

個人適応型 WWW におけるユーザモデル構築法

三浦 信幸[†] 高橋 克巳[†] 島 健一[†]

本論文では、個人適応型 WWW におけるユーザモデル構築法を提案する。個人適応型 WWW とは、従来一般的な WWW が、基本的には、すべてのユーザに対して同一の情報内容を同一の情報提示方式で提供してきたのに対し、ユーザモデルとコンテンツモデルの擦合せにより、ユーザ 1 人 1 人の特性に応じて提供する情報内容・情報提示方式をユーザごとに動的に変更しながら動作する WWW である。本提案手法の特徴を以下に示す。1) ユーザに余計な負荷となる事前アンケートは用いず、WWW へのアクセス履歴のみからユーザの特性を把握する。2) ユーザモデルをグラフで表現することにより、ユーザのアクセス行動の遷移をモデルに反映できる。3) WWW コンテンツ自身のハイパーリンク構造やコンテンツの意味的な関係を表現したコンテンツモデルとユーザのアクセス履歴を組み合わせてユーザモデルを構築することにより、より多様な個人適応が可能なモデルを構築できる。本論文ではユーザモデルとして、状態モデル・行動モデル・学習モデル・メタモデルの 4 つを提案し、主に状態モデルについて、実際に外部に公開中の 2 つの WWW サイトのアクセス履歴に適用し、約 4 万人分のユーザモデル構築を行い、構築結果の評価実験を行った結果、ユーザ間の相違点・共通点を適度に表現できること、従来法では見つかりにくいユーザ特性を見つけられること等を確認し、提案手法の有効性を示した。

A User-models Construction Method for Personal-adaptive WWW

NOBUYUKI MIURA,[†] KATSUMI TAKAHASHI[†] and KEN'ICHI SHIMA[†]

This paper proposes a user-models construction method for personal-adaptive WWW. Though Web system has provided same information in same layout to all users, personal-adaptive WWW adapts contents of information and how to display them to users' preferences. Features of our proposal method is as follows. 1) User-models are constructed from access histories, not using questionnaires. Questionnaires are excessive works for users. 2) User-models are represented as graphs in order to express users' activities. 3) User-models are constructed from access histories and WWW contents-models. As WWW contents-models express hyper-link structures of contents and semantic relationships among contents. These user-models enable more various personal-adaptations. We define four user-models, user-status model, user-activity model, user-learning model and user-meta model. We show experimental results of constructing user-models, especially user-status model, from access histories of real two public WWW sites. We constructed fourty thousand users' models and the results show effectiveness of our method in the viewpoints that we can get more various and more interesting users' profiles than what are produced by usual construction methods and that constructed user-models have both different points and common points.

1. はじめに

昨今の WWW の広がりにつれて、より質の高い、より高度で知的な WWW が求められる時代となってきた。我々は、そのような要求に応える WWW として、個人適応型 WWW を検討している。従来一般的な WWW が基本的には、すべてのユーザに対して、同一の情報内容を同一の情報提示方式で提供してきたのに対し、個人適応型 WWW は、ユーザの特性を表

現したユーザモデルと WWW コンテンツのハイパーリンク構造やコンテンツの意味的な関係を表現したコンテンツモデルとを擦り合わせるにより、提供する情報内容・情報提示方式等をユーザ 1 人 1 人の特性に応じてユーザごとに動的に変更し、ユーザに適応しながら動作する WWW である。本論文では、この個人適応型 WWW の実現に必要な要素技術の 1 つとして、WWW にアクセスするユーザのユーザモデルの構築法を提案する。

提案するユーザモデル構築法の特徴を以下に示す。

- (1) ユーザに余計な負荷となる事前アンケートは用いず、WWW へのアクセス履歴のみからユー

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

ザの特性を把握する

現在、パーソナライズド Web などという名前で、個人別対応を行いながら動作する WWW がいくつか導入されている[★]が、その多くが事前にユーザにアンケートを求め、その内容を利用する形式である。事前アンケート形式には、次のような問題点がある。

- ユーザに対して、通常の WWW へのアクセス以外に、アンケートに回答するという余計な負荷をかけてしまう
- ユーザが主体的にアンケート内容を変更しない限り、ユーザ特性の時間的な変化に追従していくことができない

提案手法では、これらの問題を解決するため、WWW へのアクセス履歴からユーザの特性を把握する。

WWW に限定しなければ、ソフトウェアのユーザインタフェースを動的に個人別対応するために、アンケートからユーザモデルを作るのではなく、ユーザの操作履歴からユーザモデルを作ることは古くから行われてきた(文献1など)。提案手法は、この後の特徴の(2)や(3)で述べるように、WWW が持つ最大の特徴、ハイパーリンク構造に着目してユーザモデルを構築する点が特徴である。

- (2) ユーザモデルをグラフで表現することにより、ユーザのアクセス行動の遷移をモデルに反映できる

いくつかの研究^{2),3)}では、ユーザ特性の把握をアンケートではなく、アクセス履歴から抽出する試みが行われているが、その方法としては、それぞれのコンテンツ(URL)にキーワードを付与し、ある URL を見たら、そのキーワードに関する重みを操作するもので、ユーザの1つ1つの行動をそれぞれ単独でモデル化しているにすぎない。

提案手法では、さらに、ある URL を見たという行動の次にどのような行動を起こしたか(どの URL を見たか)というユーザの行動の遷移もモデル化するために、ユーザモデルをグラフで表現する。

従来法^{2),3)}では、新聞記事の情報フィルタリングに代表されるような、比較的単純な仕組みの

ものや、用途の限定された個人適応しか行うことができないが、提案手法を用いることにより、ユーザの行動パターンに合わせた、種々の情報提供・ナビゲーションなど、従来法よりも、多様な、複雑な個人適応を行うことができる。

- (3) WWW コンテンツ自身のハイパーリンク構造やコンテンツの意味的な関係を表現したコンテンツモデルとユーザのアクセス履歴を組み合わせて、ユーザモデルを構築することにより、より多様な個人適応を可能とするモデルを構築できる

従来法^{2),3)}でも、ユーザモデル中に出現するキーワード群と、各 URL に対応づけるキーワード群とに、同一のキーワード群を用いることで、提案手法と同様のことを行っているが、キーワード相互の関係などは考慮されていない。提案するコンテンツモデルには、コンテンツ自身のハイパーリンク構造やコンテンツの意味的な関係などが表現されているため、それらを踏まえて作成されるユーザモデルを使うことにより、より多様な個人適応が可能となる。

本論文では、まず、2章で我々が検討している個人適応型 WWW の基本的な概念について述べる。次に、3章で個人適応型 WWW に必要とされるユーザモデルを検討し、4つのユーザモデル、状態モデル・行動モデル・学習モデル・メタモデルを提案する。4章でその具体的な構築方法を述べる。さらに、5章で提案するユーザモデル構築法の評価として、4つのユーザモデルのうち、状態モデルについて、実際に外部に公開中の2つの WWW サイトのアクセス履歴から、提案手法を用いて、合わせて約4万人分のユーザモデルを生成し、作成されたユーザモデルを評価する評価実験を行う。最後に6章で、まとめを述べる。

