DEWS2003 5-P-04

PVL のための個人向け推薦機構に関する検討

上田　真由美，上島　紳一
関西大学大学院総合情報学研究科

〒 569-1095 大阪府高槻市霊仙寺町 2-1-1
mueda@dw.kutc.kansai-u.ac.jp, ueshima@res.kutc.kansai-u.ac.jp

近年，インターネット上の情報資源を個別化して利用するための枠組みが重要視されている．我々は，利用者がインターネット上の情報資源を個人の好みで利用するための枠組みとして，プライベート仮想ライブラリ PVL を提案している．

本稿では．P2P 型ネットワークを基盤とした PVL の個人向け推薦機構を提案する．PVL は，単一利用者の利用を目的とし，メタ検索エンジン機構と，XML 形式の情報ベースで構成される．利用者はメタ検索エンジン機構が提供した情報の中から，自分によって有用な情報を本棚に保存する．これに対し，本推薦機構は，各利用者が本棚に保存している情報を PVL コミュニティで共有し，他の利用者が個別化を進め入手し，本棚に保存している情報を提供することを可能にする．推薦データの決定には，提供された推薦データに対する利用者の対話的操作を用い算出した各利用者の得点傾向の類似度によって，推薦要求の伝播先を決定する．

本手法の有効性を検証するため，プロトタイプシステムを構築し評価実験を行った．実験結果により，格納に関する適合率が 50.4%，表示に関する適合率が 94.3%となり，PVL コミュニティに属する不特定多数を対象に推薦要求を行った場合と比較し，それぞれ約 1.9 倍，約 1.5 倍上昇した．

キーワード　推薦機構，P2P 型ネットワーク，得点付け，エージェント，仮想ライブラリ

Personalized Recommendation Mechanism for PVL

Mayumi UEDA, Shinichi UESHIMA
Graduate School of Informatics, Kansai University

Recently, personalization has been one of central issues for utilizing large volume of Web contents over the Internet. The authors have proposed a Private Virtual Library PVL, which is a single-user system that enables a user to personalize his/her BookShelf using agent scoring.

In this paper, we discuss personalized recommendation mechanism for PVL for the purpose of sharing personal BookShelves among users in a PVL community. We consider a criterion for collection of BookShelves using cosine similarity between agent-characteristic vectors of each user to reflect the tendency of each user's preference. In implementing recommendation mechanism, the system has been built as principal over peer, and deploys the realm of P2P communication for gathering personalized and refined BookShelves from other members.

As a result, we have high precision factors w.r.t saving, which is 1.9 times higher than the case when request are propagated to unspecified and many users without using similarity among agents-characteristic vectors. The precision factors w.r.t browsing is 1.5 times higher.

Key words　recommendation, P2P network, scoring, agents, virtual library

1　はじめに

近年，高度情報化の進展により，膨大な情報がインターネット上に存在する．これらの情報の中から，個人にとって有効な情報を選別するのが困難となり，インターネット上の情報資源を個別化して利用するための枠組みに関する研究が盛んになってきている [1]．我々は，インターネット上の情報を本に見立てて本棚に格納し，本棚の個別化を進めるプライベート仮想ライブラリ PVL を提案している [2]．PVL では，“本棚” のメタファーを用いて，現実の本棚と仮想の本棚の利点を融合した夢の本棚の実現を目指している．エージェントを用いたメタ検索エンジン機構と，情報ベースを用いて実現しており，アクセスエージェントの得点付けによって，個別化した情報の提供を可能にしている．

通常，Web サイトを閲覧する際，利用者は自分にとって重要なサイトだけを Bookmark に保存することが多いと考えられる．本システムにおいても同様に，利用者は自分にとって重要な情報だけを本棚に保存しており，本棚には PVL を繰り返し利用することによって個別化された，専門的な情報が格納されていると考えられる．

本稿では，P2P 型ネットワークを基盤とした PVL の個人向け推薦機構を提案する．PVL の通信手段として，P2P 型ネットワークを用いることにより，PVL コミュニティ間での情報共有を可能にし、各利用者が本棚に保存した情報の共有を実現する [3]．すなわち，P2P パラダイムによって不特定多数の PVL 利用者を対象に情報収集を行うことが可能となり，個人レベルで洗練された情報を容易に入手することできる．これにより，単一利用者が繰り返し PVL を利用することによって起りうる，過度の個別化を修正することが可能になる．また，他の利用者が繰り返し PVL を利用することによって得た，洗練された専門的な情報を得ることや，PVL の利用回数が少なく，個別化の進んでいない利用者に対し，洗練された情報の提供を可能にする．

