

ニュース記事のパーソナライズサービス方式の提案と評価

橋高博行 佐藤直之 鈴木英明 曾根岡昭直

NTTソフトウェア研究所

概要

近年インターネットの普及に伴って、WWWを用いて発信される情報の量も増加の一途をたどっている。しかし一方で、膨大な情報の中からユーザが自分の興味のある情報を選択的に獲得することは難しくなっている。このため、WWWで情報提供を行う際、画一的な情報の提供を行うだけでなく、個々のユーザの興味を考慮して情報の提供を行うことが求められている。本稿では、個々のユーザの興味に応じて選択的に情報を提供する情報提供方式を提案し、ニュース記事を提供するWWWサイトに適用した評価結果について報告する。

A Proposal and Evaluation of News Personalized Navigation Service

Hiroyuki KITAKA Naoyuki SATOH Hideaki SUZUKI Terunao SONEOKA

NTT Software Laboratories

abstract

On popularity of the Internet, an amount of information becomes large on the Web. It is difficult for users to again information, in which they take interest, from many information. Therefore, it becomes more important to provide methods of filtering information to reflect the necessity of every user. In this paper, we present an adaptive personal information filtering system, and evaluate the performance of our system in case of applying it to a newspaper providing service.

1 はじめに

近年インターネットの普及に伴って、WWWを用いて発信される情報も増加の一途をたどっている。しかし一方で、膨大な情報の中からユーザが自分の興味のある情報を選択的に獲得することは難しくなっている。このため、WWWで情報提供を行う際、画一的な情報の提供を行うだけでなく、個々のユーザの興味を考慮した情報の提供を行うことが求められている。

これまで、個々のユーザの興味に応じて選択的に情報を提供する情報提供サービスとしては、WebWatcher[1][2]、FireFly[3]等が提案されている。しかし既存の方式では、ユーザの興味の表現方法、情報の特徴づけ、ユーザの興味と情報の特徴の照合等をどのように行うかという点で課題を残している。

本稿では、上で述べた課題を考慮した情報提供方式(InfoBroket方式[4])を提案する。また、ニュース記事を提供するWWWサイトに本方式を適用した評価結果について報告する。

2 情報提供方式の要求条件

個々のユーザの興味に応じた選択的な情報の提供とは、例えば、オペラと安売りチケットに興味を持つユーザには、オペラの安売りチケットに関する情報を優先的に提供することである。すなわち、ユーザの興味や個人情報等と、情報の持つ特徴や性質等を参考にして、より優先されると思われる情報を選択的にユーザに提供することである。したがって、このような情報提供方式では、ユーザの興味の表現方法、情報の特徴の表現方法、これらの照合方法を検討する必要がある。また、ユーザの興味、情報の特徴は変化するものである。このため、変化に追従してユーザの興味、情報の特徴を適切に表現するためには、これらの更新方法も併せて検討しなければならない。以下に検討すべき点をまとめ、各節でそれらの要求条件を詳述し既存方式の問題点を明らかにする。

- ・ユーザの興味の表現、更新方法
- ・情報の特徴の表現、更新方法
- ・ユーザの興味と情報の特徴の照合方法

2.1 ユーザの興味の表現、更新方法

ユーザの興味の变化に追従した情報提供方式では、ユーザの興味の表現、および更新をユーザの行動等の情報に基づいて自動的に行うことが望ましい。ここで要求条件は以下の3点である。

- (1)ユーザの労力の低減
- (2)興味の变化への動的な追従
- (3)新規項目の追加が容易

現在提案されている方式では、ユーザの興味を把握するためにアンケートを利用することが多い。これは、ユーザがアンケートに答える形で、提示されたカテゴリの中から興味のあるカテゴリを選択するものである。ここで、ユーザの興味は各カテゴリに対しての興味の有無で表現される。この方式では、興味の登録が煩雑であるためユーザの労力が大きい。また、ユーザの興味が変わった場合には、ユーザ自ら再び興味を登録し直さなければ対応できない。すなわち、(1)、(2)を満足していない。

