

# 閲覧履歴に基づく情報検索の相互支援

根本 潤<sup>†</sup> 遠山 元道<sup>††</sup>

<sup>††</sup> 慶應義塾大学理工学部情報工学科 〒223-8522 神奈川県横浜市港北区日吉 3-14-1

E-mail: †jun@db.ics.keio.ac.jp, ††toyama@ics.keio.ac.jp

あらまし 近年、Web で発信される情報は飛躍的に増加している。一方で、ユーザが検索・閲覧することができる情報量には限界がある。そのため、本来欲しい情報であっても発見できないといった状況がしばしば起こりうる。そこで本稿では、組織などで文書の閲覧履歴を共有することでユーザの情報検索を相互支援する手法を提案し、実装および評価を行う。実装したシステムにおいては、Web の閲覧履歴に基づいて興味・関心を抽出し、それに合った情報を他のユーザが得た場合に、そのユーザの主観的な判断を仰いだうえで即座に推薦を行う。これにより、従来自分自身だけでは獲得不可能であった情報を組織内で相互補完的に提供することができるようになり、情報検索の効率が上昇すると考えられる。

キーワード Web 検索, 閲覧履歴, リコメンデーション

## Mutual Support for Information Retrieval based on Browsing History

Jun NEMOTO<sup>†</sup> and Motomichi TOYAMA<sup>††</sup>

<sup>††</sup>Department of Information and Computer Science, Faculty of Science and Technology,  
Keio University

Hiyoshi 3-14-1, Kouhoku-ku, Yokohama-shi, Kanagawa, 223-8522 Japan

E-mail: †jun@db.ics.keio.ac.jp, ††toyama@ics.keio.ac.jp

**Abstract** Recently, information on the Web is remarkably increasing. However, users may be often unable to find the documents they want because of too much information. In this paper, therefore, we propose, implement, and evaluate a mutual supporting method for information retrieval by sharing browsing history of documents in organization, workgroup and so on. By our proposal technique, user's interest and concern are extracted based on their browsing history, and when other users get the information matched for it, recommendation is committed after subjective decision of the user. As a result, our system makes mutual support for information retrieval possible and the efficiency of it will be improved.

**Key words** Web Retrieval, Browsing History, Recommendation

### 1. はじめに

近年、Web で発信される情報は飛躍的に増加している。しかしながら、その情報量の多さが時に弊害となる場合がある。いわゆる、情報過多と呼ばれる状況である。例えば、一人のユーザが閲覧することのできる文書数には限りがあり、本来必要としている文書であっても発見することができなかつたり、見逃してしまつたりという事態がしばしば起こりうる。

こうした状況を避けて必要な情報だけを効率的に得る手段として、本稿では小規模な組織やワークグループ内での相互補完的な情報検索支援系を提案する。まず、各ユーザの興味・関心を閲覧履歴から抽出しておく。そして、ユーザは何か特別なことを意識することなく、普段通りに Web の閲覧を行う。このと

き、もし他のユーザが興味のあるようなページを閲覧したら、システムは相手がまだそのページを閲覧していないことを確認した上で、それが推薦すべき候補であることを通知する。そこでユーザはそのページが相手にとって有益であるかを判断し、有益でありそうなら実際に相手に推薦を行う。

このように、従来自分自身だけでは獲得不可能であった情報を小規模な組織やワークグループ内で相互補完的に提供することで、知識の幅が広がったり、検索を行う際の効率が向上したりすると考えられる。

本稿の構成は以下の通りである。まず、2. 章では関連研究を述べる。3. 章では提案手法について、4. 章では実装したシステムの概要について説明する。そして 5. 章で実験と評価を行った上で、最後に 6. 章でまとめを述べる。

## 2. 関連研究

### 2.1 情報フィルタリング

大量の情報からユーザの要求に見合ったものだけを得るための手段として、情報フィルタリングがある。

酒井らは、新聞社などから日々提供される電子記事から個々のユーザの興味に合ったものを選出し、メールなどで配信する情報フィルタリングシステム NEAT [1] を開発している。

また、情報フィルタリングを用いたグループウェアとしては、Malone ら [2] の Informaion Lens がある。これは、グループ内での情報共有を目的として、半構造化したメッセージを利用することで、メッセージの優先度付けや自動分類を行うシステムである。メッセージの処理規則は容易なインターフェースで定義できる。

これらの情報フィルタリングシステムは、本研究と同様に、各ユーザの興味に合ったコンテンツだけを提供することを目的としているが、フィルタリングという性格上、対象がメールであったり、配信されてくるニュース記事であったりと、いずれも情報のストリームを対象としている。したがって、ユーザが自由に閲覧する Web ページを対象として、興味・関心に合致するものを得ようとする本研究とは異なる。