本機構では，推薦情報を利用者に提供する際，各利用者の持つアクセスエージェントの得点傾向を各利用者の特徴ベクトルとして扱い，利用者間の類似度によって推薦要求の伝播先を決定する．

2 章では，PVL と P2P 型推薦機構の特徴について述べ，3 章で P2P 型個人向け推薦機構の構成を述べる．4 章でプロトタイプシステムを構築し，評価実験により本システムの有効性を検証する．5 章で関連研究との比較を行い，最後に 6 章で本稿のまとめを行う．

2　PVL と P2P 型推薦機構

2.1　PVL とは

PVL では，インターネット上の情報を本に見立てて本棚を構築する “本棚メタファー” を用いている．我々が考える本棚メタファーは，インターネット上にある自分好みの情報を，自分の本棚にある本のように自由に扱うことができるような仮想の本棚を用いたメタファーで，PVL は現実の本棚と仮想の本棚の利点を融合した個人向けの夢の仮想ライブラリである．

このような夢の仮想ライブラリの実現には，検索機構が重要である．PVL は，エージェント群を用いたメタ検索エンジン機構と XML 形式の情報ベースで構成されている．PVL では既存の検索エンジンが提供する情報を統合利用するメタ検索エンジンを構築し，(1) 各検索エンジンの偏りを生かすことで網羅性を高め，(2) 重複性を生かすことで提供情報の信頼性を高めることを目指している．メタ検索エンジン機構では，インターネット上に存在する複数の検索エンジンを利用して問い合わせを行い，情報を集約・個別化して利用者毎に順序付けて情報を提供する．

PVL では，既存の検索エンジン毎にアクセスエージェントを備えており，各検索エンジンが提供する情報を PVL で利用する形式に統一するためのラッパーの働きを備えている．アクセスエージェントが収集した情報は，メディエータが集約し，個別化処理を行い利用者に提供する．利用者は，提供された情報の中から要・不要を判断し，必要と判断した情報を本棚に格納し後日利用することが可能である．

2.2　P2P 型推薦機構

2.2.1　必要性

PVL では，繰り返し利用し個別化を進めることによって過度の個別化が起こる可能性があり，修正する機構が必要である．また，他の利用者が繰り返し利用した後に得た洗練された情報を得ることや，PVL の利用回数の少ない個別化の進んでいない利用者に対し，洗練された情報を提供するために推薦機構は重要である．

