

Web 閲覧行動の学習に基づく 情報推薦システムに関する評価実験

鷹野孝典[†] Kin Fun Li^{††}

本稿では、Web 閲覧行動、および利用者の評価フィードバックにより構築される情報嗜好シソーラスに基づいて、利用者の情報嗜好に適応する情報推薦システムを提案する。本システムは、個々の利用者の Web 閲覧行動を監視することにより、個々の利用者の情報嗜好を、情報嗜好シソーラスとしてデータベースに蓄積し、さらに、推薦情報に対する評価フィードバックにより、継続的な情報嗜好シソーラスの更新する機能を有する。本システムは、このように Web 閲覧行動等に基づいて学習される情報嗜好シソーラスを用いて、刻々と変化する利用者の閲覧行動や、その時点の利用者の興味・関心に適応した情報推薦を可能とするものである。本研究では、プロトタイプとして構築した実験システムを用いて、提案方式の実現可能性を検証する。

Experiments of an Adaptive Recommender by Learning Web-Browsing Behavior

Kosuke Takano[†] and Kin Fun Li^{††}

In this study, we propose an adaptive recommender system based on a preference-thesaurus constructed by learning a user's Web-browsing behavior. The adaptive nature of the proposed system allows the personalization of recommendation and the identification of the ever-changing influential Web-browsing behaviors. A prototype system is implemented and experiments are performed to verify the system's capability to personalize recommendation and to learn through user's feedbacks.

1. はじめに

World Wide Web (WWW) 上での豊富な文書データ、および多様な情報サービスが利用可能となっている状況において、Web 上での情報閲覧・獲得行動はますます重要となっている。利用者が興味・関心を持つと判断される Web 文書のナビゲーションや商品情報の提示を行うことにより、利用者が要求する情報を効率的に取得することを目的として、多くの情報推薦方式が提案されている。

協調フィルタリング方式[1]は、利用者グループから収集した嗜好情報を分析することにより、利用者の興味や関心を抽出するものであり、e-commerce [6]、ネットニュース[5]、音楽や映画といった趣味情報の共有サイト[9]などで、個々の利用者の情報嗜好に合致した情報を推薦するために、広く用いられている。協調フィルタリング方式では、推薦情報に対する利用者の評価フィードバックなどから明示的に得られる嗜好情報や、システムの操作履歴などから暗黙的に得られるものに基づいて、興味・関心の抽出や情報フィルタリングを行う。システム操作履歴などを対象としたプロファイルリング技術の進歩に伴い、情報推薦システムにおいて、個々の利用者の暗黙的な情報嗜好を効率的に収集することが可能となっている。The Stuff I've Seen (SIS) system [3]では、利用者がその情報を既に見ているという情報に基づいて、著者、サムネイル画像などの、コンテキスト・キューが関連情報の検索に利用される。The SEARCHY system [4]は、Web 閲覧履歴から得られる利用者プロファイルを利用することにより、Web 検索によって得られた Web 文書集合を対象としたフィルタリングや再ランキングを行う。Chirita らは、利用者の情報嗜好に合致した Web 文書を検索するため、利用者の PC に蓄積されているテキスト文書、email、キャッシュされた Web ページ等に基づいた、クエリ拡張方式を提案している[2]。Web ブラウザの閲覧履歴や個人が所有する文書集合に加えて、閲覧時間、マウスクリック、スクロール、テキスト選択、検索クエリおよび検索結果などの Web 閲覧行動を対象とした個人プロファイルが、個人の情報嗜好を満たす情報推薦を行うために活用されている。

本研究では、Web 閲覧行動、および利用者の評価フィードバックにより構築される情報嗜好シソーラスに基づいて、利用者の情報嗜好に適応する情報推薦システムを提案する。本システムは、個々の利用者の閲覧行動をプロファイルすることにより構築される情報嗜好シソーラスを用いて、個々の利用者の情報嗜好にパーソナライズされた情報推薦を可能とするものである。情報嗜好シソーラスは、推薦された情報に対する利用者の評価フィードバックにより継続的に更新される。本システムは、このように利用者の Web 閲覧行動や評価フィードバックに基づいて継続的に学習される情報

[†] 神奈川工科大学 情報学部 情報工学科
Department of Computer and Information Sciences, Kanagawa Institute of Technology

^{††} Department of Electrical and Computer Engineering, University of Victoria

嗜好シソーラスを用いて、刻々と変化する利用者の閲覧行動やその時点の利用者の興味・関心に適応した情報推薦を実現する。本研究では、プロトタイプとして構築した実験システムを用いて、提案方式の実現可能性を検証する。

