からもわかるように、合意しているペアの比率が全体的に高めの結果となっている。よって、

閾値 θ_1 と θ_2 はともに 0.51 以上が妥当であり, $\theta_1 > \theta_2$ が適当であることがわかる. 表 4 の極端な例として, 被験者 C と被験者 E の比率に注目したい. この 2 名の同意しているペア数は他の被験者に比べて圧倒的に多い. これは半順序関係 R がうまく抽出できておらず, R の最上位タグに同じタグを構築した結果, そのタグに関する上位下位関係がすべて合意しているペアとしてカウントされたものである.

このように数値に多少の例外はあるものの, 各個人がもっているタグ間の上位下位関係の違いを総合的に考慮することによって, 知識や嗜好に対する視点の差が自然に抽出できることが確かめられた.

6. 結 論

本研究では, ユーザの嗜好を概念の関係にとらえ, グラフで表現することで Web ページをパーソナライゼーションすることを提案した. また, 適切な半順序関係 R を提示しユーザのプロファイル作成をサポートすることで正しい類似度を算出し, ユーザを視点に従ったクラスに分類できることを示した. 今後の課題としては, 類似度の算出において $tf * idf$ を利用した評価式を用いることとより密な嗜好をもつユーザ間での提案手法の評価が挙げられる.

謝 辞

アンケートに協力していただいた研究室の皆様にお礼申し上げます. ありがとうございます.

参 考 文 献

- 1) Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999. Previous number = SIDL-WP-1999-0120.
- 2) GooglePersonalized Search. <http://www.google.com/>.
- 3) Minko Dudev, Shady Elbassuoni, Julia Luxenburger, Maya Ramanath, and Gerhard Weikum. Personalizing the search for knowledge. In *PersDB 2008*, 2008.
- 4) 岩崎 周造 and 太田 学. 構造化プロファイルを用いた個人化 web 検索システム. *DEWS2008*, 2008.
- 5) JonM. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46:668-677, 1999.
- 6) 高橋 翼 and 北川 博之. ソーシャルブックマークを利用したユーザ嗜好に基づくペー

ジの評価. *DEWS2008*, 2008.

- 7) 百田 信 and 伊東 栄典. ソーシャルブックマークに基づく情報発見. *DEWS2008*, 2008.
- 8) 丹羽 智史 and 土肥 拓生. Folksonomy マイニングに基づく web ページ推薦システム. *情報処理学会論文誌*, 47(5):1382-1392, May 2006.
- 9) Shuhei Hayashi, Yuuki Inoshita, and Satoshi Fujita. An efficient web page recommendation based on preference footprint to browsed page. **第 5 回情報科学ワークショップ**, 2009.