

F-10

ウェブサイトの定期訪問における状況に応じた推薦システムの試行

Trial of the Recommendation System according to the Situation in a Periodical Visit of Several Websites

堀之内 公亮† 辻 光宏†
Kousuke Horinouchi Mitsuhiro Tsuji

1. はじめに

我々は、ブラウジングしたり、他人から教えてもらうなどして、気になるウェブサイトや興味深いウェブサイトに出会う。そうしたウェブサイトに再び訪れ、新しく更新されたコンテンツを得ようとするのは、ごく一般的なことである。

そうして発見されたさまざまなウェブサイトは、ユーザが良質な情報源だと判断したウェブサイトであり、優先度の違いによる訪問頻度の差こそあれ、どのウェブサイトも定期的に訪問されるべきものである。しかし、情報源となるウェブサイトの数が増加するにつれ、優先度や訪問頻度に配慮した上で訪問することは困難になってくる。加えて、ユーザの状況によって訪問すべきウェブサイトは変わってくるため、その困難さは増加する。

本研究では、ユーザが容易に定期訪問を行えるようにするためのシステムを考察し試行したので報告する。

まず、定期訪問における基本的な指針についてまとめ、次に、その指針に加えて考慮すべきユーザの状況について述べる。その後、既存のツール、関連する研究について紹介した後、試作したシステムについて報告する。

2. 定期訪問での推薦システムの指針

インターネット上には、多くのウェブサイトが存在している。我々は、それらをブラウジングする中で、気になるウェブサイトや、興味深いウェブサイトを発見するだろう。更新が停止してしまわない限り、発見以降もウェブサイトのコンテンツは変化し続ける。このため、特定のウェブサイトに定期的に訪問し、そのコンテンツを確認することは、有益な情報を得るために必要不可欠なことである。

定期的なウェブサイトへの訪問を支援する研究として、アクティブブラウジングシステム¹⁾がある。このシステムは、定期訪問を支援する機能として、ユーザが最近頻繁に訪れているウェブサイトの中から、情報源として有用なものを自動で抽出し、推薦する機能を持っている。これにより、ユーザから情報源となるウェブサイトを管理する手間を省きつつ、興味の高いウェブサイトに素早くアクセスできる。

定期的なウェブサイトへの訪問を支援するという目的に関しては、本研究でも同様である。しかしながら、定期訪問すべきウェブサイトを推薦する場合に、優先度と未訪問の時間の二つを採用し、ウェブサイト間で訪問頻度に極端な偏りを生じないように配慮した。その理由を説明するため、まず、定期訪問の理想的なケースを想定し、その後、

現実的なケースを考察する。

理想的なケースでは、ユーザは定期訪問の機会があるたび、定期的に訪問したいと考え登録したウェブサイトすべてに訪問し、その内容を確認する。それならば、すべての登録されているウェブサイトの情報を得られることは自明である。

だが、これは不可能な場合がほとんどであろう。現実的には、定期訪問に利用できる時間は有限である。このため、登録したウェブサイトが多ければ、その時間内に収まらなくなる。また、それほどウェブサイトを登録していなかったとしても、定期訪問に利用できる時間が短ければ、すべてのウェブサイトを訪問することは不可能である。

こうした現実的な条件で定期訪問を行うためには、訪問するウェブサイトと訪問しないウェブサイトに分ける必要がある。そのために用いる基準として最初に考えられるのが、ウェブサイト間の優先度の違いである。登録されているすべてのウェブサイトは、ユーザが訪問したいと考え、登録されているものではある。とはいえ、それらすべてがユーザにとって同じ価値を持っているとは考えにくい。これを背景として、基準として優先度を採用し、優先度が高いウェブサイトを訪問しやすくすることは妥当であると言える。

ただし、優先度だけを基準にするのでは、適切な定期訪問は行えない。すなわち、もしその基準だけで訪問するウェブサイトを選べば、選ばれるウェブサイトは高い優先度を持つウェブサイトだけになってしまう。優先度についての考察で述べたように、登録されているすべてのウェブサイトは、ユーザが訪問したいと考えているものであるから、そのように偏った訪問が行われることは良いことではない。