2. 利用者の情報嗜好に適応する情報推薦システムの構成

図 1 は、提案システムの構成を示している。提案システムは、3 つの反復的なフェーズより構成される。最初のフェーズの間、利用者の Web 閲覧行動が監視され、個々の閲覧行動より重要な単語群が抽出される。初期の除法嗜好シソーラスは、個々の閲覧行動により得られる単語群と単語スコアにより構築される。第 2 フェーズにおいて、情報嗜好シソーラスの単語群と Web 文書間の類似性に基づいて、Web 文書がランキング形式で推薦される。最後の学習フェーズでは、推薦情報に対する利用者の評価フィードバックに基づいて、情報嗜好シソーラスが更新される。

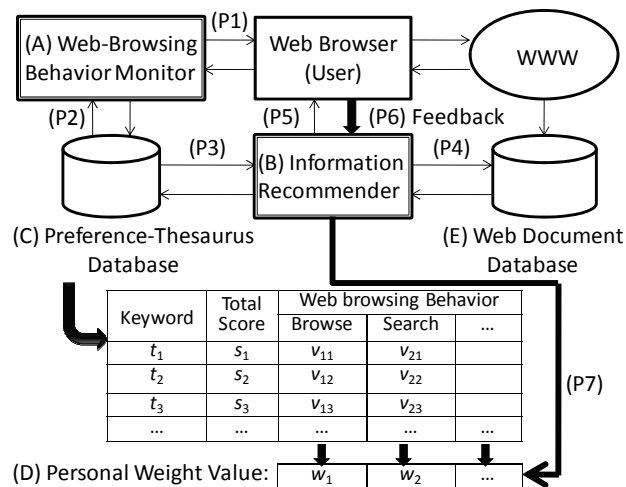


図 1 提案システムの構成

Figure 1 Recommender Architecture

2.1 Web 閲覧行動モニタ

典型的な Web 閲覧行動の例を表 1 に示す。利用者の Web 閲覧行動が継続的に監視され、図 1 (P1) に示されるように、個々の閲覧行動から重要な単語群が抽出される。

抽出された単語群とその単語スコアは情報嗜好シソーラス・データベースに蓄積される (P2)。例えば、表 1 に示されるように、閲覧される Web ページに対して、その Web ページに出現する単語群が利用者の情報嗜好に関連する重要単語としてみなされる。この場合、それぞれの抽出単語のスコアは、その単語を含む Web ページの閲覧した累積時間である。クリップボード・コピー操作に関しては、クリップボードへコピーされた単語群が抽出され、コピーした頻度が単語スコアとして格納される。

2.2 情報推薦システム

推薦候補となる Web 文書は、Web 文書データベースに蓄積される。これらの Web 文書は、Web クロール、RSS 配信、検索エンジンの検索結果など、様々な手段によって収集される。推薦候補 Web 文書は、重み付けされた単語集合より構成される情報嗜好シソーラスと Web 文書データベース上の個々の文書の類似度を算出することによりランキングされ (P3, P4)、利用者に提示される (P5)。

表 1 典型的な Web 閲覧行動
Table 1 Typical Web-browsing behaviors

ID	Web 閲覧行動	抽出される単語集合
I_1	Web ページを閲覧する	Web ページに出現する単語
I_2	Web ページ上でマウスクリックにより単語を選択する	選択された単語
I_3	Web ページ上で、単語をクリップボードにコピーする	クリップボードにコピーされる単語
I_4	Web ページ内で検索する	検索単語
I_5	Web ページを保存する	保存された Web ページに出現する単語
I_6	Web ページを印刷する	印刷された Web ページに出現する単語
I_7	Web をブックマークする	ブックマークされた Web ページに出現する単語
I_8	Web 検索エンジンで検索する	検索エンジンに入力された検索単語
I_9	Web 検索の結果として得られた Web ページを閲覧する	Web 検索の結果より、閲覧した Web ページに出現する単語

2.3 評価フィードバック

このフェーズでは、利用者は、推薦された Web 文書が自分の興味に関連があるかどうかを評価する (P6)。利用者が指定した関連文書に関連付けられる上位 n 件の Web 閲覧行動が、個々の閲覧行動に属する単語集合と関連文書間の類似度により抽出される。ここで、 n は類似度スコアが 0 より大きい閲覧行動の数であり、これらの n 個の

閲覧行動は、利用者の情報嗜好に影響があると考えられる。その時点の Web 閲覧行動と情報嗜好を反映するように、上位 n 件の Web 閲覧行動に関連付けられる重みが増加される。

3. Web 閲覧行動の学習による情報推薦方式

3.1 情報嗜好シソーラスの構築

表 1 に示されるように、典型的な Web 閲覧行動、およびそれらに相当する単語集合は、利用者の閲覧行動を監視することにより抽出され、個人の情報嗜好シソーラスを構築するために用いられる。Web 閲覧行動を I_x とし、閲覧行動 I_x から抽出される m 個の単語群を含む単語集合を T_x とする。

$$T_x = \{ t_{x1}, t_{x2}, t_{x3}, \dots, t_{xm} \} \quad (1)$$

ここで、 $t_{xi} (i=1, 2, \dots, m)$ は、 T_x に含まれる単語である。各 T_x に出現する全単語からなる総単語集合 T は、次のように表わされる。

$$T = \bigcup_{i=1}^x T_i \quad (2)$$

単語集合 T に含まれる各単語、および各閲覧行動 I_x を用いて、単語-閲覧行動行列が図 2 に示すように構築される。この単語-閲覧行動行列を情報嗜好シソーラスと呼ぶ。図 2 において、各要素 s_{ij} は、閲覧行動に対する単語 t_j のスコアである。

