

# 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

向井 誠<sup>†</sup> 青野 雅樹<sup>‡</sup>

<sup>†</sup> 豊橋技術科学大学 工学研究科 情報工学専攻

<sup>‡</sup> 豊橋技術科学大学 情報工学系

<sup>†</sup> <sup>‡</sup> 〒441-8580 愛知県豊橋市天伯町 1-1 E-mail: <sup>†</sup> mukai@kde.ics.tut.ac.jp, <sup>‡</sup> aono@ics.tut.ac.jp

**あらまし** 本研究では、RSS で記述された Web コンテンツ群からユーザの嗜好に沿ったコンテンツの取得にかかる労力を最小限に抑えることを目的に、Web コンテンツの推薦を実現するプロトタイプシステムの設計、実装を行う。既存の検索エンジンでは見つけにくい最新ニュース記事や Blog などの更新頻度の高い Web サイトに対し、その概要の記述に RSS を用いることが多くなった。しかし、RSS フィードが増加するにつれこれまで同様“情報の洪水”が問題となっている。提案手法では RSS に記述されたコンテンツに対して内容に基づくフィルタリングを適用し、ユーザの興味に沿ったコンテンツの推薦提示を実現する。

**キーワード** 情報推薦, 内容に基づくフィルタリング, ユーザプロファイル, RSS

## A Prototype of Content-based Recommendation System based on RSS

Makoto Mukai<sup>†</sup> Masaki Aono<sup>‡</sup>

<sup>†</sup> Department of Information and Computer Sciences, Graduate School of Engineering, Toyohashi University of Technology

<sup>‡</sup> Department of Information and Computer Sciences, Toyohashi University of Technology

<sup>†</sup> <sup>‡</sup> Hibirigaoka, Tempaku-cho, Toyohashi, Aichi, 441-8580, Japan

**Abstract** Recently, we can find a large amount of contents on the WWW. However, it has been difficult to extract some valuable information for us. In this paper, we propose a recommendation system which is based on Content-based Filtering. Our method aims at helping users to find information from RSS Feed that are in accordance with their personal interests. A method of generating User Profiles based on user preference is a very important task for providing recommendation systems. Therefore, we propose two method of generating User Profiles: Dynamic User Profile and Static User Profile. These Profiles corresponds to some kind of user preference.

### 1. はじめに

近年、WWW 上のコンテンツの質に変化が見え始めている。その代表的なものに Blog エントリ及びニュースコンテンツが挙げられる。個人ユーザは Blog の爆発的な普及により盛んに情報発信を行うことが可能になった。また、ニュースサイトでは様々な分野のニュース記事が逐次更新されている。このような更新頻度の高いコンテンツは内容とともにコンテンツの新鮮さが重要となるが、一般のロボット検索エンジンでは新鮮さの高いコンテンツを得ることは困難である。

このような更新頻度の高い Web コンテンツが増加する中、RSS (RDF Site Summary) が注目を浴びている。RSS は Web サイトの概要を記述するためのフォーマットであり、Web サイトの更新状況やニュースのヘッドライン、Blog エントリの概要などの記述に利用されている。RSS で記述された情報 (RSS フィード) は RSS リーダ (RSS アグリゲータ) と呼ばれるソフトウェアを用いて閲覧する。しかしながら、既存の RSS リーダは RSS フィードに記述された Web コンテンツのタイトル・概要等を一覧表示するだけのものが多く、ユーザは自分の好みに合ったコンテンツを自力で見つけなければなら

ない。そのため、閲覧する RSS フィードが多いほどこの労力は大きな問題となる。

これらの問題点を考慮し、本研究は RSS 情報推薦システムを提案する。提案手法では、RSS フィードに記述される Web コンテンツにかかる手間を削減することを目的とする。システムを用いることで、ユーザはコンテンツ群から興味のあるコンテンツを苦勞なく取得することが可能になる。

以下、2 節で関連技術、3 節で提案手法、4 節でシステムの核となるユーザプロファイルについて述べ、5 節で提案手法の有効性を確認するためのプロトタイプシステム実装実験、6 節で今後の展望について述べる。

