

# 5U-10 ウェブブラウジングの統計情報にもとづく 情報推薦システムに関する研究

阿部康一

武田利浩

布川博士†

† (財) 仙台応用情報学研究振興財団 山形大学工学部電子情報工学科 岩手県立大学ソフトウェア情報学部

## 1 はじめに

ユーザが WWW (World-Wide Web) という広大な情報空間から必要とする情報を見つけ出すためには、サーチエンジンと呼ばれる検索サービスを利用し、自ら検索結果を選別する必要がある。また、それらの内容を確認しながら目的の情報を見つけ出す作業は、ユーザにとってたいへん手間のかかる行為でもある。

現在、「マイページ」や「PUSH 型」サービスのようシステム側からユーザ側に情報を提供する WWW 上のサービスが増えてきている。これは、WWW の利用形態が「個」を対象としたサービスに変化してきていることの現れと考えられる。すなわち、今後、ユーザが必要としている情報を提供するシステムやアプリケーションの需要が大きくなると予想される。

本文では、(1) ユーザが興味をもっている分野の情報の推薦、(2) ユーザが潜在的に必要としている情報の推薦、を行うユーザのウェブブラウジングの統計情報を用いた情報推薦システムについて述べる。

## 2 情報推薦システムの概要

本システムは、図1に示すように (a) ユーザ、(b) ゲートウェイ、(c) エージェント、(d) 情報資源、の4つの要素から構成される単純なモデルで表される。

### (a) ユーザ

情報推薦を受ける主体。ユーザは情報推薦に必要なプロフィールを作成し、エージェントに送る。

### (b) ゲートウェイ

ユーザ毎に URL (Uniform Resource Locators) とコンテンツ (アクセスデータと呼ぶ) を分類し、それらを保存する。情報推薦システムを利用するユーザのブラウザは、本ゲートウェイを通じてウェブサイトへアクセスする必要がある。

### (c) エージェント

ユーザに情報を推薦するプログラム。ユーザの

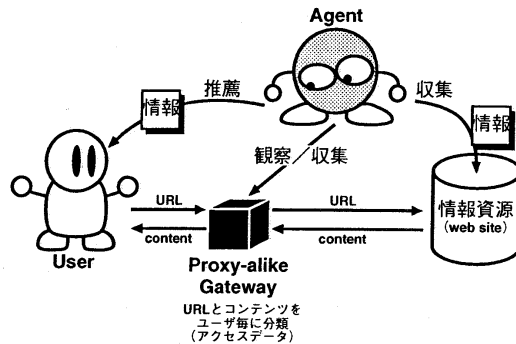


図 1: 情報推薦モデル

プロフィールとゲートウェイが管理しているユーザのアクセスデータを分析し、ユーザの好みを推測する。推薦情報は、いくつかのポータルサイトの情報を分析したデータにもとづいて収集および加工し、ユーザにウェブページとして提供する。

### (d) 情報資源

ユーザがアクセスするウェブサイト。エージェントが推薦用の情報を収集するためのポータルサイト。

図1のモデルにおける、情報推薦の流れを以下に示す。

1. ユーザはゲートウェイを通じてウェブサイトへアクセスする。
2. ゲートウェイはユーザのアクセスデータをユーザ毎に分類、収集、管理する。
3. エージェントはユーザのプロフィールとアクセスデータから好みを推測する。
4. エージェントはポータルサイトで情報収集し、ユーザへの推薦情報の選別を行う。
5. エージェントはユーザに情報を推薦する。

A Study for Information Recommending System using the Statistical Information of Web Browsing

Kouichi ABE

Sendai Foundation for Applied Information Sciences

Toshihiro TAKETA

Department of Electrical and Information Engineering, Faculty of Engineering, Yamagata University

