

RSS に基づく内容型情報推薦システムの提案

向井 誠 青野 雅樹

豊橋技術科学大学

1 はじめに

WWW の発展とともにニュースや Blog に代表される更新頻度の高い Web コンテンツが増加している。このような動的 Web コンテンツより必要な情報を効果的に得るため、最近では RSS を用いた概要情報の配信が盛んである。RSS に記述された情報は RSS リーダと呼ばれるソフトウェアを用いて表示されるが、既存の RSS リーダでは概要情報を一覧表示するだけのものが多い。同種のサービスとして Web 上のニュース記事を元にダイジェストを自動生成し、定期的に更新するサービスも存在するが、ユーザ個人の興味を反映した記事の提示には至っておらず、目的とする内容の記事情報を得るのが困難となっている。

本稿では Web 上のニュース記事などの動的 Web コンテンツの概要情報を記述した RSS を用いて、個人の興味を反映できる内容型の情報推薦システムを提案し、あわせて実験結果を報告する。

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を用いて Web コンテンツを推薦する手法^[1]やユーザの興味に基づいた動的 Web コンテンツの提示手法^[2]は存在するが、前者はニュースなどの動的 Web コンテンツの推薦には適しておらず、後者は興味の時間的変化に対応していない。

一般に人間の興味は、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が挙げられる。興味に基づいて情報推薦を行う際、この二つの興味を反映させることが重要となる。ニュース推薦に情報フィルタリング手法を用いる際、協調フィルタリングでは時事性の強い記事に対して他者の評価が集まりにくく適切な推薦が行えない。一方、内容型フィルタリングは、ユーザプロフィールを作成すれば動的に推薦を行える。

ニュース推薦に内容型フィルタリングを用いることを考える。ユーザが持つ興味のうち、時間変動する興味の対象となる時事性の強い記事は、閲覧する前後の時刻でのみ興味を示す。一方、時間変化しない興味の対象となる記事は、時期などにかかわらず

閲覧することが多い。ユーザの閲覧情報をユーザプロフィールの作成に利用するとき、記事をベクトル空間モデルによる特徴ベクトルで表現し、ユーザが閲覧した記事の特徴ベクトルを足し合わせることで、時間変化する興味と時間変化しない興味の双方を併せた興味を表現できると考えられる。

2.2 RSS

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。バージョンにより多少の違いはあるが、基本的には図 1 に示す通りである。

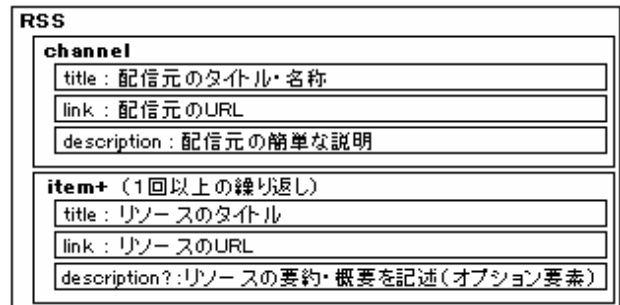


図 1 RSS の基本構造

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 エントリなど配信したいリソース情報を記述) 複数からなる。一般にニュース記事を記述した RSS においては item 要素ひとつひとつが個々の記事を表し、item 内の title 要素は記事の見出し、link 要素は記事が掲載されている Web ページの URL、description 要素は記事内容テキストが記述されている。

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。

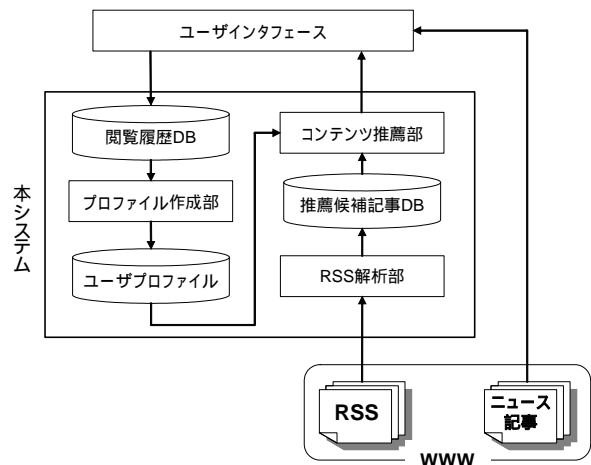


図 2 システムの構成

A Content-based Recommender System based on RSS
 Makoto MUKAI, Masaki AONO
 Tovohashi University of Technology

(1) RSS 解析部

RSS 解析部では、巡回先として登録された WWW 上の RSS より概要情報を取得する。item 要素内の description 要素に記述される記事内容テキストを形態素解析し、記事の特徴付けるキーワードからなる記事特徴ベクトルを作成する。description 要素が存在しない場合、link 要素の記事 URL に基づいて記事内容テキストを独自生成して記事特徴ベクトルの作成を行う。なお、単語の重み付けには正規化 tf を用い、作成した特徴ベクトルは推薦候補記事 DB に格納する。