### 2. 関連技術

#### 2.1. 情報フィルタリング

個人の興味を反映した情報推薦を実現する代表的な手法に情報フィルタリングがある。情報フィルタリングは一般に複数ユーザの興味情報を利用する「協調フィルタリング (Collaborative Filtering)」と 1 ユーザの興味情報を利用する「内容に基づくフィルタリング (Content-based Filtering)」に分けられる。

協調フィルタリングは「興味・嗜好の近い人は同じ情報を求める」という仮定に基づいた手法である。具体的には、ユーザ A の嗜好情報（コンテンツに対する評価など）とユーザ A に似た嗜好傾向を持つユーザ B を特定し、さらにユーザ B の嗜好情報からユーザ A の嗜好に一致するコンテンツを推定する。この手法では他ユーザの評価を判断基準とするため、文書に限らず、マルチメディアコンテンツなど内容の解析が困難なデータにも適用することができる。しかし、精度の高い推薦を実現するためには多数のユーザと多数の評価情報が必要であり、ニュース記事のような評価の少ない新規コンテンツに対しては有効な推薦が行えない（コールドスタート問題）。そのため、本研究の対象である更新頻度の高い Web コンテンツの推薦に用いるには課題が残る。

内容に基づくフィルタリングはコンテンツそのものの内容とユーザの嗜好情報（ユーザプロフィール）の比較に基づく手法である。一般にコンテンツの内容、ユーザプロフィールともにキーワードベクトルで表現され、このキーワードベクトル同士の類似度からコンテンツに対するユーザの好みを推定し、フィルタリングを行う。この手法では1ユーザの嗜好情報とコンテンツ内容を表すキーワードベクトルさえ存在すればニュース記事のような更新頻度の高い Web コンテンツにも適用可能である。しかし、ユーザの評価履歴を元にフィルタリングを行うため、過去に高い評価をしたコンテンツに類似したもののばかりが偏って提示されてしまう。そのためユーザの嗜好変化に対応した推薦や新たな嗜好の発見を促す情報提示には向いていない。

情報フィルタリングを用いた研究には、文献 [1][2] などがある。[1] はインターネットを用いた情報取得が困難な初心者ユーザを対象にした情報推薦システムであり、直感的に扱えるインタフェースなど情報取得時におけるユーザの負担軽減を目的とする点は本研究と一致している。しかし、推薦候補コンテンツが固定であるため限られた推薦結果しか得られず、ニュース記事など更新頻度の高い Web コンテンツの推薦には課題が多い。また、[2] ではユーザプロフィールの作成にユーザが日常使うブックマークの階層構造を用いて効果的な推薦を実現しているが、フィルタリング手法に協調フィルタリングを用いており、内容に基づくフィルタリングを用いる本手法とはアプローチが異なる。

## 2.2. RSS 関連技術

### 2.2.1. RSS

図 1 は RSS の基本構成である。RSS はサイト概要をメタデータとして記述する XML フォーマットである [3]。

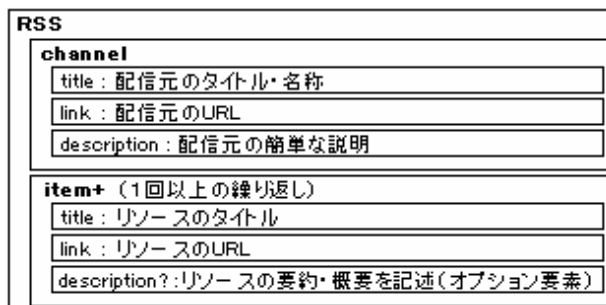


図 1 RSS の基本構成

RSS は、channel 要素（ニュースサイトや Blog サイトなど配信元となる Web サイトの情報を記述）1 つと、item 要素（ニュース記事や Blog エントリなどリソース情報を記述）複数からなる。一般に Web コンテンツを記述した RSS フィードにおいては item 要素ひとつひとつが個々のコンテンツを表し、item 内の title 要素はコンテンツのタイトル、link 要素はコンテンツの URL、description 要素はコンテンツ概要が記述されている。