この問題を解決するために、もう一つの基準として、ウェブサイトへ未訪問の期間を採用する。そして、この期間が長ければ長いほど、訪問されやすくなるようにする。こうすれば、優先度の低いウェブサイトでも、定期的に訪問されるようにできる。なぜならば、訪問するウェブサイトには選ばれない場合が続けば続くほど、未訪問の期間は増加していくからである。最終的に、未訪問期間の基準の影響が優先度の基準の影響を上回れば、優先度の低いウェブサイトが訪問するウェブサイトには選ばれる。

さらに、この二つの基準と、何らかの重みを用いて適切な評価値を求めることができれば、優先度の高いウェブサイトを訪問しやすくしつつ、定期的に優先度の低いウェブサイトを訪問できるようになる。

図 1 に、訪問すべきウェブサイトを推薦する基準について示した。定期訪問の支援を行う場合は、この基準を考慮して評価値を計算し、訪問すべきウェブサイトを提示することで、極端な偏りが生じない定期訪問が実現できる。

† 関西大学大学院 総合情報学研究科, Graduate School of Informatics, Kansai University

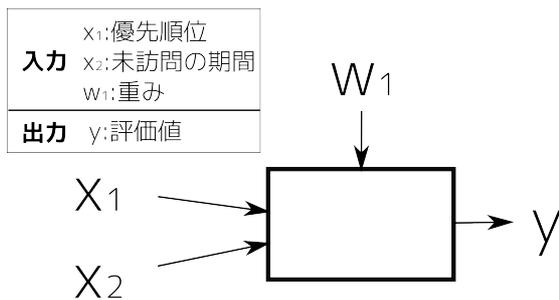


図 1:評価値計算の概要

Fig 1:Evaluation about recommendation of periodical visit

3. 状況に応じた推薦

2.で述べたように、優先度と未訪問の時間、そして重みを利用してウェブサイトを選ぶことができれば、訪問するウェブサイトに極端な偏りを生じさせないようにできると考えられる。この時、優先度と未訪問の時間をそれぞれの程度評価値に影響させるかを求めるため、何らかの方法で重みを決定する必要がある。

この重みは、状況に応じて変えるべきだと考えられる。訪問したいウェブサイト間の訪問頻度の極端な偏りをなくすため、優先度の低いウェブサイトも定期的に訪問することがユーザにとって利益となることは、2.で考察した通りである。しかし、それはあくまで、大局的な視点で見た場合でのユーザの利益であり、ユーザの主観的な感情を考慮していない。

例えば、ユーザがここ数日間多忙な日々を送っている場合を考える。この場合、ユーザがしばらく満足に定期訪問を行っていないと考えられる。ユーザからすれば、重要だと思っている情報源からあまり情報を得られていないと感じているはずである。そうした状況にある時に、優先度の低いウェブサイトを支援システムに提示されたとしても、ユーザが素直にその提案を受け入れるとは思えない。このように、提案が利益になると分かっているにもかかわらず、ユーザがその提案を受け入れにくい状況にあればストレスを感じてしまう。場合によっては、その推薦システムを利用しなくなってしまうことさえ考えられる。

逆に、ユーザがここ数日間通常どおり定期訪問をしており、優先度の高いウェブサイトを適度に訪問していれば、システムが優先度の低いウェブサイトを推薦してきても、それを受け入れる余裕があると考えられる。そうしたことを考えると、ユーザの状況によって重みを調整することは、定期訪問を支援するシステムにとって、重要な機能の一つである。

4. 既存のツールの特徴

定期訪問に関する既存のツールの傾向は、ウェブサイトへの定期訪問を支援する方法を基準として、二つに分類することができる。一つは更新チェックを主体とした支援、もう一つはルール（スケジュール）を主体とした支援である。

更新チェックを主体とした支援は、ウェブサイトの更新

日時を調べ、それを利用することで、ユーザのウェブサイトへの定期訪問を支援する。通常、ユーザがウェブサイトの確認を行う場合、事前にそのウェブサイトが更新されているかどうかを知ることができない。そのため、もしウェブサイトが更新されていなければ、ユーザは新しい情報を得ることはなく、訪問は徒労に終わってしまう。しかし更新チェックを主体とした支援の場合には、そうしたユーザの無駄な行動を抑えることができる。まず、ツール側は事前にウェブサイトの更新日時を調べる。次に、ユーザが前回訪問した日時と比較する。もし、前回訪問した日時が更新日時より後であれば、ユーザの訪問後にコンテンツの更新は起きていないということになる。すなわち、そのウェブサイトについては訪問する必要があることが分かる。これによって、余計な訪問が抑えられ、効率が良くなる。また、無駄な行動が減ることで、ユーザの徒労感も減少するだろう。

