

個人適応型観光情報収集支援システムのためのユーザモデルの構築

02005381 北海道大学大学院工学研究科 *大野貴司 OHNO Takashi
01004631 北海道大学大学院工学研究科 大内 東 OHUCHI Azuma

1. はじめに

北海道観光情報 WWW サイト数の急速な増加に伴い、ユーザが効果的に目的とする情報を収集することが可能なシステムの必要性が高まっている。

本研究では、WWW を利用して実際に旅行を計画しようとするユーザを対象に、個々のユーザの特性に適応し、目的とする観光情報の効率のよい収集を支援する個人適応型観光情報収集支援システムに必要となるユーザモデリングについて述べる。

2. 個人適応型観光情報収集支援システム

本研究では、三浦らの、「ユーザになるべく余計な作業負担をかけずにユーザ特性を把握し、ユーザ特性の時間変化に追従でき、かつ、ユーザの行動パターン等にも適応可能なシステム」という個人適応型 WWW の基本的な考え方に基づき、システムの構築を目指す [1]。

2.1 システムの構成

現在構築中の個人適応型観光情報収集支援システムは、ユーザモデル、カテゴリ辞書、Web ブラウザから構成される。

ユーザモデルは、ユーザの興味あるカテゴリやそのカテゴリの閲覧回数、カテゴリの遷移、カテゴリ間の関係などを表現したものである。ここで、カテゴリとは観光情報のジャンル（例えば、花を訪ねて、温泉めぐり、グルメ、ショッピングなど）を指す。カテゴリ辞書は、18種類の各カテゴリにキーワードを付与したものである。この辞書は、ユーザの閲覧した Web ページ (URL) をカテゴリに分類する際に用いる。Web ブラウザは、Web ページを分類する機能やグラフを構築する機能などを実装したブラウザである。

2.2 ユーザの興味の把握

一般的に、ユーザの興味をとらえる方法としてアンケートなどが考えられるが、この場合、ユーザは複数の質問に答える必要があり負担となる。また、最新のユーザの興味を把握するためには定期的なアンケートの実施が必要である。

本研究では、最近閲覧した情報や時間的要素を含む、インターネット一時ファイル (Temporary Internet Files) を用いてユーザの特性を獲得する方法を用いる。インターネット一時ファイルは、Web ブラウザがロー

カルハードディスクに保存した Web ページや画像などのファイルのことである。インターネット一時ファイルはクライアント側に保存されるため、これを解析することにより、ユーザの閲覧した Web ページの情報を獲得できる。このように、本システムはユーザの挙動をクライアント側で観察する方法を採用する。

本研究では、テキスト情報からユーザの興味分野などの特性を獲得することを目的とするため、インターネット一時ファイルのうち、HTML ファイル (Web ページのファイル) のみを利用する。

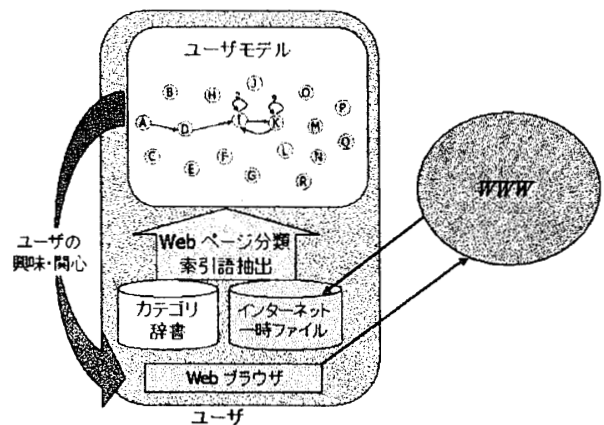


図1 個人適応型観光情報収集支援システムの概念図

3. ユーザモデル構築法

本研究でのユーザモデル構築の概要を以下に示す。

- ・ 入力情報は、ユーザの閲覧した Web ページ (URL) である。
- ・ HTML テキストから <title> や <h1> などの HTML タグに挟まれたテキスト情報を抽出する。
- ・ 抽出されたテキスト情報から形態素解析により名詞のみを抽出し、さらに文書の特徴付ける上で役に立たない「北海道」や「情報」といった不要語 (stop word) を削除したものを索引語 (index term) とする。
- ・ カテゴリ辞書に記述されているキーワードと抽出された索引語とのマッチングを行い、Web ページ (URL) を (観光 GIS 標準フォーム [2] を参考にあらかじめ作成した) 「グルメ」や「建築ウオッチング」といった 18 のカテゴリに分類する。但し、本研究で用いる分類方法は、1 つの Web ページ (URL) が複