RSS で記述される文書（RSS フィード）は一般にニュースサイトや Blog など更新頻度の高い Web サイトで配信されている。RSS を配信していない Web サイトでも、RSS を自動生成する “myrss” [4] や “なんでも RSS!” [5] 等のサービスが存在するため、ユーザが求める情報の RSS を自由に取得することができるようになりつつある。最近では RSS フィードと携帯音楽プレーヤを応用した Podcasting と呼ばれる技術が注目を浴びるなど、RSS の応用の幅も広がりつつある。このように RSS フィードの数が増加しつづける現在、RSS フィード情報の取捨選択は非常に重要な問題である。

### 2.2.2. 既存手法と問題点

RSS フィードは RSS リーダ（RSS アグリゲータ）を用いて閲覧する。RSS リーダには一般のメーラと同等の外観を持つメーラ型、電光掲示板のような外観を持つティッカー型、Web ブラウザにて操作する Web 型など様々なタイプが存在する。これらを用いて RSS を閲覧することでユーザは嗜好に合った Web コンテンツのみを取得することができる。しかし、いずれのタイプの RSS リーダも、コンテンツのタイトルや概要を一覧表示するだけのものが多く、必要な情報を得るためにはユーザ自身が取捨選択しなければならない。様々な RSS フィードが存在するため、興味のある RSS フィードの数が多くなることがあり、コンテンツの取捨選択にかかる労力は大きな問題となっている。RSS リーダによっては、ユーザが入力した任意のキーワードを含むものを提示する機能をもつものもあるが、キーワードを含むコンテンツの数が多ければ同様の問題が発生する。また、RSS フィードの情報を取得するには RSS リーダの他に、RSS フィードを元に RSS 検索エンジンを利用する方法も存在する。

RSS を使用せずに更新頻度の高い Web コンテンツを効率良く閲覧する方法には Google News[6] や My Yahoo![7] に代表されるパーソナライズド Web サービスがある。通常の Web サイトでは全てのユーザーに対して同じコンテンツを表示する静的なものだが、パーソナライズド Web ではユーザーごとに専用の Web ページを自動生成して表示する。専用 Web ページではユーザーの好みに合わせて Web ページのレイアウトを変更したり、好みのジャンルのニュース記事を表示させたりすることができる。RSS フィードを表示させることが可能なものも存在し、手軽な Web の個人化手法として注目されているが、ユーザー個人の興味に沿った情報の提示には至っておらず、RSS リードと同様、必要なコンテンツを優先的に得るのは困難となっている。

RSS に関する研究として、RSS フィードを自動生成する“なんでも RSS!” [5] が挙げられる。これは RSS を持たない通常の Web サイトの構造を解析し、これに対応した RSS を自動的に生成するシステムである。“なんでも RSS!” は RSS の生成を目的とするものであり、RSS の有効利用を目的とする本研究とは趣旨が異なるが、こういった RSS 自動生成システムと提案手法を併用することにより、更なる情報取得の効率化が望めるものと思われる。また、RSS に関連する技術として最近 Blog に関する研究が盛んに行われている[8][9]。しかしこれらは領域が Blog に限定される。本研究では RSS フィードを元にフィルタリングを行うことで、Blog 記事に限定せず、ニュース記事など更新頻度の高い Web コンテンツの推薦提示が実現できる。

