

# ブックマークと文書ディレクトリの協調

鵜飼孝典, 三末和男

富士通研究所

EMAIL: ugai@jp.fujitsu.com, misue.kazuo@jp.fujitsu.com

## 概要

WEB ページ(URL)を階層化されたカテゴリに分類して、提供する文書ディレクトリは、情報を探すためのツールとして非常に有効である。しかし、こういった文書ディレクトリの維持管理では、優良な URL を探し出して、適切なカテゴリに分類を行なうことに非常にコストがかかる。本稿では、サーバに保存されているユーザのブックマークと文書ディレクトリに対して協調フィルタリングを行なうことで、ディレクトリに対して URL の推薦を行なうシステムについて述べる。内容分類を用いた手法との比較実験で、企業内の部門情報など内容ではなく文書の所属組織などで分類しているカテゴリでも有効な URL を推薦できることがわかった。

キーワード: ブックマーク, 文書ディレクトリ, 協調フィルタリング, 知識共有

## Interaction between a Large Directory and Bookmarks

Takanori Ugai, Kazuo Misue

Fujitsu Laboratories Ltd.

EMAIL: ugai@jp.fujitsu.com, misue.kazuo@jp.fujitsu.com

## Abstract

Document directory is a very effective tool to look around useful information or one of informational common systems because directory consists of classified useful information (URL) into its hierarchical categories. However, the maintenance of directory requires cost to explore useful URLs, and to classify them into appropriate categories. We describe a system that recommends URL to a directory with the social filtering among users' bookmarks in a server and the directory. The experiments' result shows that the system could recommend useful URLs to the category classified by the meta-information like enterprise section information.

Keywords: Bookmark, Directory, Social filtering, Knowledge sharing

## 1. はじめに

インターネットばかりでなくインターネットにおいても近年大量の WEB ページが作成されるようになった。そのため、本当に優良な情報が何処にあるのか探すことが難しくなった。このような状況において Yahoo! [7]のような文書ディレクトリは、情報を探すためのツールとしてつぎの 2 つの点で非常に有用である。優良な WEB ページ(URL)が集められている。階層化されたカテゴリに分類してある。しかし、こういった文書ディレクトリの維持管理では、優良な URL を集めて、適切なカテゴリに分類を行なうことに非常にコストがかかる。例えば Yahoo ではこのような作業のために数百人の「サーファ」と呼ばれる専門家がいる。

このようなコストのかかる作業を自動化あるいは半自動化できれば、文書ディレクトリの維持管理コストの軽減化が期待できる。その際、優良な URL の収集やカテゴリへの分類を高精度で行なえることが技術的な課題となる。これまでに我々は、Google[6]の PageRank を改良したアルゴリズムを選別に用いて、優良な URL を自動的に収集し、内容の類似性による自動分類を用いたカテゴリへの自動割当を行なうシステムを開発し、実験、運用を行なってきた[2]。ただし、この手法では、新しく作成された WEB ページは、たとえ優良なものであっても収集されない、部門情報など内容ではなく文書の所属組織などで分類しているカテゴリでは、分類誤りが増えという問題があった。

本稿では、サーバに保存されているユーザのブックマークと文書ディレクトリに対して協調フィルタリングを行なうことで、ディレクトリに対して URL の推薦を行なうシステムについて述べる。本システムでは、他からリンクされていないページで口コミによって知られているような URL でもユーザのブックマークに登録されれば推薦でき、協調フィルタリングを用いているので、文書の内容とは異なった観点で分類されているカテゴリに対しても推薦を行なうことができる。またユーザはブックマーク管理システムを利用するだけで自動的にディレクトリに情報提供を行なうことができ、自然と情報共有が行なわれる。

ブックマークに対して協調フィルタリングを用いる研究、サービスはいくつか存在するが、この技術を大

規模なディレクトリの維持管理コストの削減に利用したものはない。

以下第 2 章では既存の技術について述べ、第 3 章では我々が開発しているシステムとシステムで用いている協調フィルタリングのアルゴリズムについて述べる。4 章では我々が開発した手法の実験結果を示し、5 章で考察を加える。最後に 6 章でまとめを述べる。