### 2.2 協調フィルタリング

協調フィルタリングを用いた推薦システムとしては、Resnick ら [3] の GroupLens がある。これは、当初はネットニュースの記事に対して与えられた点数に基づいて読者を分類し、個人の嗜好を予測するというものであったが、後に Net Perception という名称で商用化されて Amazon.com [4] をはじめ多くのサイトが採用している。

協調フィルタリングの問題点は、フィルタリングのための基準としてはユーザの評価や主観などが反映されているが、実際に推薦する項目に関してはシステムが自動的に選択してしまうことである。つまり、推薦の精度は協調フィルタリングのアルゴリズムに大きく依存していて、ユーザの意図しないものが推薦されてしまうということもありうる。そこで、本提案手法のようにあえて最終的な推薦の決定を同じ組織やワークグループのユーザに行わせることで、そういった事態を減らすことができると考えられる。

### 2.3 閲覧支援

閲覧支援に関する研究としては品川らの研究がある [5]。これはユーザが入力したキーワード群と閾値をユーザプロフィールとして、それをもとに閲覧ページをユーザの視点に沿った仮想ページに再構成して提供するものである。閲覧履歴を利用していない点で異なっているが、このような手法を本提案手法と組み合わせることは有効であると考えられる。

### 2.4 閲覧履歴

閲覧履歴を利用した研究も広く行われている。中島らは閲覧履歴を反映したコンテキスト依存型 Web ブックマークを提案している [6]。これはブックマークを単なるユーザの興味の反映としてとらえるのではなく、ブックマークするまでのプロセスにもユーザの意図やブックマーク行為の価値があるという考え

表 1 TF・IDF 法による履歴ページの特徴抽出例

	$w_1$	$w_2$	$w_3$	$w_4$	$w_5$	...
$H(1)$	2.5	0	0	0	0	1.8
$H(2)$	0	0.6	1.6	0	0	0
$H(3)$	0	0	0	1.2	0.2	0.1
$H(4)$	0	0	0.6	0	3.3	0.2

に基づいている。ユーザが必要な情報を検索エンジンを利用して検索する際に、検索結果として大量の Web ページがヒットすることがあるが、このときこのブックマークを問い合わせとして用いることで、全てのページをチェックすることなく、要求に合致したページをピックアップすることができる。未閲覧のページから有益なページを得ることが可能な点で本研究と類似する点もあるが、自由な閲覧行為を対象としていない点で本研究と異なる。

## 3. 情報検索の相互支援

本章では提案する相互補完的な情報検索支援系について説明する。ここでは簡略化のために、推薦を行うユーザと推薦されるユーザの二人だけが検索を行う状況を想定する。なお、前者の立場のユーザを、相手ユーザに助言を行うという意味でアドバイザーと呼ぶことにする。同様に後者の立場のユーザをレシーバと呼ぶ。ただし、アドバイザーとレシーバはある意味対等で、ユーザは状況に応じて双方の立場になりうる。

### 3.1 閲覧履歴の収集と特徴抽出

推薦の基準となる各ユーザの興味・関心の抽出手法について述べる。まず、各ユーザは特別なことを意識することなく、Web の検索・閲覧を行う。そして、閲覧されたページについてそれらの特徴を抽出する。特徴抽出の方法としては、TF・IDF 法 [7] を採用する。

ここで履歴ページ  $H(i)$  におけるキーワード  $t_j$  の TF・IDF 値を  $w_{ij}$  とすると、TF・IDF 値の算出式は以下ようになる。

$$w_{ij} = tf_{ij} \cdot \log \frac{n}{df_j}$$

ただし、 $tf_{ij}$  は履歴ページ  $H(i)$  におけるキーワード  $t_j$  の出現回数、 $df_j$  はキーワード  $t_j$  が出現する履歴ページ数、 $n$  は全履歴ページ数である。

TF・IDF 法を用いて、各履歴ページの特徴抽出を行った場合の処理の例を表 1 に示す。

$H(1) \sim H(4)$  は、履歴ページであり、 $w_1 \sim w_5$  は、履歴ページ内のキーワードである。表中の数値は、各履歴ページにおける各キーワードの TF・IDF 値である。これらの TF・IDF 値を要素として、各履歴ページをベクトルで表現すると次のようになる。

$$\vec{h}_i = (w_{i1}, w_{i2}, \dots, w_{ij})$$

### 3.2 ユーザの興味・関心

ユーザ  $k$  の興味ベクトル  $I(k)$  を以下のように定める。この式は様々な形が考えられるが、今回は単純に履歴ページの特徴ベクトルの和とした。なお、ここで  $H_k$  はユーザ  $k$  の閲覧した