多くの単語スコア s_{ij} は閲覧行動の頻度により決定される。例えば、閲覧行動であるクリップボード・コピーの単語スコアは、利用者が単語 t_j を何回クリップボードにコピーしたかを示している。また、Web ページ閲覧においては、Web ページの閲覧時間が単語スコアとして用いられる。このように、各閲覧行動 I_i における単語スコアは、基本的に異なっている（頻度、時間等）ので、次章に示される (5), (8), および (11) の各式に従って正規化される。

	I_1	I_2	I_3	...	I_x
t_1	s_{11}	s_{21}	s_{31}	...	s_{x1}
t_2	s_{12}	s_{22}	s_{32}	...	s_{x2}
t_3	s_{13}	s_{23}	s_{33}	...	s_{x3}
t_4	s_{14}	s_{24}	s_{34}	...	s_{x4}
...	...	...	...	...	...
$t_{ T }$	$s_{1 T }$	$s_{2 T }$	$s_{3 T }$	...	$s_{x T }$

図 2 情報嗜好シソーラス

Figure 2 Personal preference-thesaurus matrix

3.2 Web 文書の推薦

Web 文書の推薦は、利用者個人の情報嗜好シソーラスと各 Web 文書の類似度に基づいて行われる。

最初に、総単語集合 T を用いて文書ベクトル空間 S が構築される。Web 文書集合 D における各文書 d_i は、文書中に含まれる単語の出現頻度に基づいて、文書ベクトル $\mathbf{d}_i$ として、次式のように表現される。

$$\mathbf{d}_i = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_{|T|} \end{bmatrix} \quad (3)$$

ここで、 e_k は文書 d_i に含まれる単語 $t_k (t_k \in T)$ の出現頻度であり、 $|T|$ は総単語集合 T に含まれる単語数である。

次に、文書集合 D に対して、個人の情報嗜好に適した文書獲得を実現するためのパーソナル・ランキングが、以下のステップで実行される。

Step-1: 最初に、各閲覧行動 I_k をベクトル $\mathbf{I}_k$ として表現することにより、利用者の閲覧行動に基づいた問い合わせが生成される。各ベクトル要素 s_{kj} は、情報嗜好シソーラスにおける各閲覧行動の単語スコアを示している。

$$\mathbf{I}_k = \begin{bmatrix} s_{k1} \\ s_{k2} \\ \vdots \\ s_{k|T|} \end{bmatrix} \quad (4)$$

Step-2: 閲覧行動ベクトル $\mathbf{I}_k$ を足し合わせることににより、問い合わせベクトル $\mathbf{q}$ を、次式のように生成する。

$$\mathbf{q} = \sum_{k=1}^x w_k \cdot \frac{\mathbf{I}_k}{|\mathbf{I}_k|}, \quad \sum_{k=1}^x w_k = 1 \quad (5)$$

ここで、 x は情報嗜好シソーラスで定義されている Web 閲覧行動の数を示している。また、 w_k は各閲覧行動 I_k に対する重み値であり、1 ノルムで正規化される。各閲覧行動の相対的な重要性は、あらかじめ決定しておくことができないため、重み値 w_k の初

期値は、次式のように設定される

$$w_k^{init} = \frac{1}{x} \quad (6)$$

Step-3: 問い合わせ q と D 中の Web 文書 d の類似度が算出される. 類似度の算出には, 非対称尺度 (asymmetric measure), Jaccard 尺度, 拡張 Jaccard 尺度といった, 各種の類似度計量尺度が適用可能である. 本研究では, 次式に示されるコサイン尺度を適用する.

$$sim(q, d) = \frac{(q \cdot d)}{|q||d|} \quad (7)$$

算出された類似度スコアに応じて, 各 Web 文書がランキングされる.

3.3 推薦 Web 文書の評価フィードバック, および, Web 文書の再推薦

式(5)において, Web 閲覧時の習慣には個人差があるため, 各閲覧行動 I_k の重み値 w_k は個人ごとに異なると仮定される. このため, 個々人の Web 閲覧行動の特性に適応して, 重み値 w_k を設定する必行がある.

利用者が, ランキングされた推薦 Web 文書のリストから, 興味・関心のある Web 文書を選択すると, 選択された Web 文書の類似度の算出に強く影響のある閲覧行動 I_k が抽出される. 抽出された閲覧行動 I_k に対応する重み値が増加され, 新しい問い合わせ q^{new} が形成される. このフィードバック・プロセスは, 下記のステップにより実行される.

