

# 電子ニュースの記事に対する操作と 関心の有無との関連性について

川崎 孝介 北 英彦 林 照峯

三重大学工学部

現在、電子ニュースにおいて多くの記事が投稿されており、自分にとって関心のある記事を効率的に見つけることが困難になってきている。この問題を解決するために情報フィルタリングに関する研究が行われている。情報フィルタリングでは、新たにやって来る情報とユーザの関心とを比較してフィルタリングを行うために、ユーザの関心を記述したプロフィールを用いる。我々は、ユーザが過去に読んで関心があった記事を解析することによって、プロフィールを生成・更新するという手法を用いる。しかし、どの記事に対して関心があったかを、ユーザが一々システムに指示することは煩雑である。本稿では、ユーザにとって関心があった記事をシステムによって自動的に判別するために、ユーザの記事に対する操作と記事への関心の有無との関連性について検討する。

## **A relationship between user's behavior toward and interest in articles on NetNews**

**Kosuke Kawasaki Hidehiko Kita Terumine Hayashi**

**Faculty of Engineering, Mie University**

Recently, since many articles are posted on NetNews, it is difficult to look for articles effectively in which an user is interested. Information filtering is researched to solve the problem. On information filtering, a profile that is described user's interest is used to filter new information by comparing with user's interest. In our approach, the profile is created or rebuilt by analyzing articles which an user read before and were interested in. But it is a burden on the user to input what article the user was interested in. In this paper, we examine a relationship between user's behavior toward and interest in articles to distinguish the interested articles.

## 1.はじめに

インターネットの普及とユーザの増加にともない、インターネット上で情報交換を行うシステムとして広く使われている電子ニュースに投稿される記事の数が増加している。また、様々なユーザの参加により、記事の内容も多様化している。このため、大量な記事の中から自分にとって関心のある記事、興味のある記事を効率的に見つけだすことが困難になってきている。

この問題を解決するために、現在情報フィルタリングに関する研究が行われている[1]。情報フィルタリングでは、一般に、ユーザの関心を記述したプロフィールが用いられる。我々は、過去にユーザが読んで関心があった記事を解析することによって、ユーザのプロファイルを生成・更新するという手法を用いる。しかし、どの記事に対して関心があったかを、ユーザが一々システムに指示することは煩雑である。そこで、プロフィールの生成・更新において、ユーザの負担をできるだけ軽減するために、ユーザにとって関心があった記事をシステムによって自動的に判別する。

本稿では、ユーザにとって関心があった記事を自動的に判別する際の具体的な基準を見つけるために、ユーザの記事に対する操作と関心の有無との関連性について検討する。

## 2.電子ニュースの現状と問題点

現状の電子ニュースシステムでは、電子ニュースの記事は話題や役割別に設けられたニュースグループと呼ばれるグループごとに記事が分類されている。電子ニュースシステムのユーザはこのニュースグループの名前を頼りに、自分の関心のあるニュースグループを選び、その中で自分にとって関心のある記事を探して読む。しかし、ニュースグループの数は世界中で10,000以上、1日に投稿される記事数は全体で約400,000記事もあり、1日に投稿される記事の数が50を越えるニュースグループも数多くある[2]。このため、電子ニュースのユーザは、この膨大な量の記事の中から自分にとって関心のある記事を効率良く見つけることが困難になってきている。

## 3.情報フィルタリング

上記の問題を解決するための手法として、情報フィルタリングに関する研究が行われている。ここで、情報フィルタリングとは、多くの情報の中からユーザが必要とする情報を、コンピュータを使って効率的に抽出する技術のことである。情報フィルタリングのなかで、情報の内容とユーザの関心とを比べ、ユーザの関心に合った情報を選び出す手法のことを、特にcognitive filteringという。cognitive filteringでは、一般に、情報の内容とユーザの関心とを比べるために、ユーザの関心を記述したプロフィールを用いる。フィルタは、このプロフィールをもとに、新たにやって来る情報のフィルタリングを行う(図1)[3]。