ページ集合である。

$$I(k) = \frac{\sum_{i \in H_k} h_i}{|I(k)|}$$

興味ベクトルのその他の構成方法としては、閲覧時刻を用いて、より現在に近い時刻の履歴ページの特徴ベクトルに重みを付けて足し合わせることで、時間的な要素を取り入れた興味ベクトルを生成する方法がある。また、ユーザの負担は増えてしまうが、あらかじめユーザが指定したキーワードに対して重み付けをすることも考えられる。これらの手法に関しては、議論の余地があると考えられる。

### 3.3 リコメンデーション

レシーバが推薦を受けるまでには大きく二つの段階がある。まず、アドバイザーが Web の閲覧を行っているものとする。その際に、レシーバの興味に合致するページを閲覧した場合に、最初の段階としてそれが推薦候補であることが通知される。そこで次の段階として、アドバイザーの判断によりレシーバに実際に推薦を行う。

推薦候補を選択する際の基準としては、レシーバの興味ベクトルとアドバイザーの閲覧ページの特徴ベクトルとの類似度を用いる。この類似度は以下のようにコサイン相関値によって求める。

$$\text{Similarity}(\vec{i}_R, \vec{h}_A) = \frac{\vec{i}_R \cdot \vec{h}_A}{|\vec{i}_R| |\vec{h}_A|}$$

ここで  $\vec{i}_R$  はレシーバの興味ベクトル、 $\vec{h}_A$  はアドバイザーの閲覧ページの特徴ベクトルを表す。

$\text{Similarity}(\vec{i}_R, \vec{h}_A)$  が閾値  $\theta$  を超えた場合、そのページはレシーバへの推薦候補となり、アドバイザーへ通知される。ここで、アドバイザーとレシーバは同じワークグループに存在することから、アドバイザーは推薦候補がレシーバにとって有益かどうかをある程度判断できると仮定される。この仮定に基づいて、アドバイザーの主観的な判断により、実際にレシーバに推薦を行うものとする。

これまで述べてきた手法を取り入れて Web 検索を行うことにより、ユーザは自分が閲覧しそびれていたページや、発見できなかったページでかつ有益なもののみを知ることが可能になると考えられる。

### 3.4 相互支援による効果

図 1 は、本提案手法により得られると思われるページの集合を示すものである。左側の塗りつぶされている大きな円が、レシーバが興味・関心のあるページ集合であり、その中の網掛けの小さい円がレシーバが実際に閲覧したページ集合である。これは、情報過多の状況においては興味のある文書であっても全て閲覧することができないという状態を表している。左側と同様に、右側の大きな円、小さな円はそれぞれアドバイザーの興味・関心のあるページ集合、閲覧したページ集合を示して。なお、左右のレシーバ、アドバイザーはあくまで役割上の区別であり、状況に応じてその立場が入れ替わることに注意されたい。本研究で注目するのは、濃く塗りつぶされた部分である。こ

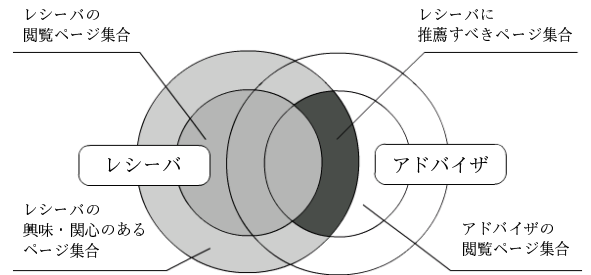


図 1 相互補完的な情報検索で得られるページ

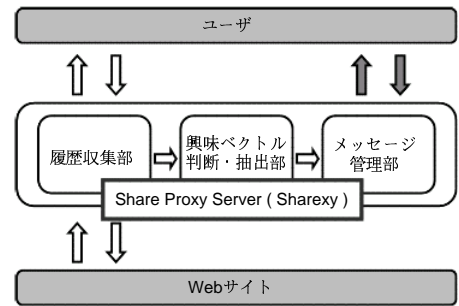


図 2 システム概要

の部分レシーバが興味・関心はあるけれど、まだ閲覧されていないページで、かつアドバイザーの閲覧したページである。つまり、この部分がレシーバに対して推薦すべきページ集合ということになる。この概念は複数のユーザについても適用することができ、濃く塗りつぶされた部分だけを適切に推薦すれば、提案する小規模な組織内における相互補完的な情報検索が実現可能であると考えられる。