### 3. 提案システムの構成

本稿で提案するシステムの構成を図 2 に示す。

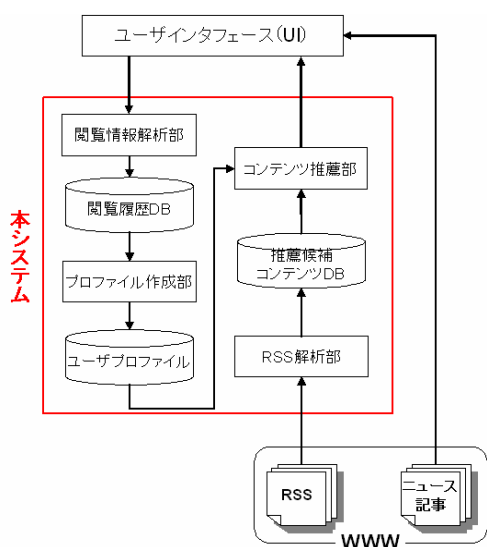


図 2 システムの構成

#### 3.1. RSS 解析部

RSS 解析部では、ユーザーが巡回先として登録した

WWW 上の RSS フィードから概要情報を定期的に巡回・取得を行い、item 要素に記述されるコンテンツごとに特徴ベクトルを作成する。

まず巡回先の RSS フィードを解析し、item 要素に記述される title, link, description 要素を取得する。次に、各 item 要素内の link 要素を参照し、コンテンツ URL を取得する。そしてコンテンツ URL を元にコンテンツ本文を取得する。取得したコンテンツ本文は茶釜[10] を用いて形態素解析し、コンテンツを特徴付ける名詞句を用いてキーワードベクトル (コンテンツ特徴ベクトル  $C$ ) を作成する。

$$C = (w_1, w_2, \dots, w_l)$$

この際、単語の重み付けには正規化した  $tf$  を用いる。ここで、キーワード  $i$  の重み  $w(i)$  は、

$$w(T, D) = tf(T, D) = \frac{\log(tfreq(T, D+1)}{\log(tnum(D))}$$

と表せる。ここで  $i$  はキーワード番号、 $tfreq(T, D)$  は文書  $D$  におけるターム  $T$  の出現頻度、 $tnum(D)$  は文書  $D$  に含まれるキーワード数である。生成したコンテンツ特徴ベクトルは推薦候補コンテンツ DB に格納する。

#### 3.2. 閲覧情報解析部

閲覧情報解析部ではユーザープロフィールを作成する前段階として閲覧したコンテンツ情報の抽出・ベクトル化を行う。

閲覧履歴情報は、閲覧したコンテンツから作成したコンテンツ特徴ベクトルと閲覧日時からなる。ユーザーはインタフェースを通して、システムの推薦によって提示されたコンテンツやその他 WWW 上のコンテンツ中から興味のあるコンテンツを閲覧する。このとき、RSS 解析部と同じようにコンテンツ特徴ベクトル  $I_h$  を作成し、日時と共に閲覧履歴 DB に格納する。

$$I_h = (w_1, w_2, \dots, w_l)$$

$$w(T, D) = tf(T, D) = \frac{\log(tfreq(T, D+1)}{\log(tnum(D))}$$

#### 3.3. ユーザプロフィール作成部

ユーザープロフィール作成部では、閲覧履歴 DB に格納するコンテンツ特徴ベクトルを元にユーザーの興味モデルであるユーザープロフィールの作成を行う。ユーザープロフィールの作成に関しては 4 節にて詳説する。

#### 3.4. コンテンツ推薦部

コンテンツ推薦部では、ユーザープロフィールと推薦候補コンテンツ DB に格納されているコンテンツ特徴ベクトルの類似度を算出し、コンテンツの推薦結果を作成する。本研究では、ユーザープロフィールとコンテンツ特徴ベクトルの類似度の算出にベクトル

ル空間モデルにおいて一般的なコサイン類似度を用いる。ユーザプロファイルを  $\mathbf{P}$ 、推薦候補コンテンツの特徴ベクトルを  $\mathbf{C}_k$  とすると、類似度  $sim$  は次式で表せる。

$$sim(\mathbf{P}, \mathbf{C}_k) = \frac{\mathbf{P} \cdot \mathbf{C}_k}{\|\mathbf{P}\| \cdot \|\mathbf{C}_k\|}$$

