パーソナライズ情報提供方式の提案と評価

橘 高 博 行† 佐 藤 直 之†
鈴 木 英 明† 曽根岡 昭直†

本稿では，WWWにおける情報提供において，サーバで保持する情報群からユーザの興味に応じて適切な情報を順次選択し提供する情報提供方式を提案する．本方式では，更新にともなう計算の簡便化および容易に理解可能な表現を目指し，従来手法の多くで用いられる単語をベースとしたユーザの興味や情報の有する特性の相対的な表現を，より抽象化した概念をベースとした絶対的な表現とした．さらに，従来のCognitive Filtering，Social Filteringの2つのフィルタリングの機構を組み合わせて1つの表現形式で実現している．これにより，ユーザの興味の変化に自動的に追従でき，かつ，情報提供者による誤った特性づけの修正および特性の時間経過に対する変化への自動追従がともに可能となる．また，興味と情報の特性の両方について，概念をベースとした絶対的な位置を表現しているため，情報の傾向を示す質だけでなく，その重要性/価値といった量についても考慮し，絶対的に情報量の多い情報を優先してユーザに提示することが可能になる．また，本方式を，ニュース記事を提供するWWWサイトに適用し評価を行った．その結果，上記特徴の有効性が確認できた．

A Proposal and Evaluation of News Personalized Navigation Service

Hiroyuki Kittaka,† Naoyuki Satoh,† Hideaki Suzuki† and Terunao Soneoka†

We propose a personalized information navigation system that filters specific information from the range of that available on the World-Wide-Web servers, based on individual user's requirements. Many existing systems describe user's interests and information characteristics using vectors composed of a set of words and their weights. We use concepts instead of words. Concepts are more abstract than words, and their weights can be represented absolutely. This allows users and information providers to analyze and update them easily. Using this conceptual representation, we combine two well-known filtering mechanisms into one system. In our system, the user vector is learned from user behavior, and is updated dynamically to reflect changes in the user's interests. At the same time, the information vector is updated based on feedback from the user vector, and inappropriate characteristics established by an information provider are corrected. The information vector always shows accurate characteristics, even if they change. The most valued information is selected from the information that has similar directivity, by comparing the user vector with the information vector and considering directivity and the amount of information. We evaluate our system's performance as applied to a newspaper-providing service, and confirm the validity of the system's features.

1. はじめに

近年インターネットの普及にともなって，WWWを用いて発信される情報の量も増加の一途をたどっている．しかし一方で，膨大な情報の中からユーザが自分の興味にあう情報を選択的に獲得することは難しくなっている．このため，WWWで情報提供を行う際，画一的な情報の提供を行うだけでなく，個々のユーザの興味を考慮して選択的に情報の提供を行うことが求められている．

この要求に対して，ディレクトリサービスやニュースサービスなどで，これまでに様々なパーソナライズサービスが実行されている[1]．しかし，その大半のサービスでは，ユーザにアンケートや情報の評価などの投入作業を要求する．しかし，これらの投入作業や興味の変更時の再登録は，ユーザに過大な負担を要求することになる．これに対して，ユーザの情報の参照履歴などから，自動的にユーザの興味を把握し，興味に応じた情報を提供する情報フィルタリング方式[5),7),8)]や

† NTTソフトウェア研究所
NTT Software Laboratories

情報リコメンド方式[2),4),6)]が提案されている．しかし，これらの方式でも，2章で詳しく述べるように，ユーザの興味や情報の特性の表現方法，両者の比較・照合方法などで多くの問題が残されている．ここで，情報の特性とは，その情報がどのような興味を有するユーザに適しているかを示したものである．

本稿では，WWWを用いた情報提供において，ユーザの興味を自動的に把握し，WWWサーバ側で保持する情報群からユーザの興味に応じて適切な情報を順次選択し，ユーザに紹介するパーソナライズ方式を提案する．特に，ニュース記事の提供に代表されるような，膨大な情報の蓄積が必要であり，情報の追加および流行となる話題の変化が頻繁に起ることが想定されるパーソナライズ情報提供を支援する．

本方式では，ユーザの興味，情報の特性の表現を，従来手法の多くで用いられてきた，単語をベースとする情報間の差異に着目した相対的な表現でなく，単語をさらに抽象化した概念をベースとする絶対的な表現とする．この表現形式は，ユーザおよび情報提供者に理解しやすく，それぞれの更新にともなう計算を簡便化することができる．

さらに，この表現方法により，1）あらかじめ情報から抽出された特性のみを基に，ユーザに適した情報を紹介すること，2）多くのユーザからの参照により情報の特性を付与し，これを基にユーザに適した情報を紹介すること，の従来の2つの手法を同時に1つの表現形式で扱うことができる．本手法では，3章で述べるように，ユーザの参照した情報の特性のベクトルをユーザの興味のベクトルに反映させることで，ユーザの参照行動から自動的にユーザの興味を把握する．同時に，情報を参照したユーザの興味のベクトルを情報の特性のベクトルに反映させることで，ユーザの観点から情報の特性を更新することができる．これにより，誤った特性づけの修正や時間の経過にともなう情報の意味の変化に対応できる．また，情報の特性を情報量に応じた絶対的な位置づけで表現することで，情報の内容の方向性を示す質だけでなく，重要性/価値といった量についても考慮した情報選択を可能としている．すなわち，同じ性質を持った情報でも，情報量の差で順序をつける情報選択を行うことができる．

以降，2章では，現状方式の問題点と本提案方式のアプローチ，本提案方式と関連する研究について述べる．3章では，本提案方式の構成ついて詳細に述べる．4章では，実サービスにおける本提案方式の有効性を確認した評価結果を報告する．

2. 既存システムの問題点と解決法

2.1 問題点と提案手法でのアプローチ

現在，ニュース記事に代表されるような，随時蓄積されていくような情報を扱う情報フィルタリング方式や情報リコメンド方式では，Cognitive Filtering（以下，CF）機構を有するものが主力となっている．これらの方式では，情報からその特性を抽出し，この特性とユーザの参照履歴からユーザの興味モデルを作成，更新し，ユーザの興味に適した情報を選択する．本稿で提案する方式も，基本的にはCFの考え方に基づいている．以下，既存のCF機構を用いた方式における問題点と，本提案方式での問題解決へのアプローチについて論じる．

1）興味/情報モデルの表現

CF機構を有するいくつかの方式では，ネットワークモデルや論理式などを用いて，より詳細なユーザの興味モデルの構築を試みている．しかし，このような複雑かつ論理的な興味モデルは，ユーザの興味の最終的な到達点が一意に決定できることを前提としている．したがって，ある時点を境にユーザの興味が以前とは相反する興味に変化した場合，興味モデルを無矛盾に保つことが難しく，さらに，無矛盾に保つためには興味モデルの更新に時間がかかるという問題がある．また，ユーザが次回のアクセスから，今までとまったく異なった興味で情報を参照したくなることも考えられる．このような場合，ユーザが自分の興味モデルを手作業で更新しようとしても，この複雑な興味モデルの解釈や誤りのない修正は非常に困難であるという問題がある．