数個のカテゴリに分類される場合がある。カテゴリ辞書の作成方法、及び Web ページの分類処理の詳細については、参考文献 [4] を参照されたい。

- ・ ユーザモデルのグラフを構築する。

ユーザモデル、インターネット一時ファイルを次のように定義する。

ユーザモデル

$$M = \langle V, A \rangle$$

M のノード集合 $V = \{c_1, c_2, \dots, c_i, \dots, c_{18}\}$

カテゴリ c_i のキーワード集合 $c_i = \{k^i_1, k^i_2, \dots\}$

※ カテゴリにおけるキーワード数は有限でありその数はカテゴリにより異なる。

M のアーク集合 $A = \{a_1, a_2, \dots, a_i, \dots, a_n\}$

アーク $a_i = (c_i, c_j) \in A$

※ 関連度 af_i をアーク a_i の属性値とする。ここで関連度は、ノード c_i から c_j へのリンクの回数のこととする。

インターネット一時ファイル

$W = (w_1, w_2, \dots, w_i), (w_i \in U) \quad U = \text{URL 集合}$

$i = \text{ユーザのアクセスした URL 数}$

w_i に含まれる索引語集合 $H_i = \{h^i_1, h^i_2, \dots\}$

※ 索引語集合における索引語の数は有限でありその数は索引語集合により異なる。

まず、ユーザの閲覧した Web ページ (URL) のうち、1つのカテゴリに分類されたものだけを抽出し、ユーザの閲覧順にソートする。1つのカテゴリに分類された Web ページ (URL) のみに着目する理由は、「ユーザの本当に欲しい (有用な) 情報は1つのカテゴリに分類されるのではないか」と考えられることによる。例えば、複数のカテゴリに分類されるようなページは概説的なページが多く、詳細な情報の獲得が期待できない可能性が高いと考えられる。一方、1つのカテゴリに分類されるようなページは、1つのジャンルに関して詳細な情報が提供されている場合が多く、有用な情報を獲得できる可能性が高いと考えられる。

これをもとにカテゴリをノードとしたグラフを作成する。但し、アーク (p, q) が存在するときは、カテゴリ p に相当する Web ページを閲覧した後、カテゴリ q に相当する Web ページを閲覧したことを意味する。このようなグラフの構造を分析することにより、例えば、あるカテゴリに対する興味の強さやまとまりなどが明らかになると考えられる (図2)。また、このモデルを用いることで、カテゴリの参照回数の他に、カテゴリの遷移やカテゴリ間の関係なども表現可能となる。

4. ユーザモデルの評価実験

今回提案したユーザモデルの有効性を示すために、

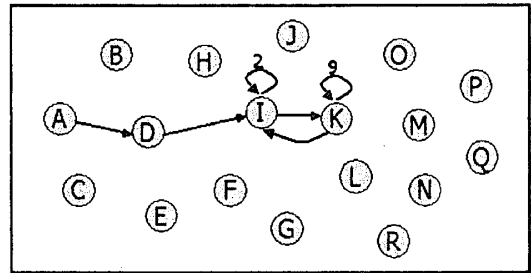


図2 ユーザモデルの例

システム情報工学を専攻する大学院生 10 人を被験者に次のような実験を行った。それぞれのユーザに対して、北海道観光大全 [3] をポータルサイトとして、興味のある Web ページを 50 ページ程度ブラウジングするように依頼した。

結果、ある被験者に対して、図2に示すようなユーザモデルが構築された。この被験者は、閲覧した Web ページの総数が 72 ページで、そのうち 1つのカテゴリに分類されたページが 16 ページであった。図2のグラフより、この被験者が「建築ウオッチング」のカテゴリに相当する Web ページを何度も閲覧していることがうかがえ、主に建築ウオッチング (ノード K) に興味があると推測できる。また、ノード I (まちあるき) とノード K (建築ウオッチング) のカテゴリに相当する Web ページを相互に閲覧したことにより、ノード I とノード K は強連結 (strongly connected) となっている。このようなグラフの構造的な特徴がユーザのどのような特性を表しているのかについての詳細な検討は今後の課題である。問題点として、ユーザの閲覧した Web ページのうち、実際にユーザモデルのグラフに反映されたページ数が比較的少ないということが明らかになった。

5. 今後の展開

今後の展開として、Web ページの分類方法や構築したユーザモデルのグラフ構造を分析し、ユーザの興味をとらえる方法について検討する。また、Web ページの分類手法についても検討する。

参考文献

- [1] 三浦, 高橋, 島 : 個人適応型 WWW におけるユーザモデル構築法, 情報処理学会論文誌, Vol.39, No.5, pp1523-1535, 1998.
- [2] 観光 GIS 標準フォーム, <http://www.nihon-kankou.or.jp/gis/form/index.html>.
- [3] 北海道観光大全, <http://www.onitoge.org/>.
- [4] 大野, 大内 : 北海道観光情報における個人適応型情報収集システムの構築, 電気学会研究会資料, IS-02-7~19, pp51-57, 2002.