#### 4. ユーザプロファイル

人間の興味は一般に時間の経過と共に変動するものと時間経過に影響されにくいものに分けられる。例えば、前者では時事性の強いニュース記事、後者では趣味などに対する興味などである。興味に基づいて情報推薦を行う際、この二つの興味をいかに反映させるかが重要となる。

内容に基づくフィルタリングを用いて推薦を行う場合、ユーザプロファイルをどのように生成するかが推薦結果を大きく左右する。通常、ユーザプロファイルはユーザの閲覧情報と閲覧したコンテンツに対する評価を元に生成する。しかしながら、本研究においては RSS フィード情報取得時の労力軽減を目的としており、ユーザに評価の手間を強いることは望ましくない。そこで、「ユーザが閲覧したコンテンツは全てユーザの嗜好に沿ったコンテンツである」という仮定に基づき、閲覧したか否かを興味の有無に対応付け、それを評価値とする。さらに、人間の二つの興味、「時間の経過と共に変動する興味（動的興味）」と「時間経過に影響されにくい興味（静的興味）」に着目し、それぞれを表現するユーザプロファイルを生成する手法を提案する。これら2つのユーザプロファイルを併用することで更新頻度の高い Web コンテンツに対して動的興味・静的興味のいずれにも対応する推薦を行うことが可能となる。

##### 4.1. 動的ユーザプロファイル

流行に関するコンテンツやニュース記事のような時事性の強い話題に対する興味は、それまでどのようなコンテンツを閲覧し、どのような話題に興味を持っていたかにかかわらず突発的に発生し、時間が経つにつれて興味が薄れていくと考えられる。本研究ではこのような時間の経過に伴い変化する興味を動的興味、動的興味を表すユーザプロファイルを動的ユーザプロファイルと定義する。

ここで、コンテンツに対する興味は1日で全く変わってしまうほど早い変化をするものではなく、 $n_d$  日前に興味があったコンテンツは現在でも興味があるものと仮定する。更に、過去  $n_d$  日間に閲覧したコンテンツでも、より現時刻に近い閲覧情報のほうが現在の興味を表す上で有効であると仮定する。これらの点を踏まえ、動的興味を表現する動的ユーザプロファイル  $\mathbf{P}_d$  を次のように定義する。

$$\mathbf{P}_d = \sum_{i=1}^m k_i \mathbf{I}_{di}$$

ここで、 $m_d$  は過去  $n_d$  日間に閲覧したコンテンツの数、 $\mathbf{I}_d$  は閲覧したコンテンツの特徴ベクトルである。また、 $k_i$  は現在日時とコンテンツ  $i$  を閲覧した日時から算出する時系列係数である。

既存の情報推薦の研究において、興味の減衰はガンマ分布に従うという仮定がなされているが[11]、本研究では、興味は指数関数に従うと仮定する。過去  $n_d$  日間に閲覧したコンテンツのうち、コンテンツ  $i$  の閲覧を行ったのが  $t_i$  日前であるとき、時系列計数  $k_i$  は、

$$k_i = \lambda e^{-\lambda t_i} \quad (0 \leq t_i \leq n_d)$$

と表すものとする。

##### 4.2. 静的ユーザプロファイル

動的興味に対して、ユーザの趣味に関するコンテンツなど、時間経過に影響されにくい興味を静的興味と定義する。静的興味を表すユーザプロファイルは静的ユーザプロファイルと呼ぶ。

静的ユーザプロファイルを作成するためには、ユーザの趣味のうち、時間経過に影響されにくいものを特定する必要がある。そこで、静的ユーザプロファイルを過去  $n_s$  日間に閲覧したコンテンツ情報を元に作成するとき、コンテンツを閲覧した日時に関わらず平均的に高い頻度で出現するキーワードが静的興味を表現するキーワードであると考えられる。そこで静的ユーザプロファイル  $\mathbf{P}_s$  は過去  $n_s$  日間に閲覧したコンテンツの特徴ベクトルの総和を日数  $n_s$  で割ったものとする。