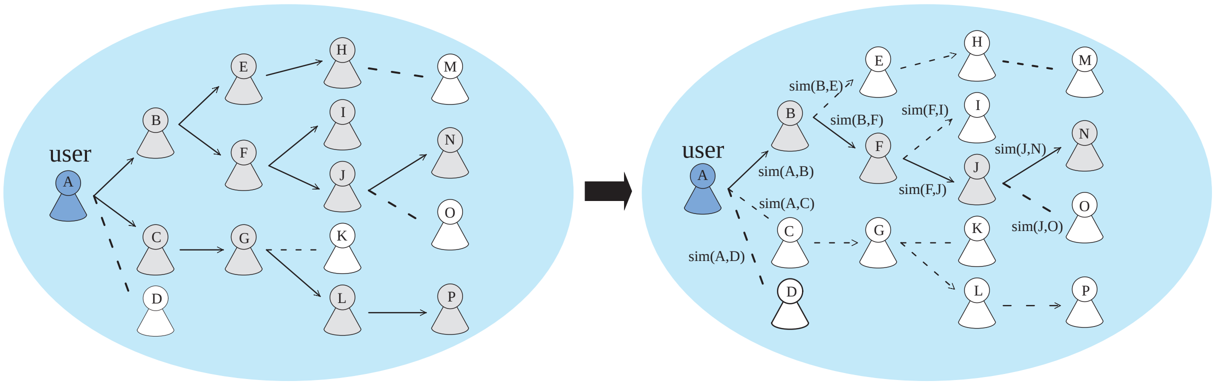

図 1: 推薦要求の伝播

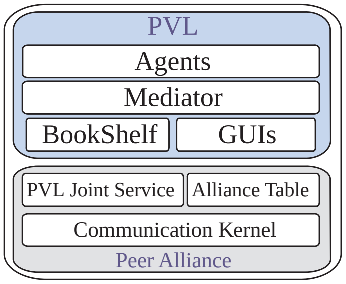

図 2: 各ピアの構成

そこで本システムでは，他の利用者がPVLを繰り返し利用し個人用に蓄えた本棚の情報を扱う推薦機構を実現する．ネットワーク上に分散した不特定多数のPVL利用者を対象に情報収集を可能にするため，P2P型ネットワークを用いる．P2P型ネットワークを通信手段に用いることにより，図1左に示すように，他のPVL利用者に対し推薦要求を伝播することが可能となり，各利用者によって個別化された情報を容易に入手することが可能となる．

2.2.2 PVLとピアの関係

PVLは，単一利用者がインターネット上の情報資源を個別化して利用することを目的とした枠組みである．しかし，ここではP2P型ネットワークを通信手段に用いることで，PVLコミュニティを形成し，不特定多数のPVL利用者を対象として情報収集を行い，推薦機構を実現する．

各ピアはPVLとピア･アライアンス･アプリケーション(以下PAと記す)によって構成される．通信手段として用いるPAは図2下部に示すように，ピア間接続に用いるコミュニケーション・カーネルと，PAのサービスの一つであるPVLJointサービス，リクエスト文の伝播先を記述したアライアンス表によって構成する．

PVLJointサービスには，各ピアに存在するPVLの本棚ファイルを検索するPVLコマンド，操作結果を指定したファイルに書き出すPVLRecommコマンド，各利用者が持つアクセスエージェントの得点を収集するための，PVLScoreコマンドなどが含まれる．

2.2.3 P2P型推薦機構の特徴

PVLは繰り返し利用することによって個別化を進めるフィードバック型システムである．しかし，個人にとって好ましくない偏りが生じる可能性がある．そこで，本システムでは，他のPVL利用者が個人的に蓄えた本棚内の情報を用いた推薦機構を実現することにより，偏りによって生じた情報の補填を行う．

また，P2P型ネットワークを用いているため，他の利用者に負担をかけずにPAが自動的に処理を行う．

3 個人向け推薦機構

3.1 個別化推薦機構の特徴

他のPVL利用者が本棚に持つ情報を用いて推薦機構を実現することにより，洗練された情報を容易に入手することが可能となる．しかし，PVL利用者が多くなるにつれ，興味の範囲が広がり，ある利用者が「カテゴリ=“computer”・キーワード=java」と保存した情報が，他の利用者にとって全く役立たない情報である場合が生じる．そこで，利

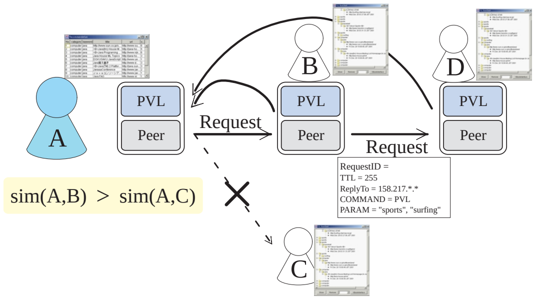

図 3: リクエストの伝播と各ピアの動作

用者間の類似度を考慮した個人向けの推薦機構を実現する．本推薦機構では，PVL 利用時に生じるアクセスエージェントの得点で構成される各利用者の特徴ベクトルを用いて，コサイン尺度により算出した利用者間の類似度を用いる (式 (1))．

$$sim(v_A, v_B) = \frac{(v_A, v_B)}{|v_A||v_B|} \tag{1}$$

ここでは，v_A，v_B は利用者 A，B の特徴ベクトルを表し，$sim(v_A, v_B)$ は利用者 A，B の類似度を表す．

本機構は図 1 右の形式で，図 3 に示すように，各利用者間の類似度によって，推薦要求の伝播先を決定する．類似度の高い利用者に推薦要求を送ることで，先に述べた認識のズレが生じる可能性が減少し，カテゴリとキーワードを入力するだけで，洗練された情報のみが入手可能となる．

3.2 推薦機構の動作

図 3 にリクエスト伝播と各ピアの動作を示す．PVL から PA へのデータの受け渡しは，ローカルホストの所定のポート番号 (ここでは 9153) へ UDP 形式でパケットを送信する．また，PA から PVL への推薦結果の受け渡しは，XML ファイルへの書き出しによって行う．