## 4. システム概要

3. 章における提案手法に基づき、閲覧履歴を利用した Web 検索の相互支援システムのプロトタイプを実装した。図 2 は、その概要図である。

プロトタイプシステムは、プロキシサーバとして実装されている。また、主として履歴を収集する部分、ユーザの興味・関心を抽出する部分、アドバイザーやレシーバとメッセージのやりとりをする部分から構成される。以下で、各部分について順に説明していく。

### 4.1 履歴収集部

履歴収集部ではプロキシサーバを通して閲覧されたページを各ユーザごとに保存していく。なお、保存するページはテキスト形式で記述されたものに限定する。また、この履歴収集部でユーザが過去にどういったページを閲覧したかという一貫的な情報の管理も行う。

### 4.2 興味ベクトル判断・抽出部

興味ベクトル抽出部では履歴収集部で得られた各ユーザの閲覧履歴からユーザの興味ベクトルを生成する。具体的には、形態素解析システムの茶筌 [8] によって名詞に分類された単語をキーワードとする。このキーワードを用いて 3. 章で述べた手法を適用し履歴文書の特徴ベクトルおよびユーザの興味ベクトルを生成する。

また、ここではアドバイザーが閲覧しているページの特徴ベクトルを生成し、レシーバの興味ベクトルとの類似度を求める役割も担っている。ここでの処理概要を説明すると、まずアドバイザーが閲覧しているページは常に形態素解析され、キーワードが抽出される。そして、履歴ページの特徴ベクトルを生成する際に求めた DF 値を用いて  $TF \cdot IDF$  値を計算し、閲覧ページの特徴ベクトルとする。このベクトルとレシーバの興味ベクトルとの類似度を求め、閾値を超えた場合はメッセージ管理部へ次の処理を依頼する。

#### 4.3 メッセージ管理部

メッセージ管理部にはアドバイザーがレシーバがシステム側と推薦や推薦候補の通知メッセージなどをやりとりするための機能が実装されている。メッセージのやりとりにはマイクロソフト社の MSN Messenger [9] [10] を使用する。MSN Messenger は、インスタントメッセージのやりとりを行うソフトウェアで、オンライン中ならオンラインの相手に即座にメッセージを送ることができる。この機能を利用することにより、オンライン中のユーザに関しては即座に推薦を行うことが可能となり、検索の相互支援をより活発なものにする。なお、ここではソフトウェアの普及度などを考慮した上で MSN Messenger を選択したが、上記のような機能を満たすインスタントメッセンジャであれば、本システムを組み込むのは容易である。

メッセージ管理部で扱われるメッセージは主として二つある。

一つはアドバイザーが閲覧しているページの特徴ベクトルとレシーバの興味ベクトルの類似度が閾値  $\theta$  を超えたときに、アドバイザーに送信される推薦候補のメッセージである。これは、『ページ はユーザ  $x$  の興味を反映しています。推薦しますか?』といった内容である。

もう一つはアドバイザーが推薦候補の通知を受けて、対象レシーバへの推薦を承諾したときに送信される、実際の推薦メッセージである。これは、『ユーザ がページ を推薦しています。』といった内容である。

また、レシーバがオフラインでメッセージを送信できない場合の対処として、同様の内容をメールで通知する機能も実装されている。なお、アドバイザーとなるユーザはオンラインであることを仮定している。

#### 4.4 利用例

本システムを利用するにあたっては、まず MSN Messenger の利用を開始しておく。続いてシステムにログインする。これは本システムを通して最初にページを閲覧する際に、認証ダイアログが現れるのでそこでログインする。

ログイン後は、通常通りのブラウジングを行う。すると、あるとき図 3 のようにシステムからメッセージを受け取る。これは閲覧中のページがレシーバへの推薦候補であることを示している。ここでシステム側に YES か NO で返事することで、実際に推薦を行うことができる。図 4 は、アドバイザーが YES と返答し、レシーバもオンラインだったため推薦が成功した様子を示している。