このような複雑な興味モデルに対して，次節に述べる既存方式の多くで用いられている，単語とその重みによるベクトルでユーザの興味モデルを表現する形式がある．このような，簡単な表現形式では，情報の特性ベクトルをユーザの興味ベクトルに反映させることで，自動的にユーザの興味の把握・更新を行う．ユーザの手作業による興味モデルの修正は，各単語に対する重みを増減させることでに行うことができる．この表現形式は，次の2点を前提としている．1）ユーザが興味を持って参照する記事には，ある共通の単語が含まれている，2）情報の特性は，その情報に高い頻度で含まれ，かつ，他の情報にはあまり含まれない単語によって代表される．この2つの前提に従い，情報ベクトルは，TFIDF[9)]☆計算によるメトリックで計測

☆ $\mathrm{TFIDF}(w,d) = TF(w,d) \times \log(N/DF(w))$：$TF$ は情報 d における単語 w の出現頻度，DF は単語 w が出現している情報の数，N は提供されている情報の数．

した重みを要素とすることが多い．この方法は，情報を代表する単語の出現頻度の違いを基にして，それぞれの情報が全体の情報の中でどの位置にあるか，相対的に表現するものである．

したがって，情報の特性として単語の出現頻度を用いた場合，多くの情報で共通して使用される単語については，重みをほとんど持たないことになる．たとえば，ニュース記事に代表される，流行となる話題が頻繁に登場するような種類の情報を扱った場合，提供される情報のすべてにおいて，流行語やトレンドとなる単語はその重みの違いが微少なものとなる．このため，流行の話題に着目していたユーザには，必ずしも最適な情報が紹介されないという問題がある．その他にも，単語レベルで興味と情報の特性を表現するため，単語間の類似や関連の処理をあわせて行う必要があるという，よく知られた問題もある．

本方式でも，3.1 節で述べるように，既存のシステムと同様に，情報の特性とユーザの興味を同一のベクトルで表現する．ただし，本方式では，情報をそのまま出現する単語のベクトルで解釈するのではなく，さらに抽象的なフィルタをかけて，概念の軸と重要度/価値の重みのベクトルで情報を解釈する．ここでの概念は，概念辞書にあるような個々の概念であり，たとえば，提供する情報が「政治」に関係した新聞記事ならば，「郵政民営化」や「選挙」などが，それぞれの概念である．本方式では，概念となる軸は，情報提供者ごとに，彼らが提供していく情報の種類に基づき，適度な粒度で選択する．階層構造などによる概念どうしのつながりは，選択の際の粒度として吸収される．また，本方式は，カテゴリ化のように，ある情報を 1 つの概念にのみに帰属させず，情報提供者が設定した個々の概念に対する帰属度合である重要度に基づいた重みづけを行う．これによって，原点からの距離が重要度として表現される概念空間の中での，絶対的な位置づけが可能である．

本方式では，表現が今後提供する情報群も含めた絶対的な位置をベクトル表現しているため，すべての情報におけるトレンドの単語を含む概念の重みは，情報の追加では変わらない．このため，トレンドの単語もしくはそれを含む概念に着目しているユーザに最適な情報が紹介される．また，単語の関連性や類似性に関しても，3.2 節で例示するような半自動的な特性づけの仕組みなどであらかじめ解決され，パーソナライズシステムに必要な機能だけに絞り込んで議論できる．

2）特性の補正と変化

既存方式における単語の軸のベクトルでは，情報の意味が陽に文字列で表現されていない，すなわち，先の前提の 2）がうまく成り立たっていない場合がある．この情報の特性は，設定に誤りがあったことになり，この情報についてのみ，なんらかの修正を行う機構が必要になる．情報提供者が，手動で情報の特性を修正するには，使用する単語に注意して情報の内容を記述し直すなど，労力が大きくなりあまり好ましくない．

さらに，時間の経過とともに，情報の持つ意味や価値が変化することがある．たとえば，古い論文やニュースが，現在になって脚光をあびるようなことがある．これは，この情報だけが，他の情報の特性に影響せずに，単独にその特性の重みを変えることである．先に述べた TFIDF を用いた場合，新たな情報の追加とともに，既存の情報の特性も更新される．しかし，すべての情報の相対的な位置関係が更新されただけで，情報が単独にその特性の重みを変えているわけではない．

Social Filtering（以下，SF）機構では，情報を紹介すべき興味モデル群の更新を，ユーザから見た情報の特性づけにより行う．このため，特性の補正や更新は，SF 機構を応用し，情報を参照するユーザ群からの修正を加えることで行うことが期待できる．既存の SF 機構を持つシステムは，ユーザから見た情報の特性としてのフィールドだけを持っているが，ユーザが評価を投入する必要があるため，ユーザの労力の問題がある．実際，既存の方式では，SF 機構などによる修正機能を取り入れず，情報の特性に基づいてユーザに情報を紹介する CF 機構のみのシステムになっている．

本提案方式では，ユーザの興味も概念空間の中での絶対的な位置を示すことになる．したがって，3.4 節で述べるように，情報を参照したユーザの興味を再び情報の特性に反映させることで，ユーザの入力などの労力を要する別フィールドなしで，ある情報についてだけ情報の特性を更新する SF 機構を実現できる．これによって，情報の特性の設定の誤りや，時間の経過にともなう情報の重要度や価値の変化に対応できる．

本方式では，この他に，特性が正しく設定されていてほとんど更新されない情報や，新規に追加された情報に関しては，CF 機構も実現している．したがって，SF と CF の 2 つの機構を同時にあわせ持つ形で実現しているという点で特徴的である．このため，既存のシステムに比べて，単語の軸で情報に関する特性を設定できない情報に関しても，ユーザの参照により自動的に特性が設定されるため，単純なテキスト情報だけでなく様々なマルチメディア情報を統一的に扱うことができる．

3）情報の選択

単語の頻度ベクトルで興味や情報の特性を表現した場合，両者が相対的な位置づけを表現しているため，ユーザの興味と情報の特性の2つのベクトルの方向性にのみ着目することが多い．具体的には，比較・照合で，2つのベクトルを正規化してベクトルの内積を求め，ベクトルの方向の類似度を計算する．そして，より類似する方向性を持つ情報を選択することで，ユーザの興味に従った情報の選択を行っている．しかし，この照合方法では，情報の特性の方向性にのみ着目しており，情報量については考慮されない．ここでの情報量とは，方向性とは異なった，情報が持つ重要性や価値などを指す．このため，方向性の同じ情報群では，各情報に優先順位をつけることができず，情報量が多い情報も他の情報と同等に扱われる問題がある．

本提案方式では，3.5節で詳述するように，情報量にも着目した情報の選択を行う．本方式では，情報の特性は概念空間における絶対的な位置を表現しているため，方向性を比較するだけでなく，その量（重要性/価値）についても比較することができる．これにより，方向性と量の両方に着目した，きめ細かい情報の選択が可能である．

2.2 関連研究

以下に，既存の情報フィルタリング方式や情報リコメンド方式について関連するものを述べる．最初にCF機構を有する方式について述べる．

Stanford大学のLira[5]では，ユーザの興味と情報の特性の両方に，単語を軸としTFIDFで求めた重みによるベクトルをそのまま使用する．この方式では，リンクに対する評価値をユーザが手作業で投入する．情報の特性のベクトルを，投入した評価値に準拠した度合いで，ユーザの興味のベクトルに反映させ，ユーザの興味を把握する．逆に，興味のベクトルの特性ベクトルへの反映は行わない．

