WWW サーバ・ログからのユーザモデル構築実験*

三浦 信幸† 島 健一‡

NTT ソフトウェア研究所

E-mail: {miura , kshima}@slab.ntt.jp

1 はじめに

昨今の WWW の広がりにつれて，WWW も質を問われる時代となってきた．我々は現在，質の高い WWW の構築の一貫として「個人適応型 WWW」を検討している [1] ．個人適応型 WWW は，個々人のアクセス状況に応じて，提供する情報の提示方式，提供する情報の内容などをユーザ毎に動的に変更しながら動作する WWW である．

本稿では，個人適応型 WWW において重要な役割を果たす，ユーザモデリングについてその方法を概説し，ユーザモデル構築の予備実験結果を報告し，実施予定のユーザモデル構築実験の概要について述べる．

2 個人適応型 WWW と WWW サーバ・ログ

従来，いくつかの WWW サイト (例えば，[2, 3, 4]) で個人対応を実現しているが，それらと我々が目指す個人適応型 WWW は以下の点が異なる．

- ユーザは通常通りアクセスするだけで適応
  従来の個人対応 WWW サイトのほとんどは，事前に，単純なキーワード群の中から自分の興味等を選択してアンケートとして提出する方式，もしくは，ユーザへの何らかのテストを行う方式である．我々が目指す WWW では，サーバ側でユーザの特性を自動的に取得し，サーバ自身が個人に適応するため，そのような作業は不要である．
- アクセスの繰り返しによるユーザの変化に適応
  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．我々が目指す WWW では，サーバは常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などにも適応
  アンケート形式では，ユーザの興味・関心等のいわば，ユーザの状態のみにしか対応ができない．我々が目指す WWW では，ユーザの行動・学習の特性についても個人適応を行う．

個人適応型 WWW を構築するためには，ユーザ一人一人の特性の把握，すなわち，ユーザ一人一人のユーザモデルを構築する必要がある．ユーザモデル構築のための情報の取得方法には，先に述べたアンケート形式を始め，いくつかの方法が考えられるが，本研究では上記の目的のために，WWW サーバ・ログからユーザモデルを構築する方針をとる．

3 個人適応型 WWW のためのユーザモデル

提供する情報の提示方式，提供する情報の内容などを動的に変更する個人適応方式を実現するためには，次のようなユーザに関する情報が必要と考えられる．

- ユーザの興味・関心等のユーザの状態の情報
- 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
- 該当 WWW の構造・提供している情報の種類等をユーザがどの程度，理解しているかの情報
- ユーザがどのような時にどのような行動を取り，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

本研究ではこれらそれぞれに合わせて，図 1のような状態モデル・行動モデル・学習モデルの 3 つのモデルとこれらのモデル間の関係を表現するメタモデルの計 4 つのモデルを構築する．個々のモデルの詳細は検討中であるが，当面考えているモデルの内容は次のようなものである．

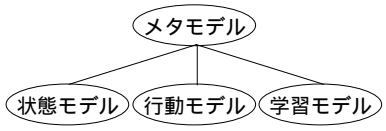

図 1: 4 つのユーザモデルの関係

- 状態モデル
  各 URL[1] 毎に 1 個以上のキーワードを設定し，各 URL へのアクセス 1 回を各キーワードへのアクセス 1 回と数えた結果作られる，キーワードとそれへのアクセス回数との組合せ．
- 行動モデル
  各ユーザが行ったハイパーリンクの遷移列のうち，頻出する遷移列をユーザの行動パターンとして抽出した遷移列の集まり．この遷移列には，どれくらいの時間間隔で行動を起こしたかの平均値なども属性値として付く．
- 学習モデル
  該当 WWW が持つ特徴的な構造・機能や提供している情報の種類などの該当 WWW にアクセスする上で学習していくと考えられる項目と，それらを使っているか否か・理解しているか否かとを対応づけたもの．なお，各項目毎にいつ学習が行われたかの時間も記録される．
- メタモデル
  いつ，どういう状態モデルの時に，どのような行動パターンが発生し，どのような学習が行われたかの記録の集まり．

4 ユーザモデルの構築方法

3章で述べたモデルを構築するには，該当 WWW のコンテンツに関する情報が必要になる．本研究ではそれらの情報をコンテンツ・モデルとしてモデル化する．例えば，職業別電話帳のサーバのユーザの状態モデル構築のためには，業種別検索の URL に対して業種のキーワードが付与されるが，これらのキーワードは相

---

* An Experiment of Composing the User Model from the Access Logs of the WWW Server

† Nobuyuki Miura, NTT Software Laboratories

‡ Ken'ichi Shima, NTT Software Laboratories

[1] ここでは CGI 等の引数の query string を考慮に入れ，query string を含めた URL それぞれについてキーワードを設定する

互に多様な関係を持っており，明確にモデル化しておく必要がある．コンテンツ・モデルは一般に，ある程度複雑な構造を持つことが予想され，グラフで表現しておくことが必要である．

一方，サーバ・ログは，個々のログ間の時系列の関係の他に，同一ユーザのログという関係や同一 URL の関係など，複雑な構造を持っており，サーバ・ログもグラフ表現することができる．

そこで，グラフ表現された，コンテンツ・モデルおよびサーバ・ログに対して，適切なグラフ操作を施すことにより，各ユーザ毎のユーザモデルを構築することができると考えられる．

