

ブラウジングにおける動的興味の推定手法

4K-13

平嶋宗 松田憲幸 辻本昇平 野本豊裕 豊田順一

大阪大学産業科学研究所

1. はじめに

ハイパーテキスト構造を対象とした情報収集形態として、最もよく見受けられるものがブラウジングである。このブラウジングにも様々な形態が存在するが、収集した情報の影響を受けることにより、興味の対象を徐々に変化してゆくブラウジングが注目されている[1, 2, 4]。このようなブラウジングにおいては、ある時点での興味はその時点の近傍において収集された情報に強く影響されることになる。筆者等はこのようなブラウジングを文脈依存性ブラウジング (context-sensitive browsing) と呼んでいる。この文脈依存性ブラウジングが行われるためには、自由度の大きな情報収集が可能でなければならないが、WWWは大規模かつ多様な興味の下での利用が可能であるだけでなく、オーサによって設計されたリンクを辿る以外に、サーチエンジンなどを用いる手段が提供されている。したがって、WWWの普及に伴って、文脈依存性ブラウジングの頻度はますます大きくなっていくことが予想される。本研究では、この文脈依存性ブラウジングの支援を指向して、文脈に依存して動的に変化するユーザの興味を推定する手法を検討している。さらに、この推定結果に基づく支援手段としての情報フィルタリングについても述べる。

2. 興味の推定手法

2.1 概要

ブラウジングにおけるユーザの興味を履歴より推定する手法は、これまでにいくつか提案されている。しかしながら、これらの手法では時系列を考慮してないため、十分過去に収集された情報と直前に収集された情報とで、現時点での興味に与える影響を等価に見積もっている(*1)。これに対して、本研究では、過去に収集された情報の影響を減衰させる手法(重み減衰手法: Weight Fading Method)を提案する。

ここでは、ブラウジングにおいて収集される情報は、各ノードを一単位として考え、各ノードの内容は、そのノードに含まれているキーワードの集合として表現することができると仮定する(*2)。このとき、ユーザがあるノードに興味を持ったということを、そのノードの内容を表現しているキーワードに興味を持ったということであると解釈する。履歴としては、ユーザがこれまでに訪れたノードを訪れた順に並べたものを用い(*3)、各ノードには過去のものほど減衰する重みを与える。履歴中のあるノードに含まれるキーワードは、そのノードの重みを持つこととする。あるキーワードの現在の重みは、履歴中のそのキーワードの属しているノードの重みから計算されることになる。現在の興味は、キーワードとその重みの組みの集

合として表現される。この表現を興味モデルと呼んでいる。

2.2 重みの定義

この手法では、ユーザが新しいノードを訪れる度に過去のノードの重みが増加する。したがって、単純に考えれば再計算が必要になり、また履歴が長くなるに従って多くの計算が必要となる。実効的な支援を実現するためには、計算量を減らす工夫が不可欠となる。また、興味の動的な変化に追従するためには、履歴の長さに関わらず現在の興味における各キーワードの重みは可変であり、発散してはいけない。以上の条件を満足する関数として、ここでは、等比数列を用い、

$$Nw(k) = r^{n-k} \quad (0 < r < 1)$$

と定義する。ここで、 $Nw(k)$ ($1 \leq k \leq n$) は、 n 個のノードを見た時点での k 番目に通ったノードの重みを表す。 r の値は、ノードの重みの減衰率である。 n 個のノードを訪れた時点でのキーワード i の重み $Iw(i, n)$ は、この式を用いて、以下のように表せる。

$$Iw(i, n) = \frac{\sum_{k=1}^n Nw(k)}{\sum_{k=1}^n Nw(k)} = \frac{S(i, n)}{T(n)}$$

ここで、 $Nw(k)$ はキーワード i が含まれていたノードの重みであり、 $S(i, n)$ は n 個のノードを見た時点でのキーワード i の重みの総和を表す。 $T(n)$ は n 個のノード全ての重みの総和であり、規格化のために用いられている。このような式を用いることにより、 $S(i, n)$ と $T(n)$ の更新には次の漸化式を用いることができる [3]。

$$\begin{aligned} T(n) &= \sum_{k=1}^n Nw(k) = \sum_{k=1}^n r^{n-k} \\ &= 1 + rT(n-1) < \frac{1}{1-r} \end{aligned}$$

$$\begin{aligned} S(i, n) &= s + rS(i, n-1) \leq T(n) \\ s=0: & n \text{ 番目のノードに } i \text{ がいないとき} \\ s=1: & n \text{ 番目のノードに } i \text{ があるとき} \end{aligned}$$

3. 興味推定の結果を用いたフィルタリング

3.1 フィルタリング

2. で述べた興味の推定手法の有効性を確認するためには、CD-ROM化された事典(*4)を対象としたブラウザを作成し、そのブラウザを用いたブラウジング

*1: ユーザの興味が安定している場合にはこの手法が有効であり、実際に、マニュアルなどの使われ方が比較的にはっきりしているハイパーテキスト構造を対象としたブラウジングは、興味の安定したものが多いことが報告されている。

*2: ただし、現段階ではキーワードの有無だけが用いられており、頻度は考慮されていない。

*3: ただし、ユーザが興味があると判断したノードのみを対象としている。

の支援機能としての有効性を検証した。このブラウザを用いることにより、各ノードに用意されたキーワードを介してハイパーテキスト的に他のノードを訪問できる。この際、一つのキーワードには2~2, 276 (平均7) 個のノードが結びつけられており、これらの中から興味あるものを見つけることがユーザの負担になることが多い。そこで、2. で述べた手法によってユーザの興味を推定し、次候補のフィルタリングを行う。まず、次候補に含まれているキーワードを抽出する。次に、各キーワードに対するユーザの興味の度合いを重みとして2. で述べた手法を用いて計算する。各候補の重みは、その候補の含んでいるキーワードの重みの合計とし、大きな重みを持つものほど現在のユーザの興味に合致しているとして、選択候補リストの上位に提示する。