よって、静的ユーザプロファイル  $\mathbf{P}_s$  は、

$$\mathbf{P}_s = \frac{\sum_{j=1}^{m_s} \mathbf{I}_{sj}}{n_s}$$

で表される。ここで、 $m_s$  は過去  $n_s$  日間に閲覧したコンテンツの数、 $\mathbf{I}_s$  は閲覧したコンテンツの特徴ベクトルである。

## 5. システムの実装

### 5.1. 実装システムの概要

システムの実装画面を図3に示す。システムは推薦提示部とブラウジング部の2つからなる。推薦提示部では、システムに登録した RSS フィードを巡回して得られたコンテンツを一覧表示する機能と推薦結果を一覧表示させる機能を持つ小型のブラウザである。推薦提示部に一覧表示された RSS フィードの情報やシステムの推薦結果を選択してクリックする (①) ことでブラウジング部にそのコンテンツ

が表示される。このときの閲覧情報（閲覧コンテンツの URL、時刻情報など）はシステムに送られ、閲覧履歴として記録され、ユーザプロフィールを作成するのに使用される。

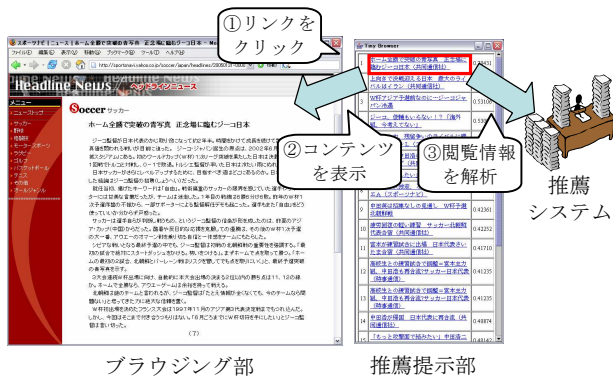


図3 システムの実装イメージ

## 5.2. 実装実験

3 節、4 節で述べた提案手法の有効性を確かめるため、提案手法に基づいたプロトタイプシステムを用い、実装実験を行った。実装実験の条件は以下の通りである。

- ・ 試験期間：2005/7/1～7/31
- ・ 今回の実験では WWW 上のニュースコンテンツに関する RSS フィードのみを取り扱うものとし、Blog 等の RSS は対象としない
- ・ ユーザプロフィールは過去7日間 ( $n_d = 7$ ,  $n_s = 7$ ) に閲覧したコンテンツを元に作成し、動的ユーザプロフィール作成時の時系列係数は  $\lambda = 1.0$  とする
- ・ コンテンツ推薦部は「コンテンツのタイトル」「コンテンツへのリンク」「ユーザプロフィールとの類似度」の組からなる推薦結果を生成する。このとき、推薦するコンテンツの数は類似度の上位 25 件とする
- ・ 推薦候補ロボット型ニュース検索エンジン ceek.jp news [12] にて公開されている RSS フィードに基づいて生成する

また、実験の手順は以下の通りである。

- ① 実験対象となるユーザは、WWW 上のニュース記事を自由に閲覧する。この際の閲覧情報はシステムが保持し、ユーザプロフィールの作成に用いる
- ② 1 日に 1 度、生成したユーザプロフィールを用いて Web コンテンツの推薦を行う
- ③ ユーザは推薦されたコンテンツを全て閲覧し、コンテンツに対し、興味の有無を 5 段階で評価を行う

## 5.3. 実験結果

実験の結果、推薦されたコンテンツに対して、ユーザが「i. 興味がある」もしくは「ii. やや興味がある」と評価をした場合、ユーザはそのコンテンツに興味があったと見なせる。よって、本プロトタイプシステムの推薦適合率は次式で表せる。