Step-1: 推薦された Web 文書のランキング・リストから, 利用者は興味・関心のある Web 文書 d_f ($\in D$) を選択する.

Step-2: 閲覧行動 I_k と Web 文書の d_f 間の類似度スコアの算出に強く影響する閲覧行動 I_k を抽出するために, 下記のコサイン尺度を用いる.

$$score(I_k, d_f) = \frac{(I_k \cdot d_f)}{|I_k||d_f|} \quad (8)$$

閲覧行動 I_k は, 式(8)により算出される類似度スコアによりランキングされる.

Step-3: 式(8)により算出される類似度スコアが 0 より大きい場合, 閲覧行動 I_k の重み値が増加される.

$$w_k = w_k + \alpha_k \quad (9)$$

ここで, α_k は w_k の増加値である. Step-2 における閲覧行動 I_k のランキング順位を r とすると, ランキング順位 r に応じて, 各閲覧行動 I_k に対する増加値 α_k が下記のように設定される.

$$\alpha_k = w_k^{init} \cdot \frac{1}{r} = \frac{1}{x \cdot r} \quad (10)$$

また, アプリケーションに応じて, 式(10)の代わりに式(11)を使用する.

$$\alpha_k = 1 - r \cdot w_k^{init} = 1 - \frac{r}{x} \quad (11)$$

Step-4: 新しい問い合わせベクトル q^{new} が次式のように表わされる.

$$q^{new} = \sum_{i=1}^x w_i^{new} \cdot \frac{I_i}{|I_i|}, \quad \sum_{i=1}^x w_i^{new} = 1 \quad (12)$$

ここで, w_i^{new} は各閲覧行動 I_i に対する重み値であり, Step-3 における式(9), (10)

または(11)の処理後に 1 ノルムに正規化される.

新しい問い合わせ q^{new} を用いて, 3.2 節で述べたランキング処理が繰り返して実行され, 類似度スコアの上位 n 件の Web 文書が, 利用者に推薦される. この評価フィードバック処理は, 継続的に繰り返して実行されるものであり, たとえ時間の経過により, 利用者の情報嗜好や閲覧行動が変化したとしても, 提案システムはそれに適応することにより, 適切な情報推薦を行うことができる.

4. 実験

本実験では, 提案システムが, 利用者の Web 閲覧行動と情報嗜好に基づいた適切な情報推薦を行い, さらに, 評価フィードバックによる学習によって, 個々の利用者の情報嗜好に適応可能であることを確認する.

4.1 実験環境

3 章で示した方式により, 実験システムの構築を行った. Web 閲覧行動モニタは,

Firefox Web ブラウザの拡張として実装されている。情報嗜好シソーラスは、3.1 節に示されている方式で生成した。単語集合は、表 1 に示されている 9 つの Web 閲覧行動を監視することにより抽出した。また、2 つの情報嗜好シソーラス PT_1 , PT_2 を、2 人の利用者の閲覧行動を監視することにより構築した。ユーザ 1 は、しばしばニュースや音楽に関する Web ページにアクセスし、この情報嗜好は PT_1 に反映されている。また、 PT_2 はユーザ 2 のニュース、コンピュータ、テニス、音楽といった話題についての情報嗜好をプロファイルしている。

表 2 RSS フィードの内容
Table 2 Description of RSS Feeds

RSS feed 名	ジャンル	記事数
The New York Times	Business, Politics, Art, Technology, Entertainment.	729
BBC	Business, Politics, Art, Technology, Entertainment.	911
Reuters	Business, Politics, Art, Technology, Entertainment.	439
CNN	Business, Politics, Art, Technology, Entertainment.	342
Google News	Business, Politics, Art, Technology, Entertainment.	676
New Musical Express	Music	158
Rolling Stone Magazine	Music	63
MIT	Education	258
Stanford University	Education	84
IEEE	Computer Science	626
ACM	Computer Science	93
Nature	Natural Science	525
NASA	Space Engineering	131
Art Gallery of Ontario	Art	18
Royal Ontario Museum	Art	502
Apple Inc.	Computer Technology	119
JavaWorld Magazine	Computer Technology	22
US OPEN	Tennis	10
Wimbledon	Tennis	3
Premier League	Soccer	44
Total		5753

Web 文書データベースは、表 2 に示されるように、ビジネス、経済、政治、健康、自然科学、工学、コンピュータ、音楽、芸術、スポーツといったジャンルを対象とした複数の RSS フィードから取得された 5,753 件の文書から構成される。RSS 配信にお