CMU大学のWebWatcher[2]やPersonal WebWatcher[3]では，情報中に出現する単語群だけでなく，ユーザがキーワードとして投入した単語と，情報およびそのリンク中に出現する単語をベクトルの軸として使用する．ただし，このベクトルの要素数は，あらかじめ情報提供者が限定している．WebWatcherは，Webページをユーザの興味に従ってナビゲート（ツアー）するシステムで，多くのユーザがツアーをしたときの履歴やハイパーテキストの構造などの情報を基に，現在参照しているページにおいて，ユーザの興味にあっているリンクをリコメンド（マークアップによる）する．この方式では，ツアーの最初にユーザがキーワードを投入する必要がある．このキーワードは，ユーザがリンクを選択したとき，情報の特性として付加される．この方式では，ユーザが投入したキーワードは，各情報の単語の頻度ベクトルとは異なるフィールドで管理している．このため，キーワードAを投入したユーザと，キーワードBを投入したユーザが同じリンクをたどれば，これらのキーワードに関連があることになる．すなわち，情報の特性自体を更新しているというより，特性の中の単語やキーワードの関連を学習しているとも考えられる．Personal WebWatcherは，ユーザ側にプロキシを置き，ユーザが参照した情報に含まれる単語を軸とするベクトルで，ユーザの興味と情報の特性を表現する．ユーザが参照しているページのリンクを先読みするなどして，興味にあったページを提示する．WebWatcherと違い，他のユーザの参照履歴などは利用されない．

CMU大学のWebMate[7]では，参照した情報の単語の頻度ベクトルを，類似度にあわせてカテゴライズしたM個の集合で，ユーザの興味を表現する．したがって，ユーザの興味は，単語の頻度ベクトルで代表されたM個の概念から構成されている．これは，本提案方式における概念の表現を用いた方式に近い．しかし，参照した情報の集合を，代表的なM個の概念で切り分けているため，ユーザが次々と興味のターゲットを変えていった場合，ユーザの興味を構成する概念が，概念辞書などにおける非常に抽象的なM個の概念へと収束してしまう可能性がある．

SF機構を備えた代表的なシステムとして，Firefly (HOMR)[4]がある．この方式では，興味の類似するユーザの履歴を参照して，あるユーザがまだ評価していないが興味がありそうな情報をリコメンドする．この方法では，ユーザの評価点の投入労力がかかるほか，多くのユーザの評価が集まらないとうまくリコメンドができない．また，Stanford大学のFab[10]では，基本的にSF機構を用いて情報の紹介を行い，ユーザの評価の蓄積が十分でない場合をCF機構で補っている．したがって，CF機構，SF機構の両方を備えているといえる．しかし，このCF機構はSF機構が働かない情報に関しての代用であり，CF機構における情報の特性のフィールドはSF機構におけるそれとは別フィールドである．このため，ユーザから見た情報の特性づけは，CF機構におけるフィールドには反映されない．

3. InfoBroket方式の構成

本章では，2章で述べた要求を満たす情報提供方式

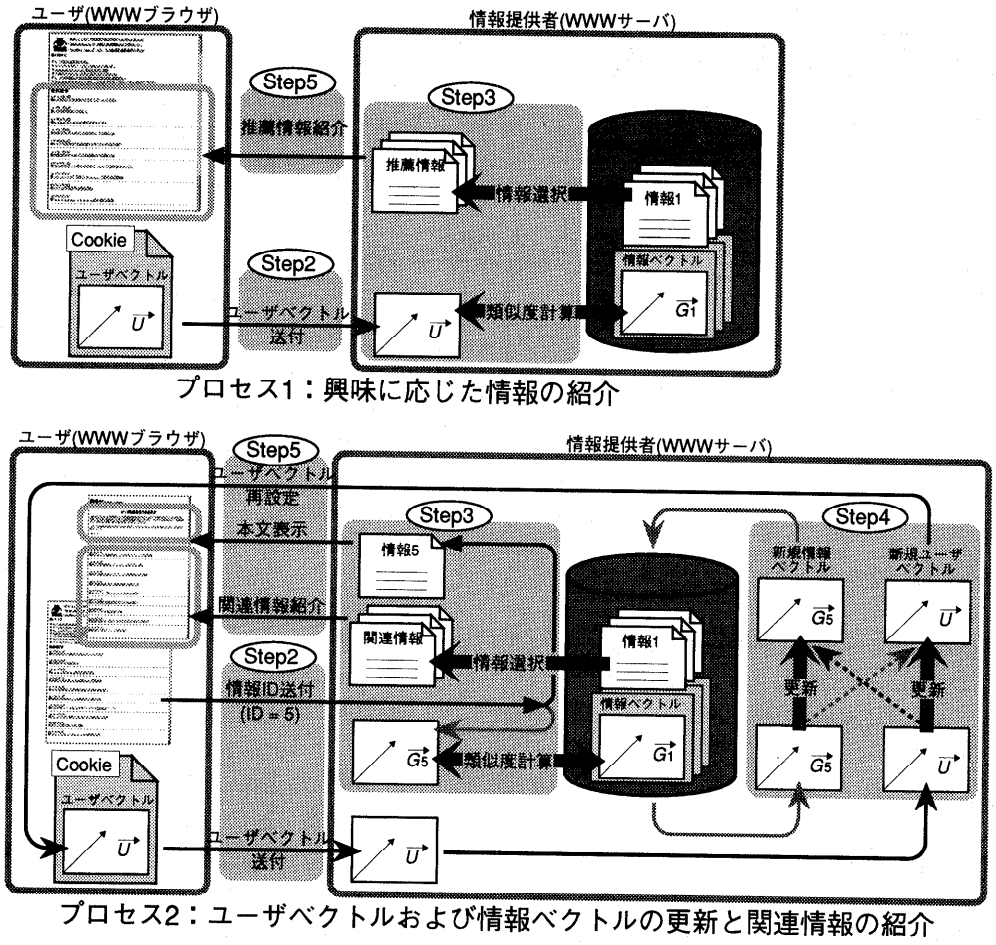


図1 InfoBroket 方式のシステム構成
Fig. 1 An overview of InfoBroket system.

（以下，InfoBroket 方式）を提案し，その特徴を詳細に述べる．以下，3.1 節では提案する InfoBroket 方式のシステム構成とその概要について述べる．以降の節で，情報ベクトルの初期値の設定方法，ユーザの興味の更新方法，情報の特性の更新方法，情報の選択方法の各々を詳細に述べる．

3.1 InfoBroket 方式の概要

WWW を用いたパーソナライズサービスに，InfoBroket 方式を用いた場合のシステム構成を図1 に示す．情報提供者は WWW サーバ上に情報を公開し，ユーザは WWW ブラウザで情報を参照する．情報提供者とユーザの間の通信は，すべて HTTP で行われる．InfoBroket のシステムは，CGI アプリケーションとして実装されている．本方式では，ユーザの興味と情報の特性を，概念を軸とした同一の次元からなるベクトルで表現する．ここで，ユーザの興味をベクトル表現したものをユーザベクトル $\vec{U}$，情報の特性をベクトル表現したものを情報ベクトル $\vec{G}$ とする．ユーザベクトルをユーザ端末に，それぞれの情報の情報ベクトルを WWW サーバに保存する．両ベクトルは，以下に示すように，w_j，w'_j の重みを要素とする N 次元のベクトルである．

$$\vec{U} = (w_0, w_1, \ldots, w_N),\ \vec{G} = (w'_0, w'_1, \ldots, w'_N)$$

システム構成における特徴は，ユーザベクトルをユーザ端末で保存し，情報を参照する際にサーバへ送信，サーバから受信する点である（Cookie を使用）．ユーザ端末に保存する情報は，ユーザベクトルだけであり，ユーザ ID などの識別情報を含む必要はない．本システムでは，ユーザ ID などのユーザを識別する情報と，マーケティング・リサーチやパーソナライズに必要となるユーザの興味を示す情報を分離して扱うことを可能としている．また，Cookie は WWW サーバごとに独自に設定することが可能であるため，特別