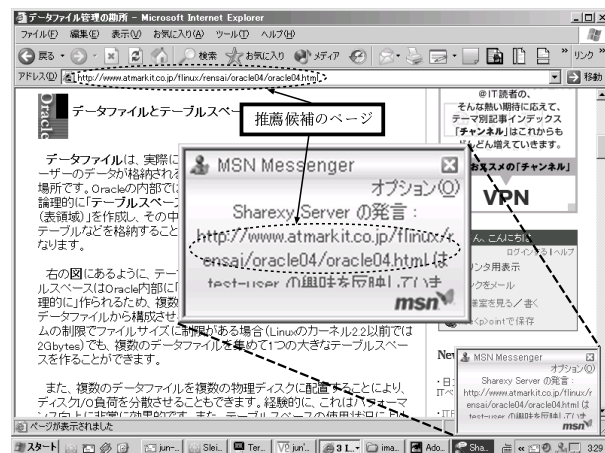


図 3 推薦候補を通知された状態

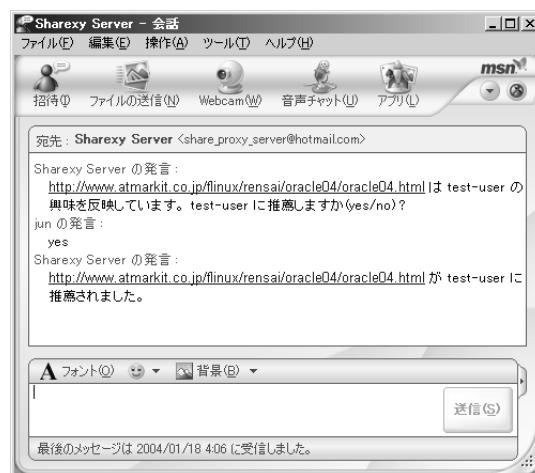


図 4 推薦メッセージの送信に成功した様子

## 5. 実験と評価

評価を行うにあたって、以下の点に注目して実験を行う。

- (1) システムのみに依存した場合の推薦精度
- (2) アドバイザの補助がある場合の推薦精度
- (3) ワークグループに属さないユーザがアドバイザーとなる場合の推薦精度

まず一つ目であるが、これはアドバイザーによる支援を活用せずに、レシーバの興味・関心に従って、システムのみに依存して推薦を行う場合、どれだけ正確に推薦を行えるかというものである。

そして二つ目は、アドバイザーが本システムを使用して主観的な判断をくわえてから推薦を行う場合、それがどれだけ正確かどうかというものである。

最後の三つ目は、ワークグループに属さないユーザが推薦行為にたずさわった場合、ワークグループ内のアドバイザーと比べてどれだけ推薦精度が異なるかという観点である。

### 5.1 実験

#### 5.1.1 対象

実験は、同じ研究室に所属する 5 名を対象として行った。そのうち、1 名にはアドバイザーの役割を、残りの 4 名にはレシー

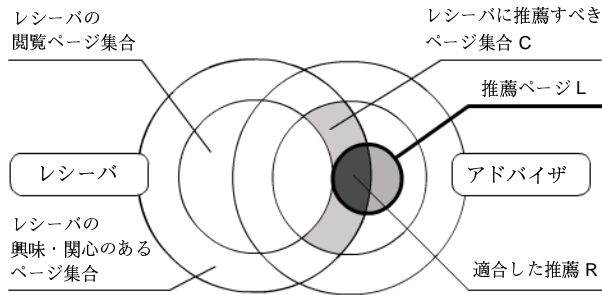


図 5 推薦ページと正解ページの関係

バの役割を担当してもらった。

また、閲覧してもらう Web ページはあらかじめ指定された範囲に限定した。本研究は、ユーザの自由な Web ブラウジングを仮定しているが、閲覧範囲を限定することで簡易的に評価を行った。

閲覧対象として@ IT - アットマーク・アイティ [11] におけるデータベースチャンネルの記事約 150 件を利用した。@ IT は情報技術関連の記事が多数紹介されているサイトである。データベースチャンネルは、それらの多数の記事のうち、データベース関連のものをまとめ、再編成したページである。

@ IT におけるデータベースチャンネルの記事を対象としたのは、これらの記事のはじめには概要が述べられており、ユーザはそれを見ることでその記事に興味・関心があるかどうかの判断を行いやすいためである。また、データベースは被験者の興味・関心のなかで比較的共通したトピックであることも理由の一つである。

### 5.1.2 手順

実験の手順は以下の通りである。

(1) レシーバを対象に約 150 件の記事の中から、興味・関心のあるものを選択してもらう。なお、全ての記事の全文を閲覧してから選ぶというのではなく、記事のタイトルを参考に短時間で選んでもらう。ここで選択された記事は、図 5 におけるレシーバの閲覧ページ集合に該当する。図 5 は、図 1 にシステムを用いて推薦を行ったページ  $L$  とそのうち実際にレシーバの要求に適合していたページ  $R$  の関係を加えた図である。