これに対して、FireFly等では、提供された情報が興味に合っていたかどうかユーザに随時評価してもらう。この方式では、対話的に興味の登録と更新を行うことで、(1)、(2)を考慮しているが、完全な解決には至っていない。

WebWatcher[1]、NewsWeeder[5]等のサービスでは、TF-IDF(Term Frequency - Inverse Document Frequency)を用いた確率計算によりユーザの興味を自動的に把握する方式を取っている。この方式では、ユーザの興味を、カテゴリに相当する項目を独立した軸として、各項目に対する興味の強さを重みとしたベクトル(以下ユーザベクトル)で表現する。ユーザが情報を参照するたびに各項目の重みを再計算することで、ユーザによる興味の登録作業を不要とし、興味の変化に対しても自動的に対応することが可能となっている。すなわち、要求条件の(1)、(2)を満たしている。しかし、この方式では新規項目を追加した場合、ベクトル空間に新たな軸が加わるため、ユーザベクトルの再構築が必要である。この作業は提供される情報の数だけ必要であり、負荷が高い。すなわち(3)を満足しない。

2.2 情報の特徴の表現、更新方法

情報の特徴は、ユーザの興味と同様にベクトル(以下情報ベクトル)で表現することが多い。すなわち、情報の特徴を、先の項目を軸として項目毎に重みを持ったベクトルで表現する。この作業は情報提供者によってTF-IDF等の確率計算を用いて任意に行われることが多い。また、現在提案さ

れている多くの方式では、一度設定された項目の重みは不変である。このような場合、各項目に対する重みがユーザと情報提供者の間で異なったり、情報提供者が誤った項目に対して重みを設定してしまうことが考えられる。したがって、ユーザ全体からみた情報の特徴づけと、誤りの修正が行えることが望ましい。ここで要求条件は以下の2点である。

- (1)ユーザからみた特徴づけ
- (2)間違った特徴から正しい特徴への修正が可能

WebWatcher[2]、FireFly等のサービスでは、SIF(Social Information Filtering)の考え方を導入し、(1)を解決している。この方式では、提供情報の分類に情報を参照するユーザの興味を用いることで、ユーザから見た情報の分類を可能としている。特に、WebWatcherでは、ユーザがキーワードとして投入した単語を新たな項目として追加することができ、(2)についてもある程度満足している。ただし、各項目の重みの増加はできるが、減少はできない。すなわち、情報が持つべき正しい項目の重みを増加させること、誤った項目に対する重みを減少させることの両者を同時に行うことができない。

2.3 ユーザの興味と情報の特徴の参照方法

2.1節、2.2節で述べた各ベクトルを参照することで、ユーザの興味に応じた情報の選択を行う。この際、ベクトルの方向では、興味あるいは情報の指向性(性質)が表現される。また、ベクトルの長さでは重要性(価値)が表現される。したがって、ユーザの興味に応じた情報の選択では、興味や情報の、性質と価値の両方に着目した情報選択を行えることが望ましい。ここで要求条件は以下の2点である。

- (1)情報の性質を考慮
- (2)情報の価値を考慮

これまで述べてきた、ベクトルを利用した情報選択方式として、ベクトル空間法[6]が知られている。ベクトル空間法による選択の概要を図1に示す。この方式は、ユーザベクトルと情報ベクトルの方向に着目し類似度を求め、類似度の高いものから順に選択するものである。あらかじめ、共通のカテゴリに対してTF-IDFを用いた確率計算等から重み付けを行いユーザベクトル U 、各情報ベクトル G_i を求めておく。次に、ユーザベクトル U と各情報ベクトル G_i の類似度を、ベクトル間の角度 θ_i を計算することによって求め、角度 θ_i が小さい順に選択する。

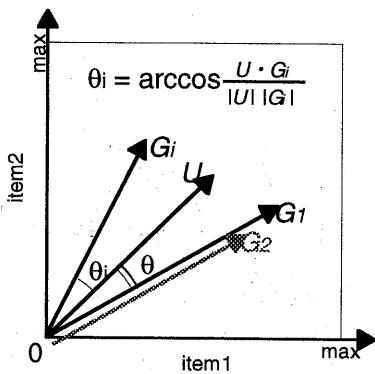


図1 ベクトル空間法による情報の選択

この方式の問題点としては、類似度の計算にあたってベクトルの方向(=情報の性質)しか考慮していない。このため、図1に示すG1、G2のベクトルをもつ情報は同じ類似度とみなされる。すなわち、ベクトルの長さが示す情報の価値については全く考慮されていない。