この種の支援は比較的多くの場所で採用されている。また、更新をチェックするという意味では、フィードリーダーもこの区分に入るだろう。フィードリーダーは、Atom²⁾のような更新情報を記述したフィードを解析し、ユーザに更新を通知するツールである。入力はフィードに限られるものが多いが、フィードからは更新日時だけではなく、コンテンツの概要や内容も得られるため、表現力が高い。実際のツールとしては、WWWC³⁾や、Brief⁴⁾などが挙げられる。

もう一つのルール（スケジュール）を主体とした支援では、何らかの条件を満たした場合に、ウェブサイトをユーザに提示する。こうした支援方法は、定期的に更新されるウェブサイトを確認する際など、習慣的に訪問するために使う場合に有用であろう。この種の支援を行うツールはあまり見られないが、例えば、Morning Coffee⁵⁾というツールでは、「毎週月曜日」「平日」のように、スケジュールのような形で条件を設定できる。

こうした既存のツールでは、ユーザが訪問すべきウェブサイトについて、2.や3.での考察を基に訪問を行うことは難しい。

どちらの種類ツールでも、直接2.で考えたような基準でユーザにウェブサイトを提示させることはできない。また、人の手のある程度借りるとしても、基準を考慮して訪問頻度が偏らないようにしながら、訪問すべきウェブサイトを選び出すのは困難であろう。更新チェック型のツールでは、更新日時でウェブサイトを並べ替えることができるものが多い。この機能を用いることで、更新されてから時間の経ったウェブサイトを探すことはできる。だが、そこからさらに自身にとっての優先度を考慮し、訪問頻度を偏らないようにウェブサイトを選び出すのは骨の折れる仕事である。また、人手による判断では、例えば優先度の高いウェブサイトを選びやすくなったり、優先度と未訪問の時間を考慮する基準が変化する、といった問題も考えられる。

5. 関連研究

フィードリーダーに関する研究として、ニュース記事の多数の更新情報の中から、ユーザの嗜好に沿ったものを選び出し、推薦する研究がある⁶⁾⁷⁾。嗜好に沿ったものを選

び出すことで、ユーザは不要な情報を見ずに済み、効率良く情報を収集することができる。これらは、ニュースサイトのように、ユーザにとって関係の無い、ほぼ不要な情報が多く含まれる場合に非常に有効だと考えられる。だが、こうしたアプローチでは、今回の定期訪問のように、単純に不要だと除外しにくいコンテンツがある場合には適用しにくいことがある。本研究では、そうしたコンテンツを省みる仕組みを用意しているという点で異なる。

6. 試作システム

優先度と未訪問の時間、そしてユーザの状況を鑑み、ユーザが訪問すべきウェブサイトを推薦するシステムを試作した。なお、本システムは Firefox の拡張機能として実装を行った。2010年7月30日の時点での実装環境は Firefox 3.6.8 と Ubuntu Linux 10.04 である。

6.1 更新の監視

4.で更新チェック型の利点として述べたように、ウェブサイトの更新日時を調べることで、ユーザが更新されていないウェブサイトへの無意味な訪問を防ぐことができる。

更新チェックの方法として利用する方法として、主にフィードや HTTP レスポンス・ヘッダーの Last-Modified

フィールドを用いる。フィードを用いる理由の一つとして、近年、ブログの台頭などにより、フィードを提供するウェブサイトは増加してきていることが挙げられる。また、フィードを提供していないウェブサイトからフィードを生成するサービス⁸⁾も存在することも理由の一つである。更新日時が正確に取得できることなどからも利用する意義は大きいと判断し、フィードを更新チェックに利用している。

ただし、フィードが生成できない場合なども考え、Last-Modified フィールドを用いる方法や、更新チェックを行わない場合も選択できるようにしている。

6.2 評価値の計算

評価値を計算する式として、以下の式を用いる。

$$estimate_i = \frac{unvisited_i}{LimitTime} + Weight \times priority_i \quad (式1)$$