2. 個人適応型 WWW の基本的な概念

本章では、個人適応型 WWW の基本的な概念を、基本的な考え方、構成、例の順に述べる。

2.1 個人適応型 WWW の基本的な考え方

個人適応型 WWW は、次のような特徴を持つものである。

- ユーザになるべく余計な作業負担をかけない
従来の個人別対応 WWW サイトのほとんどは、事前に、キーワード群の中から自分の興味等を選択して事前アンケートとして提出する方式である。この方式は、ユーザに余計な作業負担を強いるものであり、あまり望ましくない。我々の個人適応

★たとえば、My Yahoo (<http://my.yahoo.com/>)、MSN (<http://www.jp.msn.com/>) など

型 WWW では、そのような事前アンケートを基本的に不要にするため、ユーザのアクセス履歴からユーザの特性を自動的に取得し、サーバ自身に個人に適応させる。

- ユーザの時間変化に追従できる
事前アンケート形式では、特にユーザが自主的にアンケート内容を変更しない限り、個人対応形式は変化しない。個人適応型 WWW では、サーバはつねにユーザの最新の特性に合わせて個人適応を行う。
- ユーザの行動パターンや学習状況等にも適応
アンケート形式では、ユーザの興味・関心等の、いわば、ある特定の瞬間のユーザの状態のみにしか対応ができない。個人適応型 WWW では、ユーザの行動パターンやユーザの該当 WWW のコンテンツ・機能に関する学習状況の特性についても個人適応を行う。

2.2 個人適応型 WWW の構成

2.1 節で述べたような特徴を満たすための個人適応型 WWW の構成を以下に示す。

WWW の世界では、サーバのコンテンツ提供者とコンテンツをブラウズする利用者が存在する。両者はそれぞれ、一般に複数人であり、各個人はそれぞれ異なる背景知識や興味を持つ。つまり、複数人のコンテンツ提供者と複数人の利用者の間には、提供されているコンテンツの認識に関して、何らかのギャップが存在している可能性がある。そのギャップを埋めることこそが、個人適応である。したがって、まず、コンテンツ利用者のモデル（ユーザモデル）とコンテンツ提供者のモデル（コンテンツモデル）を作成することが必要である。この両者のモデルを擦り合わせ、どのように個人適応するか、すなわち、コンテンツをどのように加工してユーザに提示するかルールを作成することにより、個人適応が可能になる。個人適応ルールの生成には、特定のある 1 人のユーザのユーザモデルとコンテンツモデルを擦り合わせてそのユーザ用のルールを生成する方法と、複数のユーザのユーザモデルどうしの擦合せ結果とコンテンツモデルとを擦り合わせる方法、の 2 種類が考えられる。後者は、文献 6) などでも採用されている Social Filtering,あるいは、Collaboration Filtering と呼ばれる考え方を利用したもので、他者の履歴を利用してコンテンツを特定のユーザ用に適応させる（情報 filtering する）方法である。ここでは、後者の方法と区別するために前者の方法を Personal Filtering と呼ぶ。図 1 に個人適応型 WWW の概念図を示す。

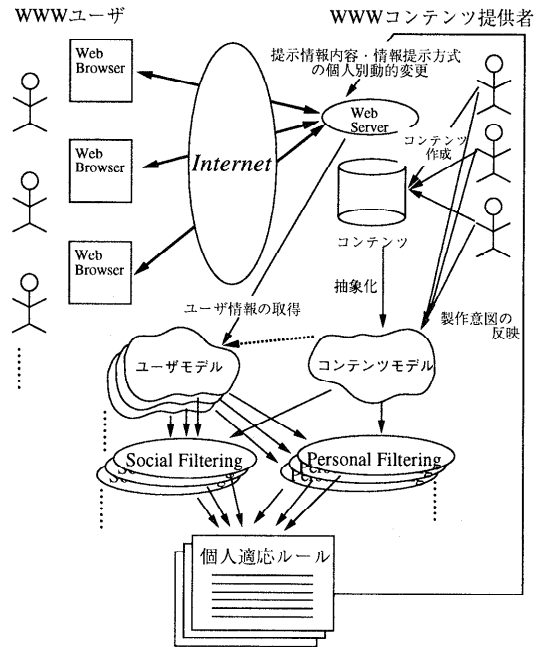


図 1 個人適応型 WWW の概念図

Fig.1 Outline of Personal-Adaptive WWW.

ユーザモデルの構築に必要なユーザ情報の取得は、なるべくユーザに余計な作業・コスト等の負担をかけないこと、つまり、基本的には WWW サーバ側の作業のみで情報の抽出が可能なこと、また、ユーザの時間変化に追従できる情報がとれること、などの必要性から、ユーザの挙動をサーバ側で観察する方法を採用する。WWW の場合、サーバ側で観察できるユーザの挙動のほとんどは、WWW サーバ上にアクセス履歴が保存されており、これを利用することにする。

個人適応に際しては、文献 7) などで行われているように、クライアント側でユーザ情報を取得し、クライアント側で個人適応を行う方法も考えられるが、ユーザモデルの構築および WWW コンテンツの個人適応にあたっては、コンテンツの内容・構成・機能といった情報が不可欠であり、それらの最新の情報が豊富に蓄積されているサーバ側でユーザモデルの構築や個人適応を行う方が効果的であると考えられる。

2.3 個人適応型 WWW の個人適応の例

個人適応型 WWW で行う、個人適応の例を示す。題材として、我々が現在試験提供している English TOWNPAGE⁴⁾および Internet TOWNPAGE⁵⁾を想定する。これらの WWW は、職業別電話帳の内容を検索することができるとともに、広告情報、地図情報、各種観光情報、流行情報などを統合的に提供している。これらの WWW を題材とする場合、以下の

ような個人適応が可能な WWW の構築を考えている。

- ユーザ特有の関連業種を表示

ある業種について検索を行った後に、関連する業種一覧を表示する機能がある。この一覧は、一般的に関連があると思われる業種のリストである。これに加えて、ユーザの検索履歴の情報を基に、そのユーザ独自の関連業種を優先的にリストアップする。

- 行動パターンとユーザの興味を考慮した広告表示

多くのユーザが、引越業者の検索の後に、家具屋の検索を行っていたとする。また、あるユーザは輸入物に関する検索をよく行っていたとする。このような状況において、このユーザが引越業者の検索を行った際に、輸入家具屋の広告を表示する。

- ユーザのコンテンツに関する理解度に応じた機能ガイド

あるユーザが、この WWW サイト特有の機能である、検索結果を地図で表示できる機能をまったく使っておらず、この機能の存在を知らないと予想されるとする。地図機能が有効に機能すると考えられる場合にこのユーザがたどり着いたとき、地図機能の存在をアピールする文面等を画面に追加する。

3. 個人適応型 WWW に必要とされるユーザモデル

2章で述べた個人適応を行うためには、作成されるユーザモデルは、次のような条件を満たす必要がある。

条件 1 ユーザの興味や関心といった、ユーザの状態を表現していること

条件 2 ユーザの行動パターンを表現していること

条件 3 ユーザの該当 WWW のコンテンツ・機能に関する理解状況・使用度合などを表現していること

条件 4 コンテンツモデルとの擦合せにより、Social Filtering, Personal Filtering の個人適応ルールの生成が可能な表現形態であること。また、そこでできるルールが有効に働く可能性があること。

条件 1~3 に基づいて、状態モデル・行動モデル・学習モデル・メタモデルの 4 つのユーザモデルを提案する。これらのモデルは図 2 のような関係にある。条件 4 については、5 章において評価を行ってその条件を満たしていることを確認する。

4 つのユーザモデルはそれぞれ、次のことを表現する。

- 状態モデル：ユーザの興味・関心等のユーザの状

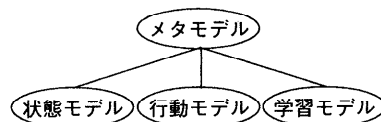


図 2 4 つのユーザモデルの関係

Fig. 2 Relationship among 4 user-models.