3 情報提供方式(InfoBrocket方式)の提案

本章では、前章で論じた要求条件を満足する情報提供方式を提案する。この方式をInfoBrocket方式と呼ぶ(図2)。

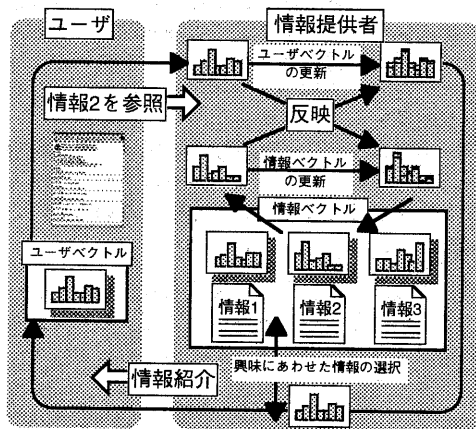


図2 InfoBrocket方式の概要

本方式では、ユーザーの興味および情報の特徴の表現と更新を、ユーザーベクトルと情報ベクトルを相互に反映させることで行う。これにより2.1節、2.2節の要求条件を満足している。また、ユーザーの興味と情報の特徴を参照するにあたって、ベクトルの方向だけでなく長さを考慮することで、

2.3節の要求条件を満足している。以下の節では、本方式の特徴を整理し、本方式が要求条件を満たすことを示す。

3.1 ユーザーの興味の表現、更新方法

ユーザーの興味を、前章で述べたユーザーベクトルで表現する。ユーザーのベクトルの更新は、ユーザーベクトルをユーザーが参照した情報の情報ベクトルに近づけることで行う。以下、図3を用いて説明する。

ユーザーがある情報を参照した場合、その情報の特徴を表現した情報ベクトルGはその瞬間のユーザーの興味を表わしたものと考えられる。したがって、ユーザーベクトルUの方向を、参照した情報の情報ベクトルGの方向に回転させ新規ユーザーベクトルU_{new}を設定する。図中で、nは1以上の値の定数で、回転させる度合を調整するため導入する。nの値は、提供する情報の分野により調整する。例えば、興味の持続が長期に及ぶような分野では、nを大きくとりユーザーベクトルの変化を小さくすることができる。

この方式では、ユーザーが情報を参照することで自動的に重みが更新されていくため、ユーザーが予め興味を登録する必要がない。また、ユーザーの興味が変わった場合、参照する情報の傾向も変わるため、自動的にユーザーベクトルの重みが再設定されユーザーの興味の変化に追従する。また、新規項目の追加も、登場頻度の再計算を行う必要がなく、ユーザーベクトルに新項目として加えるだけでよい。このように、2.1節で述べた要求条件をすべて満足する。

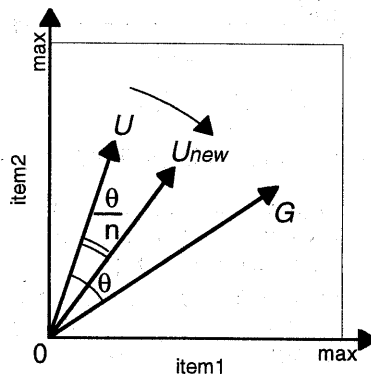


図3 ユーザーベクトルの更新

3.2 情報の特徴の表現、更新方法

情報の特徴を、前章で述べた情報ベクトルで表現する。情報ベクトルの更新は、情報を参照した

ユーザのユーザベクトルを情報ベクトルに反映させることを行う。以下、図4を用いて説明する。

情報があるユーザによって参照された場合、その情報は参照したユーザが属しているクラスに分類されるべきであると考えられる。ユーザが属すクラスとは、全ユーザ中でのあるユーザの特徴(=どのような項目にどの程度興味を持っているか)を示すものである。ここでユーザの特徴を、全ユーザのユーザベクトルの平均(平均ユーザベクトル M)とユーザベクトル U の差として、ベクトル C で表現する。この特徴ベクトル C を情報ベクトル G に加算し、新規情報ベクトル G_{new} とする。図中で、 m は1以下の定数で、参照されるたびに情報ベクトルが大きく変化する不安定な状態となることを避けるために導入する。

