社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2008－MPS－72 (41)
2008－BIO－15 (41)
2008／12／18

対話的なキーワード抽出によるブログ推薦システム

澁谷 翔吾[†], 廣安 知之[††], 三木 光範[‡], 横内 久猛[††]

[†] 同志社大学大学院 [††] 同志社大学生命医科学部 [‡] 同志社大学理工学部

近年，オンラインショッピングサイトでは，ユーザの注文データ，趣味，属性，クリック履歴などの様々な利用情報を分析し，一人ひとりの嗜好に合った商品を適宜案内することで，ユーザの商品購買率を高める商品推薦を行っている．そこで，本稿では，情報推薦をブログリーダーに用いることを提案する．近年，自身のブログを容易に作成することが可能となったことから，ブログを書くユーザが増加してきている．また，ユーザは自身と面識のないユーザのブログを読むこともあり，友人の枠を超えて興味深いブログを求めている．そこで，ユーザがそれらブログを読んでいく過程で，システムからユーザの嗜好にあったブログを推薦させると，ユーザのブログを読む楽しみは増加すると期待できる．本稿では，ユーザの嗜好に基づいて，ブログを推薦するシステムを提案する．

Blog Recommendation System using Interactive Keyword Extraction

Shogo SHIBUTANI[†], Tomoyuki HIROYASU[††], Mitsunori MIKI [‡]and Hisatake YOKOUCHI[††]

[†] Graduate School of Engineering, Doshisha University
[††] Department of Life and Medical Sciences, Doshisha University
[‡] Department of Science and Engineering, Doshisha University

Recently, the system in the online shopping site analyzes the various information such as the order data, a preference, an attribute and click history. Then, it recommends information to encourage users to purchase products by guiding the article which fits each user's preference appropriately. In our research, we suggest that we use recommendation in blog reader. In these years, users who write blogs increase because they can make them easily. In addition, users can read the blogs which they does not know who wrote them and they can search interesting blog. Then, we expect that users enjoy reading blogs and their pleasure increases if system recommends blogs based on users' preference when they look for blogs. In this research, we suggest a system to recommend blogs based on the users' preference.

1 はじめに

インターネットの普及により，amazon.com[1]や楽天市場[2]といったオンラインショッピングサイトを利用するユーザが増えている．オンラインショッピングサイトを利用することで，ユーザは商品などの注文をオンラインで行うことができる．近年，それらのオンラインショッピングサイトでは，個々のユーザの嗜好にあった商品を推薦するリコメンドサービスを行っている．リコメンドとは，ユーザの注文データ，趣味，属性，クリック履歴などの様々な利用情報を分析し，一人ひとりの嗜好に合った商品を適宜案内し，ユーザの情報収集を支援する手法である．ユーザにとっては，自分の欲しい情報に素早くアクセスできる可能性が高まり，一方，商品，サービスを提供する企業にとってはユーザの商品購買率を高められるなど，双方にメリットが大きい手法として急速に普及している．情報推薦するためには情報をフィルタリングする必要がある．情報のフィルタリングには，キーワードや単語出現頻度などの情報を用いた内容に基づくフィルタリング，また情報に対する他のユーザの評価に基づく協調フィルタリング[1) 2)]がある．いずれの手法もユーザのプロファイリングが重要である[3)]．

本稿では，上述した情報推薦をブログリーダーに用いることを提案する．近年，自身のブログを容易に作成することが可能となったことから，ブログを書くユーザが増加してきている．また，ユーザは自身と面識のないユーザのブログを読むこともあり，友人の枠を超えて興味深いブログを求め

[1] http://www.amazon.co.jp/
[2] http://www.rakuten.co.jp/

ている．そこで，ユーザがそれらブログを読んでいく過程で，システムからユーザの嗜好にあったブログを推薦させると，ユーザのブログを読む楽しみは増加すると期待できる．本稿で提案するシステムは，ユーザの嗜好に合ったブログを推薦するシステムである．本システムにおける，ユーザの興味抽出はシステムとユーザが対話的に行う．ユーザの興味抽出をユーザ自身の操作により行うことで，ユーザの興味を的確に抽出することが可能である．

2 関連システム

リアルコム[3]は，ユーザーごとに最適化した記事を推薦するRSSリーダー「SocialFeed」のサービスを行っている．SocialFeedでは，同じフィードを登録しているかどうかでユーザ間の関連度を算出し，関連度でユーザをクラスタリングする．同じクラスタに属するユーザが購読している記事がユーザに推薦される仕組みである．また，推薦された記事に対して閲覧，登録，および保存などの操作を行なうと，行動ごとに関連度の計算が行われる．この繰り返しにより，推薦の精度が高められる．

本稿で提案するシステムも，SocialFeedと同様，情報の推薦とユーザの評価を繰り返していくことで，精度向上が期待されるシステムである．

3 ブログ推薦システム：Drag Based(DB) Empfehlung

DB Empfehlungはユーザの嗜好にあったブログを推薦するシステムである．近年，ブログを読む習慣のあるユーザは友人の枠を超えて，興味深いブログを求めている．しかし，ブログの数は膨大であり，その中から自身の興味あるブログを見つけ出すことは困難である．本システムを用いることにより，一人ひとりのユーザに応じたブログ推薦を行うことが可能となる．