ある」と評価をした場合、ユーザはそのコンテンツに興味があったと見なせる。よって、本プロトタイプシステムの推薦適合率は次式で表せる。

$$precision = \frac{\text{評価 } i, ii \text{ のコンテンツ数}}{\text{全推薦コンテンツ数}}$$

動的ユーザプロフィールおよび静的ユーザプロフィールに基づいた推薦を行った結果、それぞれの適合率  $precision$  を平均すると以下ようになった。

表 1 各ユーザプロフィールの推薦適合率

ユーザプロフィール	平均適合率 $precision$
動的ユーザプロフィール	0.54
静的ユーザプロフィール	0.64

両ユーザプロフィールともに、同じ Web サイトのコンテンツが類似度上位に偏って推薦されることがあり、動的ユーザプロフィール、静的ユーザプロフィールに関わらず、推薦適合率は 0.6 前後とあまり高くはない。本手法ではユーザプロフィールを作成する際、Web コンテンツをページ単位で取得・解析するが、その際にコンテンツ本来の内容に関係のない部分も含まれるため、その不要部分がコンテンツ特徴ベクトル生成時にノイズとして残ってしまうことが原因とみられる。更に、類似度上位 25 件という推薦の閾値も関係していると思われる。インタフェースに一度に表示できる件数が 25 件程度であったために 26 件目以下は結果から排除したが、実際には 26 件目以降にもユーザの興味に沿ったコンテンツが含まれていることも多く、この推薦漏れが適合率の低下に繋がったと考えられる。

推薦適合率を向上させるには、閲覧したコンテンツから本文のみを抽出し、それをコンテンツ特徴ベクトルの作成に使用することが考えられる。また、推薦結果として提示する件数も、類似度に閾値を設定したり全候補記事数に連動させるなどの工夫が必要である。

ここで、あるユーザ U の 2005 年 7 月における閲覧傾向に着目する。ユーザ U の閲覧傾向を考慮し、閲覧したコンテンツを日ごとにまとめ、「サッカー」および「ロンドンテロ」に関するコンテンツを手手で分類、算出した。その結果、ユーザの閲覧傾向として図 4 のグラフが得られた。

図 4 より、ユーザ U は普段、サッカーに興味を持っており、調査期間を通して「サッカー」に関するコンテンツを多く閲覧している。また、7 月 7 日にロンドンで発生したテロの影響を受け、それ以降「ロンドンテロ」に関するコンテンツを閲覧するようになるものの、7 月 8 日の 11 件をピークに「ロンドンテロ」に関するコンテンツはほとんど閲覧しなくなる。これらはそれぞれ、本研究における静的興味と動的興味であるといえる。したがって、

静的ユーザプロフィールによる推薦では「サッカー」に関するコンテンツを、動的ユーザプロフィールによる推薦では7月7日前後で「ロンドンテロ」に関するコンテンツが結果として得られればよい。

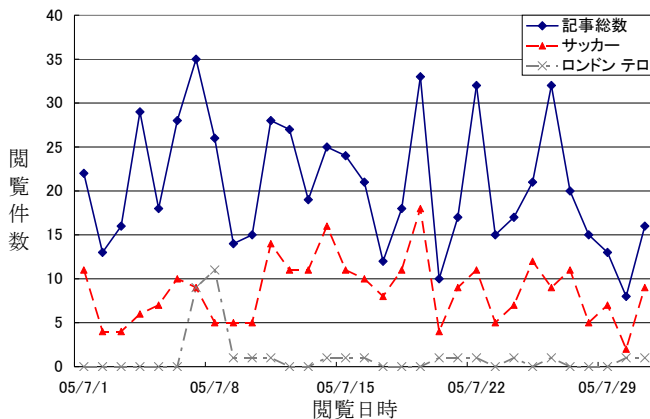


図4 ユーザUの閲覧傾向

そこで、図4に示したユーザUに対して、動的ユーザプロフィール  $P_d$  および静的ユーザプロフィール  $P_s$  を用いて推薦を行ない、動的興味・静的興味を反映した推薦が行われているかを確認した。その結果を表2に示す。

表2 動的・静的プロフィールと推薦結果

	7/5	7/8	7/11	7/14
「テロ」: $P_d$	0 [件]	8 [件]	2 [件]	1 [件]
「サッカー」: $P_d$	8 [件]	4 [件]	6 [件]	5 [件]
「テロ」: $P_s$	0 [件]	4 [件]	4 [件]	2 [件]
「サッカー」: $P_s$	6 [件]	5 [件]	7 [件]	4 [件]