## 2. 既存技術と関連サービス、研究

本章では、我々がこれまでに開発したシステム[2]、およびそれに関連する技術について述べる。

### 2.1. Web リンク解析を応用した優良 URL の収集

これまでわれわれが開発したシステムでは Google の PageRank を改良したリンク解析アルゴリズム[5]を用いて優良 URL を選別する。Web のリンク解析において、リモートリンク(異なるサーバからのリンク)とローカルリンク(同じサーバ内からのリンク)を区別し、それらの重みを変えた点が特徴であり、メニューのようなディレクトリに適したページの選別に有効である。

ただし、リンク解析では、新しく作成された URL はたとえ優良であっても収集されないという問題がある。これは新しく作成された Web ページは、他からのリンクが少ないために高い PageRank が与えられない、つまり優良と判断されないためである。

### 2.2. 内容の類似性による自動分類

このシステムでは URL が指す文書の内容に基づいて分類を行なう。自動分類の手法としては、文書の特徴を示すキーワードベクトルの余弦や距離を分類に利用するベクトル空間法、カテゴリとの関連度を数値化した分類規則をあらかじめ用意するルールベース法、単純な分類方法を繰り返し実行して分類結果の多数決を分類に利用するブースティング法の 3 種類の方法を提供する。

カテゴリに含まれている文書の内容が似ている場合は高い率でそのカテゴリにふさわしい URL が得られる。しかし部門情報など内容ではなく文書の所属組織などメタ情報で分類しているカテゴリでは、

分類誤りが増え、自動推薦の効果が低くなるという問題がある。

### 2.3. 協調フィルタリングを用いたインターネット上のサービス

リンク[8]では、ブックマークをサーバに保存し、他人のブックマークと比較して、同じURLを持つ人を探し出して、その人のブックマークとの差分を示すことで個人のブックマークを新鮮に保つサービスを提供している。

この方法では、インターネットで十分に多くのユーザー(数万人以上)のブックマークを管理している場合には有効であるが、イントラネットのユーザを想定した場合(数十人から数百人)ではうまく機能しない。

### 2.4. ブックマークエージェント

ブックマークエージェント[10]は、ブックマークファイルを参照し、ブックマークの URL が指す HTML ファイルから抽出するキーワードによってユーザ間で情報の共有を行なうシステムである。小規模で興味の似たユーザからなるグループにおいてこのエージェントを用いることで既存の検索エンジンよりも適合率の高い Web ページの提示が可能になる。

このシステムは、リンクのようなサーバに蓄えるシステムと比較して少ないリソースで実現できるが、スケーラビリティに問題がある。興味が多種多様に及ぶインターネットやイントラネットのユーザでは、十分な量の検索結果を提示できない。

## 3. システムの概要

本章では、我々が開発している文書ディレクトリの維持管理コストの軽減化のためにシステムが提供する機能について述べる。まず優良な URL を推薦する手法について述べ、その後、それ以外にシステムが提供する機能について述べる。

### 3.1. ブックマークとディレクトリの協調フィルタリング

本システムではサーバ上にカテゴリ毎に分類されているユーザのブックマークとディレクトリの各カテゴリを比較して、同じ URL を持つカテゴリを探し出して、そのカテゴリとの差分を推薦 URL としてディレクトリに推薦する。ディレクトリ管理者はシステムから推

薦される URL をチェックして、優良であればディレクトリに登録する。

ユーザのブックマークと協調フィルタリングを行うことには次の2つの利点が期待される。

1. 内容ではなく文書のメタ情報で分類しているカテゴリでも正解率が高くなる。そのためディレクトリ管理者の手間が軽減される。
2. ブックマークを利用しているユーザはブックマーク管理システムを利用するだけで自動的にディレクトリに情報提供を行なうことができ、自然と情報共有が行なわれる。そのためディレクトリ管理者にとっては、陽にボランティアを募ったり、数多くの専門家に頼ったりする必要がない。