ける Web 文書のタイトルおよび内容要旨 (description) が Web 文書のメタデータとして設定した。

コサイン尺度 (7) を用いて、情報嗜好シソーラスと Web 文書間の類似度を算出することにより、全 5,773 件の文書がランキングされる。

初期の PT_1 を用いてランキングした結果、上位 30 件の推薦文書のうち、25 件が政治に関するもの、5 件が音楽に関するものであった。これから、ユーザ 1 は、政治や音楽に関心があると考えられる。同様に、初期の PT_2 を用いてランキングした結果、上位 30 件の推薦文書のうち、9 件がコンピュータに関するもの、6 件がテニスに関するもの、11 件が政治に関するものであった。これより、ユーザ 2 は、テニス、コンピュータ、政治に関心がある可能性が高い。

4.2 実験方法

提案システムの利用者の情報嗜好への適応性を検証するために、評価フィードバックに関する実験を行う。 PT_1 に関しては、3 つの Web 文書 d_{558} (音楽), d_{4118} (音楽), および d_{4234} (政治) を指定することにより、ユーザ 1 が各々の文書に関心があることをシステムにフィードバックする。他の 3 つの Web 文書 d_{3118} (テニス), d_{4894} (政治), および d_{5559} (コンピュータ) を PT_2 へのフィードバック文書として用いる。各フィードバックは情報嗜好シソーラス PT_1 , PT_2 の初期状態において開始され、それぞれ独立に実行される。フィードバックの累積的な効果を評価するために、FB7, FB8, および FB9 を PT_2 に対して実行する。

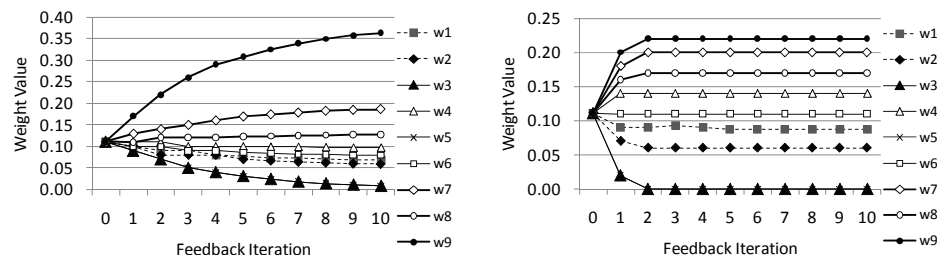
- [FB1] d_{558} を 10 回指定する。 (PT_1)
- [FB2] d_{4118} を 10 回指定する。 (PT_1)
- [FB3] d_{4234} を 10 回指定する。 (PT_1)
- [FB4] d_{3118} を 10 回指定する。 (PT_2)
- [FB5] d_{4894} を 10 回指定する。 (PT_2)
- [FB6] d_{5559} を 10 回指定する。 (PT_2)
- [FB7] d_{4894} および d_{3118} を交互に 10 回指定する。 (PT_2)
- [FB8] d_{3118} および d_{5559} を交互に 10 回指定する。 (PT_2)
- [FB9] d_{5559} および d_{4894} を交互に 10 回指定する。 (PT_2)

また、評価フィードバックによる学習に関して、3.3 節に示される式(10), (11)の両方の式を用いて、重み値を増加させた場合の結果を比較する。

4.3 実験結果

まず、図 3 に評価フィードバックによる Web 閲覧行動の重み値の変化の例として、FB2 において、3.3 節に示した式(10), (11)をそれぞれ適用した場合の結果を示す。こ

れより、式(10)を適用した場合、1回のフィードバックにおいて、FB2で指定したWeb文書に関付けられるWeb閲覧行動として1位にランクされたもの(w_9 , Web検索結果の閲覧)が、特に強い重み値が割り当てられることがわかる。一方、式(11)を適用した場合は、2位以下にランクしたWeb閲覧行動に対しても、ほぼ等しい間隔の減少分で、重み値が割り当てられていることがわかる。



(A) 式(10)を適用した場合 (B) 式(11)を適用した場合

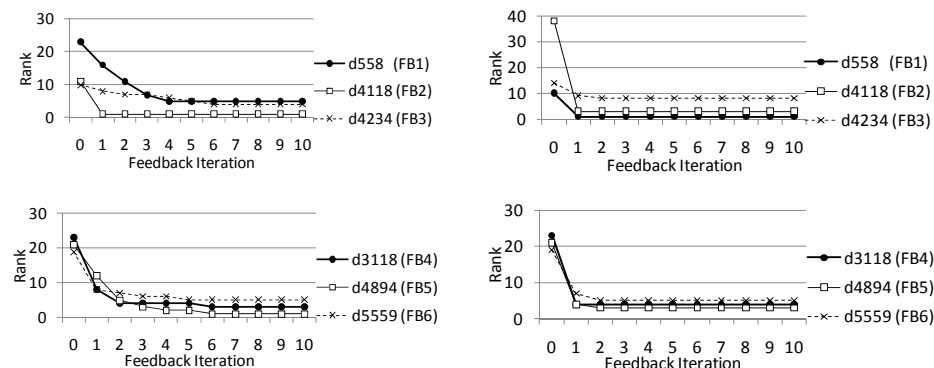
図3 評価フィードバックによるWeb閲覧行動の重み値の変化 (FB2)

