

## Web ナビゲーションのためのブラウジングパターンの学習

## Learning User's Browsing Patterns for Web Navigation

椰野 憲克\*1 山田 誠二\*2  
Norikatsu NAGINO Seiji YAMADA

\*1 東京工業大学大学院 総合理工学研究科 知能システム科学専攻  
Tokyo Institute of Technology

\*2 国立情報学研究所  
National Institute of Informatics

In this paper, we propose a Future View system that assists user's Web browsing in everyday life. A Future View will prefetch Web pages based on user's browsing strategies and present them to a user in order to assist Web browsing. To learn user's browsing patterns, Future View uses two types of learning classifier systems: content-based classifier system for contents change patterns and action-based classifier system for action patterns. The results of learning is applied to crawling by Web robots, and the gathered Web pages are presented to a user through a Web browser. We experimentally show effectiveness of navigation using a Future View.

## 1. はじめに

WWW は、情報収集のために頻りに利用されるようになってきた。WWW 上の情報の量は膨大であり、ユーザの経験的なブラウジング戦略に基づくナビゲーションを行なうことは重要である。探索目的が明確であり変化しないブラウジングを支援する方法 [Joachims 97] や関連情報を収集する方法 [Bra 94]、サイト内ナビゲーションのためのアクセスパターンの学習 [Spiliopoulou 00] はいくつか研究されているが、新たなサイトに訪れたり、ブラウジングパターンを考慮して、日常のブラウジングを支援する研究はほとんどない。

本論文では、ユーザのブラウジングタスクを支援するシステム、Future View を提案する。Future View は、ユーザが現在の戦略を続けたときに収集される可能性が高い Web ページを収集、ユーザに提示するために、ユーザのブラウジングパターンを学習し、Web ロボットによる Web ページ収集に適用することにより、ユーザのブラウジング戦略に基づいた Web ページの先読みを行なう。Future View は、Web ページのコンテンツ情報と、ユーザの行動情報に基づいた、ブラウジングパターンの学習を行なう。収集した Web ページは、アクセスする可能性が高いと考えられるページがより上位になるように順序付けし、一覧として表示する。ユーザは、表示された結果が興味を反映したものであれば、そのユーザインタフェースを通して、目的の Web ページに直接アクセスすることができる。また逆に、目的の結果と異なっていれば、自分のブラウジング戦略を変更するかもしれない。

## 2. Future View の構成

Future View は、クラシファイアシステム [Holland 78] をベースにした学習器 Learner、Web ページを先読みする Prefetcher で構成される (図 1)。

## 3. ブラウジングパターンの学習

## 3.1 ページクラス

Learner は、内容ベースクラシファイアシステム (CCS) と行動ベースクラシファイアシステム (ACS) の 2 つで構成され、それぞれ並行して動作する (図 2)。クラシファイアシ

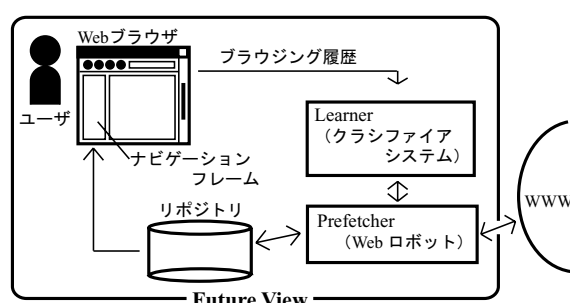


図 1: システムの概観

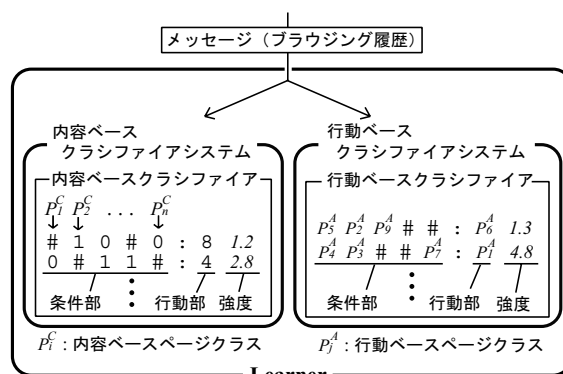


図 2: Learner の構成

システムにより、時系列パターンを学習するために、条件部と行動部を Web ページの URL を用いて次に訪れる Web ページを学習する方法が考えられるが、ユーザが訪れる可能性のある URL は無限であることから、Future View は Web ページの内容 / ユーザの行動情報に基づいて Web ページ集合を抽象化した「ページクラス」を用いて学習する。ページクラスを決定する際、以下の観点に従って定義すると良い。

- 内容ベースページクラス  
ブラウジング履歴を同一トピックについての一連のアクセスシーケンス毎に区切った、「セッション」単位を考慮して決定する。
  - 内容の連続性：セッション中の tfidf 値が高い単語を、タイトルや本文に多く含むページ
  - 同一トピックにおける内容の差異性：セッション中のページ集合で計算した tfidf 値の高い単語が、他のページに含まれていないページ

● 行動ベースページクラス

- 新しい情報への興味：新しく追加されたリンクを辿ってアクセスされた Web ページ
- 戦略的探索：ブラウザに直接 URL を入力してアクセスされた，検索エンジンのトップページ

ACS と CCS は，クラシファイアの強化部で，競りに基づいた各クラシファイアの強度の更新，発見部で GA といった，クラシファイアシステムの基本的な動作を行なう。さらに ACS では，クラシファイアの学習に，ページクラスだけでなく，Web ページの URL をインスタンスとして用いた学習も行なう。