### 3.2. 本システムで用いたアルゴリズム

本節では、本システムで用いた協調フィルタリングのアルゴリズムを定義し、例を用いて説明する。

**前提:** URL の全体集合を  $U$ 、カテゴリの全体集合を  $C$  とする。カテゴリ  $c \in C$  は  $c \subseteq U$  とする。

**定義 1:** [類似度] カテゴリ  $c_1$  と  $c_2$  の類似度

$F(c_1, c_2)$  を  $F(c_1, c_2) = |c_1 \cap c_2| / |c_1 \cup c_2|$  と定義する。

**定義 2:** [人気度] URL  $u \in U$  の人気度  $P(u)$  を  $P(u) = |\{c \in C \mid u \in c\}| / |C|$  と定義する。

**定義 3:** [推薦候補] 閾値  $r$  としたとき、カテゴリ  $c$  に対する推薦候補  $R(c, r)$  を  $R(c, r) = \{u \mid \forall c_1 \in C \langle u \in c_1 \wedge F(c_1, c) \geq r \wedge u \notin c \rangle\}$  と定義する。

**定義 4:** [絞込み] 推荐候補  $V$  に対して個数  $m$  としたとき 絞り込み候補  $M(V, m)$  を  $M(V, m) = \{u \in U \setminus U_1 \subseteq V \wedge |U \setminus U_1| = m \wedge \forall u_1 \in U_1, \forall u_2 \in V \setminus U_1 (P(u_1) \geq P(u_2))\}$

ただし、 $|V| \leq m$  ならば、 $M(V, m) = V$  とする。

と定義する。これは  $V$  の要素から人気度が大きい順に  $m$  個集めた集合を意味する。簡単のため  $M(R(c, r), m)$  を  $M(c, r, m)$  と書くことにする。

### 3.3. ディレクトリへの URL の推薦

本システムでは、他のユーザのカテゴリ、および他のユーザと一緒に用いる大規模ディレクトリの各カテゴリに含まれる URL の集合を  $U$  とする。そして類似度の閾値  $r$  と推薦する数  $m$  を定数としてディレクトリのカテゴリ  $c$  に対して、 $M(c,r,m)$  を算出して推薦する。

表 1を例に説明する。例えば、ディレクトリにカテゴリ A と異なる 2 人ユーザのブックマークにカテゴリ B, C があり、それぞれのカテゴリに がついている URL(u1 ~ u6)が登録されているとする。また閾値  $r$  を 0.5、最大推薦数  $m$  を 2 とする。

表 1: カテゴリの例

カテゴリ	u1	u2	u3	u4	u5	u6
A		x				x
B	x					
C						x

カテゴリ A と B は 3 つの URL が一致し、A と C は 4 つの URL が一致する。類似度は それぞれ  $0.50=3/6$ ,  $0.60=4/6$  になる。A の推薦候補 URL  $R(A,0.5)$  は B と C の URL のうち A に含まれない URL{u2, u6}となる。u2 と u6 の人気度はそれぞれ  $0.67=2/3$  と  $0.33=1/3$  となる。 $M(A,0.5,2)$  は{u2, u6}となり、システムは A に対して u2 と u6 を推薦する。(m=2 のため、ここでは人気度は特に使われない。)

### 3.4. ブックマークへの URL の推薦

リンクと同様に、本システムはサーバ上でカテゴリ毎に分類されているブックマークを他人のものと比較して、同じ URL を持つカテゴリを探し出して、そのカテゴリとの差分を推薦 URL としてユーザに示す。ただし、他のユーザの各カテゴリと同様に、Open Directory Project[9]や、社内ポータルなどの大規模ディレクトリの各カテゴリとも比較する。

大規模ディレクトリの各カテゴリを他のユーザと同様に協調フィルタリングに用いることにはブックマークをサーバに蓄えるユーザにとって次の 3 つの利点が期待される。

1. 協調フィルタリングを用いることで文書の内容の類似性に基づいて分類する方法に比較して、部門情報など内容ではなく文書の

所属組織などで分類しているカテゴリでもふさわしい URL がえられる。

2. カテゴリ単位での協調フィルタリングを行なうので、ディレクトリのカテゴリ構造とブックマークのカテゴリの構造が異なっていてもかまわない。
3. 数十人から数百人の小人数でも網羅的な大規模ディレクトリとの比較によって有効な URL の推薦を得ることができる。