態の情報 (条件 1 より)

- 行動モデル：ユーザの該当 WWW における行動パターン等のユーザの行動に関する情報 (条件 2 より)
- 学習モデル：該当 WWW の構造・提供している情報の種類・機能等をユーザがどの程度、理解しているかの情報 (条件 3 より)
- メタモデル：ユーザがどのような状態のときにどのような行動をとり、どのようなことを学習したかといったエピソード的な情報 (上記 3 つのモデル間の時間的関係を保持するため)

また、ユーザモデルの構築およびその解釈 (個人適応) は、次の仮定に基づいて行う。

- アクセス回数の多いコンテンツ (URL) は、そのユーザにとって興味や関心の高いコンテンツ (URL) である
- 連続してアクセスされた任意のコンテンツ (URL) は、そのユーザにとって相互に興味や関心の高いコンテンツ (URL) の組である
- 前後のコンテンツ (URL) にアクセスした時刻の差が短いほど、そのユーザにとって相互に興味や関心の高いコンテンツ (URL) の組である。時間間隔が広ければ、そのユーザにとって前後のコンテンツ (URL) 間の関連性は低いものと予想される。逆に、時間間隔が短ければ関連性は高いものと予想される。

4. ユーザモデル構築法

2章で述べたように、個人適応を行う場合、ユーザモデルとコンテンツモデルとの擦合せを行う。その際、擦合せを効率的に行うためには、ユーザモデルを、コンテンツモデルの構造に近い形で作成する必要がある。そこで、まず、WWW のコンテンツモデルについて検討する。

WWW のコンテンツはハイパーリンクという構造をすでに持っていること、コンテンツどうしの意味的な関係は、一般に、ある程度複雑な構造を持つことが予想されることの 2 点から考えて、コンテンツモデルは、グラフで表現しておく必要があると考える。このグラフは、次の 2 つの要素から構成する。

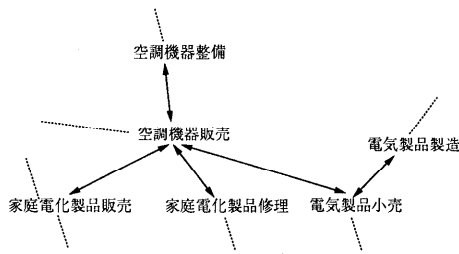


図3 電話帳 Web の業種別検索のコンテンツモデルの例

Fig. 3 An example of contents model in case of category search of telephone directory.

ノード URL を要素とする集合

リンク 任意の上記ノード間を結んだもの

このようなグラフ表現のコンテンツモデルの構築法の概略は次のとおりである。

- (1) 以下のような基準に従って、URL 群であるノードを構成する。
 - コンテンツ提供者が何らかの意味的なまとまりを持っていると判断できる URL 群
 - その URL によって提供される情報が同一のキーワード、ないしは、シソーラスなどにより類似していると判断されるキーワードを含んでいる URL 群
 - その他、ある意味的なまとまりを持っていると判断可能である URL 群
- (2) 上記で構成したノード間に以下のような基準に従ってリンクを張る
 - あるノード内の要素の URL と別のノード内の要素の URL とにハイパーリンクが張られている場合
 - URL の集合であるノードに対して、その URL 群が提供している情報内容に関するラベルを付与した場合に、そのラベルどうしに意味的な関連性がある場合
 - その他、互いに意味的な関係が存在すると判断可能である場合

たとえば、2 章で題材にした職業別電話帳の WWW の場合を考えると、図 3 のようなグラフが考えられる。これは、検索条件中にある業種名が含まれている、すなわち、ある業種に関する検索を実行したという URL 群をノードとして、関連のある業種をリンクしたものである。

一方、WWW サーバ上に保存されているアクセス履歴に関しても、コンテンツモデル同様、以下のよう

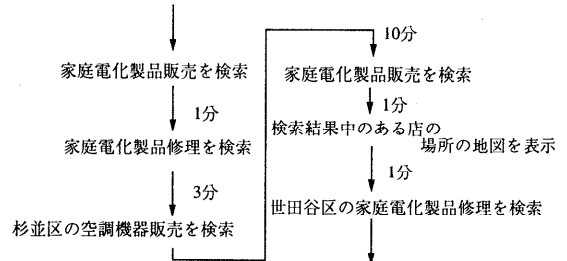


図4 職業別電話帳 WWW のアクセス履歴の例

Fig. 4 An example of access histories in case of category search.

な要素からなるグラフとして表現できる。

ノード ユーザがアクセスした URL

リンク 上記ノード間を時系列で結んだもの。リンク

には、アクセス時間間隔を属性値として付与する。

たとえば、職業別電話帳の WWW の場合を考えると、図 4 のようなものである。

このように、ユーザモデル構築に必要なコンテンツモデル・アクセス履歴はともにグラフで表現することができる。また、ユーザモデルはこれら 2 つから構築されることから、ユーザモデルもグラフで表現することが適切である。ゆえに、ユーザモデル構築方法としては、コンテンツモデル・アクセス履歴をグラフ表現したものに対し、必要なグラフ操作を施し、グラフ表現されたユーザモデルを構築する方法を採用する。また、このことにより、コンテンツモデルとユーザモデルとの擦合せによる個人適応ルールの生成も、コンテンツモデルのグラフとユーザモデルのグラフとに対するグラフ操作によって行うことができる。ユーザモデルのグラフは以下の要素から構成される。

ノード コンテンツモデル中のノードの全部または一部。

リンク ユーザが上記ノード間をアクセスした場合に、上記ノード間を結んだもの。

構築方法の概要は図 5 のように表せる。

では、4 つのユーザモデルをどのように構築するかについて、個々に述べる。

4.1 状態モデルの構築法

3 章で掲げた仮定に基づいて、ノード (URL の集合) に対する興味や関心の高さをノードに対するアクセス頻度、前後のノード間の興味や関心の高さをリンクに対するアクセス頻度と定義し、アクセス回数が多いノードはノードのアクセス頻度が高いとし、たどられた回数が多いリンクや短時間の間にたどられたリンクはリンクのアクセス頻度が高いとするようなモデル構築を行う。ユーザモデルを使って個人適応を行う際

* URL には、CGI の引数も含めて考える。

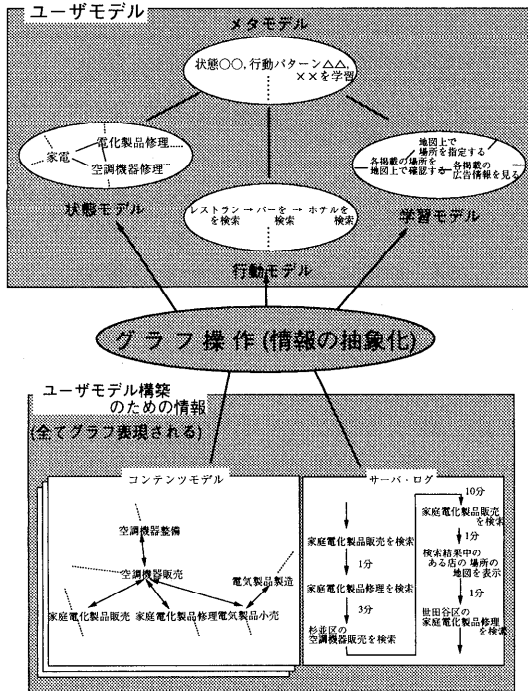


図5 ユーザモデル構築方法の概要

Fig. 5 Outline of a methodology of the user-models construction.