利用者 A が利用者インタフェースにカテゴリとキーワードを入力し，推薦要求を行うと，利用者 A のローカルホストに常駐する PA は，PVL コマンドを記述したリクエスト文を PVL から受け取り，利用者 B の PA に伝播させる．リクエスト文には，利用者 A のローカルホストの IP アドレスや推薦要求時に利用者インタフェースに入力されたカテゴリ，キーワードなどを記述しておく．リクエストの伝播先は 3.1 節の式 (1) によって算出された類似度の値によって決定する．図 3 では，$sim(v_A, v_B) > sim(v_A, v_C)$ により，利用者 B にリクエスト文を伝播する．ただし，各利用者間の類似度の値は，あらかじめ PA が各利用者からアクセスエージェントの得点を収集しており，PVL のプロファイル操作エージェントが式 (1) により算出しておく．

利用者 B の PA は PVL コマンドが記入されたリクエスト文を受け取ると，PVL の本棚ファイルを検索し，リクエスト文に記述されたカテゴリとキーワードに一致する情報の抽出を行い，応答リクエストを生成する．応答リクエストには，PVLRecomm コマンドと抽出した情報などを記述し，受けとったリクエスト文に記述された IP アドレスが示すピア (利用者 A) に渡す．

また，利用者 B は受け取ったリクエスト文 (PVL コマンドが記述されたもの) を利用者 D の PA に伝播し，利用者 D の PA は利用者 B と同様の操作を行う．

推薦要求を発信した利用者 A の PA は，PVL-Recomm コマンドを記述したリクエスト文を受け取ると，各利用者の本棚ファイルから抽出した情報を推薦情報として指定のファイルに書き出す．

利用者 A の PVL は，指定ファイルに書き込まれた情報を利用者インタフェースに表示する．

複数の利用者から同じ情報が提供されることがある．本システムでは，複数の利用者が重要と判断し保存していた情報は評価が高い情報と判断し，上位に表示する．また，利用者が洗練された情報を入手する手助けとするため，フレーム内に，重複回数を得点として表示する．

4 プロトタイプシステムの実装と評価

本節では，提案手法を用いたプロトタイプシステムを実装し，評価実験を行うことによって，本手法の有効性を考察する．

4.1 プロトタイプシステム

P2P 型ネットワーク上に 3 章で述べた個人向け推薦機構を実現した．

メタ検索エンジン機構を実現するため，既存の検索エンジン (Yahoo! JAPAN，Infoseek，Excite，Lycos) に対応するアクセスエージェントを作成した．さらに，利用者プロファイルなどの PVL 独自の情報資源を操作するためのエージェントを実装し，メディエータによって統合している．

利用者は，提供された情報に対し，(1)“選択” し

表 1: 実験参加者のエージェントの得点より算出した類似度 ($cos\theta$) の値

	利用者 A	利用者 B	利用者 C	利用者 D	利用者 E	利用者 F	利用者 G	利用者 H
A	-	0.725386	0.751066	0.82372	0.919016	0.87096	0.749483	0.747246
B	0.725386	-	0.97456	0.952877	0.937109	0.883453	0.886778	0.770843
C	0.751065	0.97456	-	0.990646	0.928735	0.93587	0.938667	0.854326
D	0.823719	0.952877	0.990646	-	0.952389	0.955522	0.95634	0.895236
E	0.919016	0.937109	0.928735	0.952389	-	0.936909	0.874033	0.803278
F	0.87096	0.883453	0.93587	0.955522	0.936909	-	0.842354	0.782696
G	0.749483	0.886778	0.938667	0.95634	0.874033	0.842354	-	0.976438
H	0.747246	0.770843	0.854326	0.895236	0.803278	0.782696	0.976438	-

て実データを表示する，(2) 本棚に "保存" する，といった操作を行うことによって，アクセスエージェントに対する得点付けを行うことが可能である．本棚ファイルへの保存は，本棚操作エージェントによって自動的に XML 形式でタグ付けされる．

推薦機構では，他の利用者が本棚に保存したデータを情報資源として扱う．ここでは，研究室の学生約 30 人，合計約 550 件のデータを対象として扱い，利用者間の類似度によって，推薦要求の伝播先を決定する．推薦要求は，式 (1) によって算出した類似度の値を用いて，上位 2 人に対して伝播する．図 3 で，利用者 A から推薦要求を受け取った利用者 B は，他の利用者との類似度を記述したファイルから要求の伝播先を特定し，推薦要求を伝播する．