この方式では、情報ベクトルの更新を情報を参照したユーザのユーザベクトルを反映させることを行うため、SIFによる情報の特徴づけが可能となる。また、ベクトルの加算によって誤ったカテゴリの重みを減少させることができる。このように、2.2節で述べた要求条件をすべて満足する。

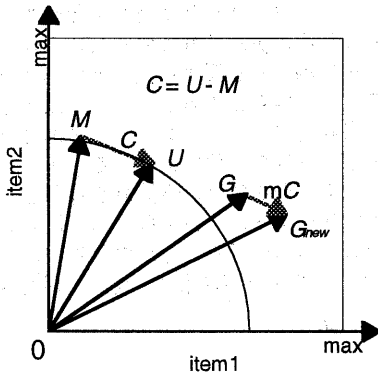


図4 情報ベクトルの更新

3.3 ユーザの興味と情報の特徴の照合方法

ユーザの興味と情報の特徴の比較は、ユーザベクトルと情報ベクトルを照合することにより行う。以下、図5を用いて説明する。

ユーザベクトル U に対する各情報ベクトル G_i の写像ベクトル G_i' を求め、写像ベクトル G_i' の長さが長いものから順に選択する。

この方式では、ベクトルの方向だけでなく長さも考慮しているため、情報のもつ価値に基づいた選択を行うことができる。特に、本方式では情報の価値を優先させているため、方向が同じでも長さの長いものが優先されるが、長さが同じ場合は方向に近いものが選択される。

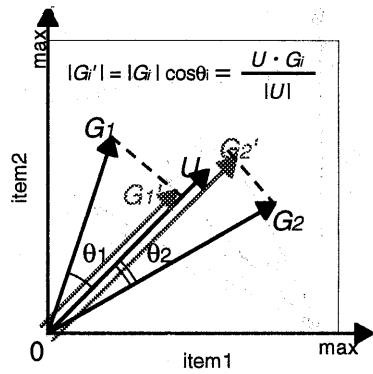


図5 情報の価値を考慮した選択

4 InfoBrocket方式の性質

本方式では、ユーザの参照を通してユーザベクトル、情報ベクトルの両者を同時に更新する。したがって、どちらか一方のベクトルが不安定な状態となると、全体が影響を受ける可能性がある。本章では特に、情報ベクトルが安定するために要求される条件を整理し、本方式が要求条件を満たすことを示す。

・情報ベクトルの非均一性

一般に、情報ベクトルの更新にユーザベクトルを反映させる方式では、場のもつ平均的なユーザベクトルに全ての情報ベクトルが収束し易い。言い替えば、全ての情報が、平均的な興味を持つユーザが属すクラスに分類され易い。

本方式では、各ユーザの特徴ベクトルを導入し、平均ユーザベクトルを除いたベクトルを反映させることでこれを防いでいる。

・各項目の重みの収束性

本方式では、十分な参照回数を経た情報ベクトルの各項目の重みは最大値もしくは最小値のどちらかへ収束する。以下、図6を用いて説明する。

情報ベクトル G_i をもつ情報に対して、この情報を参照するユーザのユーザベクトルの平均を M_{Gi} とする。この M_{Gi} が全体の平均ユーザベクトル M と完全に一致しない限り、平均特徴ベクトル M_{Gi} は0ベクトルとはならない。ユーザがこの情報を参照するたびに M_{Gi} に M_{Gi} が加算されることになるため、十分な参照回数を重ねると各カテゴリの重みは最大値もしくは最小値のどちらかへ収束する。