ユーザUの静的興味である「サッカー」に関するコンテンツについて、静的ユーザプロフィール  $P_s$  を用いた場合、「サッカー」に関するコンテンツの推薦件数は実験期間を通してほぼ均一となっていることがわかる。しかしながら、動的ユーザプロフィール  $P_d$  を用いた場合においても、静的興味「サッカー」に関する推薦結果が得ることができた。

一方、ユーザUの7/8時点の動的興味である「ロンドンテロ」に関しては、動的ユーザプロフィール  $P_d$  を用いて推薦を行った場合、推薦件数は7/8に最大となり、その後時間が経つにつれて収束した。また、静的ユーザプロフィール  $P_s$  を用いた推薦でも7/8では動的ユーザプロフィール  $P_d$  と同様に推薦が行われたが、その後推薦件数が収束するまでに動的ユーザプロフィールよりも長い時間を費やした。これは動的ユーザプロフィールによる推薦では「ロ

ンドンテロ」に関するコンテンツを一時的なものであると判断したのに対し、静的ユーザプロフィールによる推薦では静的興味の一つであると判断したためであると思われる。これらのことより、動的ユーザプロフィール・静的ユーザプロフィールは、動的興味・静的興味それぞれに対して有効な推薦を行うことができると言える。

## 6. おわりに

本研究では、RSS フィードに記述される更新頻度の高い Web コンテンツに対し、2つの興味モデルを用いたコンテンツ推薦システムを提案し、プロトタイプシステムの実装実験を行った。そして提案手法を用いることによりユーザの興味の種類に沿ったコンテンツの推薦が可能であることを示した。

今後は、推薦適合率の向上や2つのユーザプロフィールの併用、更に短い興味変化に対応したプロフィールなどを検討するとともに、システムの評価を行う必要がある。

## 参考文献

- [1] 久津見洋, 内藤榮一, 荒木昭一, 江村里志, 新居薫治, “ユーザ適応型ホームページ推薦ソフト“ウェブナビゲーター”の開発”, 電子情報通信学会論文誌, D-II, Vol.84-D-II, No.6, pp.1149-1157, 2001
- [2] 佐保田圭介, 波多野賢治, 宮崎純, 吉川正俊, 植村俊亮, “ブックマークの階層構造情報を組み込んだ協調フィルタリングによる Web ページの推薦手法”, 電子情報通信学会第15回データ工学ワークショップ (DEWS2004), 6-B-4, 2004
- [3] Douglas W. Bass, “RSS, OPML and Weblog Ecosystems: A Survey of New Technologies in Internet Publication”, Conference on Advances in Internet Technologies and Applications (CAITA 2004), 2004
- [4] myrss.jp, <http://myrss.jp/>
- [5] 南野朋之, 奥村学, “なんでも RSS! - HTML 文書からの RSS Feed 自動生成”, 人工知能学会第10回セマンティックウェブとオントロジー研究会 (SIG-SWO-A501-03), 2005
- [6] Google ニュース 日本版, <http://news.google.jp/>
- [7] My Yahoo!, <http://my.yahoo.co.jp/>
- [8] 武田英明, “Weblog 研究の現状”, 第7回セマンティックウェブとオントロジー研究会 (SIG-SWO-A402-06), 2005
- [9] 南野朋之, 鈴木泰裕, 藤木稔明, 奥村学, “blog の自動収集と監視”, 人工知能学会論文誌, Vol.19, No.6, pp.511-520, 2004
- [10] 形態素解析システム茶釜, <http://chasen.naist.jp/hiki/ChaSen/>
- [11] 土方嘉徳, “情報推薦・情報フィルタリングのためのユーザプロフィール技術”, 人工知能学会誌, Vol19, No 3, pp. 365-372, 2004
- [12] CEEK.JP NEWS, <http://news.ceek.jp/>