表 1 にプロトタイプシステムの利用により得た各利用者の類似度の値を一部示す．以下の実験では，この値を用いた．

4.2 評価実験

ここでは，4.1 節で構築したプロトタイプシステムを用いて，評価実験を行い，本手法の有効性を検証する．

4.2.1 評価パラメータ

4.1 節で述べたプロトタイプシステムを用いて，以下の実験を行った．

1. PVL 単一利用時の適合率α，βの推移
2. PVL コミュニティより推薦された情報に関する適合率α，βの推移
3. PVL コミュニティより個別化して推薦された情報に関する適合率α，βの推移

ただし，格納に関する適合率を適合率α，表示に関する適合率を適合率βとし，それぞれ次式とする．

$$\text{適合率α} = \frac{\text{本棚に格納した数}\ (C)}{\text{検索結果の数}\ (A)} \qquad (2)$$

$$\text{適合率β} = \frac{\text{実データを表示した数}\ (B)}{\text{検索結果の数}\ (A)} \qquad (3)$$

式 (3) は，提供された情報の中から利用者が興味を示す割合の調査に用い，式 (2) は，実際にデータを見て，利用者が後日必要であると判断し，本棚に保存すると判断する割合の調査に用いる．

実験では，あらかじめカテゴリを "computer" に設定し，キーワードは自由に入力してもらった．ただし，PVL 利用時と検索エンジン（Infoseek，Lycos）利用時では同じキーワードを用いて繰り返し検索を行った．

検索結果の中から，利用者独自の判断で，要・不要を判断し，表示する実データの決定と，本棚に保存するデータの決定を行った．ただし，キーワードによって大量の検索結果を返すものがあるため，上位 30 件の結果を対象として実験を行った．

4.2.2 実験結果

【実験 1：PVL の有効性】
PVL の有効性を示すため，格納に関する適合率α，表示に関する適合率βを用いて既存の検索エンジンと比較した．図 4 左に 10 人の利用者の格納に