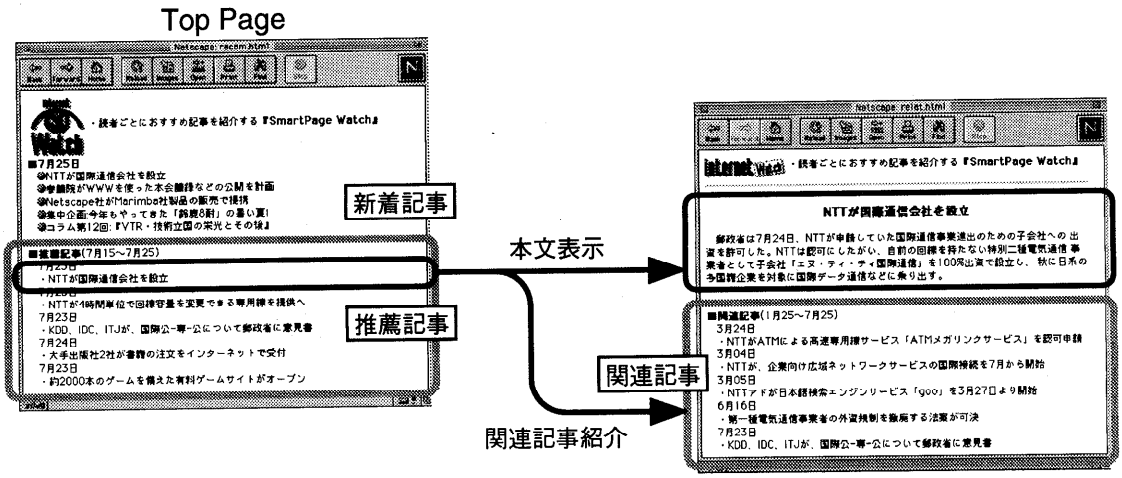


図2 実装サービス例
Fig. 2 An example of InfoBroket service.

な機構を別途用意することなく，各情報提供者ごとに独自のユーザベクトルを用いることができる．このため，各情報提供者は，彼らが提供する情報の種類に基づいてユーザに適した情報を提示したりマーケティングを行ううえで最適となる概念をベクトルの軸に用いることができる．

以下，図1における各ステップでの処理について説明する．Step1では，ユーザベクトル，情報ベクトルの初期値を設定する．この処理は，ユーザが初めてサーバにアクセスした時点，および，情報を新規に掲載した時点で行われる．ユーザがある情報を参照した際のユーザベクトルの更新，情報ベクトルの更新，および参照した情報に関連する情報の選択（図1のプロセス2）については，Step2～5で行う．ここで，Step3では参照した情報の情報ベクトルと，それ以外の情報の情報ベクトルの類似度を計算する．これにより，参照した情報と関連する度合いが高い情報の選択を行う．さらに，ユーザの興味に応じた情報の選択（図1のプロセス1）では，Step2，Step3，Step5のみを行う．このとき，Step3では，ユーザベクトルとすべての情報の情報ベクトルとの類似度を計算し，ユーザの興味に応じた情報の選択を行う．本システムでは，これらのプロセスを自由に組み合わせることで，**図2**に示すように，ユーザの興味に応じた情報（図2の推薦記事），および，参照した情報と関連した情報（図2の関連記事）とを情報提供者が望む見せ方で紹介することを可能にしている．

Step1. ユーザベクトル，情報ベクトルの初期値の設定

ユーザベクトルの初期値は，すべての軸に同じ重みを設定した単位ベクトルとする．情報ベクトルの初期値の設定方法については，3.2節で詳細に説明する．

Step2. 情報要求

プロセス1では，ユーザベクトルをサーバに送付する．プロセス2では，ユーザベクトルと，本文を参照する情報の情報IDをサーバに送付する．

Step3. 情報の選択

ベクトルの類似度を計算し，類似度の高い情報から順に順序をつける．そして，この順序に従い，情報提供者が指定した個数の情報を選択する．類似度の計算方法については，3.5節で詳細に説明する．プロセス1では，ユーザベクトルと提供情報の情報ベクトルとの類似度を計算することで，ユーザの興味に応じた情報を選択する．プロセス2では，ユーザが本文を参照した情報の情報ベクトルと，その他の提供情報の情報ベクトルとの類似度を計算することで，参照した情報と関連した情報を選択する．同時に，ユーザから送付された情報IDに対応する情報の本文を取り出す．

Step4. ユーザの興味および情報の特性の更新

ユーザが本文を参照した場合，すなわちプロセス2の場合のみ，以下の処理を行う．ユーザから送付された情報IDに対応する情報の情報ベクトルを取り出す．そして，ユーザベクトルに情報ベクトルを反映させることでユーザの興味を更新する．逆に，情報ベクトルにユーザベクトルを反映させることで，情報の特性を更新する．これらの更新方法については，3.3節および3.4節で詳細に説明する．

Step5. 情報紹介

プロセス1では，Step3で選択された情報をユーザに送付する．プロセス2では，Step3で選択された情報，Step4で更新されたユーザベクトル，Step2で送付された情報IDに対応する情報の本文を送付する．

3.2 情報ベクトルの初期値の設定

情報ベクトルの初期値は，人間（情報提供者）が手入力で設定するほかに，半自動的に設定する方法がある．ここでは，半自動的に設定する方法について説明する．以下，半自動設定における，各ステップにおける処理の詳細を記す．(1)〜(3)と(7)の処理は，情報提供者が手動で行う必要があるが，(1)〜(3)の処理は原則として，サービス開始時に1度行うだけでよい．(4)〜(6)の処理は自動的に行うことができる．また，新たな概念が必要となった場合には，この新しい概念に対して(1)〜(3)の処理を行う．新規の概念の追加は，既存の概念に対して影響を与えないので，既存の概念に対しては特別な処理は必要ない．この場合は，既存の情報群は，新規の概念（軸）に対する重みを持たないが，3.4節で述べる情報の特性の更新により，自動的に重みが付加される．あらかじめ新規の軸に対して初期値を設定したい場合には，既存の情報群に対して(4)〜(6)の処理を行う．

(1) 提供する情報に応じて，概念空間を形成する軸を決定し，名前（ラベル）をつける．この軸が，情報ベクトルの軸になる．また，軸となる概念の個数は任意であるが，それぞれが独立していて，直交することが望ましい．

(2) それぞれの概念について，その概念に強く帰属すると考えられる情報をいくつか選択する．これを，各概念におけるサンプル情報の集合とする．

(3) 2章で述べた方法を用いて，サンプル情報に対してTFIDFを実行し，各概念ごとに単語の頻度ベクトルを求める．この単語の頻度ベクトルを，概念空間における各軸の代表元とする．