には、アクセス頻度が高いノードやリンクはユーザにとって興味・関心が高いものであると解釈して、個人適応ルールを生成する。

ここでは、電話帳 WWW の業種別検索に関する状態モデル作成を例として説明する。コンテンツモデルの例として図3を、アクセス履歴の例として図4を考える。この2つのグラフに対して、次のような手順に従って、電話帳 WWW の業種別検索の状態モデルを作成する。

まず、コンテンツモデル、アクセス履歴、ユーザモデルのグラフを次のように定義する。

- コンテンツモデル
ノード $C_i (i = 0 \dots n)$, ノード C_i から C_j へのリンクを $lc_{i,j}$ とする。
- アクセス履歴
時系列順に並んだノード $a_i (i = 0 \dots k)$, ノード a_i から a_{i+1} へのリンクを $la_{i,i+1}$, ノード a_i から a_{i+1} までの時間間隔 $t_{i,i+1}$ をリンク $la_{i,i+1}$ の属性値とする。
- ユーザモデル
ノード $u_i (i = 0 \dots n)$, ノード u_i から u_j へのリンクを $lu_{i,j}$ とする。また、ノード u_i のアクセス頻度を nf_i , リンク $lu_{i,j}$ のアクセス頻度を

$lf_{i,j}$ とする。

これに従って、次のようにモデルを構築する。

- (1) ノードのアクセス頻度の計算

アクセス履歴のグラフ上のノード $a_p (p = 0 \dots k)$ のうち、コンテンツモデルのグラフ中のノード C_i に属するノード数をカウントし、その合計で平均化したものをアクセス頻度 nf_i とする。

$$nf_i = \frac{\sum_{p=0}^k \text{if } (a_p \in C_i) \text{ then } 1 \text{ else } 0}{\sum_{q=0}^n \sum_{p=0}^k \text{if } (a_p \in C_q) \text{ then } 1 \text{ else } 0}$$

たとえば、図4では、家庭電化製品販売の検索が2回、家庭電化製品修理の検索が2回、空調機器販売の検索が1回であるので、家庭電化製品販売 $2/(2+2+1) = 0.4$, 家庭電化製品修理 $2/(2+2+1) = 0.4$, 空調機器販売 $1/(2+2+1) = 0.2$ となる。

- (2) リンクのアクセス頻度 $lf_{i,j}$ の計算

まず、アクセス履歴のグラフを次のように、今、着目しているコンテンツモデルに合わせて変形する。

アクセス履歴のグラフ中のノードのうち、コンテンツモデルのノードに含まれるノードのみを残す。これにより、途中のノードが欠落してしまったリンクは、それらを1つのリンクに接続する。リンクを接続した際には、属性値であるノード間のアクセス時間間隔の和を接続されたリンクの属性値とする。

変形後のアクセス履歴のグラフを以下のように定義する。

時系列順に並んだノード $a'_i (i = 0 \dots m) (m \leq k)$, ノード a'_i から a'_{i+1} へのリンクを $la'_{i,i+1}$, ノード a'_i から a'_{i+1} までの時間間隔 $t'_{i,i+1}$ をリンク $la'_{i,i+1}$ の属性値とする。

この変形後のアクセス履歴のグラフ上のリンクの両端のノードをコンテンツモデル中のノードに対応づけて、アクセス履歴のグラフ上のリンクについての属性値であるアクセス時間間隔の逆数の総和をリンクのアクセス頻度 $lf_{i,j}$ とする。

$$lf_{i,j} = \sum_{p=0}^{m-1} \text{if } ((a'_p \in C_i \wedge a'_{p+1} \in C_j) \vee (a'_p \in C_j \wedge a'_{p+1} \in C_i)) \text{ then } \frac{1}{t'_{p,p+1}} \text{ else } 0$$

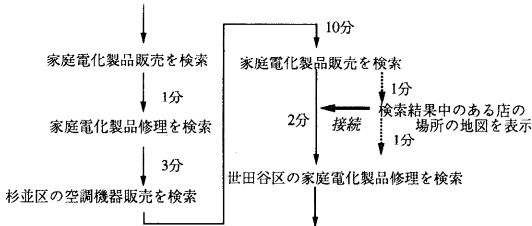


図 6 アクセス履歴のグラフの変形例

Fig. 6 An example of changing access history graphs.

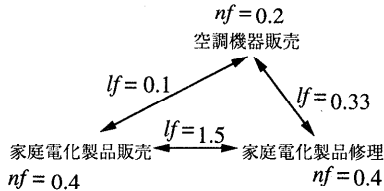


図 7 抽出された状態モデルの例

Fig. 7 An example of user-status models.

$lf_{i,j}$ は、そのリンクがたどられた回数が多いほど、また、たどられた際のアクセス時間間隔が短いほど、大きな値をとる。

たとえば、図 4 では、図 3 のようなコンテンツモデルに対してユーザモデルを作る場合、このコンテンツモデルに合わせてユーザモデルを変形すると、「検索結果中のある店の場所の地図を表示」というノードが欠落して、図 6 のようなグラフに変形される。変形済みのグラフを利用して、リンクのアクセス頻度を計算すると、家庭電化製品販売—家庭電化製品修理 $1/1 + 1/2 = 1.5$ 、家庭電化製品修理—空調機器販売 $1/3 = 0.33$ 、空調機器販売—家庭電化製品販売 $1/10 = 0.1$ となる。

(3) ユーザモデルのグラフの構築

以下の要素からなるグラフをユーザモデルとする。以下の条件に適合しないノードやリンクはユーザモデルに含めない。

ノード u_i : $nf_i > 0$ のノード

リンク $lu_{i,j}$: 両端のノードが $nf_i > 0$ かつ $nf_j > 0$ であるリンク

以上の作業で抽出されたユーザの状態モデルは、図 7 のようになる。

このようにして構築された状態モデルは、コンテンツ提供者が想定しているコンテンツ・モデルの中から、各ユーザの特性と一致する部分を抽出し、さらにコンテンツモデルには存在しなかったユーザ独自の特性を追加したグラフであり、状態モデルの構築はコンテン

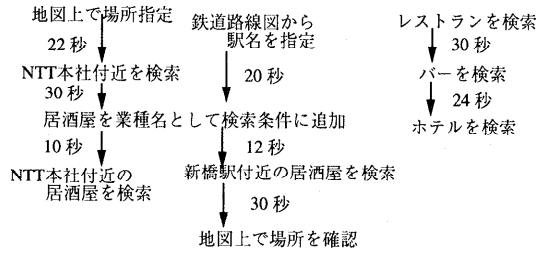


図 8 抽出された行動モデルの例

Fig. 8 An example of user-activity models.