### 3.5. その他の機能

本システムはディレクトリの維持管理コスト軽減のために、優良な URL 推薦以外に次の機能を提供する。

1. 登録されている URL に定期的にアクセスしてリンク切れを起こしていないかチェックすること。URL に変更があり、リダイレクトされているときは、自動的に URL を変更する。
2. URL を登録するときに、コンテンツから自動的にタイトル、キーワードを取り出し、入力を省力化する。
3. 情報の編集は、フォームを通して行ない、各カテゴリの画面はテンプレートを利用して作成されるために、デザイン上の統一性を保つことができる。
4. 更新されたカテゴリや URL の一覧を“新着情報”として自動的に作成する。

### 3.6. 動作画面例

本節では、推薦された URL を編集者が処理するための画面の様子を示す。

TOP / Perl			
操作/編集/移動	登録URL	備考	アクセス数
	perlの私的技術		1
	COL/Perl入門		1
	Perlの山林	Perlをパーコの日本語訳	2
	みじく今日のひとこと - 2000/09/27中Daylightについて 2000/09/27	2000/09/27中Daylightについて	0
	EAP	std::stringが無いというエラ ーが出てるんですけど	0
	Perl王		5

図 1: 管理者向け編集画面例

本システムは、推薦された URL は図 1 のようにそのカテゴリの編集画面ですでに登録されている

URL と比較しながら適当かどうかを判断して推薦されたものを採用するかどうか決めることが出来る。システムによって推薦されたものは他と異なる色(図1では背景が灰色)で表示され、一目で区別できるようになっている。また図2のように推薦されたすべての URL を一覧してまとめて処理することも出来る。

TOP	更新/登録/削除	登録日時	URL	件名
● ● ●	2002-09-14 2002-08-03	結界土壁	/ODPJ/部門技術/各部門のセイ	
● ● ●	2002-09-28 2002-08-03	日本語版の技術書籍	/Perl/	
● ● ●	2002-08-28 11/09/01/0	ODPJ入門	/Perl/	
● ● ●	2002-08-28 11/02/4/0	Perl初心者	/Perl/	
● ● ●	2002-08-28 12/08/03/0	Java入門書	/Java/	
● ● ●	2002-08-18 2002-08-18	Java言語 パーティクル操作 説文	/Java/	
● ● ●	2002-08-28 12/08/03/0	Javaで始める JavaServer Pages サーバー	/Java/JSP/	
● ● ●	2002-08-28 12/08/13/0	JavaServlet技術書	/Java/JSP/	
● ● ●	2002-08-27 12/08/03/0	Windows Server	/Perl/	

図 2:URL 推薦画面例

## 4. 評価実験

### 4.1. 実験に用いたデータ

表2は実験に用いたブックマークの数、それに含まれるカテゴリの大きさと URL の数である。ユーザ数は、サーバにブックマークを保存している延べ数、全カテゴリ数は、全ユーザの延べカテゴリ数、全 URL 数は全カテゴリに登録されている延べ URL 数である。

表 2:全ブックマークとカテゴリ

ユーザ数	73
全カテゴリ数	448
全URL数	3172

表 3:実験に用いた大規模ディレクトリ

データ名	対象ディレクトリ	URL数	カテゴリ数
ODPJ	Open Directory Japanese/コンピュータ/プログラミング	211	11
INTRA	社内独自ディレクトリ	1560	210

表3は協調フィルタリングに用いた大規模ディレクトリである。Open Directory[9] はボランティアに

よって維持管理されている大規模ディレクトリで、そのデータは商用、非商用を問わず自由に利用することができる。今回用いたのは、そのうちの一部で、2002年1月のデータである。社内独自ディレクトリは、社内全体を広く浅く網羅するポータルで登録 URL1560 の内、約9割が社内の URL である。