構築されたユーザ･モデルの情報を使って個人適応を行う．個人適応型 WWW の構成は図 2のようになる．

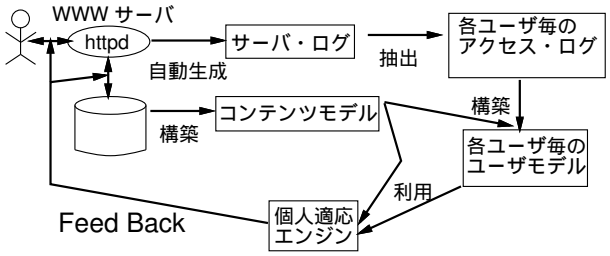

図 2: 個人適応型 WWW の構成

5　ユーザモデル構築の予備実験

ユーザモデル構築の予備実験として，あるユーザのアクセス・ログの簡易な分析を行った．ここでは，職業別電話帳サーバ[2] の職業別検索に着目し，ユーザが検索した職業の分布を分析した．対象ユーザは任意に抽出した 2 人である．結果は，それぞれ，表 1，2 の通りである．

いずれの分析結果も，個々の職業分類単位で見ると，検索回数が非常に多いいくつかの職業を除いては，あまり検索回数の割合に偏りがない．しかしながら，コンテンツ・モデルを利用し，職業分類を抽象化してみると，ユーザの職業別検索における，職業の嗜好性がより一層はっきり見えてくることがわかった．A さんの場合，興味の対象は，主に，電気・電子関係の業種である．また，B さんの場合，ホテルと自動車に関する業種である．本分析結果は，コンテンツ・モデルとサーバ・ログを統合的に扱うユーザモデリングという手法が有効に機能する例と言える．

6　ユーザモデル構築実験

以下のような実験設定でユーザモデル構築実験を行う予定である．

- 実験対象 WWW
  Japan Telephone Directory[2] ：
  本サーバは，英語版職業別電話帳サーバであり，次の点から，本実験に適したサーバである．
  - 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．
  - このサーバには，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．

[2] Japan Telephone Directory
http://www.pearnet.org/jtd/

表 1: 職業別検索の検索した職業の分布 (A さん)

抽象化カテゴリ名	合計 (割合)	カテゴリ名	回数	割合
電気・電子	47 (67 %)	電気製品製造	29	41 %
		電気用品	7	10 %
		電子部品・機器	4	6 %
		電気製品小売	4	6 %
		電力会社	2	3 %
		電気製品中古／輸出	1	1 %
実験器具	6 (8 %)	計量／試験器具／実験器具	3	4 %
		科学機器・装置	3	4 %
コンピュータ	5 (7 %)	コンピュータ製造	5	7 %
事務用品・事務機器	4 (6 %)	コピー機	2	3 %
		計算機	2	3 %
通信	2 (3 %)	通信業	1	1 %
		通信機器	1	1 %
医療用機器	2 (3 %)	医療用機器・用品	2	3 %
化粧品	2 (3 %)	化粧品	2	3 %
化学製品	2 (3 %)	化学製品	2	3 %

(総検索回数 70 回 (1996.1.15 ~ 1996.3.11))

表 2: 職業別検索の検索した職業の分布 (B さん)

抽象化カテゴリ名	合計 (割合)	カテゴリ名	回数	割合
ホテル	28 (51 %)	ホテル	27	49 %
		ホテル (大阪)	1	2 %
自動車	21 (38 %)	自動車販売 中古／輸出	11	20 %
		自動車部品	8	15 %
		自動車部品 中古／再生	1	2 %
		タイヤ販売	1	2 %
電気・電子	3 (5 %)	電子部品／機器	3	5 %
レコード類	2 (4 %)	レコード･テープ･ディスク 卸	1	2%
		レコード･テープ･ディスク 小売	1	2%
音響	1 (2 %)	音響機器	1	2 %

(総検索回数 55 回 (1996.2.16 ~ 1996.3.11))

  - 1995 年 10 月より実験提供しており，多数のユーザのログを利用することが可能である．
  - 予備実験結果において，確かにユーザの特性の把握ができている．
- 実験目的
  - グラフ操作によるユーザモデリングが可能かつ有効であることを確認する．
  - モデルのうち，どの部分がユーザ毎の相違が大きいかを調べる．相違の大きい部分こそ個人適応すべき部分である．また，この結果から，ユーザモデルの再検討および個人適応の具体的内容の検討を行う．
- 実験方法
  モデル構築ツールを実装し，それをサーバ・ログに適用して，実際にモデル構築を行う．

7　今後の課題

- cost と benefit を考慮に入れたユーザモデルの検討．例えば，ユーザモデルから個人適応方法を考えるのではなく，ある個人適応方法にはどのようなユーザモデルが必要かを検討し，必要最小限のユーザモデルを導出することなど．
- 心理学や教育学的見地からの個人適応法の検討．

参考文献

[1] 三浦信幸, 島健一. サーバアクセス履歴からのユーザモデルの構築. 情報処理学会 第 52 回 全国大会, Vol. 1, pp. 153–154, Mar. 1996.

[2] 仲林清, 小池義昌, 丸山美奈, 福原美三. WWW を用いた知的 CAI システム CALAT. NTT 技術ジャーナル, Vol. 95, No. 8, pp. 72–75, Aug. 1995. http://calat.min.ntt.jp/

[3] Gyro Scoop. http://www.gyroscoop.com/

[4] Fire Fly. http://www.ffly.com/

[5] 島健一, 高橋克巳, 三浦信幸. インターネット版マルチメディア電話帳の構築. In *Japan World Wide Web Conference '95*. 日本インターネット協会, Nov. 1995. http://www.pearnet.org/jtd/paper/jtd-overview.html.