Figure 3 Change of behavior weights (FB2)

図4により、フィードバックFB1からFB6に応じて、Web文書 d_{558} , d_{4118} , d_{4234} , d_{3118} , d_{4894} , および d_{5559} の順位が改善されていることがわかる。例えば、 d_{558} , d_{4118} , および d_{4234} はそれぞれ、23位、11位、および9位から、5位、1位、および4位と順位を向上させている。また、図4より、特定の文書を数回指定した後、フィードバックにおいて指定したWeb文書の順位が変動しなくなることがわかる。これは、上位にランクされている文書は、重みが増加したWeb閲覧行動に対してより強い関連があるためと考えられる。また、下位にランクされている文書は、フィードバックにおいて指定したWeb文書よりも、Web閲覧行動に対する関連性が低いため、フィードバックの繰り返しにより、フィードバックにおいて指定したWeb文書の順位は下がることはない。また、図4(A), (B)の結果において、式(10), (11)をそれぞれ適用した場合の差異は見られなかった。

図5は、フィードバックFB7, FB8, およびFB9における10回の繰り返しにおいて、各Web文書の順位がジグザグに変動している様子を示している。これは、異なるジャンルのWeb文書を指定することにより、その指定したWeb文書に関連付けられる異なるWeb閲覧行動が抽出されるためである。例えば、FB7において、 d_{3118} の順位が上昇すると、 d_{5559} の順位が下降しており、その逆もパターンも同様である。このような順位の変動は、Web文書の推薦順位に影響するWeb閲覧行動が異なること、および抽

出されたWeb閲覧行動の、Web文書の順位に影響する度合いが異なるために起こる。また、図5(A), (B)の結果において、式(11)を適用した場合は、式(10)を適用した場合よりも、大きな順位の振幅が起こることが確認できた。



(A) 式(10)を適用した場合

(B) 式(11)を適用した場合

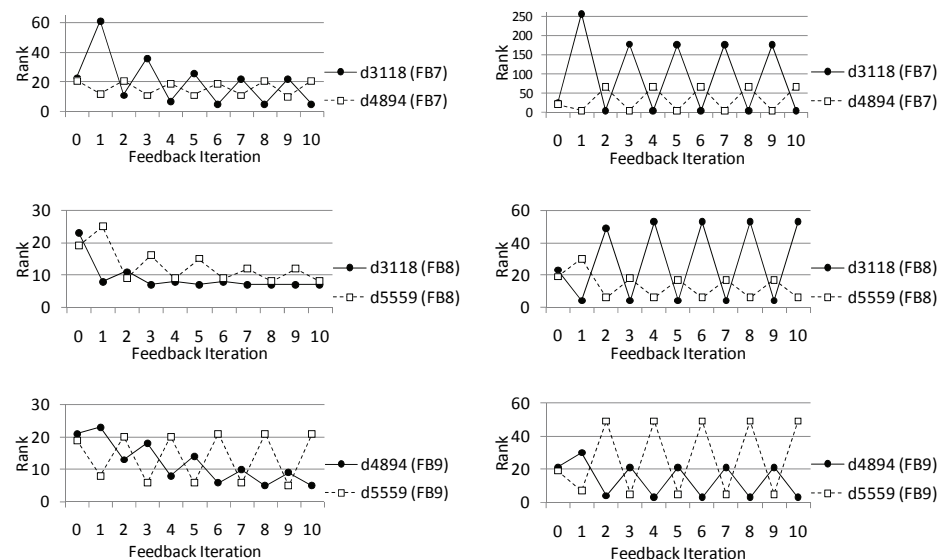
図4 ランキング順位の改善 (FB1~FB6)

Figure 4 Improvement of ranking results (FB1 to FB6)

図6と図7は、 PT_1 と PT_2 における上位10件の推薦Web文書のジャンル分布を示している。これより、評価フィードバックにより、上位10件の結果において、指定したWeb文書が属するジャンルの割合が増加することが確認できる。これは、評価フィードバックにおいてWeb文書を指定することにより、指定したWeb文書が属するジャンルに関連するWeb文書の順位も同様に上昇するためである。例えば、図7(A), (B)の両方において、初期の状態においては、政治、音楽、テニス、およびコンピュータに関連するWeb文書が上位10件に含まれている。FB5により、政治に関連するWeb文書 d_{4894} を10回指定することにより、上位10件中の結果において、政治に関連するWeb文書の割合が増加していることがわかる。また、式(10), (11)をそれぞれ適用した場合の比較において、式(10)を適用した場合の方が、上位10件中、指定したWeb文書と同ジャンルのWeb文書が占める割合が大きくなる傾向にあることが確認できる。