(2) コンテンツ推薦部

コンテンツ推薦部では、推薦候補記事 DB の特徴ベクトルとユーザプロファイルの類似度（コサイン相関値）を求め、得られた類似度を用いて推薦候補記事の順位付けを行い、上位のものを推薦記事としてユーザに提示する。

(3) 閲覧履歴 DB

ユーザはコンテンツ推薦部にて提示された推薦記事やその他の WWW 上のニュース記事の中から興味のある記事を開覧する。この際、ユーザは閲覧した記事に対して評価を行う。評価値と閲覧記事の特徴ベクトルより作成した閲覧特徴ベクトルを閲覧履歴 DB に格納する。閲覧特徴ベクトルは閲覧した日時を元に 1 日単位で作成する。その日閲覧した記事数が m 個あり、それらのベクトルを \vec{I} とすると、閲覧記事ベクトル \vec{H} は次式で表される。

$$\vec{H} = \sum_{i=1}^m k_i \vec{I}_i$$

ここで、 k_i は閲覧した記事 i への興味度合を表す重みである。本稿では、ユーザが閲覧した記事は全てに興味があったと仮定し、 $k_i = 1$ とする。

(4) プロファイル作成部

コンテンツ推薦部にて推薦候補を決定するため、閲覧履歴 DB に格納された閲覧特徴ベクトルを元にユーザプロファイルを作成する。

人間の興味や嗜好は時間経過により変動するものの、急激な変化をすることはない。そこでニュース記事に対する興味は n 日間継続するものと仮定し、過去 n 日分の閲覧特徴ベクトルを元にユーザプロファイルをベクトル形式で作成する。日時 i の閲覧特徴ベクトルを \vec{H}_i とするとユーザプロファイル \vec{P} は次式で表せる。

$$\vec{P} = \sum_{i=1}^n w_i \vec{H}_i$$

ここで、 w_i は時間経過に対する興味の変化を表す重みである。本稿では過去 n 日に閲覧した記事全てに等しく興味を持つものとし、 $w_i = 1$ とする。

4 実装実験

本提案システムを実装し、特定のユーザ 1 名に対してユーザプロファイルの作成および推薦の実験を行い、推薦結果とユーザの興味推移の傾向について調査した。

ユーザプロファイルは、既読記事に対する興味がなくなるまでの期間 n を 7 日と設定して作成し、コンテンツ推薦部で算出された類似度の上位 20 記事を推薦記事とした。また、推薦候補記事として現在 WWW 上に公開されている RSS^[3] のうちニュースサイトに関するものを使用した。

今回実験対象となったユーザは、「サッカー」に関するニュース記事に強い興味を持っており、「サッカー」関連記事を開覧することが多い。しかし、12/26 に発生したスマトラ沖大地震の影響で「地震」に関する記事に興味を持ち、「地震」関連記事を多く読むようになった。その結果、システムが「地震」関連記事を多く推薦するようになった（図 3）。なお、推薦記事の 77.5% がユーザの興味に沿った記事であった。

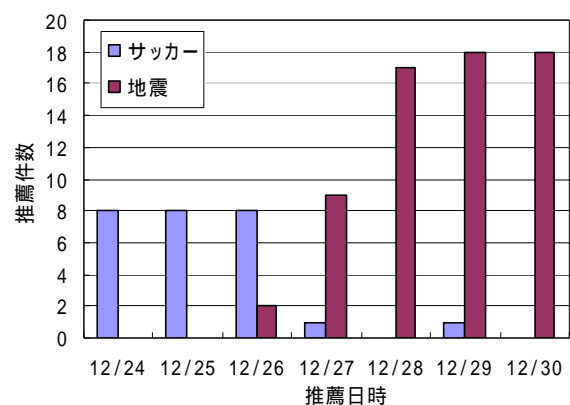


図 3 ユーザに対する推薦結果の内訳

5 おわりに

RSS に基づいたニュース推薦システムを提案し、ユーザの興味推移と推薦結果の傾向について調査を行った。本システムによりユーザの興味推移を反映したニュース記事の推薦が行える。しかし、今回の実験では時事的な興味に対する推薦が強く、時間変化しない興味「サッカー」に対する推薦が不十分であった。今後は推薦条件、パラメータの調整と共に実験ユーザの拡大、及び推薦結果の評価を行う。

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- , Vol.84-D- , No.6, pp.1149-1157
- [2] mySiteOutliner, <http://www-6.ibm.com/jp/jpccinfo/clubibm/msol/>
- [3] rss-jp.net, <http://rss-jp.net/>