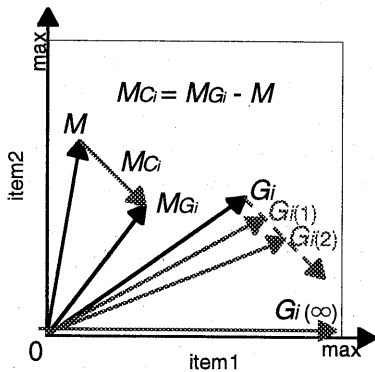


図6 各カテゴリの重みの収束性

5 実験結果

本方式を、ニュース記事を提供するWWWサイトに適用した評価結果について報告する。

5.1 実験方法

ユーザへのニュース記事の提供画面の例を図7に示す。画面下部に各ニュース記事のタイトルを10件表示する。タイトルがクリックされた場合、該当する記事の本文を画面上部に表示する。今回の実験では、本方式で推測したユーザの興味と、実際のユーザの参照行動がどれだけ一致するか観察する。したがって、ユーザが興味以外の要因で記事を参照することは望ましくない。このため、興味に応じた選択的な情報の提示や、それらの情報に明示的に印をつけるといったことは行わないこととした。

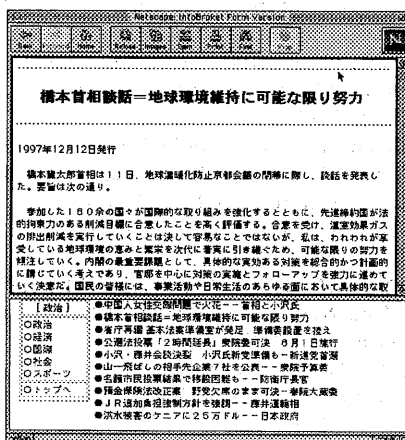


図7 ニュース記事提供画面例

実験条件を表1に示す。

記事数/日数	850記事 / 17日 (50記事 / 日)
ユーザ数	24
総参照回数	1351

表1 実験条件

5.2 ユーザベクトルの収束性

ユーザベクトルの時系列での変動の例を図8に示す。縦軸は、最終的なユーザベクトルと各参照回数時でのユーザベクトルとの相関値である。

ユーザベクトルの変動要因としては、ユーザベクトルの更新に用いる定数 n が支配的である。情報ベクトルの更新に用いる定数 m はさほど影響しない。

$n < 2$ の場合、相関値に収束性が見られず、各参照回数時でのユーザベクトルは単にその瞬間のユーザの興味を表現したものである。 $n = 3$ の場合、約150回を周期とした相関値の変動が見られる。これは、実験に用いた提供情報がニュース記事であり、提供情報のカテゴリに周期性が存在するためと考えられる。長期的なユーザの興味を把握するためには $n > 3$ とし、短期的なユーザの興味を把握するには $n \leq 3$ とする。

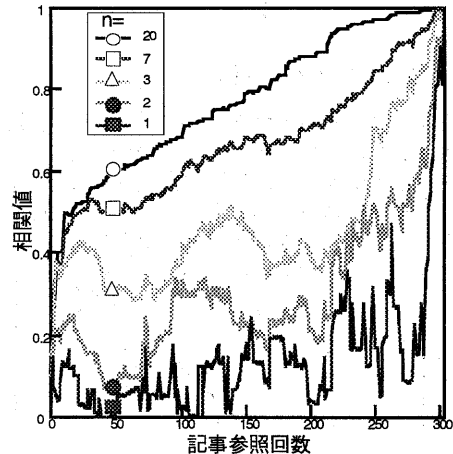


図8 ユーザベクトルの変動 (user ID = 5)

5.3 適合率

本方式で把握したユーザの興味と実際のユーザの参照行動がどれだけ一致したかを示す指標として、適合率の概念を導入する。あるユーザの適合率を図9に示す。図において、横軸は本方式で推測した興味順位である。縦軸はその興味順位の記事を参照した割合である。ただし、この参照割合

は興味順位が1からその興味順位までの累積で表示する。わずかながらも本方式を用いた場合の適合率がベクトル空間法を用いた場合の適合率を上回っていることがわかる。したがって、ユーザの興味と情報の特徴の照合には、情報のもつ価値を考慮して行う必要があると考えられる。適合率は、 $n=1$ の場合に最良の結果となる。これは、ユーザがある記事を参照して次の記事を参照する際に、事前に参照した記事と関連がある記事を参照するからと考えられる。