Hiroshi NUNOKAWA

Faculty of Software and Information Science, Iwate Prefectural University

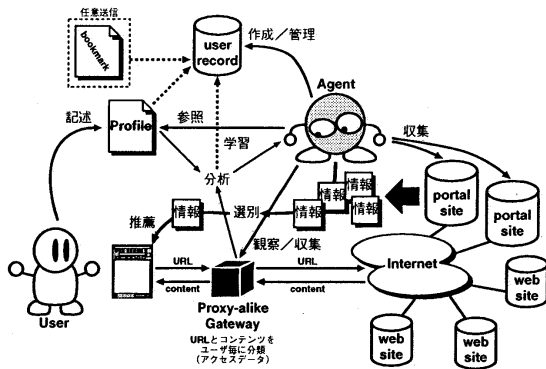


図 2: 情報推薦メカニズム

プロフィール設定

名前:

パスワード:

興味/関心のあるカテゴリをチェックしてください

<input type="checkbox"/> 芸術と人文	<input type="checkbox"/> メディアとニュース	<input type="checkbox"/> ビジネスと経済
<input type="checkbox"/> 趣味とスポーツ	<input type="checkbox"/> コンピュータとインターネット	<input type="checkbox"/> 各種資料と情報源
<input type="checkbox"/> 健康と医学	<input type="checkbox"/> エンターテインメント	<input type="checkbox"/> 生活と文化
<input type="checkbox"/> 社会科学	<input type="checkbox"/> 自然科学と技術	<input type="checkbox"/> 政治

情報収集の対象となるポータルサイトを選択してください

<input type="checkbox"/> Yahoo!	<input type="checkbox"/> goo	<input type="checkbox"/> Hotmail	<input type="checkbox"/> fresheye	<input type="checkbox"/> ZDNet
<input type="checkbox"/> Infoseek	<input type="checkbox"/> Lycos	<input type="checkbox"/> MSN	<input type="checkbox"/> Excite	<input type="checkbox"/> ISIZE

送信するブックマークファイル名

図 3: プロファイル設定画面例

### 3 情報推薦システムの設計

図 2 は、システム全体の情報推薦メカニズムを示している。本章では、システムの主要部分を占めるエージェントの情報推薦における機能を説明する。また、ゲートウェイの機能とプロフィール設定や情報推薦に伴うインタフェースについて簡単に述べる。

#### 3.1 エージェントによる情報推薦

エージェントは、ユーザー一人につき一つ存在する。また、ユーザーに情報推薦を行うために、エージェントはユーザープロフィールとアクセスデータからユーザーの好みを推測し、それをデータベースとして管理する(図 2 中の“user record”)。

ユーザーの好みの推測には、(1) アクセスデータからキーワードを抽出し、その頻度を調べる、(2) 頻繁にアクセスする URL 先のコンテンツを対象として、キーワードを抽出し、その頻度を調べる、という方法をとる。

情報推薦は、出現頻度の高いキーワードを含むコンテンツを見つける作業から始まる。本システムでは、ポータルサイトを推薦情報の探索の場として活用する。すなわち、ポータルサイトのトップページから出現頻度の高いキーワードを含むリンク先をたどり情報を探し出していく。

推薦情報として収集したコンテンツは、さらに高頻度キーワードを中心に解析され、ダイジェストページとして加工される。このページがユーザーへ推薦される情報となる。

#### 3.2 ゲートウェイの機能

ゲートウェイは、ユーザーのリクエストとサーバからのレスポンスを中継する Proxy 機能、情報推薦を受けるユーザーを識別する認証機能、ユーザーのアクセスデータを管理する機能をもつ。また、エージェントとの通信もゲートウェイを通じて行われる。

#### 3.3 インタフェース

システムとユーザーとの通信用インタフェースは、CGI (Common Gateway Interface) を用いたウェブページである。図 3 は、プロフィール設定ページの画面例を表している。

図 3 のページを通じて設定されたプロフィールは、エージェントがユーザーの好みの分野と推薦情報を探すポータルサイトを決定するのに用いられる。また、ユーザーはブラウザのブックマークをエージェントに送ることもできる。これは、エージェントがユーザーの好みを推測する際に利用される。

## 4 おわりに

本文では、ウェブブラウジングの統計情報を用いた情報推薦システムについて、そのモデルとメカニズム、設計について述べた。今後は、プロトタイプモデルを実装し、実験を行う。また、どのようなキーワードが推測に有効かを考察する。