3. 2 評価

12名のユーザに上記のブラウザを用いたCD-ROM事典のブラウジングを行ってもらったことにより評価を行った。この際、(1)本手法、(2)直前に見たノードに含まれているキーワードの重みだけを考慮した場合、(3)単なる検索によって提示される順序、の3通りの選択候補の提示方法を用意しておき、各人が3回、それぞれ異なった選択候補提示順序でのブラウジングを行ってもらった。ただし、ユーザには選択候補の提示方法が変化していることは知らせていない。また、ユーザに対しては単にブラウザを使って事典を読むように依頼しただけである。

ここでは、ある程度のユーザのブラウジングが進んだ段階(5個以上のノードを訪れた後)で、かつ、次の選択候補がある程度数が多くなる場合(30~60個)に、ブラウジングを中断し、各選択候補の内容を読んでもらい、次に見る内容として適当であるかどうかを適・不適の2値で評価してもらった。そして、「適」とされた選択候補が上位に配置されているかどうかで各選択候補の提示法の得点を計算した。この得点の大小を順位尺度として扱い、3条件のうち2条件についてそれぞれサイン検定により両側検定を行った。結果として、本手法が他の手法に対して有意に高い点をとることを1%の危険率でいうことができた。このことより、本手法は、他の二つの提示法に比べて、ユーザによって興味深い内容の選択候補を、リストの上位に配置できているといえる。

実験後にシステムについての使い勝手についてのコメントを求めたところ、3回のブラウジングにおいて使い勝手が変化した旨のコメントは得られなかった。このことは、本手法が他の単純な手法と大差のない実用的な速度で動作していることを示すと同時に、ユーザにとって直接実感できるほどの支援とはなっていないことを示している。本評価実験では、ユーザに対して全く自由にブラウジングしてもらっており、各試行毎にユーザの行っているブラウジングの内容は異なっている。このことが、本手法による選択候補の提示が他の手法よりも遅い(*5)ことを隠しているとともに、支援の効果についても認識できなくしていると考えら

*4:各ノードはある用語についての解説になっており、全部で18,892ノード含まれている。各ノードには検索用に予めキーワードが与えられている(平均70個)。

*5:単純に検索された順序の提示に比べて、約1.5倍の時間がかかっている。

れる。

4. まとめ

本稿では、文脈依存性ブラウジングの支援を試行した、重み減衰法を用いたユーザの興味推定手法と、推定された興味に基づくフィルタリング手法について述べ、実際のユーザによる実験評価について報告した。この結果は本手法が有望であることを示しているが、本研究はまだ初歩的な段階にあり、多くの問題を含んでいる。以下に、今後解決すべき問題点を挙げる。

(1)文脈依存性ブラウジングについての調査:文脈依存性ブラウジングの存在はいくつかの研究ですでに示唆されているが、いまだ十分な調査が行われているとはいえない。筆者等は、ブラウジング過程でユーザが訪れたノード間の関連性の変化を調べることにより、文脈依存性ブラウジングの存在を明らかにしようとしている。また、対象となるハイパーテキスト構造の規模や性質によって、よく現れるブラウジングの形態も異なってくると予想される。これらについても調査することを計画している。

(2)キーワードの自動抽出と頻度情報の利用:本手法では、予め与えられているキーワードを利用し、また、頻度については考慮しない。しかしながら、ベクトル空間法をはじめとする情報検索に関する研究では、どのような単語をキーワードとし、また頻度情報をどのように利用するかについて多くの提案がなされている。本研究でもこれらの成果を利用することが必要と考えられる。

(3)WWWへの適用と直接的評価:本研究の目標は、ユーザに快適なブラウジング環境を提供することである。したがって、ユーザによって直接的に有効と判断できるレベルの支援を目指している。対象となるハイパーテキストが大規模になればなるほど本手法による支援効果が明らかになることが期待できる。筆者等は、最も大規模なハイパーテキスト構造であるWWWを対象とした支援の実現を目指している。

(4)減衰しない場合との比較:本手法の最大の特徴は、過去の履歴の影響を減衰させる点にある。しかしながら、本稿での評価実験においては、減衰しない場合との比較を行っていない点で不備がある。本手法で減衰させない場合は r の値を1とした場合に相当する。この場合に加えて、既に提案されているブラウジング支援に関するいくつかの手法との比較検討を行うための実験を行う予定である。

参考文献

- [1] Armstrong, R., Freitag, D., Joachims, T., Mitchell, T.: WebWatcher: A Learning Apprentice for the World Wide Web, Proc. of AAAI Symposium on Information Gathering from Distributed, Heterogeneous Environments (1995) (available at <http://WWW.isi.edu/sims/knoblock/sss95/info-gathering.html>).
- [2] 蜂谷憲一, 平嶋宗, 柏原昭博, 豊田順一: 文脈情報を用いたハイパーテキストナビゲーションシステム, 人工知能学会全国大会(第9回)論文集, pp.605-608 (1995).
- [3] 平嶋宗, 柏原昭博, 豊田順一: 文脈情報を利用したハイパーテキストのブラウジング支援とその実験的評価, 人工知能学会研究会資料 SIG-HIDSN-9602-04 (1996).
- [4] Lieberman, H.: Letizia: An Agent That Assists Web Browsing, Proc. of IJCAI95, pp.924-929 (1995).