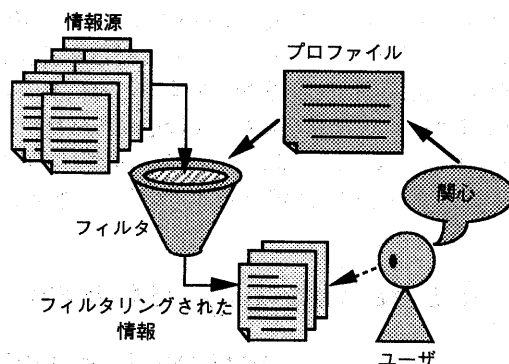


図1:情報フィルタリングの概念図

本研究においても、cognitive filteringを用いたシステムを考えており、過去に読んだ記事のうち、ユーザにとって関心があった記事を解析することによってユーザの関心を学習し、ユーザにとって関心のある記事を探すことにする。

## 4.プロフィール

先にも述べたように、情報フィルタリング(cognitive filtering)には、一般に、ユーザがどのようなことに関心があるのかという事を記述したプロフィールを用いる。プロフィールの生成法には、次の2つの方法がある。

- (1)ユーザ自身がプロフィールを記述
- (2)システムによって自動的に生成

(1)の方法は、あらかじめやって来る情報が推測できる場合に対しては有効であるが、時事刻々と話題が推移していく電子ニュースにおいては、

新たな話題が生じる度にプロフィールをユーザ自身が追加，修正しなければならず面倒である。また，自分がどのようなことに関心があるかをあらかじめ全て記述することは困難である。

(2)の方法は，システムによって自動的にプロフィールを生成・更新するものであるが，ユーザの関心がわかり，システムに取り扱うことができる，プロフィールの元となるデータが必要である。

## 5.プロフィールの自動生成

本研究では，システムによって自動的にプロフィールを生成する方法を用いることにし，ユーザが過去に読んで関心があった記事をもとに，プロフィールの生成・更新を行っていく。しかし，ユーザがどの記事に対して関心があったかを一々システムに指示することは煩雑である。そこで，ユーザにとって関心があった記事には，記事に対する行動に何らかの特徴があるということ仮定して，記事に対する操作から関心があった記事を判別することにする。具体的には以下のような手法をとる(図2)。

- (1)記事を読んでいる間に，記事に対して行われた操作，記事を読んだ範囲，時間を記録する
- (2)ユーザにとって関心があった記事を，ユーザの記事に対する操作記録から判別する
- (3)ユーザにとって関心があった記事群からユーザの関心を解析して，プロフィールを作

成・更新する

## 6.記事に対する操作と関心の有無

電子ニュースの記事に対してユーザが行う操作には何らかの意味がある。例えば，「保存」という操作はユーザにとって重要な情報が含まれている記事である可能性が高く，「フォロー」という操作はその話題に関してさらに議論を深めたいという意志の表れといえる。

また，上記のような記事に対する直接的な操作以外にも，記事を読んだ範囲，時間はその記事がユーザにとって関心があったものかどうかの判断材料となる。例えば，長い記事であるにもかかわらず，記事の最初の部分しか読んでいなかった場合などは，その記事がユーザにとって関心が無かった記事である可能性が高い。このように，ユーザの記事に対する操作から，ユーザのその記事に対する関心の有無を推測することができる。

## 7.試行実験

先の仮説を立証し，ユーザにとって関心があった記事をシステムによって自動的に判別する際に必要となる，具体的な判別の基準を調べるために，ユーザの各記事に対する操作とその記事に対する関心の有無との関連性を調べる。そこで，以下のような試行実験を行った。