(2) 次に同じくレシーバを対象に、実際に記事の概要や本文を閲覧して、興味・関心のある記事を選んでもらう。ここで意図は、本当は興味があったにも関わらず (1) では見落とししていた記事や優先順位がやや低く閲覧を見送っていた記事などを知ることにある。なお、ここで選択された記事は、図 5 におけるレシーバに推薦すべきページ集合に該当し、後述する 5.2 節の評価式における全正解ページ  $C$  となる。

(3) レシーバが (1) で閲覧したページをもとに興味ベクトルを生成し、全ての記事の特徴ベクトルと比較し、類似度を求める。

(4) 類似度閾値ごとに推薦候補となるページを求める。

(5) 後述する 5.2 節の評価式に基づいて、システムのみ依存した場合の推薦精度を求める。なお、ここではシステムが推薦候補として挙げたページ全てをレシーバに推薦するものと

して、推薦精度を求める。その上で、最適と思われる閾値を決定する。

(6) 閾値を (5) で求めた値に設定した場合に得られる推薦候補ページをアドバイザーに閲覧してもらい、実際に推薦するかどうかを判断してもらう。そして、その場合の推薦精度を求め、(5) の精度と比較する。

### 5.2 評価方法

評価尺度としては、情報検索で一般的に用いられる再現率、適合率、 $F$  値を用いることとする。また、評価尺度における各パラメータは以下のように定める。

- 全正解ページ数  $C$

レシーバが未閲覧のページのうち、興味・関心をもっているページ数。5.1.2 節の (2) で選択されたページが該当する。

- 推薦ページ数  $L$

アドバイザーがレシーバに推薦したページ数。5.1.2 節の (6) において最終的に推薦されたページが該当する。ただし、システムのみ推薦精度を求める場合は推薦候補ページを推薦ページとする。

- 推薦ページ中の正解ページ数  $R$

推薦ページのうち、レシーバが興味・関心をもっていて、かつ未閲覧のページ数。

これらのパラメータを用いて、再現率、適合率、 $F$  値は次のように表せる。

$$recall = \frac{\text{推薦ページ中の正解ページ数 } R}{\text{全正解ページ数 } C}$$

$$precision = \frac{\text{推薦ページ中の正解ページ数 } R}{\text{推薦ページ数 } L}$$

$$F\text{-measure} = \frac{(1 + b^2) \times precision \times recall}{b^2 \times precision + recall}$$

言い換えると再現率は推薦漏れ  $C - R$  がどれだけ少ないかを示し、適合率は推薦ミス  $L - R$  がどれだけ少ないかを示しているといえる。

なお、 $F$  値におけるパラメータ  $b$  は、適合率に対する再現率の相対的な重みを表す。本研究ではどちらも対等の重みとし、 $b = 1$  として評価する。

### 5.3 実験結果・考察

#### 5.3.1 システムのみ依存した推薦の精度

図 6 は、5.1.2 節における手順に従って実験を行い、システムのみ依存した場合の推薦精度を求めたときの各評価パラメータの推移を示している。評価パラメータは各レシーバに対して、閾値ごとに推薦候補数、正解数を求め、その平均をとった値を用いている。

この図が示しているように、閾値が高くなるにつれ、推薦候補の数が非常に少なくなっている。また、それが原因となって必然的に推薦候補に含まれる正解の数も少なくなっている。この問題点のそもそもの原因は、履歴ページの特徴ベクトル、レシーバの興味ベクトル、アドバイザーの閲覧中のページの特徴ベクトルといった、特徴抽出関連の手法にあると思われる。