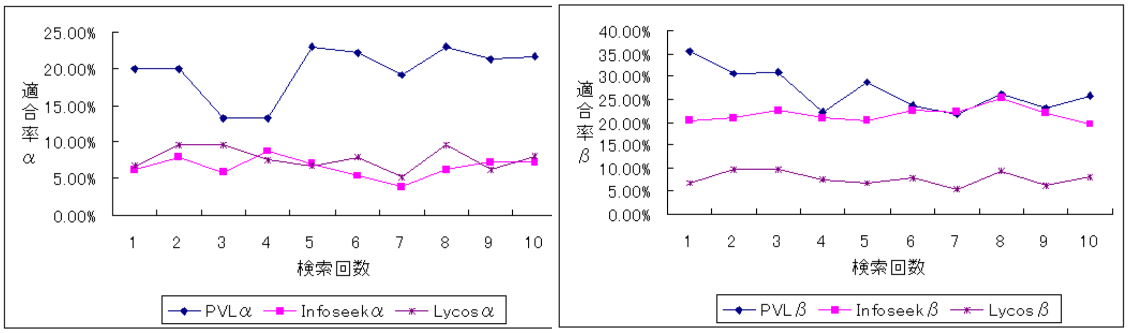

図 4: 実験 1 による適合率α，βの推移

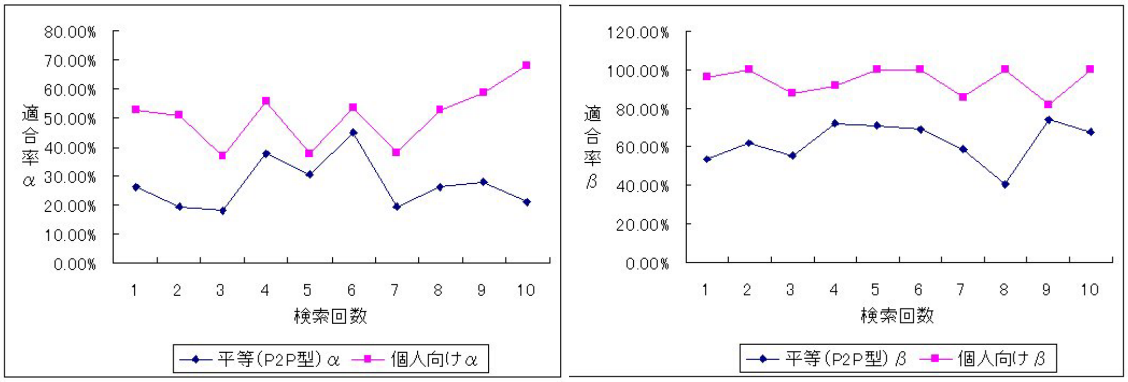

図 5: 実験 2，3 による適合率α，βの推移

関する適合率αの平均値の推移を示す．また，図 4 右では表示に関する適合率βの推移を示す．図 4 縦軸は適合率α，βを示し，横軸は検索回数を示す．また，◆は PVL 利用時，■，＊が各々検索エンジン（Infoseek，Lycos）に対応するエージェントのみを利用した時の適合率α，βの推移を示す．Infoseek，Lycos の実験は，それぞれ 1 つのアクセスエージェントだけを動作させて実験を行った．この結果より，PVL は既存の検索エンジンと比較して適合率α，βともに増加している．しかし，一部前の操作と比較して適合率の減少が見られ，利用を進めるに従って常に適合率が上昇しているとは言い難い．これらは検索キーワードにより適合率の増減があると考えられる．そこで，フィードバック効果を確認するため，各利用者が 10 回の検索終了後に，再び 1 回目の検索時に用いたキーワードによって検索を行った．1 回目と 11 回目の適合率の比較を表 2 に示す．

本実験を行うことにより，キーワードに依存した適合率の増減をなくし，本手法の有効性を確認した．本実験により，適合率βは約 4.6%減少したが，適合率αは約 28.6%増加することを確認した (表 2)．表 2 において適合率βが減少しているのは，被験者が過去の試行においてブラウズした実データを記憶しており，不適合と判断したデータを，以降の試行では表示しなかったことが原因と推測される．

表 2: 同じキーワードを用いた 1 回目と 11 回目の適合率α，βの変化

	1 回目	11 回目	変化率
適合率α	11.67 %	15.00 %	約 28.6 %増加
適合率β	30.56 %	29.44 %	約 4.6 %減少

【実験 2：P2P 型推薦機構の有効性】

P2P 型ネットワーク上に PVL のための推薦機構を実装した．予備実験として，個別化を行わずすべての被験者が提供する情報を用いて評価を行っ

表 3: 各手法による適合率α，βの比較

	適合率α	適合率β
Infoseek	6.28 %	19.39 %
Lycos	6.94 %	19.28 %
PVL(単一利用)	10.50 %	25.44 %
P2P 型推薦機構	27.03 %	62.54 %
個人向け推薦機構	50.40 %	94.3 %

た．P2P 型推薦機構の有効性を示すため，利用者全員に推薦要求を伝播して実験を行った (図 1 左の形式)．図 5 に適合率α，βの推移を示す．左が格納に関する適合率α，右が表示に関する適合率βを示し，それぞれ◆が P2P 型推薦機構を表す．

表 3 は，PVL，既存の検索エンジン，P2P 型推薦機構，個人向け推薦機構により得た適合率α，βの値を示す．ただし，値はそれぞれ 10 回検索を行った後の平均値を用いた．適合率α，βは 27.03%，62.54%となり，P2P 型推薦機構を用いた場合，従来型の PVL 利用時と比較して，適合率α，β供に約 2.5 倍上昇することがわかる．

従って，P2P 型推薦機構を利用することにより，利用回数に関わらず，洗練された情報を提供することが可能になり，有効性が示される．

【実験 3：個人向け推薦機構の有効性】
個人向け推薦機構の有効性を示すため，図 1 右の形式で要求の伝播を行った．図 5 に適合率α，βの推移を示す．左が格納に関する適合率α，右が表示に関する適合率βを示し，それぞれ■が個人向け推薦機構を表す．

図 5 により，個人向け推薦機構は常に P2P 型推薦機構より高い適合率α，βを保持していることがわかる．

また，表 3 より，10 回検索を行った後の適合率α，βの平均値は，50.40%，94.3%となり，従来型の PVL 利用時，P2P 型推薦機構利用時と比較し，適合率αが約 4.8 倍，約 1.9 倍上昇し，適合率βが約 3.7 倍，約 1.5 倍上昇することがわかる．