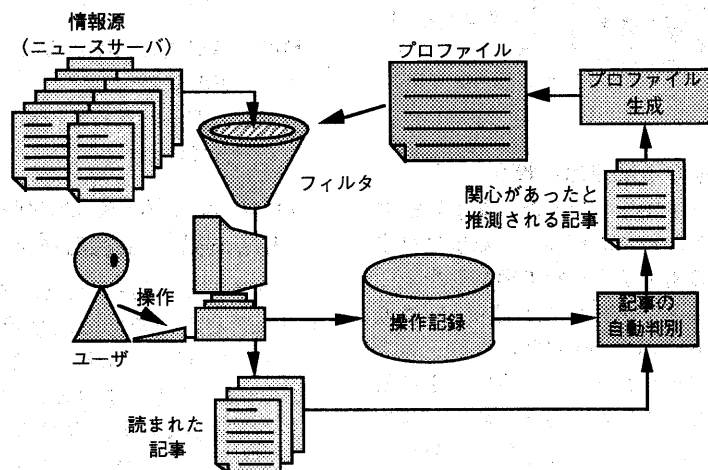


図2:プロフィールの自動生成

## 7.1 実験方法

試行実験は、本研究室の学生6人を対象とし、1997年8月下旬から11月下旬の約3ヶ月間、以下のような手順で行った。実験システムの構成図を図3に示す。

- (1)まず被験者には普段通りに電子ニュースの記事を読んでもらう。このときに、ユーザが記事に対して行った直接的な操作（保存、フォロー、リプライ）及び、記事が画面上に表示された範囲、時間を操作記録として自動的に記録する。
- (2)1つのニュースグループにおける記事を全て読み終えたあとで、今まで読んでいた記事をもう一度見直してもらい、それぞれの記事に対して関心があったか、無かったか、を答えてもらう。
- (3)各記事に対してユーザが答えた実際の関心の有無と、システムが自動的に取得したユーザの操作記録とを解析して、ユーザにとって関心があった記事を判別するための基準を見つけ出す。

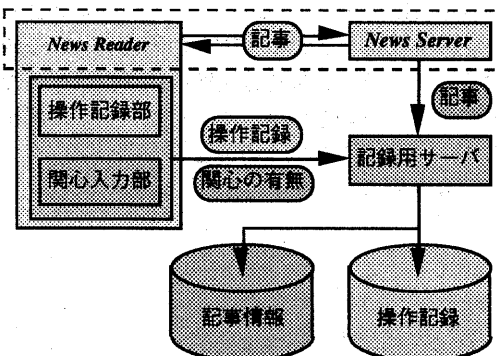


図3:実験システムのシステム構成図

## 7.2 実験結果

### 7.2.1 記事に対する直接的な操作

記事の保存、フォローといった、記事に対する直接的な操作が行われた記事と、記事に対する関心の有無との関連を調べた実験結果を表1に示す。直接的な操作が行われた記事は実験期間中27件（全記事における0.87%）あり、いずれも、ユーザが関心があったと答えた記事に対して行われていた。この結果から、記事に対する直接的な操作は、ユーザがその記事に対して関心を持っていることの表れと言える。

表1:実験結果1

総記事数 3114	直接的な操作が行われた記事数 27	保存 25	関心あり 25
			関心なし 0
	フォロー 2	関心あり 2	
		関心なし 0	
読まれただけの記事数 3087			

### 7.2.2 記事を読んだ時間および範囲

直接的な操作が行われなかった、ただ読まれただけの記事に対しては、記事を読んだ範囲、時間を測り、関心のある記事に対してはどのような特徴が見られるかを調べた。ここで、1画面内に納まってしまう記事は、記事を読んだ範囲が全て一定となり、記事を読んだ範囲から特徴が探しだすことができない。そこで、2画面以上にまたがる記事についてのみこれらの解析をおこない、図4、5のような、各記事に対する「記事を読んだ範囲vs記事を読んだ時間」のプロット図を作成した。

ここで、情報検索の分野においてよく用いられている、精度、再現率（呼出率ともいう）を用いて[4]、次のような評価式を挙げる。

$$\text{評価式} = \text{精度} + \text{再現率}$$

精度：領域内の記事のうち、関心があった記事の割合

再現率：関心があった記事のうち、領域内に含まれている割合

\*領域とは、記事を読んだ時間と範囲によって囲まれる、図4、5における網掛の部分

この評価式の値が最大となる領域をつくりだす記事を読んだ範囲および時間の境界を、ユーザ毎、2週間毎に分けて、それぞれ調べた。

ユーザ毎の解析においては、普段から電子ニュースを利用しており、今回の実験システムのベースとなっているGNUS（ニュースリーダ）の操作にかなり慣れているユーザAと、最近電子ニュ

ースを使いはじめ、GNUSの操作にはあまり慣れていないユーザDの2人の結果を、図4、5にそれぞれ示す。

ユーザA、Dの評価式の最大値およびそのときの精度、再現率、境界である記事を読んだ時間、範囲はそれぞれ以下の通りであった。

ユーザA:

評価式の最大値=1.8732

精度=0.9036

再現率=0.9696

記事を読んだ範囲=83~85 (%)

記事を読んだ時間=5.9 (sec/1Kbyte)

ユーザD:

評価式の最大値=1.2633

精度=0.3433

再現率=0.92

記事を読んだ範囲=62~80 (%)

記事を読んだ時間=3.1~3.4 (sec/1Kbyte)

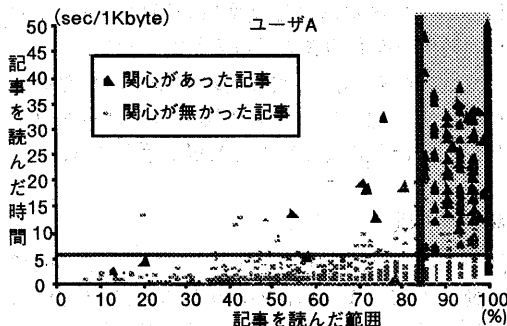


図4:ユーザAの記事を読んだ範囲vs時間のプロット図

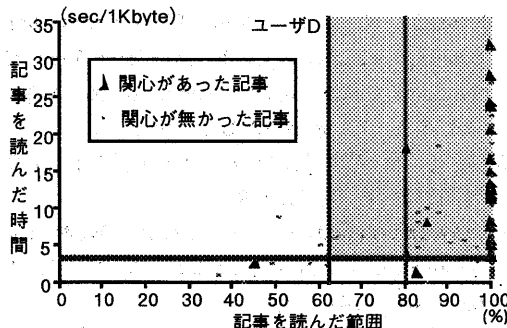


図5:ユーザDの記事を読んだ範囲vs時間のプロット図

この結果から、記事を読んだ時間、範囲において、関心があった記事と関心が無かった記事とがはっきりと分れるユーザと、そうでないユ

ーザがあり、個人差があるということがわかる。

また、データを期間毎に分けた場合における、評価式が最大となるとき境界値(記事を読んだ範囲、時間)の推移を図6、7に示す。ここで、グラフに幅があるところは、この部分においてはどこで領域を区切っても評価式が最大となることを示している。

この結果から、記事を読んだ範囲は70~86(%)の間で変動しており、記事を読んだ時間は5.6~8.0(sec/1Kbyte)の間で変動している。評価式が最大となるとき精度および再現率の変動を調べると、精度は0.5~0.9、再現率は0.6~0.9の間を変動していた。

従って、ユーザ毎の基準ではなく、一般的な基準を設けるとするなら、それぞれの中間点を取り、記事を読んだ範囲において78(%)以上、記事を読んだ時間においては6.7(sec/1Kbyte)以上という基準を与える。この基準を適用した場合の精度と再現率は、精度=0.7787、再現率=0.7844であり、この領域に含まれる記事は、関心があった記事である可能性が高いといえる。

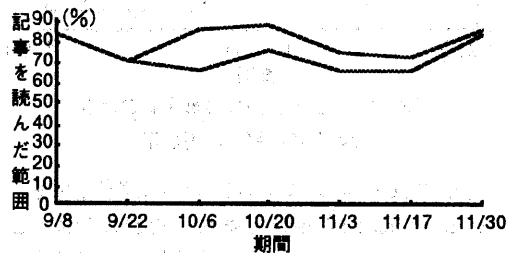


図6:記事を読んだ範囲の期間における変動

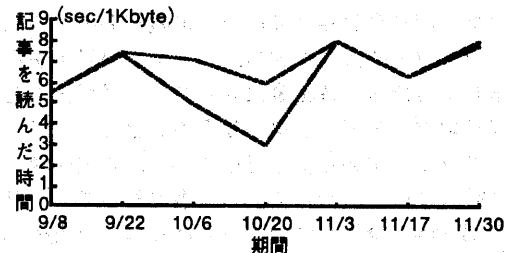


図7:記事を読んだ時間の期間における変動

## 8. 考察

ユーザAとDを比較してわかる通り、ユーザによる個人差はかなり見られる。ただし、ユーザDのデータを期間別に見てみると、最初の2週間は精度が高くなかったのに対し、実験期間の途