図7は、FB7, FB8, およびFB9の実行後に、評価フィードバックの累積的な効果により、上位10件の結果において、2種類のジャンルの占める割合が増加していることがわかる。例えば、FB9の結果は、FB5とFB6を組み合わせた結果に似ており、政治(FB5)とコンピュータ(FB6)に関するWeb文書の割合が大きく占めている。ま

た、この傾向は、式(10)を適用した場合の方が、式(11)を適用した場合よりも、顕著であることがわかる。



(A) 式(10)を適用した場合 (B) 式(11)を適用した場合

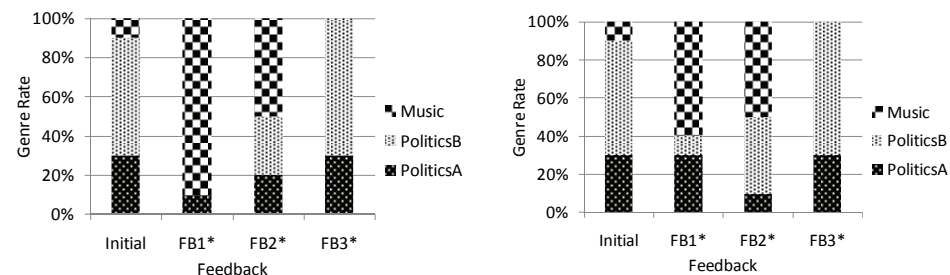
図 5 ランキング順位の改善 (FB7, FB8, FB9)

Figure 5 Improvement of ranking results (FB7, FB8, FB9)

表 3, 4 は, PT_1 と PT_2 のそれぞれに対して, 式(8)を用いて算出される Web 閲覧行動の順位およびスコアを示している。

表 3, 4 における順位に基づいて, 各 w_i は, 式(9), (10)または(11)により増加される。例えば, FB1 と FB2 に対して, I_9 (Web 検索結果の閲覧), I_6 (Web ページの印刷), および I_7 (Web ページのブックマーク) が最も影響のある Web 閲覧行動として抽出され, 図 3 に示すように, フィードバックの繰り返しに伴い, 対応づけられる重み値が増加される。FB1 と FB2 から, ユーザ 1 は, 音楽に関連する Web 文書について定期的に, Web 検索, 印刷, ブックマークする傾向にあることがわかる。さらに, FB3 より, I_7 (Web ページのブックマーク), I_6 (Web ページの印刷), および I_5 (Web ページの保存) が強い影響のある Web 閲覧行動であると考えられる。これから, ユーザ 1 は, 政治に関連する記事について, ブックマーク, 印刷, 保存する傾向にあることがわかる。

FB1, FB2, および FB3 の結果の比較において, I_6 と I_7 は, 異なるジャンルの Web 文書に対する共通の Web 閲覧行動であることがわかる。逆に, ユーザ 1 が, 興味・関心を示す Web ページについて, しばしば印刷, ブックマークする傾向があるという事実を利用して, ユーザ 1 の興味・関心を抽出することも可能であると考えられる。ユーザ 2 について, 表 4 より, 異なるジャンルの Web 文書に対して, 異なる Web 閲覧行動の影響が強いことがわかる。しかしながら, I_6 は影響のある Web 閲覧行動として共通であり, ユーザ 2 は重要な Web ページを印刷する傾向にあると考えられる。

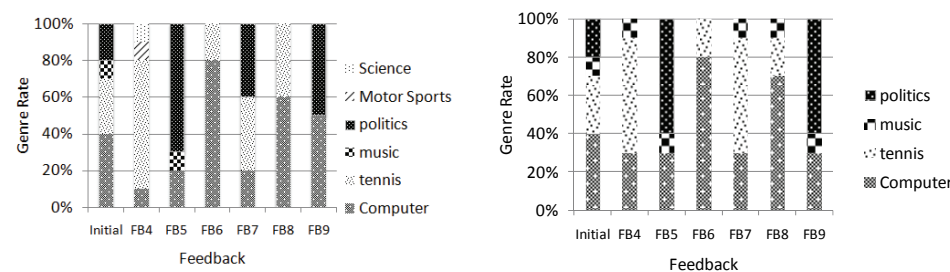


(A) 式(10)を適用した場合

(B) 式(11)を適用した場合

図 6 PT_1 を用いた場合の上位 10 件のジャンル分布

Figure 6 Genre distribution: top-10 of PT_1



(A) 式(10)を適用した場合

(B) 式(11)を適用した場合

図 7 PT_2 を用いた場合の上位 10 件のジャンル分布

Figure 7 Genre distribution: top-10 of PT_2

表 3 Web 閲覧行動の順位とスコア (PT_1)

Table 3 Rank and score of Web-browsing behaviors (PT_1)