(4) 新規に追加する情報（以下，新情報）に対してTFIDFを実行し，単語の頻度ベクトルを求める．ここでのIDFには上記(3)のステップで使用した各単語のIDFを用いる．

(5) 新情報の単語の頻度ベクトルと，各軸の代表元との類似度を計算する．ここでは，両ベクトルのなす角度に応じた類似度を求める．

(6) 類似度に基づいて，新情報の情報ベクトルの各軸に対して重みづけを行う．たとえば，情報ベクトルの各軸のとりうる重みが0〜50の整数値である場合，類似度がある閾値を超えた軸は25，それ以外の軸は0などとする．

(7) 上記(6)のステップで設定された各軸の重みを修正する必要がある場合，手動で調整する．情報の特性は，軸の名前と重みの組で表現されているため，GUIを用意するなどして簡便な作業で修正を行うことができる．

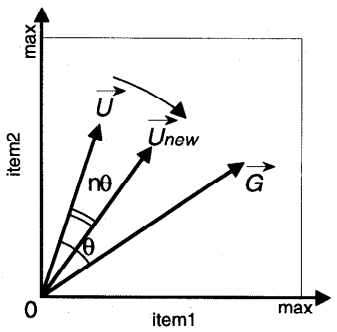


図3 ユーザの興味の把握，更新
Fig. 3 Grasp and renewal of user's interest profiles.

3.3 ユーザの興味の更新

ユーザが情報を参照する行動から，ユーザの興味を把握する．以下，**図3**を用いて説明する．図中で，あるユーザのユーザベクトルを $\vec{U}$，このユーザが参照した情報の情報ベクトルを $\vec{G}$ とする．ユーザがある情報を参照した場合，その情報の特性を表現した情報ベクトル $\vec{G}$ は，その瞬間のユーザの興味を表現したものと考えられる．たとえば，図に示すように概念2よりも概念1に重みがある情報を参照した場合には，概念1により興味があると考えられる．このようにユーザの興味が参照した情報の集合から方向づけられるものとすると，参照した情報の集合の各概念の重みにユーザベクトルの重みを近づけることで，ユーザの興味を把握することができる．そこで，ユーザベクトル $\vec{U}$ の方向を，参照した情報の情報ベクトル $\vec{G}$ の方向へ回転させ各概念の重みを更新し，$\overrightarrow{U_{new}}$ とする．ここで，n は $0 < n < 1$ の範囲の値を持つ定数で，回転させる度合いを調整するために導入する．ユーザベクトルを更新する計算式を式(1)に示す．ここで，$rad(\vec{U}, \vec{G})$ は $\vec{U}$ から $\vec{G}$ 方向への時計回りの角度 θ を出力する関数で，$rotate(\alpha, \vec{U})$ はベクトル $\vec{U}$ を時計回りに角度 α だけ回転させる関数である．

$$\overrightarrow{U_{new}} = rotate(n \times rad(\vec{U}, \vec{G}), \vec{U}) \qquad (1)$$

この方式では，ユーザに興味を登録するといった認知的負荷を要求することなく，ユーザの参照行動から

ユーザの興味を把握することが可能である．たとえば，参照した情報の情報ベクトルで概念1に高い重みが設定されている場合には，ユーザベクトルの概念1の重みも増加する．逆に，概念2の重みが低い場合には，ユーザベクトルの概念2の重みも減少する．このように，情報を参照するたびにユーザベクトルの各概念の重みが増減し，参照した情報ベクトルの各概念の重みの累積に近似した重みが設定される．さらに，ユーザの興味が変化して参照する情報の傾向が変化した場合には，ユーザ自身による興味の再調整やサーバ側で複雑な計算を行うことなく，ユーザの興味の変化に自動的に追従することが可能である．たとえば，概念1から概念2に興味が変化した場合，参照する情報の情報ベクトルの概念2の重みに高い重みが設定されている割合が高くなる．したがって，ユーザベクトルの概念2の重みも高くなり，興味の変化に追従した重みづけが行われる．

ここで，n の値を調整することで，興味の変化に追従する度合いを調整することができる．興味の持続が長期に及ぶような情報を扱う場合には，n の値を小さくしてユーザベクトルの変化を少なくすることができる．逆に，流行の話題を提供するニュース記事のような情報の場合は，n の値を大きくする．これにより，直前に参照した情報ベクトルの影響度が大きくなり，急な興味の変化に追従することができる．

3.4 情報の特性の更新

本方式では，SF 機構の考え方に基づいて，ユーザが情報を参照する行動から各情報の特性を更新する．2章で述べたように，どのような興味を有するユーザにその情報が適しているか示すものとして情報の特性を扱うため，情報の特性の更新にはその情報を参照するユーザ群の興味の傾向を用いる．ある情報を参照するユーザ群の興味の傾向は，全ユーザが有する興味からどれだけずれがあるかで傾向づけられる．そこで，全ユーザの平均的な興味と，ある情報を参照するユーザ群の興味との差分をとり，ユーザ群の興味の傾向を求める．これを情報の特性に反映させることで情報の特性の更新を行う．以下図4を用いて説明する．

図中で，平均的なユーザの興味を，平均ベクトル $\vec{M}$ とする．平均ベクトル $\vec{M}$ の定義を，式(2)に示す．ここで，$\vec{U_i}$ はi番目に情報を参照したユーザのユーザベクトルで，l は情報を参照した総参照回数である．ここで求めた平均ベクトルは，次にユーザの総参照回数が l 回を超えるまで有効とする．本来，ある瞬間の平均ベクトルは，その瞬間にサービスを利用している全ユーザのユーザベクトルから求めることが理想である．しかし，本システムにおいては，ユーザベクトルがユーザ側でのみ管理されている場合もあり，全ユーザのユーザベクトルを瞬時に収集することは難しい．そこで，ある一定時間もしくは一定参照回数で区切り，この期間にサービスを利用したユーザのユーザベクトルの平均を，その瞬間にサービスを利用している全ユーザの平均の近似値としている．このような区切りをつけずに，過去に蓄積された全ユーザベクトルから平均を求めた場合は，現在サービスを利用している総ユーザの平均から大きく異なる可能性がある．

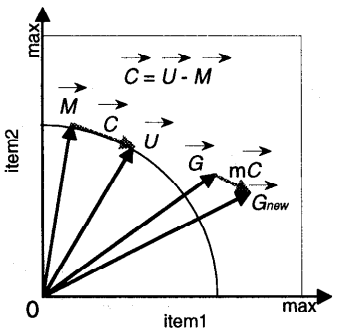


図4 情報の特性の更新
Fig. 4 Renewal of information properties.