4. ユーザインタフェースの構築

Web ロボットは，現在のページを始点とし，クラシファイアの強度を基に最良優先探索を行なう。収集された Web ページは，最後に適用されたクラシファイアの強度値でランキング表示される。ナビゲーションウィンドウには，ブラウジングの間，常に最新の情報が表示される（図 3）。

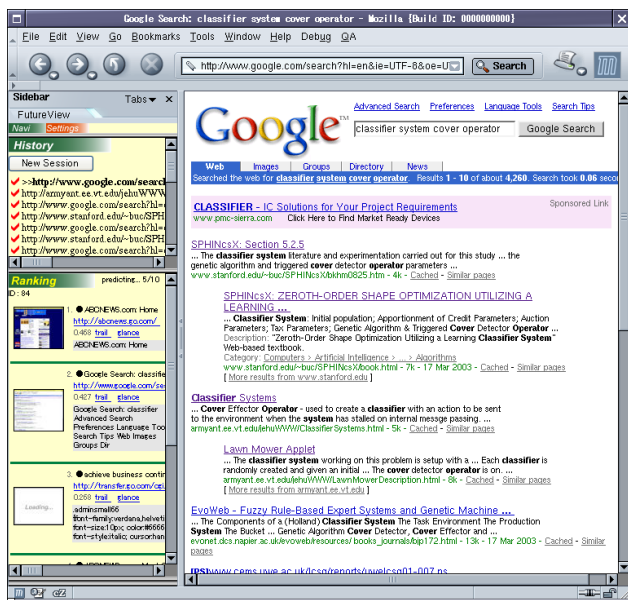


図 3: ユーザインタフェース

5. 実行例と実験

Future View を数日間使用したところ，以下のようなクラシファイアが学習された。

- 検索エンジンで用いた検索キーワードがアンカーテキストに含まれるリンクをいくつかたどる (CCS)
- リンクを選択し 1 ページ戻るといふ行動を数回繰り返した後，ニュースサイトのトップページの URL を入力する (ACS)
- 新しく追加されたリンクを上から順に選択する (ACS)

これらのクラシファイアにより，ユーザが検索エンジンを用いた場合，主にニュースサイトに追加された新たな記事のページが推薦され（図 4），これらのページがユーザにとって興味があることをアンケートにより確認した。この結果により，ユーザのブラウジングパターンを反映し，従来の方法では探索，推薦が難しいページを適切に推薦できたことが分かった。

また，我々は CCS と ACS の効果を調べるために，探索目的を持ったブラウジング（表 1 の  $w_1$ ）と探索目的を持たない

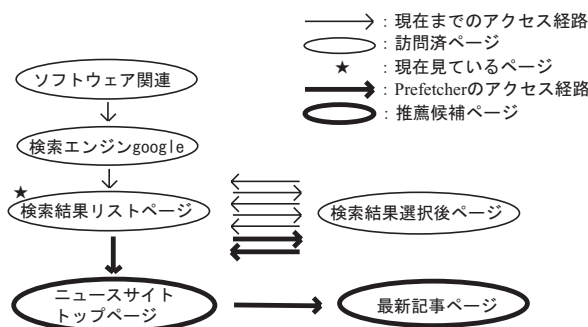


図 4: Prefetcher の訪問経路

ブラウジング（表 1 の  $w_2$ ）に対し，CCS または ACS のどちらか，または両方を用いて Web ページを探索した場合の推薦ページの評価を行なった。表 1 は，各ページを 0（まったく関係のないページ）から 4（非常に興味のあるページ）の 5 段階で評価したときの平均値を示している。実験の結果，探索目的を持ったブラウジングの場合は CCS が有効に働き，探索目的を持たないブラウジングの場合は ACS が有効に働くことが分かった。また，ほとんどの場合 CCS と ACS の両方を用いるとさらに効果的であることが分かった。

	subject A		subject B	
	$w_1$	$w_2$	$w_1$	$w_2$
$e_{c,a}$	1.45 (42)	2.06 (50)	1.47 (37)	1.28 (36)
$e_c$	1.41 (49)	0.40 (26)	1.09 (42)	1.14 (32)
$e_a$	1.40 (38)	1.87 (43)	0.89 (48)	1.82 (16)

$w_1$ : 技術論文の探索タスク  $w_2$ : 目的を持たないブラウジング

(n): n は提示された Future ページの数で，最大値は 50。

Future ページの収集に， $(e_{c,a})$  CCS と ACS の両方を使用， $(e_c)$  CCS のみ使用， $(e_a)$  ACS のみ使用。

表 1: Future ページの評価

6. まとめ

我々は，ユーザのブラウジングを支援するシステム Future View を提案した。実験により Future View は，従来の手法では提示困難な興味深いページを提示する能力を持つことを確認し，また，CCS と ACS をともに用いることが効果的であることが分かった。

参考文献

[Holland 78] J. H. Holland and J. S. Reitman: Cognitive Systems Based on Adaptive Algorithms, Pattern-Directed Inference Systems, pp.313-329, (1978).

[Bra 94] P. De Bra and R. Post: Information Retrieval in the World-Wide Web: Making Client-based Searching Feasible, Computer Networks and ISDN Systems, vol.27, no.2, pp183-192 (1994).

[Joachims 97] T. Joachims and D. Freitag and T. Mitchell: WebWatcher: A Tour Guide for the World Wide Web, Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence, pp770-0775, (1997).

[Spiliopoulou 00] M. Spiliopoulou: Web usage mining for Web site evaluation, Communications of the ACM, 43(8):127-134, (2000).