ツ・モデルとアクセス履歴のグラフの重ね合わせの操作である。

図 7 の例でいえば、得られたグラフはもとのコンテンツ・モデルのグラフの一部に相当するものであると同時に、家庭電化製品販売と家庭電化製品修理との間のリンクを加えたグラフである。新たにリンクを加えたことが、コンテンツ・モデルにユーザ特性を追加したことに相当する。

4.2 行動モデルの構築法

アクセス行動履歴のグラフ上のノードがコンテンツモデル中のどのノードの要素となるかを調べ、アクセス行動履歴のグラフ上のノードを対応するコンテンツモデル中のノードに置き換えたグラフを作成し、その遷移列から、頻出する行動パターンを抽出し、それを行動モデルとする。頻出行動パターンの抽出アルゴリズムとして、文献 8)、9) などで使われている、MRP (Maximal Repeating Pattern) 抽出アルゴリズムを用いて、頻出する行動パターンを抽出する。

抽出された行動パターンには、各行動の間に費やされている時間の平均値を付与する。

以上の作業で抽出されたユーザの行動モデルは、図 8 のようになる。

4.3 学習モデルの構築法

まず、該当サイトのコンテンツに関して、ユーザが理解を深めていくと思われる、該当 WWW が持つ特徴的な構造・機能、提供している情報の種類などを使用したという URL の集合をコンテンツモデルのノードとして表現し、ノード間に関係がある場合にリンクを張った、学習モデル構築用のコンテンツモデルを構築する。

たとえば、Internet TOWNPAGE では、地図上から検索したい場所を指定したり、検索結果の個々の掲載について地図上で場所を確認したりできる。また、この 2 つの項目はともに地図機能の利用という意味で互いに関係を持っているため、コンテンツモデル上ではリンクを張る。このようにして作った、Internet TOWNPAGE の学習モデル用のコンテンツモデルの

一部を図9に示す。

このようなコンテンツモデルとアクセス履歴のグラフを組み合わせ、状態モデル構築と同じ方法で、学習モデルを構築する。作成されたユーザの学習モデルは、図10のようになる。

4.4 メタモデルの構築法

メタモデルは3つのモデル間の時間的関係を表現するモデルである。

具体的には、行動モデル中に学習モデル用コンテンツモデルの中のノードが出現したとき、そのノードにアクセスされた時間をもって、3つのモデル間の関連づけを行う。

たとえば、図8と図10を見ると、地図上で場所を指定したというノードが両方のモデルに出現している。このとき、地図上で場所を指定した時刻をもって、3つのモデルを関連づけ、関連づけられたグラフをメタ

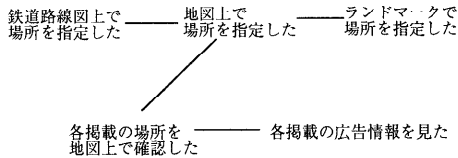


図9 An example of contents models for user-learning models.

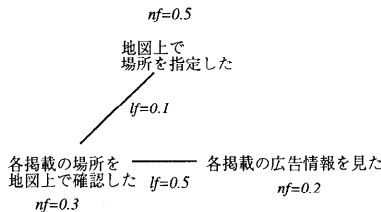


図10 抽出された学習モデルの例
Fig. 10 An example of user-learning models.

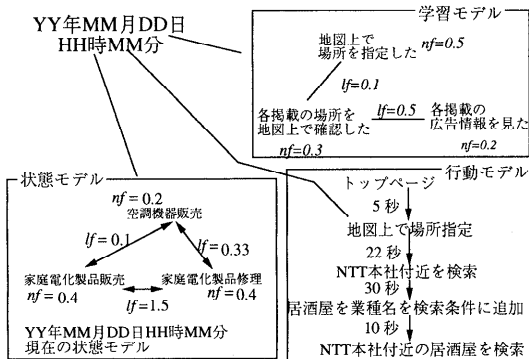


図11 抽出されたメタモデルの例
Fig. 11 An example of user-meta models.

モデルとして保持する。

作成されたメタモデルは、図11のようになる。

5. ユーザモデル構築法の評価実験

本章では、提案した個人適応型WWWのためのユーザモデル構築法の評価として、評価実験の内容とその結果について述べる。

5.1 評価実験の内容

評価実験は、外部に公開中のある2つのWWWサイトに対して、コンテンツモデルを作成し、アクセス履歴と組み合わせ、合わせて約4万人分のユーザモデルを構築し、構築結果を評価する。ここでは、4つのユーザモデルのうち、主に状態モデルについて、実験を行った。構築されたユーザモデルを3章の条件4を踏まえて評価する。実験の概要を以下に示す。

● 実験対象WWWサイト

English TOWNPAGE⁴⁾,

およびInternet TOWNPAGE⁵⁾

本WWWは、英語版および日本語版の職業別電話帳WWWであり、次の点から、本実験に適したWWWサイトである。なお、これらのWWWは、記述されている言語が異なるのみではなく、提供している電話帳の掲載情報も含め、まったく異なるサイトである。

- 提供している情報が比較的明解に分類可能であり、客観的なコンテンツモデルが作成しやすい。
- このWWWには、ユーザは、通常、具体的な目的を持ってアクセスしてくるため、そこから作られるユーザモデルは個人適応を行うための適切なモデルになりうる。
- 電話番号の掲載情報等の本来の情報の他に、広告や日本の観光案内といった副次的な情報も含まれており、多様なユーザの性質を把握するのに適している。
- English TOWNPAGEは1995年10月から、Internet TOWNPAGEは1996年12月から、Internet上で一般ユーザに対して試験提供しており、多数の実際のユーザのアクセス履歴を利用することが可能である。
- 我々の研究10)~12)におけるユーザモデル構築実験において、確かにユーザの特性のいくつかの傾向が見つかっている。

● 実験対象ユーザ

English TOWNPAGEについては1996年9月~1997年5月までにアクセスしたユーザの中か

ら 4351 人, Internet TOWNPAGE については 1996 年 12 月~1997 年 5 月までにアクセスしたユーザの中から 36423 人, 合わせて約 4 万人を無作為に抽出し, 実験対象ユーザとした. なお, 各ユーザの平均検索回数は, English TOWNPAGE のユーザが約 42 回, Internet TOWNPAGE のユーザが約 20 回であった.

● 実験目的

3 章の条件 4 に沿って, 作成されたユーザモデルを評価し, 提案するユーザモデル構築手法の有効性を示す. 具体的な評価点は次のとおりである.

評価点 1 有限かつ実用的なデータサイズでユーザモデルを表現できていること. ここでは, この評価尺度として, ユーザモデルがアクセス履歴よりもファイル容量が小さくなっていることとする.