$$\vec{M} = \frac{\sum_{i=1}^{l} \vec{U_i}}{l} \tag{2}$$

ある情報を参照したユーザの興味の傾向は，平均ベクトル $\vec{M}$ とこのユーザのユーザベクトル $\vec{U}$ との差であり，これを特徴ベクトル $\vec{C}$ とする．これを式(3)に示す．

$$\vec{C} = \vec{U} - \vec{M} \tag{3}$$

たとえば，図4に示すように，平均的なユーザは概念1にほとんど興味がなく，あるユーザが比較的強い興味がある場合，このユーザは概念1に興味があるユーザと傾向づけされる．そして，このユーザに参照された情報は，概念1に興味があるユーザに適した情報であると考えられる．そこで，このユーザの特徴ベクトル $\vec{C}$ を，情報ベクトル $\vec{G}$ に加算し，概念1の重みが増加するように情報ベクトル $\vec{G}$ を更新する．この計算式を，式(4)に示す．ここで，m は $0 < m$ の範囲の値をとる定数で，特徴ベクトル $\vec{C}$ を加算する度合いを調整するために導入する．

$$\overrightarrow{Gnew} = \vec{G} + m\vec{C} \tag{4}$$

この方式では，情報ベクトルは方向だけでなく長さも変化する．情報ベクトルの方向は，情報をどのような興味を持つユーザ群に適したものであるかという方

向性を示すものである．一方，情報ベクトルの長さは，複数の情報の中でその情報がどのような人気，重要度といった価値を持つものであるか示すものである．たとえば，図4に示すように，ある情報が概念1に興味があるユーザ群に高い頻度で参照された場合，この情報の概念1の重みは参照される回数に比例して増加する．したがって，複数の情報の中で概念1に最も高い重みが付加された情報は，概念1に興味があるユーザ群が最も参照した価値のある情報であることを示している．逆に，概念2に興味がないユーザ群に高い頻度で参照された場合，この情報の概念2の重みは参照される回数に比例して減少する．したがって，この情報は概念2に興味があるユーザ群が参照することが少ない，すなわち，概念2に興味があるユーザ群には価値がない情報であることを示している．

このような方式を用いることで，ある情報を紹介すべきユーザ群が変化した場合でも，自動的に最適となるユーザ群に紹介することが可能である．たとえば，概念1に興味があるユーザ群に高い頻度で参照されていた情報が，概念2に興味があるユーザ群に高い頻度で参照されるように変化した場合，この情報は概念2に興味があるユーザ群に紹介すべき情報に変化したといえる．この場合，自動的に情報ベクトルの概念2の重みが高くなり，この情報は概念2に興味があるユーザ群に紹介されるようになる．

3.5 情報の選択

本方式では，情報の選択の際に情報の傾向（方向性）だけでなく価値（情報量）についても考慮し，より価値のある情報を優先的に選択する．情報の選択は，CF機構の考え方に基づいて，ユーザベクトルと各情報の情報ベクトルの類似度を計算し，類似度の高い情報を選択することで行う．以下，図**5**を用いて説明する．

式(5)により，ユーザベクトル $\vec{U}$ に対する各情報の情報ベクトル $\vec{G}_i$ の写像ベクトル $\vec{G}'_i$ の長さを求め，写像ベクトル $\vec{G}'_i$ の長さが長いものから順に選択する．情報ベクトルの方向が同じ場合には，長さが長いものが選択され，情報ベクトルの長さが同じ場合には，方向が近いものが選択される．ここで，$\vec{U}\cdot\vec{G}_i$ は，$w_1 \times w'_1 + \ldots + w_N \times w'_N$ である．

$$|\vec{G'_i}| = |\vec{G_i}|cos(\theta_i) = \frac{\vec{U}\cdot\vec{G_i}}{|\vec{U}|} \qquad (5)$$

この方式では，ベクトルの方向だけでなく長さも考慮しているため，情報の持つ価値に基づいた情報選択を行うことができる．

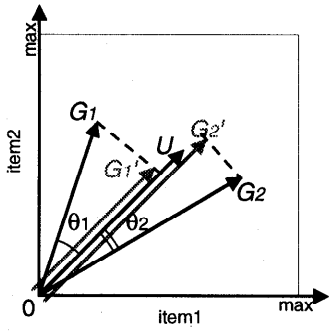


図5 情報の価値を考慮した情報選択
Fig. 5 Matching with consideration of information values.

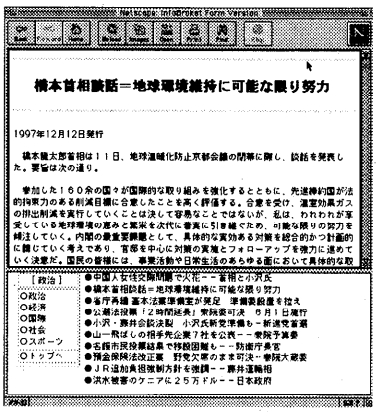


図6 実験画面
Fig. 6 A display of InfoBroket trial service.

4. InfoBroket方式の評価実験の結果

本方式をニュース記事を提供するWWWサイトに適用した評価結果について報告する．今回の評価は，2つの異なる種類のニュースを提供するサイトで行った．1つは，一般的な政治・経済・社会などの新聞と同様のニュースを提供するサイト（DailyClick：毎日新聞社）で，もう1つは，インターネットのサービスや技術のみの専門ニュースを提供するサイト（InternetWatch：インプレス社）である．

4.1 実験方法

評価実験における，ユーザに対するニュース記事の提供画面の例を図**6**に示す．ユーザから見れば，1）ニュースサイトのトップページにアクセスした場合，WWWブラウザの画面下部に記事のヘッドラインが固定で10件表示される．画面上部にはなにも表示されない．このヘッドラインは日替わりで更新される．2）WWWブラウザの画面下部に表示された記事のヘッドラインをクリックした場合，画面上部に該当する記

事の本文が表示される．画面下部は不変である．

今回の実験では，本方式で推測したユーザの興味と，実際のユーザの行動がどれだけ一致していたかを，ユーザの参照行動から観察する．したがって，ユーザが興味意外の要因で記事を参照することは望ましくない．このため，提供される記事は，すべてのユーザに対して共通であり，さらに，興味にあわせた記事の並べ替えや，明示的に印をつけるといった行為はいっさい行っていない．

実験条件として，DailyClick での実験では，記事参照回数が20回以上のユーザが10人，提供した記事の総数が850記事，ユーザの総参照回数は1351回である．InternetWatch の実験では，記事参照回数30回以上のユーザが86人，提供した記事の総数が4816記事，ユーザの総参照回数は6314回である．

本評価実験における実装は，1）WWW サーバ，2）InfoBroket システム，3）初期値設定機能，4）ログ収集機能の4者から構成されている．以下，処理の流れに従って，各機能での処理の詳細を示す．

(1) 情報ベクトルの初期値の設定
初期値設定機能を用いて，提供する情報の情報ベクトルに初期値を設定する．InternetWatch の場合は，総数256の概念に手作業で重みをつけた．重みは0から49の50段階の整数値で設定可能であるが，初期値としては，0か49かの二値的な値で設定した．DailyClick の場合は，3章で述べた方法により，総数256の概念に半自動で重みづけを行い，手作業で微調整した．この処理は，ほぼ1日朝1回，新着記事が到着するたびにまとめて行う．また，ユーザの参照によって更新された情報ベクトルの妥当性を評価するため，この初期値を設定したフィールド以外にも，ユーザベクトルを反映させる専用のフィールドを用意した．このフィールドの情報ベクトルは，すべての概念の重みを0とした初期値を持ち，純粋にユーザの参照行動によってのみ重みが付加される．また，この評価用のベクトルは，後で述べる記事の順位づけには用いない．