中からは6割以上の精度に上がってきた(図8)。

ユーザDにきいてみると、使いはじめたころは記事の読み飛ばす方法などが良くわからず、ほとんどの記事を全部見ていたが、GNUSに慣れてくるにつれて、記事の読み飛ばしや、サブジェクトを見ただけでその記事に既読マークを付けるといった操作を覚えた、ということであった。他にも、実験期間中、全ての記事を、全ての範囲読んでいたユーザもいた。

従って、記事を読むときに、ニュースリーダの操作に慣れているかどうかという点は、かなり大きな影響を与えるといえる。

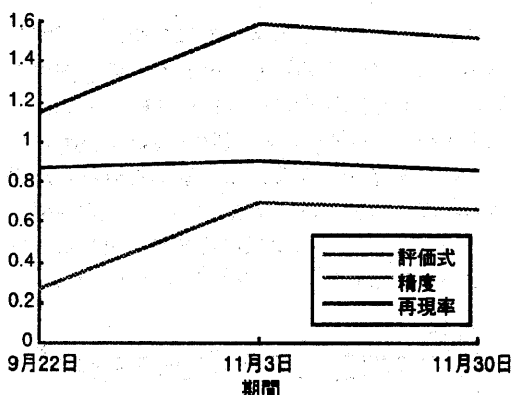


図8:ユーザDの評価式の最大値およびそのときの精度、再現率

電子ニュースシステムにおける問題点を、cognitive filteringを用いて解決した他のシステムとして、MITのNewT[5]や、北陸先端大学院大学のシステム[6]などがある。NewTでは、最初に読みたい記事や読みたいくない記事の記事例を与えてプロファイルを生成し、システムが記事を推薦してきたら、各記事に対してユーザが正または負のフィードバックを与えることによって、プロファイルの更新を行っている。北陸先端大学院大学のシステムでは、記事に対してユーザが明示的に読みたいか、読みたいくないかを指示するとともに、指示が無かった記事に対しては、読んでいた時間が長かったものは、有用な記事として判断し、これらの記事を元にプロファイルの生成・更新を行っている。

我々の考えているシステムでは、保存やフォローといった記事に対する直接的な操作と、記事を読んだ範囲、時間とを用いることによって、

ユーザが明示的に指示を出さなくても高い精度でユーザにとって関心があった記事を判別できるという点で、上記の他のシステムと異なっている。

## 9.まとめ

本稿では、一般に情報フィルタリングに用いられるプロファイルを、過去にユーザが読んで関心があった記事から生成することにした。そして、そのプロファイルの生成・更新において、ユーザの負担をできるだけ軽減するために、関心のあった記事を、ユーザの記事に対する操作から判別することとし、その関連性を検討するために試行実験を行った。

その試行実験より、保存やフォローといった直接的な操作が行われた記事は、ユーザにとって関心があった記事であるといえる。また、記事を読んだ範囲、時間については、ニュースリーダの操作に慣れているかどうかで個人差が現れるが、全体における基準値を設けても、かなりの精度で関心のある記事を判別できる。

今後の課題は、今回の実験で得られた結果を利用して、電子ニュースの記事のフィルタリングシステムを完成させることが挙げられる。

## 参考文献

- [1] 森田 昌宏:情報フィルタリングに関する研究動向.JAIST Research Report IS-RR-93-91,北陸先端科学技術大学院大学情報科学研究科,1993.
- [2] <URL:news:tnn.netnews.stats>
- [3] Belkin, N. J. and Croft, W. B.:Information Filtering and Information Retrieval:Two Sides of the Same Coin? Communications of the ACM, Vol.35, No.12, pp.29-38, Dec, 1992.
- [4] 長尾 真:岩波講座ソフトウェア科学15 自然言語処理. 岩波書店, 1996.
- [5] Pattie Maes:Agent that Reduce Work and Information Overload, CACM, Vol.37, No.7, pp.30-40, 1994.
- [6] 森田 昌宏,篠田 陽一:情報洪水の緩和のための情報フィルタリングの実現—ユーザアクティビティの分析と最適照合検索による情報フィルタリング.第1回 JAIN Consortium Symposium 論文集,pp.31-38,1994.