評価点 2 作成されたユーザモデルがコンテンツモデルと異なる部分を持つことにより, Personal Filtering を行う際に必要な, コンテンツ提供者と個々のユーザとの間に存在するギャップを表現していること

評価点 3 作成されたユーザモデルがユーザごとに共通な部分を持つことにより, Social Filtering を行う際に必要な, ユーザのクラスタリング, グループング, マッチメイキングが可能となっていること

評価点 4 作成されたユーザモデルは, 従来法で考えられるようなユーザモデルよりもユーザの多様な特性を表現できていること

評価点 5 作成されたユーザモデルとコンテンツモデルとの擦合せによって生成される個人適応ルールの生成が有効に働く可能性があること

● 実験方法

モデル構築ツールのプロトタイプを実装し, それをコンテンツモデルおよびアクセス履歴に適用して, 実際にモデル構築を行う. ここでは特に, 業種別検索および住所別検索における職業分類や住所分類に関するコンテンツモデルに着目し, 職業分類や住所分類に関するユーザの状態モデルのユーザモデル構築結果を評価する.

それぞれの WWW サイトのコンテンツモデルの規模を表 1 に示す.

5.2 評価実験結果および考察

【評価点 1】

ユーザモデルの容量とアクセス履歴の容量との比は,

表 1 コンテンツモデルの規模

Table 1 The scale of contents models.

		English TOWNPAGE	Internet TOWNPAGE
業種コンテンツ モデル規模	ノード数	532	1918
	リンク数	8213	37705
住所コンテンツ モデル規模	ノード数	65	7271
	リンク数	781	74074

表 2 ユーザモデルとコンテンツモデルの比較

Table 2 A comparison between user-models and contents models.

	English TOWNPAGE	Internet TOWNPAGE
業種別の状態モデル		
実験対象ユーザ数 [☆]	459	12607
平均リンク数/人	3.2	5.5
コンテンツモデルと重複しない リンク数の平均割合(%) / 人	73.1	78.8
住所別の状態モデル		
実験対象ユーザ数 [☆]	1141	7122
平均リンク数/人	7.6	6.2
コンテンツモデルと重複しない リンク数の平均割合(%) / 人	22.7	74.6

[☆]ユーザモデル中に 2 本以上のリンクを持たないユーザは実験対象外とした.

English TOWNPAGE で 1/50, Internet TOWNPAGE で 1/70 と, ユーザモデルの方がファイル容量が小さくなっている. 評価点 1 について, 提案手法は有効である.

【評価点 2】

コンテンツモデルと各ユーザモデルとを比較して, 相違点を調べた結果は, 表 2 のとおりである.

業種別の状態モデルでは, 7~8 割のリンクがコンテンツモデルのリンクとは一致せず, ユーザの検索行動のうちの 7~8 割はコンテンツ提供者側が一般に想定している検索行動とは異なる行動をとっていることをモデリングできた. 住所別の状態モデルでは, English TOWNPAGE と Internet TOWNPAGE で傾向が異なるが, English TOWNPAGE で 2 割強のリンク, Internet TOWNPAGE で 7 割強のリンクがコンテンツモデルのリンクと一致せず, ユーザの検索行動のうちの 2 割強ないし 7 割強は, コンテンツ提供者が一般に想定している検索行動とは異なる行動をとっていることをモデリングすることができた. また, English TOWNPAGE, Internet TOWNPAGE とで程度は異なるが, ユーザの行動とコンテンツモデルとを比較すると, 業種別検索での検索行動がコンテンツモデルに一致しない割合 (73.1%, 78.8%) よりも住所別検索での検索行動がコンテンツモデルに一致しない割合 (22.7%, 74.6%) の方が低く, 住所別検索での検索行動におけるユーザの独自性の方が薄いという傾向もモデリングすることができた. このように, コ

表 3 ユーザモデルの比較

Table 3 A comparison of user-models.

	English TOWN-PAGE	Internet TOWN-PAGE
業種別の状態モデル		
全出現ノード中、2人以上のユーザに共通するノードの種類割合 (%)	73.0	87.7
全出現リンク中、2人以上のユーザに共通するリンクの種類割合 (%)	28.9	22.3
住所別の状態モデル		
全出現ノード中、2人以上のユーザに共通するノードの種類割合 (%)	89.4	74.0
全出現リンク中、2人以上のユーザに共通するリンクの種類割合 (%)	56.8	24.7

コンテンツモデルとユーザモデルとは、確かに相違点が存在し、コンテンツ提供者とユーザとのギャップを確かに表現できており、評価点2について、提案手法は有効である。

【評価点3】

ユーザモデルどうしの共通点を比較した結果は、表3のとおりである。ここでは、全ユーザモデルに出現するノードやリンクを列挙し、それらのうち、2人以上のユーザに共通して存在するノードやリンクの割合を調べた。English TOWNPAGE, Internet TOWNPAGEともノードのうちの7割~9割弱は2人以上のユーザに共通して他のいずれかのユーザに共通して存在するノードであり、状態モデル中のノードに関してはユーザ同士の共通点が多く存在することが確認できた。一方、モデル中に出現するリンクの種類を調べると、その2割強~6割弱は2人以上のユーザに共通して存在するリンクであり、リンクについてもユーザ同士の共通点が存在することが分かった。このように、提案手法で作成されたユーザモデルはユーザ同士の共通点を持っているため、ユーザのクラスタリング、グルーピング、マッチメイキングといったことが可能であり、評価点3について提案手法は有効である。

【評価点4】

図12は、English TOWNPAGEのあるユーザの業種別検索に関する状態モデルである。このユーザは、約3カ月間に合計157回の電話番号検索を行っており、そのうち、業種名からの検索を87回行っていた。構築されたこのユーザの業種別検索に関する状態モデルには、11個のノード、コンテンツモデルと一致する15本のリンクと、一致しない8本のリンクが含まれている。

一方、提案手法のようなユーザモデリングではなく、1章で例にあげた従来のパーソナライズド Webなどが行っているアクセス履歴中のノードのみの意味的な階層構造のみを考慮した単純なユーザモデル構築法では、表4のようなモデルが構築され、このユーザは、

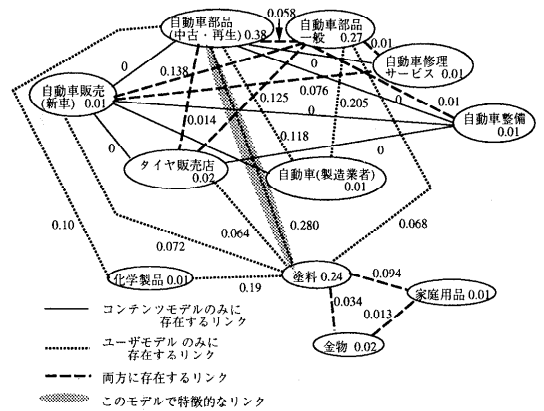


図12 ユーザモデル構築結果の一例
Fig.12 An example of user-status model of a user.

表4 従来のモデリング法によるユーザモデル例

Table 4 An example of user-model of a user by usual methods.