(2) トップページ表示
WWW サーバは，ユーザがトップページにアクセスした場合，記事のヘッドラインを10件表示する．この際，WWW サーバはユーザからユーザ ID とユーザベクトルを Cookie で取得し，InfoBroket システムに渡す．InfoBroket システムでは記事の順位づけを行う．

(3) 記事順位づけ
InfoBroket システムは，WWW サーバから，ユーザ ID とユーザベクトルを受け取る．ユーザベクトルと，提供する10件の記事の情報ベクトルとを照合し，記事に対して1番から10番までの，ユーザの興味に従った順位づけを行う．ユーザ ID，10件の記事 ID および記事の順位を，ログ収集機能に渡す．

(4) 記事本文表示
WWW サーバは，ユーザが記事のヘッドラインをクリックした場合，該当する記事の本文を表示する．この際，WWW サーバはユーザ側からユーザ ID，参照した記事 ID，ユーザベクトルを取得する．これらを InfoBroket システムに渡し，ユーザベクトル，情報ベクトルの更新を行い，更新されたユーザベクトルを記事本文とあわせて返信する．

(5) ユーザベクトル，情報ベクトル更新
InfoBroket システムは，ユーザが記事を参照した際，WWW サーバを介して取得したユーザベクトル，および，参照された情報ベクトルを更新する．同時に，評価用の情報ベクトルにユーザベクトルを反映させる．更新したユーザベクトルは，WWW サーバを介して，記事本文表示の際に Cookie によってユーザ側へ送り返す．両ベクトルの更新後，上記記事順位づけを再度行い，各記事の順位を更新する．また，ユーザ ID と，ユーザが参照した記事 ID をログ収集機能に渡す．

(6) ログ収集
ログ収集機能では，ユーザがトップページにアクセスした際，InfoBroket システムから，ユーザ ID，10件の記事 ID および記事の順位を受け取り，ログとして蓄積する．また，ユーザが記事の本文を参照した際に，ユーザ ID，記事 ID を InfoBroket システムより受け取り，ログとして蓄積する．これにより，どのユーザが，何番目の興味順位の記事を読んだか，ログから解析できる．

4.2 適 合 率

本節では，InfoBroket 方式で推測したユーザの興味と，実際のユーザの参照行動がどれだけ一致しかた評価する．4.1節で説明したように，画面下部に記事のヘッドラインを10件表示する際に，システム側で推測した興味に従って1から10までの興味順位を記事につけておく．この興味順位はユーザには知らされな

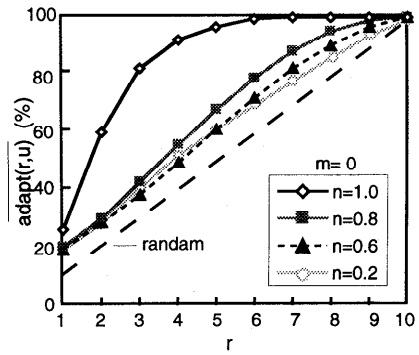


図7 適合率（DailyClick）
Fig. 7 Adaptation ratio on the DailyClick site.

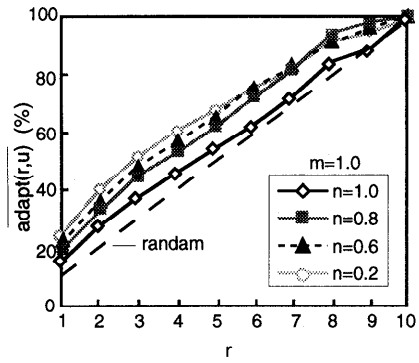


図8 適合率（InternetWatch）
Fig. 8 Adaptation ratio on the InternetWatch site.

い．ユーザ u が興味順位1番から r 番目までの記事を参照した回数の累計を総参照回数で除したものを適合率とし，式(6)に示すように $adapt(r,u)$ で表す．

$$adapt(r,u) = \frac{\text{興味順位1番から } r \text{ 番目の記事参照回数}}{\text{総参照回数}} \tag{6}$$

図7と図8にそれぞれ，DailyClick，InternetWatchのニュース記事を用いた実験における適合率を示す．ここでの適合率は，それぞれのユーザの適合率の平均をとったものである．また，情報の特性を反映させる度合いを調整する係数 n を，$0.2 < n < 1.0$ で変化させた．

DailyClickのニュース記事を用いた実験では，適合率は $n = 1.0$ の場合に最良の結果となる．すなわち，直前に参照した記事の特性を大きく反映させた方が，現在のユーザの興味を正しく表現できることになる．これは，ユーザがある記事を参照して次の記事を参照する際に，直前に参照した記事と関連がある記事を参照することを示している．ここでの，関連する記事を次々と参照する傾向は，提供する記事にその要因があると思われる．ここで提供したニュース記事は一般の新聞と同様の記事のため，比較的短い間隔で流行の話題が連続的に提供される．このため，ユーザは自然に短期的な興味にあわせて，短い間隔で流行の記事を参照するようになると考えられる．

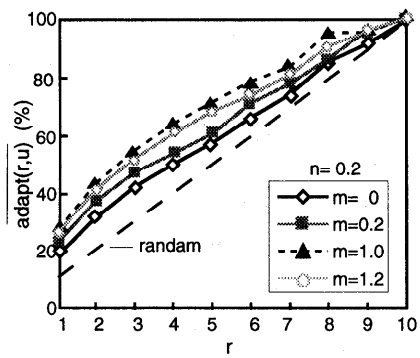


図9 適合率：m を変化させた場合の比較（InternetWatch）
Fig. 9 Adaptation ratio on the InternetWatch site: based on the parameter m.

また，InternetWatchのニュース記事を用いた実験では，適合率は $n = 0.2$ の場合に最良となり，DailyClickの場合と相反する結果となる．ここで提供したニュース記事はインターネット，コンピュータ関連の記事で，短い間隔で流行の話題が連続的に提供されることが少ない．このため，ユーザは比較的長期的な興味にあわせて記事を参照すると考えられる．

これら2つの実験から，ユーザの興味変化への追従度合いを調整することで，興味順位の上位3番目までで55～80%の割合でユーザの興味を把握できた．さらに，ユーザの長期的な興味・短期的な興味に対応した記事の選択が可能であることが確認できた．

次に，記事の特性にユーザの興味の傾向を反映させる度合いを変更した場合について，適合率の変化を確認する．図9に，InternetWatchのニュース記事を用いた実験での適合率を示す．ここで，情報ベクトルにユーザベクトルを反映させる度合いを調整する係数 m を，$0 < m < 1.2$ で変化させた．

この結果，適合率は $m = 1.0$ で最良の結果となり，記事にユーザの傾向をまったく反映させない場合（$m = 0$ の場合）に比べて適合率が向上している．これより，ユーザの興味を用いた記事の特性づけ，記事の価値を考慮した選択方式が有効であるといえる．

4.3 情報の特性の正確さ

図10に，付加された重みが正確なものであるかどうかアンケートによる調査を行った結果，およびユーザの参照行動から記事に付加された情報ベクトルの例

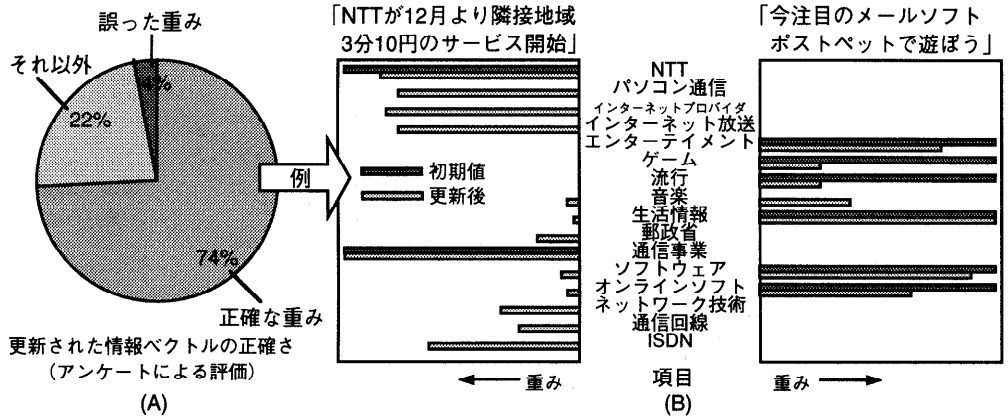


図10 更新された特性の正確さとその例
Fig. 10 Correct ratio of information properties and a sample.