表 4:推薦対象のカテゴリ

カテゴリ名	大規模ディレクトリ	登録URL数
JAVA	ODPJ	25
XML	INTRA	3
PERL	ODPJ	26
社内手続き	INTRA	136
製品担当連絡先	INTRA	36

表4は推薦の対象にしたディレクトリのカテゴリである。JAVA,PERL は ODPJ の中で技術系の文書が登録されているカテゴリである。XML は INTRA の中で社内の技術文書が登録されているカテゴリである。社内手続き、製品担当連絡先はインターネットで多くのユーザが URL を登録しているカテゴリである。

表 5: 推荐対象のブックマーク

カテゴリ名	登録URL数	社内	社外	日本語	英語
JAVA	17	0	17	3	14
XML	6	3	3	6	0
PERL	10	0	10	0	10
社内手続き	11	11	0	11	0
製品担当連絡先	12	12	0	12	0

表5は推薦の対象にしたブックマークのカテゴリである。社内手続きには社内で良く使われる URL が登録されており、同じ URL を数多くのユーザが登録している。

### 4.2. 正解の評価方法

実験は各カテゴリにおのの最大5つの URL を推薦することとし、特に断りのない場合、3.2節で定義した M を、閾値  $r=0.2$  として算出した。ブックマークやディレクトリのカテゴリにはそのユーザや編集者の方針によって URL が集められている。したがって本来正解は、ブックマークのユーザやディレクトリの編集者でなければわからない。本実験では実験者が採用しても良いと判断したものを正解として数えることにした。実験者はディレクトリの編集者ではないため、ディレクトリへの推薦結果に対して

は、そのカテゴリに登録されている URL の内容と比較して妥当と考えるものと正解とした。実験対象のブックマークは実験者のものであり、推薦結果は自らの基準で妥当と考えるものと正解とした。

#### 4.3. 実験 1: ディレクトリへの推薦

ここでは、閾値の変化に対して、推薦数、適合率がどのように変化するかを測定した。

表 6 は閾値を変えて推薦を行なった実験の結果である。閾値を大きくすると、推薦数が小さくなるが適合率は大きくなる傾向にあることがわかる。

表 6: 推薦結果

カテゴリ名	閾値0.2			閾値0.3		
	推薦数	正解数	適合率	推薦数	正解数	適合率
JAVA	5	3	0.60	4	3	0.75
XML	5	5	1.00	3	3	1.00
PERL	5	4	0.80	3	3	1.00
社内手続き	5	2	0.40	2	1	0.50
製品担当連絡先	4	3	0.75	4	3	0.75

#### 4.4. 実験 2: ブックマークへの推薦

ここではディレクトリとブックマークの協調フィルタリングを用いたブックマークへの推薦の精度を測定した。

表 7 は、全ユーザのブックマークに ODPJ と INTRA のディレクトリを加えて協調フィルタリングを行なった結果である。

表 7: ディレクトリとの協調

カテゴリ名	ODPJ			INTRA		
	推薦数	正解数	適合率	推薦数	正解数	適合率
JAVA	5	3	0.60	5	2	0.40
XML	5	0	0.00	5	1	0.20
PERL	4	0	0.00	5	1	0.20
社内手続き	2	0	0.00	5	3	0.60
製品担当連絡先	1	0	0.00	3	3	1.00

PERL に対して ODPJ からの推薦数が 0 だったのはブックマークのカテゴリが英語のコンテンツだけなのに対して ODPJ には日本語のコンテンツしか含まれていなかつたためである。また XML に対して ODPJ からの推薦数が 0 だったのは ODPJ には XML に関するコンテンツがあまり含まれていなかつたためである。社内手続き、製品担当連絡先の結果を見ると INTRA は社内コンテンツに関して有効に働いたことがわかる。

#### 4.5. 比較実験 1 : 内容分類による推薦

ここでは内容分類を用いた推薦結果を示す。4.3 節、4.4 節の協調フィルタリングによる推薦結果と比較することを目的としている。

##### 内容分類によるディレクトリへの推薦

表 8 はユーザがブックマークに登録している 3172 の URL を内容分類によって各カテゴリに推薦した結果である。

表 8: 内容分類によるディレクトリへの推薦

カテゴリ名	推薦数	正解数	適合率
JAVA	5	2	0.40
XML	5	4	0.80
PERL	5	3	0.60
社内手続き	5	2	0.40
製品担当連絡先	4	0	0.00