自動車	71%	自動車部品・修理・整備	67%	自動車部品(中古・再生)一般	38%
		自動車関連その他	4%	自動車修理サービス	27%
				自動車整備	1%
				タイヤ販売店	2%
				自動車販売(新車)	1%
				自動車(製造業者)	1%
日曜大工	27%			塗料	24%
				金物	2%
				家庭用品	1%
その他	1%			化学製品	1%

主に、自動車に関する業種と日曜大工に関する業種の検索を行っていると判断されることが多い。

しかし、今回提案する手法で得られたグラフ全体を見ると、塗料の業種を媒介にして、自動車に関する業種と日曜大工に関する業種とは強い関係を持っている。特に、“塗料”と“自動車部品(中古・再生)”との間のリンク(図12中の影付きのリンク)が最も頻りにたどられている。これらは、このユーザにとっては、この2種類の業種が何らかの関係を持つものと考えられていると解釈できる。このように、ユーザの行動を個々の行動だけでモデリングするのではなく、ある行動の次にどのような行動をとったかというリンクも含めた、本論文で提案しているモデリングの方が、より多様なユーザの情報を表現できているといえる。評価点4についても、提案手法の有効性を確認できた。

【評価点5】

提案手法で作成される状態モデルの場合、コンテンツモデルとユーザモデルに対するグラフ操作によって、

個人適応ルールを生成するための情報を抽出するには次のような方法がある。

- (a) ユーザモデルには存在しないがコンテンツモデルには存在するノードを取り出す
 たとえば、図 12 でいえば、ユーザがある地域のタイヤ販売店について検索を行ったが、該当する検索結果が非常に少なかった場合に、ユーザモデル上には存在しないノードであるが、コンテンツモデル上ではタイヤ販売店とリンクが張られている業種ノード、たとえば、自動車用品店を取り出し、その業種を検索条件に加えて検索することをユーザに勧めたり、あるいは自動的に検索条件に加えて検索を行う。
- (b) ユーザモデルには存在しないがコンテンツモデルには存在するリンクを取り出す
 たとえば、ユーザが過去にまったく検索を行ったことのない、新しい業種を検索しようとした場合、ユーザが入力する業種名はユーザの認識と異なる業種を指している場合がありうる。このとき、ユーザが主体的に関連業種を求めてきた場合には、このようなリンクを優先的にユーザに勧める。
- (c) コンテンツモデルには存在しないが、ユーザモデルのみに存在するリンクを取り出す
 たとえば、ユーザがある業種について検索を行った場合に、関連する業種の広告を表示するとき、用意していた関連業種には広告がなかったとする。このとき、特にリンクのアクセス頻度が高く、かつ、ユーザモデルのみに存在するリンクをたどって、その先のノードの業種に関する広告を表示する。図 12 でいえば、塗料が検索された際に、塗料・家庭用品・金物のいずれの業種にも広告がなかった場合に、自動車部品（中古・再生）の広告を表示する。

このような個人適応が有効に働く可能性を確認するためには、実際に上記のような検索条件の自動補完や動的な広告提示などのなんらかの個人適応を行ってその結果を検討する評価方法が考えられるが、本実験では、ある時点のユーザモデルは用意できても、その過去の時点でのユーザに対して個人適応を行うことはできないため、そのような評価はここでは困難である。また、本実験の目的は個人適応方法の評価ではなく、あくまで個人適応のためのユーザモデル構築方法の評価であるため、このような評価方法は個人適応方法の良否（ユーザモデルの使い方の良否）という外部要因が入り込んでしまうことになる。

そこで、これに代わる評価項目としてユーザのアク

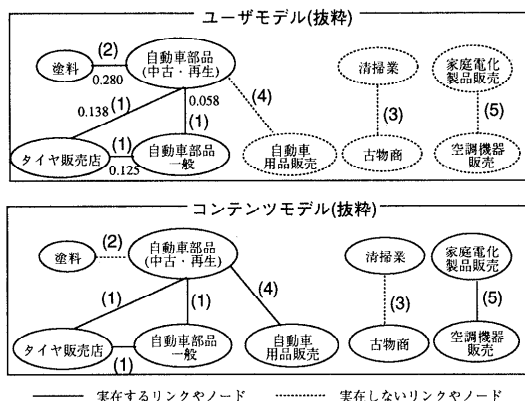


図 13 ユーザモデルとコンテンツモデルの比較結果に対応づけた、アクセス履歴の分類

Fig.13 A classification of users' access histories in the viewpoint of differences with user-models and contents models.

セス履歴の前半部分のみを使って作成したユーザモデルとコンテンツモデルとの比較結果と、残りの後半部分のユーザのアクセス履歴との相関を調べた。ユーザモデルとコンテンツモデルの比較の結果に登場するリンクと同一の行動をその後、ユーザがとっているとき、相関は高いと判断する。ユーザモデルとコンテンツモデルの比較結果は、そのユーザの特徴的な行動列や、そのユーザはまだ体験したことはないが今後行うと予測できる行動列である。そのような比較結果の行動列を基に個人適応ルールを生成し、個人適応を行う。したがって、このような行動が実際に行われたということは、行おうとする個人適応が非常に高い確率で有効に働く事例が存在したということにはほかならず、提案手法によるユーザモデルとコンテンツモデルとの比較による個人適応ルール生成の有効性を示すものである。

後半のユーザのアクセス履歴は、次の 5 種類に分類できる。図 13 はこの分類をユーザモデル・コンテンツモデル上で例示した図である。

- (1) ユーザモデルにもコンテンツモデルにも含まれるリンクと同一の行動。
- (2) ユーザモデルには含まれるが、コンテンツモデルにはないリンクと同一の行動。
 以前行った、ユーザ独自の行動と同一の行動に相当する。上記 (c) のような個人適応が有効に働く可能性を示す行動である。
- (3) ユーザモデルにもコンテンツモデルにも含まれないリンクと同一の行動。
 新しく行われたユーザ独自の行動に相当する。この段階では、個人適応が有効に働く可能性は

表5 ユーザモデルとコンテンツモデルの比較結果とアクセス履歴との相関
Table 5 A correlation between users' access histories and differences with user-models and contents models.

その後のユーザの行動に出現したリンクの分類		English TOWNPAGE	Internet TOWNPAGE	
ユーザモデルにある	コンテンツモデルにある (1)	7.4 %	1.8 %	
	コンテンツモデルにない (2)	6.7 %	3.6 %	
ユーザモデルにない	コンテンツモデルにない (3)	72.2 %	75.2 %	
	コンテンツモデルにある	片方のノードはユーザモデルにある (4)	7.5 %	8.9 %
		両方のノードともユーザモデルにない (5)	6.2 %	10.5 %
(2)+(4)+(5)		20.4 %	23.0 %	

ないが、この行動がユーザモデルに取り込まれた後の行動においては、上記(c)のような個人適応を行うことに役立つ可能性がある。

- (4) ユーザモデルには含まれないがコンテンツモデルに含まれるリンク、ただし、リンクの両側のどちらか片方のノードは、ユーザモデルに含まれているようなリンクと同一の行動。上記(a)のような個人適応が有効に働く可能性を示す行動である。
- (5) ユーザモデルには含まれないがコンテンツモデルに含まれるリンク、ただし、リンクの両側のノードとも、ユーザモデルに含まれないリンクと同一の行動。