を示す．ここでは，ユーザの参照から付加された重みが正確なものであるか評価する．このため，4.1 節で述べた，各概念の重みをすべて 0 として初期値を設定した評価用の情報ベクトルを用いる．すなわち，この評価用の情報ベクトルの重みは純粋にユーザの参照によってのみ付加されたものである．アンケートの被験者は 6 人で，対象とした記事は 29 記事である．この 29 記事は，ユーザから 30 回以上参照された記事である．被験者は，高い重みが想定される概念に実際に重みが付加されていたか，低い重みが想定される概念に実際に重みが付加されていないか，を評価基準とし，記事の情報ベクトルが妥当であるか否か評価した(図 10 の A)．本稿では，全被験者より妥当なものと判断されたものを正確に重みが付加されている記事とした．これより，74%の記事に正確な重みが付加されているという結果を得た．完全に誤った重みが付加されていた割合は 4%である．残りの 22%の記事は，重みが想定される概念に実際に重みが付加されていたが，それ以外の概念にも一部重みが付加され，妥当であるか否か被験者にも判別がつかなかったものである．

ユーザの参照によって更新された情報ベクトルの例(図 10 の B) より，情報提供者が初期値として想定した重み以外にも重みの付加される概念があることが分かる．また，重みが付加される概念も，ある程度意味的にまとまって付加される傾向がある．

5. ま と め

本稿では，ユーザの情報参照行動に着目して，ユーザの興味の把握，情報の特性づけを行う InfoBroket 方式を提案した．また，ニュース記事を提供する WWW サイトに適用し評価を行い，以下の有効性を確認した．

- 短期的，長期的な興味に対応可能
  参照した情報の特性を用いることで，ユーザに興味の登録を要求することなく，ユーザの興味を把握することが可能である．また情報の特性をユーザの興味に反映させる度合いを調整することで，ユーザの短期的，長期的な興味に対応することが可能である．
- 自動的な情報の特性づけ，および修正が可能
  情報を参照したユーザの興味の傾向を反映させる方式により，ユーザに評価を要求することなく Social Filtering による情報の特性づけが可能である．
- 情報の価値を考慮した情報の特性づけ，情報選択が有効
  情報が参照された回数を情報の価値として情報の特性に含め，情報の価値を考慮した情報選択を行うことで，よりユーザの興味にあった情報を選択することが可能である．

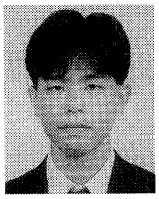
今後は，このようなパーソナライズシステムが，複数の情報提供者ごとに分散化されて存在していた場合の，システムの連携についても検討する．また，情報提供者が任意に設定していた「概念」，および，情報の特性をユーザの興味に反映させる度合いを，個々のユーザに対して最適化できるように拡張する予定である．

謝辞 本方式の評価のために，共同実験をさせていただいた InternetWatch の管理元である株式会社インプレスの皆様，DailyClick の情報を提供していただいた毎日新聞社，さらに実験に参加していただいた多くの皆様に感謝します．

参考文献

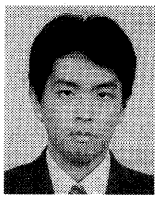
1) MY Directory: http://myd.nttnavi.co.jp.
2) Joachims, T. and Freitag, D.: WebWatcher: A Tour Guide for the World Wide Web, *Proc. IJCAI'97* (1997).
3) Mlandenic, D.: Personal WebWatcher: Design and implementation, Technical Report IJS-DP-7472 (1996).
4) Shardanand, U. and Maes, P.: Social Information Filtering: Algolithms for Automating "World of Mouth", *Proc. CHI-95 Conference*, pp.210–217 (1995).
5) Balabanovic, M. and Shaham, Y.: Learning Information Retrieval Agent: Experiments with Automated Web Browsing, *AAAI Spring Symposium on Information Gathering from Heterogeneous*, Distributed Environments (1995).
6) Lieberman, H.: Letizia: An agent that assists web browsing, *International Joint Conference on Artificial Intelligence*, Montreal (1995).
7) Chen, L. and Sycara, K.: WebMate: A Personal Agent for Browsing and Searching, *The 2nd International Conference on Automous Agent* (*Agent '98*) (1998).
8) Lang, K.: NewsWeeder: Learning to Filter Netnews, *ICML '95*, pp.331–339 (1995).
9) Salton, G.: Developments in Automatic Text Retrieval, *Science*, Vol.253, pp.974–980 (1991).
10) Balabanovic, M. and Shoham, Y.: Fab: Content-Based, Collaborative Recommendation, *Comm. ACM*, Vol.40, No.3, pp.66–72 (1997).

(平成10年 5 月 8 日受付)
(平成10年10月 2 日採録)



橘高 博行

平成5年慶應大学理工学部機械工学科卒業．7年同大学大学院機械工学専攻修士課程修了．同年，NTTに入社．情報検索等情報流通プラットフォームに関する研究に従事．

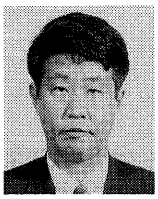


佐藤 直之

平成7年慶応大学理工学部電気工学科卒業．9年同大学大学院計算機科学専攻修士課程修了．同年，NTTに入社．OS，ネットワーク等開放型分散計算機環境に関する研究，および情報検索やコミュニティ等情報流通プラットフォームに関する研究に従事．日本ソフトウェア科学会会員．

鈴木 英明（正会員）

昭和60年早稲田大学理工学部数学卒業．同年，NTT武蔵野電気通信研究所入所．以来，知識獲得，定理証明系，リエンジニアリング，プログラムの抽象化検索，情報流通基盤技術の研究に従事．現在，NTTソフトウェア研究所主任研究員．ACM，IEEE-CS各会員．



曽根岡昭直（正会員）

昭和55年東京大学工学部電子工学科卒業．57年同大学大学院工学部修士課程修了．同年NTT（電電公社）に入社．高信頼網トポロジー，分散システムの理論の研究，通信ソフトウェア作成環境の研究・実用化，ソフトウェアアーキテクチャの推進，マルチメディアおよびECサービスの開発に従事．現在，情報流通プラットフォームInfoketの研究実用化に従事．平成元年～2年コーネル大学客員研究員．工学博士．IEEE，ACM，電子情報通信学会，TUG各会員．