3.1 概要

DB Empfehlungでは，ユーザの興味とブログの特徴からユーザの嗜好に合ったブログを推薦する．本システムでは，ユーザの興味はユーザにより示される仕組みとなっており，ブログの特徴はブログを読む全ユーザにより特徴付けられる仕組みとなっている．本などを読むとき，「興味ある箇所に線を引く」操作を行うが，本システムにおいても，同様の操作をユーザインタフェースとして提供する．システムはユーザごとの興味，および全ユーザにより特徴付けられたブログの特徴を蓄積することで，個々のユーザの嗜好に合ったブログを推薦する．システム全体の概要を図1に示す．

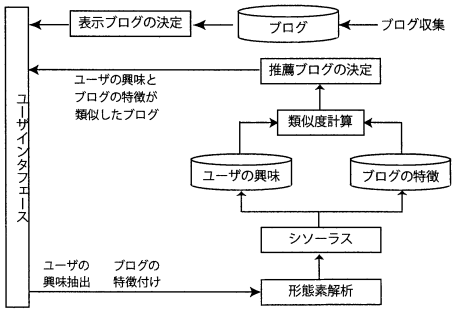

図 1: システム概要

3.2 ブログの収集

RSSを取得し，記事ごとのブログタイトル，本文，およびURLを格納している．

3.3 ユーザの興味抽出

本システムでは，ユーザの興味はユーザにより示される仕組みとなっている．つまり，ユーザはブログを読む過程で，なんらかの操作を行い，システムに興味を伝える．本システムでは，その操作をユーザのドラッグ操作（以下，ドラッグ）とする．「興味ある箇所に線を引く」操作はドラッグによって実現する．つまり，ユーザはブログを読み，興味のある単語，および文章をドラッグすることで自身の興味を示す．ユーザが文章をドラッグした場合は，MeCab[4]を用いて，文章を形態素解析し，名詞を抽出し，興味ある単語として保存する．システムは，ユーザのドラッグにより得られたユーザの興味を示す単語を蓄積することで，各ユーザの興味を把握する．さらに，システムはそれらの単語に加えて，シソーラスを用いることで，単語の上位概念を取得し，そのユーザの興味として蓄積する．つまり，図2に示すように，ユーザが”サッカー”という単語をドラッグしたときには，”サッカー”という単語に加えて，”サッカー”の上位概念である”スポーツ”という単語を蓄積する．

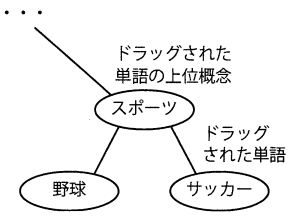

図 2: シソーラスの利用

[3] http://www.realcom.co.jp/

[4] http://mecab.sourceforge.net/

また，ユーザがある単語にどの程度の興味を示しているのかを把握するために，表1のようにユーザが興味を示した単語の出現回数を保持している．

表 1: ユーザの興味

単語	出現回数
単語 1	frequency(単語 1)
単語 2	frequency(単語 2)
…	…
単語 m	frequency(単語 m)

frequency(単語 m) は，単語 m の出現回数を示す．ユーザにより，ドラッグされた単語は，1 回とカウントし，ドラッグされた単語の上位概念の単語には，ユーザの直接的な操作ではないため，0.5 回とする．上位概念を回数に数えることで重み付けを行っている．出現回数が多い単語ほど，その単語に対する興味の度合いは大きい．システムは，この単語と出現回数から各ユーザの興味ベクトルを作成し，保持している．つまり，システムは興味のある単語が m 個の場合，m 次元のユーザの興味ベクトルを保持する．

3.4 ブログの特徴

本節では，ブログの特徴付けについて解説する．ブログを特徴付けるには，文章中の特徴的な単語を抽出するアルゴリズムである tf-idf[4] などが考えられる．しかし，ブログは一般に情報が入り交じっており，特徴的な単語を抽出することは困難であると考えられる．そこで，本システムでは，ブログはシステムを利用する全ユーザに特徴付けられる仕組みとした．ユーザは，あるブログを読むとき，興味ある単語，および文章をドラッグする．文章がドラッグされた場合，3.3 で解説したのと同様に，形態素解析を行い名詞を抽出する．この単語をブログを特徴付ける単語とする．ブログは様々なユーザに読まれるので，1 つのブログは様々なユーザから特徴付けられる．また，シソーラスを用いてブログを特徴付ける単語の上位概念である単語を取得し蓄積する．ブログを特徴付ける単語の重みを考慮するため，ユーザの興味と同様に，単語の出現回数を保持している．出現回数が多い単語ほど，ブログはその単語に強く特徴付けられている．システムは，この単語と出現回数から各ブログの特徴ベクトルを保持している．ドラッグされた単語が n 個の場合，システムは n 次元のブログの特徴ベクトルを保持する．