表 6 と比較して全般的に適合率が低く、特に製品担当連絡先の適合率が 0 である。

##### 内容分類によるブックマークへの推薦

表 9: 内容分類による推薦

カテゴリ名	推薦数	正解数	適合率
JAVA	5	2	0.40
XML	5	2	0.40
PERL	5	1	0.20
社内手続き	5	1	0.20
製品担当連絡先	5	0	0.00

表 9 は内容に基づく自動分類で分類し推薦した結果である。分類対象とした URL は社内外から文献[8]のアルゴリズムで抽出した 2000 の URL を用いた。製品担当連絡先など内容ではなく文書の所属組織などで分類しているカテゴリでは、正解が得られていない。

#### 4.6. 比較実験 2: ユーザだけによる協調フィルタリング

ここではブックマークだけを用いて、実験対象のブックマークへ推薦を行なった結果を示す。4.4 節の結果と比較することでディレクトリと協調させた場合と比較する。

表 10：ユーザ間の協調フィルタリング

カテゴリ名	推薦数	正解数	適合率
JAVA	3	2	0.67
XML	5	1	0.20
PERL	4	1	0.25
社内手続き	2	0	0.00
製品担当連絡先	1	0	0.00

社内手続きはみなが同じ URL を集めているために推薦されるものが少なかった。また推薦されたのが部署毎に異なる手続きであったために正解とならなかった。製品担当連絡先への推薦数が1と少なかったのは、他におなじ URL を持った人がいなかつたためである。

#### 4.7. 参考実験

##### プリリンクによる実験

表 5に示した URL をプリリンクに登録して推薦結果を評価した。これは、ユーザの数が多くなった場合のユーザのみによるブックマークへの推薦結果を考えることができる。

表 11：プリリンクによる推薦結果

カテゴリ名	推薦数	正解数	適合率
JAVA	5	2	0.40
XML	5	1	0.20
PERL	5	1	0.20

各カテゴリに登録されている URL をプリリンクに登録して、類似 URL の検索を行い、上位5つについて評価した。セキュリティ上の理由から社内 URL は除いた。

##### ディレクトリに重みをつけた場合

ここでディレクトリとブックマークの協調フィルタリングにおいてディレクトリに重みをかけた実験結果を示す。ディレクトリの重みを例えば 3 にした場合、URL の人気度計算の際にディレクトリに含まれる URL は3つのカテゴリが含んでいるとして計算する。つまり2人のカテゴリに含まれる URL よりもディレクトリに含まれる URL が優先的に推薦される。

表 12：INTRA の重み 3

カテゴリ名	推薦数	正解数	適合率
JAVA	5	2	0.40
XML	5	2	0.40
PERL	5	1	0.20
社内手続き	5	3	0.60
製品担当連絡先	5	3	0.60

INTRA を用いた場合に XML において正解数が増えた。これはユーザが持っていた XML に関する URL にあまり有効なものが無かったためである。

ちなみにディレクトリの重みを 10 にした場合、本実験では、ディレクトリに含まれる URL は必ず採用され、ディレクトリから推薦候補が得られない場合にのみ他のユーザの URL から推薦された。INTRA を用いた場合は、重みが 3 のときから変化が無かった。

## 5. 考察

### 5.1. ディレクトリへの推薦

4.3 節の実験結果から閾値を大きくすると、適合率(正解数 / 推薦数)が大きくなり、閾値を小さくすると適合率が小さくなることがわかる。これはディレクトリのカテゴリとの共通 URL が多いブックマークから得られた推薦は正解となりやすいことを意味する。このことは類似性の高いカテゴリ同士からえた URL が推薦されるべき URL となるべきという直感と一致するものである。