ここで、 $estimate_i$ はウェブサイト i に対する評価値、 $unvisited_i$ は未訪問の時間、 $LimitTime$ は $unvisited_i$ の限界となる時間を表す。また、 $priority_i$ はウェブサイト i の優先度、 $Weight$ は優先度に対する重みである。ただし、評価値 $estimate_i$ は 0 から 1 (計算の結果最大値を超えた場合は 1 として扱う) の値を、優先度 $priority_i$ と重み $Weight$ は 0 から 1 の値をとる。ただし、重み $Weight$ のとる値に関しては、

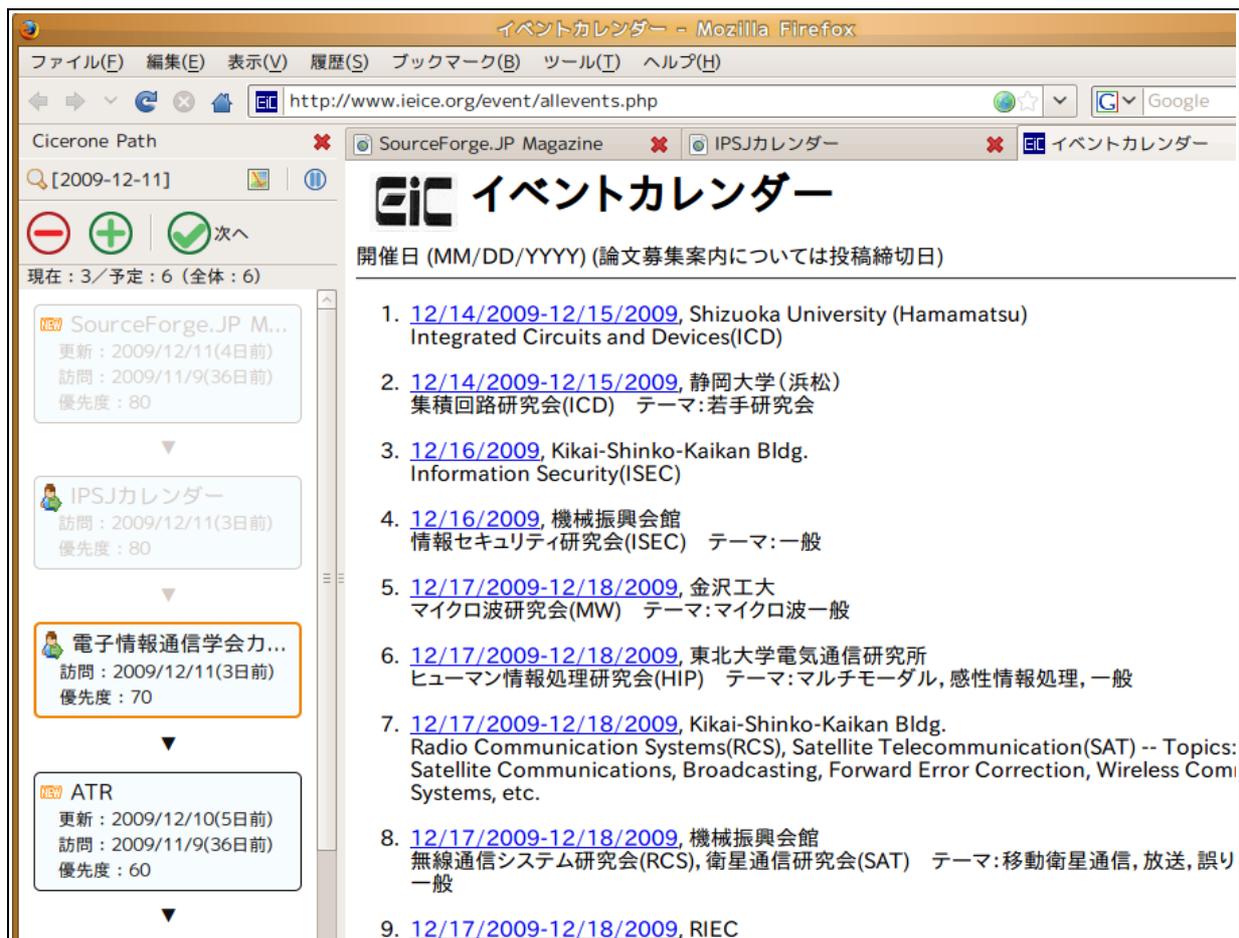


図 2:定期訪問の例

Fig 2:Example of periodical visit.