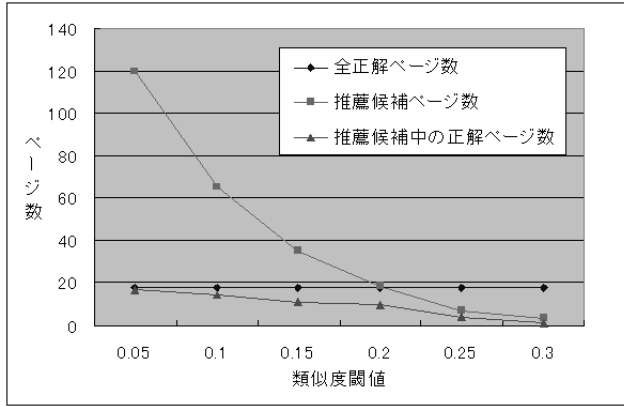


図 6 評価パラメータの推移

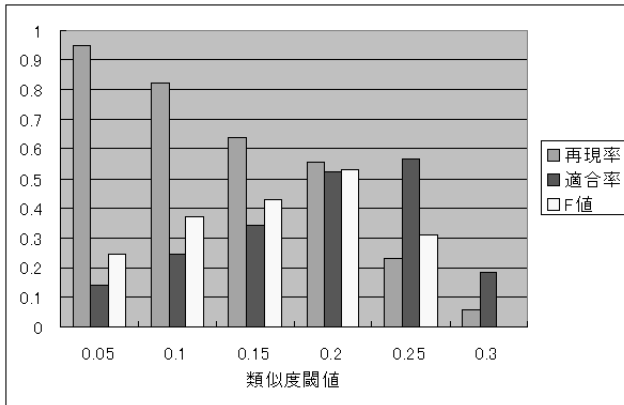


図 7 システムのみに依存した推薦の精度

図 7 は、同様の実験を行ったときの、システムのみ依存した推薦に関する再現率、適合率、F 値を示している。この図では、閾値の低いとき、再現率は高く、適合率は低い。また、閾値が高くなるにつれて再現率は低く、適合率は高くなっている。つまり、再現率と適合率がトレードオフの関係にあることが読み取れる。そこで、F 値について見てみると、F 値は閾値が 0.2 のときに最も大きな値となっている。

### 5.3.2 アドバイザが関与した推薦の精度

5.1.2 節の (6) で述べたように、アドバイザにシステムが選んだ推薦候補の中から最終的にレシーバへ推薦するページを決めてもらった。なお、このときの推薦候補選択の際の類似度閾値は 5.3.1 節の結果から、0.2 を用いた。

また、ワークグループに属さないユーザが推薦行為にたずさわった場合の推薦精度をワークグループ内のアドバイザの推薦精度と比較するために、同様の実験を研究室外の 1 名に一般ユーザとして行ってもらった。

アドバイザが推薦にたずさわった場合における各レシーバについての推薦ページ数、推薦ページ中の正解ページ数、再現率、適合率、F 値を表 2 に示す。また、同様に一般ユーザが推薦にたずさわった場合についても表 3 に示す。さらに、比較のためにシステムが推薦候補として選んだページ数、推薦候補ページ中の正解ページ数なども表 4 に示す。なお、表中の A から D は 4 人のレシーバを表している。

表 2 アドバイザが推薦にたずさわった場合の精度

	全正解数	推薦数	推薦ページ中の正解数	再現率	適合率	F 値
A	20	15	10	0.50	0.67	0.57
B	15	16	8	0.53	0.50	0.52
C	20	13	8	0.40	0.62	0.48
D	16	17	11	0.69	0.65	0.67

表 3 一般ユーザが推薦にたずさわった場合の精度

	全正解数	推薦数	推薦ページ中の正解数	再現率	適合率	F 値
A	20	14	7	0.35	0.50	0.41
B	15	17	6	0.40	0.35	0.38
C	20	15	8	0.40	0.53	0.46
D	16	19	9	0.56	0.47	0.51

表 4 システムが推薦候補を直接推薦した場合の精度

	全正解数	推薦候補数	推薦候補中の正解数	再現率	適合率	F 値
A	20	18	11	0.55	0.61	0.58
B	15	20	8	0.53	0.40	0.46
C	20	16	9	0.45	0.56	0.50
D	16	21	11	0.69	0.52	0.59

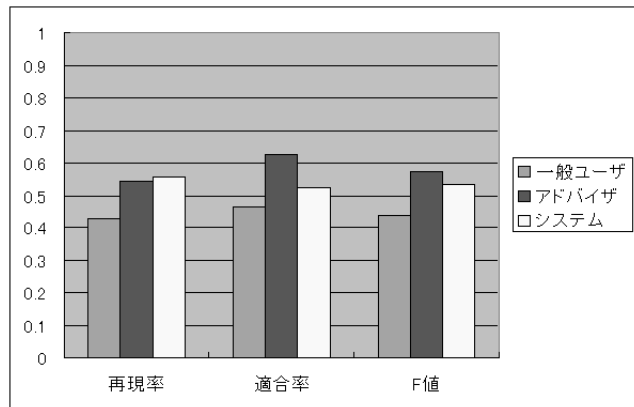


図 8 推薦精度の比較

表 2 から、アドバイザが推薦候補内の不正解ページを適切に排除していることがわかる。また、表 3 からは、ワークグループ外のユーザはレシーバへの推薦について判断材料がないために、システムの推薦候補をそのまま推薦する傾向がやや見られる。しかしながら、それでも正解ページを誤って排除してしまうケースもあり、再現率、適合率はともに低い水準となっている。

また、図 8 は、表 2・3・4 から再現率、適合率、F 値の平均値を求めて比較したグラフである。この図が示しているように、アドバイザが推薦に関与することで、再現率はほとんど下がることなく適合率が上昇している。その結果、F 値についてもシステムのみ推薦をまかせるときより上昇していることがわかる。一方で、一般ユーザについては、適合率、再現率がともに低下し、システムにのみ依存したほうが良い結果となっている。