Rank	FB1	FB2	FB3
1	I_9 (0.31)	I_9 (0.44)	I_7 (0.29)
2	I_6 (0.30)	I_6 (0.30)	I_6 (0.26)
3	I_7 (0.15)	I_7 (0.27)	I_5 (0.24)
4	I_2 (0.14)	I_5 (0.17)	I_3 (0.18)
5	I_1 (0.13)	I_1 (0.10)	I_1 (0.15)
6	I_3 (0.11)	I_4 (0.10)	I_9 (0.14)
7	I_4 (0.10)	I_2 (0.09)	I_4 (0.10)
8	I_5 (0.08)	I_3 (0.04)	I_2 (0.08)
9	I_8 (0.00)	I_8 (0.00)	I_8 (0.00)

表 4 Web 閲覧行動の順位とスコア (PT_2)

Table 4 Rank and score of Web-browsing behaviors (PT_2)

Rank	FB4	FB5	FB6
1	I_3 (0.33)	I_8 (0.26)	I_4 (0.26)
2	I_2 (0.26)	I_1 (0.18)	I_6 (0.22)
3	I_6 (0.18)	I_6 (0.17)	I_9 (0.21)
4	I_9 (0.11)	I_7 (0.16)	I_7 (0.10)
5	I_5 (0.06)	I_9 (0.13)	I_8 (0.08)
6	I_4 (0.06)	I_4 (0.08)	I_5 (0.07)
7	I_7 (0.02)	I_3 (0.03)	I_3 (0.06)
8	I_1 (0.01)	I_2 (0.03)	I_2 (0.05)
9	I_8 (0.00)	I_5 (0.00)	I_1 (0.01)

5. おわりに

本研究では、利用者の Web 閲覧行動の学習により構築される情報嗜好シソーラスに基づいて、利用者の情報嗜好に適応する情報推薦システムを提案した。プロトタイプ・システムを用いた実験により、提案方式の利用者の情報嗜好への適応機能を検証するとともに、評価フィードバックにより抽出される影響のある Web 閲覧行動に対応づけられる重み値の増加方法について、2 種類の増加関数を適用し、その考察を行った。

現在の実装では、もし、初期の情報嗜好シソーラスにおいて、実質的に影響のある

Web 閲覧行動に対応づけられる重み値が低く設定されてしまうと、評価フィードバックにおいて、それに関連付けられる Web 文書を指定する機会も減少するため、以降のフィードバック過程においても改善されない可能性がある。このような場合、提案システムは、利用者に対して適切な Web 文書を推薦することができなくなってしまう恐れがある。

今後の展開として、Web 閲覧行動に対応づけられる重み値を初期化する関数を設計することにより、このような問題に対処する予定である。また、従来の情報推薦方式との比較実験を行うとともに、より実際の Web 閲覧行動を伴った評価実験を実施することにより、提案方式の有効性を検証していく予定である。

参考文献

- 1) J.S. Breese, D. Heckerman, and C. Kadie, “Empirical Analysis of Predictive Algorithms for Collaborative Filtering,” Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, 1998, pp. 43-52.
- 2) P.-A. Chirita, C.S. Firan and W. Nejdl, “Personalized Query Expansion for the Web,” Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Amsterdam, The Netherlands, 2007, pp. 7-14.
- 3) S. Dumais, E. Cutrell, J.J. Cadiz, G. Jancke, R. Sarin and D.C. Robbins, “Stuff I’ve Seen: A System for Personal Information Retrieval and Re-Use,” Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2003, pp. 72-79.
- 4) I. Marcalis and E.D. Vita, “SEARCHY: An Agent to Personalize Search Results,” Proceedings of the 3rd International Conference on Internet and Web Applications and Services (ICIW 2008), 2008, pp. 512-517.
- 5) P. Resnick, N. Iacovou, M. Sushak, P. Bergstrom, and J. Reidl, “GroupLens: An Open Architecture for Collaborative Filtering of Netnews,” Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work Conference, 1994, pp. 175-186.
- 6) B.M. Sarwar, G. Karypis, J.A. Konstan, and J. Riedl, “Analysis of Recommendation Algorithms for e-commerce,” Proceedings of ACM Conference on Electronic Commerce, 2000, pp. 158-167.
- 7) Y.-W. Seo, B.-T. Zhang, “Learning User’s Preferences by Analyzing Web-Browsing Behaviors,” Proceedings of the 4th International Conference on Autonomous Agents, Barcelona, Spain, 2000, pp. 381-387.
- 8) L. Spalteholz, K.F. Li and N. Livingston, “KeySurf: A Character Controlled Browser for People with Physical Disabilities,” The Proceedings of the 17th International World Wide Web Conference WWW2008, Beijing, China, April 21-25, 2008, pp. 31-39.
- 9) K.H.L. Tso-Sutter, L.B. Marinho, L. Schmidt-Thieme, “Tag-aware Recommender Systems by Fusion of Collaborative Filtering Algorithms,” Proceedings of the 2008 ACM Symposium on Applied computing, 2008, pp. 1995-1999.