4.5 節で行なった内容分類の実験結果において、内容分類を用いたシステムでは有効な推薦が得られなかった、“製品担当連絡先”というカテゴリに対して協調フィルタリングを用いたシステムでは 3 つの URL を推薦し、いずれも正解であった。この結果から、今回開発したシステムが、部門情報など、内容ではなく文書の所属組織などメタ情報で分類しているカテゴリでも有効な URL を推薦できることが示されたといえる。

推薦アルゴリズムのパラメータとなる類似度の閾値と推薦数は、次のような方針で運用フェーズによって使い分けることが有効であると考えている。1) ディレクトリの立ち上げの段階では掲載 URL を迅速に増やすために類似度の閾値を小さくすることで推薦数を増やす。編集者によって適当なカテゴリに

振り分けなおすことで、適合率が低くなることを補う。2) 運用段階で十分 URL が集まっている場合は、類似度の閾値を大きくすることで、不要な URL があまり推薦されないようにし、編集者のコストを減らす。

## 5.2. ブックマークへの推薦

4.4 節の表 7 の結果からディレクトリによって得手、不得手があることがわかる。そのため、社内のポータルサイトなどと ODP のようなインターネットの大規模ディレクトリをうまく組み合わせることでより良い結果が得られると考えられる。4.4 節の表 7, 4.6 節の表 10, 4.7 節の表 11 の比較において、協調フィルタリングに大規模ディレクトリを組み合わせることの効果を期待したが、今回の実験では有意な差はみられなかった。

## 5.3. パーソナライズ版ディレクトリへの応用

協調フィルタリングにおいてディレクトリの重みを大きくすると、ディレクトリに登録されている URL が多く薦められるようになる。これによってユーザは、ディレクトリのパーソナライズ版として利用することが出来る。本手法では、カテゴリの構造、名前にかかわらず有効であるので、ユーザは、ディレクトリの一部を、自分の好きな階層構造で、自分の好みで取捨選択、追加したディレクトリを手に入れることが出来る。

# 6. まとめ

本稿では、文書ディレクトリの運用支援を目的として、URL の収集、分類を半自動化する URL 推薦システムについて報告した。本システムはサーバに保存されているユーザのブックマークと文書ディレクトリに対して協調フィルタリングを行なうことで、ディレクトリに対して URL の推薦を行なう。内容分類を用いた手法との比較実験で、部門情報など内容ではなく文書の所属組織などで分類しているカテゴリでも有効な URL を推薦できることがわかった。推薦アルゴリズムのパラメータとなる類似度の閾値と推薦数は、運用フェーズによって使い分けることが有効であると考えられる。

ブックマークへの推薦については、少ないユーザを補うという点でディレクトリを同時に用いることが有

効ではないかと考えているが有意な差が得られるまでには至っていない。

ブックマークやディレクトリのカテゴリは、そのユーザや編集者の方針によって URL が集められている。そのため、このような実験は本来、編集者や個々のユーザの意図に基づいて評価されるべきである。このことが、評価データ作成を困難にしている。このような問題に対して、いかに有意な評価データを得るかが今後の課題である。

## 参考文献

- 鶴飼、三末:人と大規模ディレクトリの協調によるブックマーク管理、情報処理学会第43回グループウェア研究会(2002).
- 鶴飼、片山、津田: 文書ディレクトリ管理のための自動収集、自動分類の利用、人工知能学会全国大会 5C-01 (2001).
- 鶴飼、イントラネット向けディレクトリ管理システム、情報処理学会第33回グループウェア研究会(1999).
- 片山: 多様な要求に対応するテキストの自動分類システム、情報処理学会62回全国大会(2001).
- Tsuda, Ugai, Misue: Link-based Acquisition of Web Metadata for Domain-specific Directories, PKAW2000, pp. 317-324 (2000).
- Google: <http://www.google.com/>
- Yahoo: <http://www.yahoo.com/>
- Blink: <http://www.blink.jp/>
- Open Directory Project: <http://dmoz.org/>
- 森、山田：ブックマークエージェント：ブックマークの共有による情報検索の支援、電子情報通信学会論文誌 D-I, Vol. J83-D-I No.5 pp.487-494 (2000).