上記により，本稿で提案する個人向け推薦機構は，既存の検索エンジン，従来型 PVL，P2P 型推薦機構と比較して，利用者にとって効果的な情報を提供することが可能であると言える．

5　関連研究との比較

インターネット上の情報量の増加により，推薦システムが重要視されている．

本システムは P2P 型通信機構を用いて，不特定多数の利用者を対象とし，利用者間の類似度を考慮してリクエストの伝播先を決定した，個人向けの推薦システムである．本システムの推薦機構による情報収集機構は，内容フィルタリング方式ではなく，協調フィルタリング方式に相当する．この点に関しても不特定多数の利用者を対象とした方式となっている．

Singh らは，P2P パラダイムにおいてピア間の相互信頼度と各ピアの特性ベクトルを定義し，類似性をコサイン尺度により評価を行い，ピア集合全体の動きを推定する研究を行っている [4]．我々は，実際にピアから得た情報を推薦情報として利用し，独自に定義した複数の適合率を用いて評価を行っている．

PowerBookmark[5] も個人のブックマークを，インターネット上の他の利用者と共有することを可能にしている．コンテンツに基づいて自動的にブックマークの分類を行う．また，利用頻度の高いページを自動的にブックマークに追加するため，プロキシにより利用者の行動を監視している．

ウェブナビゲーター [6] は，閲覧したホームページの特徴と利用者プロファイル内の情報を合成することにより，利用者の興味の推移に合わせたホームページの推薦を行うことを可能にしている．

WAVE[7] は電子コミュニティ上に口コミを再現し，人々の評価と伝播を得ることで有用な情報のみにフィルタリングされている．WAVE では，利用者の閲覧履歴や，情報提供者への評価，情報提供者のその情報に対する評価，情報伝達者数などを用いて評価を行い情報を推薦する．

6　おわりに

本稿では，P2P 型ネットワークを基盤とした PVL のための推薦機構を提案した．また，プロトタイプシステムを用いた評価実験により，本システムの有効性を検証した．

PVL は単一利用者の利用を対象としているが，本推薦機構は，不特定多数の利用者を対象としている．不特定多数の利用者の中から，利用者間の類似度を用いて自動的にコミュニティを形成し，コミュニティによる洗練された情報の提供を可能に

した．

プロトタイプシステムを用いた評価実験により，本推薦機構を利用することで，格納に関する適合率が 50.4%，表示に関する適合率が 94.3%となり，従来型 PVL(単一利用)，P2P 型推薦機構を利用した場合と比較し，適合率 α が約 4.8 倍，約 1.9 倍上昇し，適合率 β が約 3.7 倍，約 1.5 倍上昇した．

今後の課題として，(1) リクエストの伝播範囲の違いによる適合率の追跡，(2) 直接的なリクエスト伝播先の決定による高精度システムの実現などが挙げられる．

参考文献

[1] Special issue on Personalization, Communications of the ACM, Vol.43, No.8, 2000.

[2] 上田　真由美，上島　紳一，“エージェントの得点付け手法により個別化を行うプライベート仮想ライブラリ PVL の提案と評価”，情報処理学会論文誌:データベース，Vol.43, No.SIG12(TOD16), pp.102–114, 2002.

[3] 上田　真由美，近藤　育雄，上島　紳一，“プライベート仮想ライブラリ PVL のための P2P 型推薦機構”，データ工学ワークショップ 2002(DEWS2002)，B5-2，2002.

[4] Munindar P. Singh, et al, “COMMUNITY-BASED SERVICE LOCATION”, Communications of the ACM, Vol.44, No.4, pp.49–54, 2001.

[5] Wen-Syan Li, et al, “PowerBookmarks: A System for Personalizable Web Information Organization, Sharing, and Management”, Proceedings of the SIGMOD’99, pp.565–567, 1999.

[6] 九津見洋，内藤榮一，荒木昭一，江村里志，新居薫治，“ユーザ適応型ホームページ推薦ソフトウェア “ウェブナビゲータ” の開発”，電子情報通信学会論文誌，D-II Vol.J84-D-II No.6, pp.1149–1157，2001.

[7] 吉田匡志，伊藤雄介，沼尾正行，“口コミによる分散型情報収集システム”，第 10 回マルチ・エージェントと協調計算ワークショップ (MACC2001)，pp.61–68，2001.