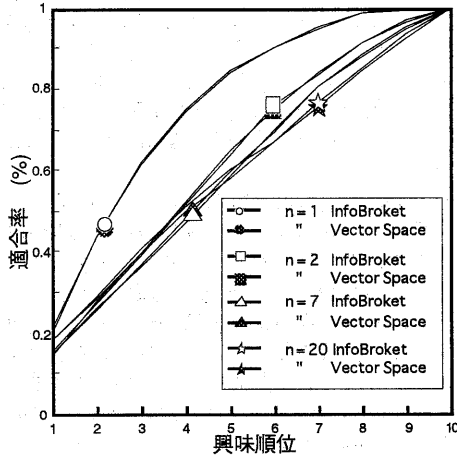


図9 適合率

5.4 適合率の時系列変化

適合率の時系列変化を図10に示す。参照回数の増加とともに適合率が向上する傾向が見られるものの変動幅が大きく、単調に向上するわけではない。これは、ユーザベクトルは50回以下の少ない参照回数である程度収束するためと考えられる。

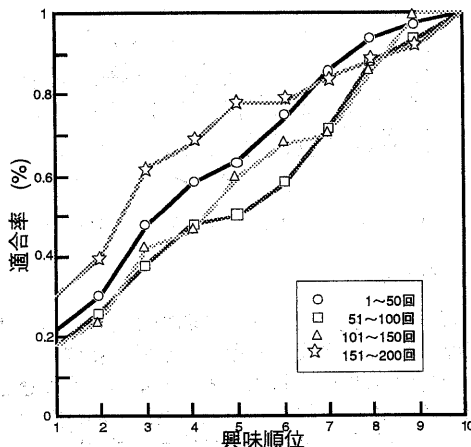


図10 適合率の時系列変化 (user ID = 5)

6 まとめ

本稿では、ユーザの興味と情報の特徴を相互に反映させ、情報の価値を考慮して選択を行う情報提供方式InfoBrocketを提案した。本方式を、ニュース記事を提供するWWWサイトに適用し評価を行った。以下に、評価を通して確認された事項についてまとめる。

- ・情報選択において情報の価値の考慮が必要
本方式はベクトル空間法と比較して同等以上の結果が得られた。情報の価値を考慮した情報選択が必要である。
- ・長期的な興味と短期的な興味に対応可能
本方式では、ユーザの興味の更新において適切な係数を選択することで、短期的な興味と長期的な興味を把握できる。

また、実験より以下の知見を得た。

- ・ユーザの興味の変化には周期性がある
- ・ニュース記事の場合は、ユーザは連続的に、事前に参照した記事と関連する記事を参照する

また、今回は以下の2点について充分評価を行うことができなかった。

- ・情報の特徴の更新
- ・ニュース記事以外の情報での評価

今後は、WWWサイト情報を提供情報とするWWWディレクトリサービスへ本方式を適用し、評価を行う予定である。

参考文献

- [1] Duja Mlandenic. Personal WebWatcher: design and implementation. Technical Report IIS-DP-7472. 1996.
- [2] Thorsten Joachims, Dayne Freitag, Tom Mitchell. WebWatcher: A Tour Guide for the World Wide Web. IJCAI '97, pp.770-775. 1997.
- [3] U. Shardanand, P.Maes. Social Information Filtering: Algorithms for Automating 'Word of Mouth'. Proceeding of the CHI-95 Conference, pp.210-217. 1995.
- [4] 橋高博行, 鈴木英明, 柴崎雅史, 曾根岡昭直. 個人適応型情報フィルタリングシステムの適用評価について. 情報処理学会第55回全国大会, vol.3, pp. 234-235. 1997.
- [5] Ken Lang. NewsWeeder: Learning to Filter Netnews. ICML '95, pp.331-339. 1995.
- [6] Gerard Salton. Developments in Automatic Text Retrieval. Science, vol. 253, pp. 974-980.