後述するようにやや制限がある。

(式1)の第一項では、未訪問の時間 *unvisited* が高いほど評価値が上昇し、限界時間 *LimitTime* に達したとき、最大値1を取る。評価値 *estimate* の最大値も1であるため、未訪問の時間 *unvisited* が限界時間 *LimitTime* に達したウェブサイトは、必ず最大の評価値が与えられる。つまり限界時間 *LimitTime* は、訪問されない時間に制限を課す役割を果たしている。*LimitTime* はユーザによって異なると考えられるが、後で述べる試作システムの初期値としては、30日を設定している。

一方、(式1)の第二項では、優先度 *priority* が高いほど評価値が上昇する。優先度 *priority* は0から1の値をとるため、重み *Weight* を調整することで、評価値 *estimate* への影響を変化させることができる。これにより、状況に応じた推薦を実現することができる。

本研究では、ユーザの状況として、3.で述べた、ユーザの忙しさを考える。忙しさを求める指標として、直近数日間の定期訪問を行っている時間の平均を用いる。これがそれまでの定期訪問の時間の平均に比べて小さければ、直近数日間は通常より定期訪問していない、つまり忙しいということになるだろう。逆に大きければ、通常よりも時間に余裕があるということである。こうしてユーザが忙しいかそうでないかを判断し、重み *Weight* を増減させる。

ただし、*Weight* は0から1までの範囲で変動するものとしているが、実際には0.3から0.7といった、やや狭い範囲で増減させるべきである。たとえば、*Weight* が0であれば、まったく評価値に優先度が反映されなくなり、逆に1であれば優先度の高いウェブサイトばかりが推薦されるようになってしまう。これは、本システムの提案する、訪問頻度の極端な偏りを生じさせない推薦システムという理念から反するため、ユーザの状況により重み *Weight* を調整する場合、下限と上限を設けておく必要がある。

6.3 基本的な動作の流れ

システムを用いて定期訪問を行っている様子を、図2に示す。ユーザが訪問すべきウェブサイトの評価値の高い順に並べ、ブラウザのサイドバーに提示している。その上部には、次のウェブサイトを表示するためのボタンや、優先度を調整するためのボタンなどを配置している。

定期訪問を行う場合の基本的な流れは、以下の通りである。

- (1) ユーザは、更新のチェックを要求する
- (2) システムは、更新のチェックを行う
- (3) ユーザは、次のサイトの表示を要求する
- (4) システムは、評価値が最高のサイトを表示する
- (5) ユーザは、内容を閲覧する
- (6) 必要ならば優先度を調整する
- (7) 続けてウェブサイトを確認する場合は(3)に戻る

このように、ユーザは基本的には、次のウェブサイトを表示するよう要求し、それを閲覧するだけであり、優先度や未訪問の時間を意識する必要はない。

7. 今後の課題と展望

今後は、より直感的なウェブサイトの管理インタフェースの構築を進めたい。今回の研究では、主に定期訪問における推薦の指針を研究対象としてきた。だが、現在のシステムでは、ウェブサイトの情報がユーザにとって管理しにくいものになっている。管理情報の視覚化などによりそれらを開示するように改善することで、より良い推薦が行えることが期待される。

参考文献

- 1) 戸田貴博, 峯恒憲, 雨宮真人: ブラウジング支援のためのアクティブブラウジングシステムの提案, 電子情報通信学会技術研究報告. KBSE, 知能ソフトウェア工学, Vol.106, No.473, pp.1-6(2007)
- 2) <http://www.ietf.org/rfc/rfc4287>
- 3) <http://www.nakka.com/wwwc/>
- 4) <https://addons.mozilla.org/ja/firefox/addon/4578/>
- 5) <https://addons.mozilla.org/ja/firefox/addon/2677/>
- 6) 向井誠, 青野雅樹: RSSに基づく個人向け内容型情報推薦プロトタイプシステム, 情報処理学会研究報告. FI, 情報学基礎, Vol.2005, No.94, pp.27-32(2005)
- 7) 高林光, 古井陽之助, 速水治夫: ニュースを選別する手間を軽減させるRSSリーダの提案, 情報処理学会研究報告. GN, グループウェアとネットワークサービス, Vol.2007, No.32, pp.133-137(2007)
- 8) <http://myrss.jp/>