上記(b)のような個人適応が有効に働く可能性を示す行動である。

である。したがって、上の(2)、(4)、(5)に相当する行動の存在が行おうとする個人適応が非常に高い確率で有効に働く事例の存在を示しており、後半のユーザのアクセス履歴の中でこれに相当する行動が確かにある程度存在していれば、評価点5について、提案手法が有効であるといえる。

ここでは、特に業種名に関する検索行動について分析を行った。評価結果は表5のとおりである。

後半のユーザのアクセス履歴のうち、2割程度の行動は、ユーザモデルとコンテンツモデルの比較結果に存在する行動列と同一のものであった。行う個人適応のうちの少なくとも2割は有効に働く可能性があるということになる。評価点2の結果にもあるように、本実験でのユーザの行動の2~7割程度は、コンテンツモデルとは一致しない、いわば、予測困難な行動であり、そのようなユーザの行動のうちの2割程度は個人適応が有効に働く可能性があるということであるので、評価点2の結果も考慮すると、この数字は提案手法によるユーザモデル構築とそれのコンテンツモデルとの比較による個人適応ルール生成の有効性を示すに足る

ものである。評価点5についても提案手法は有効であるといえる。

また、実際に明示的な検索行動としては現れない場合でも、ユーザの特性に合わせた広告表示などはユーザの潜在的な要求を満たす可能性があるため、この2割という数字は最低限、この程度は個人適応が有効に働くであろうという数字であり、実際に個人適応を行った場合には、個人適応が有効に働く場面はさらに多いと考えている。

6. おわりに

本論文では、ユーザのアクセス履歴を基に、提供する情報内容や情報提示方式をユーザごとに動的に変更しながら動作する、個人適応型WWWのために必要な要素技術の1つとして、WWWにアクセスするユーザのユーザモデルの構築法を提案した。提案手法には3つの特徴があり、1) ユーザに余計な負荷となる事前アンケートは用いず、WWWへのアクセス履歴のみからユーザの特性を把握すること、2) ユーザモデルをグラフで表現することにより、ユーザのアクセス行動の遷移をモデルに反映できること、3) WWWコンテンツ自身のハイパーリンク構造やコンテンツの意味的な関係を表現したコンテンツモデルとユーザのアクセス履歴を組み合わせることでユーザモデルを構築することにより、より多様な個人適応を可能とするモデルを構築できること、である。提案する4つのユーザモデル、状態モデル・行動モデル・学習モデル・メタモデルのうち、状態モデルについて、実際に公開している、ある2つのWWWサーバのアクセス履歴から合わせて約4万人分のユーザモデルを構築し、できあがったユーザモデルの評価を行う評価実験により、ユーザ間の相違点・共通点を適度に表現できること、従来法では見つかりにくいユーザ特性を見つけられること、作成したユーザモデルとコンテンツモデルの比較結果に類似する行動がその後のユーザの行動中に存在したことを

確認し、提案手法の有効性を示した。

今後、さらに以下の点について検討を行い、個人適応型 WWW のシステム化を行う。

- 行動モデル・学習モデル・メタモデル構築手法の評価実験
- 文献 13) などで検討されているような、ユーザの時間的な変化に、より積極的に追従するためのユーザモデル構築方法
- ユーザモデルのクラスタリング等による Social Filtering⁶⁾
- コンテンツモデルのグラフとユーザモデルのグラフに対するグラフ操作により、コンテンツモデルとユーザモデルとの擦合せを行うことによる、個人適応ルールの系統的な導出方法

謝辞 English TOWNPAGE, Internet TOWNPAGE の研究題材としての使用を快く承諾して下さいました。弊社電話帳事業推進部の皆様に感謝いたします。また、アクセス履歴データの解析に協力いただいた、東芝情報システム(株)の増田康明氏に感謝いたします。

参考文献

- 1) 永田ほか：個別利用者モデルに基づく対話的ソフトウェアのメッセージ生成，電子情報通信学会論文誌，Vol.J70-D，No.11，pp.2077-2082 (1987)。
- 2) Sagami, H., et al.: Learning Personal Preferences On Online Newspaper Articles From User Behaviors, *6th International WWW Conference* (1997)。
- 3) Morita, M., et al.: Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval, *17th ACM-SIGIR Conference*, pp.272-281 (1994)。
- 4) 島，高橋，三浦：インターネット版マルチメディア電話帳の構築，*Japan World Wide Web Conference '95*，日本インターネット協会 (1995)。
- 5) 島，高橋，三浦：インターネット・タウンページの構築 (1)～概要，第 54 回情報処理学会全国大会論文集，Vol.3，pp.481-482 (1997)。
- 6) Hill, W., et al.: Recommending and Evaluating choices in a virtual community of use, *CHI '95*, pp.194-201, ACM (1995)。
- 7) Barrett, R., et al.: How to Personalize the Web, *Conf. on Human Factors in Computer Systems* (1997)。
- 8) 岡田ほか：使いやすさ評価ツール“GUI テスタ”における共通操作ボタン抽出方式の提案と評価，情報処理学会研究報告，Vol.95，No.104，pp.37-42 (1995)。
- 9) Siochi, A.C., et al.: Computer Analysis of User

Interface Based on Repetition in Transcripts of User sessions, *ACM Trans. Information systems*, Vol.9, No.4, pp.309-335 (1991)。

- 10) 三浦，島：サーバアクセス履歴からのユーザモデルの構築，第 52 回情報処理学会全国大会論文集，Vol.1，pp.153-154 (1996)。
- 11) 三浦，島：WWW サーバ・ログからのユーザモデル構築実験，第 53 回情報処理学会全国大会論文集，Vol.4，pp.158-159 (1996)。
- 12) 三浦，高橋，島：個人適応型 WWW のためのユーザモデル構築，*Interaction '97*, pp.87-94, 情報処理学会 (1997)。
- 13) Lam, W., et al.: Detection of Shifts in User Interests for Personalized Information Filtering, *19th ACM-SIGIR Conference*, pp.317-325 (1996)。

(平成 9 年 7 月 7 日受付)

(平成 10 年 1 月 16 日採録)

三浦 信幸 (正会員)



1993 年東京工業大学工学部情報工学科卒業。1995 年同大学院理工学研究科電気電子工学専攻修士課程修了。同年、日本電信電話(株)入社。現在、同社ソフトウェア研究所に勤務。知的情報サーバ構築・運用技術、ユーザ利用履歴分析・ユーザ適応システム、モバイル環境向け情報検索システム等の研究に従事。

高橋 克巳 (正会員)



1988 年東京工業大学理学部数学科卒業。同年 NTT に入社。現在 NTT ソフトウェア研究所研究主任。情報検索システムとアルゴリズム、知的エージェント通信の研究に従事。ACM, 人工知能学会各会員。

島 健一 (正会員)



1976 年北海道大学工学部電気工学科卒業。1978 年同大学院情報工学専攻修士課程修了。同年 NTT 武蔵野電気通信研究所入所。現在、NTT ソフトウェア研究所主任研究員。主に、知識ベース構築用システムの基礎研究、ソフトウェア設計での知識獲得、学習システムなどの研究開発に従事。また、WWW でのユーザモデル研究に興味を持つ。電子情報通信学会、人工知能学会各会員。