以上の結果をまとめると、システムによる完全に自動的な推

薦よりも、ワークグループ内の興味・関心が似通っているユーザ同士で、主観的な判断を加えつつ推薦を行った方がより効果的であるといえる。

#### 5.4 今後の課題

##### 5.4.1 推薦候補決定のパフォーマンス

実験の際にはシステムを利用するユーザをアドバイザー役に1人、レシーバ役に4人と比較的小人数に限定した。しかし、実際にはより多数のユーザが利用も考えられる。多数のレシーバが存在する場合は、推薦候補を決定するための類似度計算の負担が増大してしまう。アドバイザーの閲覧行為を妨げないように、効率的にレシーバへの推薦候補を決定する手法は今後の課題の一つである。

##### 5.4.2 ユーザプロフィール

現在の実装では、ユーザの興味・関心を閲覧履歴から自動的に抽出している。これはユーザ側にプロフィール作成の負担を減らすメリットがある一方で、ユーザの意図しないキーワードが混入したり、閲覧していないページへの興味・関心は無視されてしまったりというデメリットもある。この点に関しては、ユーザにかかる負担と評価尺度を考慮しつつ最適な方法をとる必要がある。

また、新規ユーザは閲覧を行うまでプロフィールは存在していない。そこで、あらかじめそのワークグループ固有の初期プロフィールを用意し、それをプロフィールとして用いれば、新規ユーザへの教育的な推薦を意図的に増加させることが可能である。こうしたユーザプロフィールの二次的な応用に関しても検討すべきである。

##### 5.4.3 推薦後のフィードバック

レシーバがアドバイザーから推薦を受けた際に、それが実際に自分にとって有益であったかどうかは重要な要素である。つまり、推薦システムである以上、アドバイザーはレシーバにとって興味のないページを推薦してしまうことが起こりうる。このような推薦ミスを放置したまま単に推薦を行い続けるのではなく、推薦に価値があったかどうかという情報をユーザの興味ベクトル等にフィードバックすることで、それ以降の推薦がよりユーザの興味に沿ったものになると考えられる。

## 6. ま と め

本稿では、閲覧履歴に基づく情報検索の相互支援について提案した。また、提案手法を実現するプロトタイプシステムを実装し、評価実験を行った結果、ワークグループ内のユーザが情報推薦に携わることで、システムのみで推薦の決定を依存するより効果的に推薦することが示された。すなわち、他のユーザとの連携の中で、これまでは大量の情報の中に埋もれていた未発見の有益な情報を、効率的に獲得することが可能となったといえる。

### 文 献

- [1] 酒井 哲也, 梶浦 正浩, 住田 一男 “情報フィルタリングのためのブル式と文書構造を利用した検索条件生成と検索精度評価” 情報処理学会論文誌, Vol.39, No.11, pp.3076-3083, 1998
- [2] T. W. Malone, K. R. Grant and F. A. Turback, “The Infor-

mation Lens: An Intelligent System for Information Sharing in Organizations”, *Proceedings of ACM CHI '86*, pp.1-8, 1986

- [3] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, J. Riedl, “GroupLens: an open architecture for collaborative filtering of netnews”, *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pp.175-186, 1994
- [4] Amazon.com: <http://www.amazon.co.jp/>
- [5] 品川 徳秀, 北川 博之, 川田 純 “ユーザプロフィールに基づくビューページの動的生成による WWW 閲覧支援” 情報処理学会論文誌: データベース, Vol.41, No.SIG6(TOD7), pp.22-36, 2000
- [6] 中島 伸介, 黒田 慎介, 田中 克己, “閲覧履歴を反映したコンテンツ依存型 Web ブックマーク” 情報処理学会論文誌: データベース, Vol.43, No.SIG5(TOD14), pp.23-36, 2002
- [7] G. Salton, M. J. McGill, “Introduction to Modern Information Retrieval”, *McGraw-Hill*, 1983
- [8] 松本 裕治, 北内 啓, 山下 達雄, 平野 善隆, 松田 寛, 高岡 一馬, 浅原 正幸 “形態素解析システム『茶筌』 version 2.3.3 使用説明書” 奈良先端科学技術大学 情報科学研究科自然言語処理講座, 2003
- [9] MSN Messenger: <http://messenger.msn.co.jp/>
- [10] MSN Messenger Protocol: <http://www.hypothetic.org/>
- [11] @IT: <http://www.atmarkit.co.jp/>