3.5 推薦ブログの決定

本節では，推薦ブログの決定について解説する．ブログを推薦するためには，ユーザの興味に基づいて，ブログを選出する必要がある．本システムでは，ユーザの嗜好とブログの特徴が似ているものを推薦対象とする．ここで，似ているとはユーザの興味とブログの特徴との類似度が高いことを意味する．ユーザの興味とブログの特徴との類似度を算出するために，本システムではベクトル空間モデル[5] を用いて類似度を算出する．

ユーザの興味ベクトルとブログの特徴ベクトルの方向は，それぞれの特徴を示すものであるため，2 つのベクトルが同じ方向を向いていると類似していることになる．つまり，2 つのベクトルのなす角が小さいほど類似度が高いということである．ユーザの興味とブログの特徴との類似度は，以下の式で計算することができる．

$$sim(User, Blog) = cos\theta = \frac{\vec{User} \cdot \vec{Blog}}{\left|\vec{User}\right| \left|\vec{Blog}\right|} \quad (1)$$

式 (1) の θ はベクトル User とベクトル Blog のなす角であり，sim(User, Blog) は User と Blog の類似度である．θ が 0，つまり sim(User, Blog) が 1 に近づくほど類似性は高く，逆に θ が 90，つまり sim(User, Blog) が 0 に近づくほど類似性は低くなる．本システムでは，あるユーザと各ブログとの類似度を計算し，類似度の高いブログを推薦する．

上述したように，ユーザの興味抽出、およびブログの特徴付けはそれぞれドラッグされた単語の上位概念を保持している．これにより，図 3 に示すように，上位概念を持たない場合に比べて，幅広い推薦が可能となる．

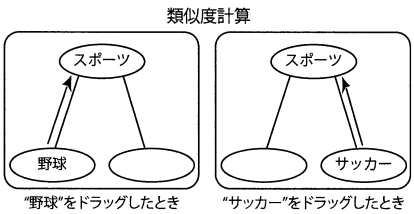

図 3: シソーラスを用いた場合の関連度計算

シソーラスを用いない場合，ドラッグされた単語のみを用いて類似度を算出する．シソーラスを用いることで，上位概念を取得し，その概念を考慮して類似度を計算することになる．図 3 では，シソーラスを用いることで，”サッカー”の上位概念である”スポーツ”を取得し，類似度の算出に用いている．シソーラスを用いなければ，図 3 の”サッカー”と”野球”は関連のないものとして扱われるが，シ

ソーラスを用いることで関連を持たせることが可能となる．

3.6 ユーザインタフェース

本システムのユーザインタフェースを図4に示す．

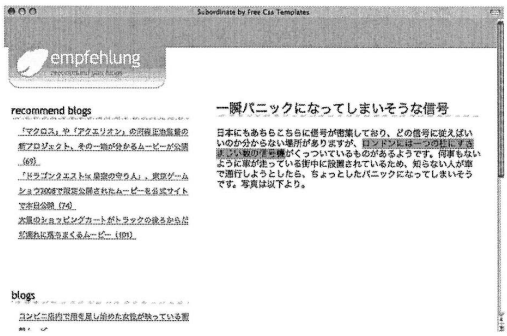

図 4: ユーザインタフェース

本システムでは，図4のように画面左に推薦ブログ，およびデータベースからランダムに提示されるブログのリストがある．ユーザはリストからブログを選択すると，画面右の領域にブログのタイトル，および内容が表示される．ユーザはブログを読み，その内容で興味ある箇所にドラッグを行うことで，システムはユーザの興味を抽出し，ユーザはブログを特徴付ける．図4において，ブログ本文に線が引かれている部分は，ユーザがドラッグした文章である．システムはドラッグ，つまりユーザの興味に基づき，ブログの特徴との類似度を算出し，ブログを推薦する．ユーザは推薦されたブログの興味ある単語，および文章にドラッグを行うことが可能である．このことから，システムはブログを推薦，ユーザはドラッグといったように対話的にシステムを利用することが可能である．

3.7 推薦結果

あるユーザへの推薦結果を図5に示す．

1．2010 年発売予定のスーパーカー「Audi R8 Spider」の流出ムービー
2．ニコニコ生放送のストリートファイターオンラインイベント
3．ミクロの世界の美麗な写真コンテスト

図 5: 推薦結果

このユーザは「イベント，ムービー，モバイル」といった単語にドラッグしていた．図5の結果は，ユーザが興味を示した単語とその出現回数でベクトル化し，ブログの特徴との類似度を算出した結果，類似度が高かったものである．1つ目のブログは，”ムービー”という単語，2つ目のブログは”イベント”という単語で特徴付けられていたため推薦された．3つ目のブログは，ユーザがドラッグした単語では特徴付けられていなかったが，ユーザがドラッグした”イベント”という単語の上位概念である”催し”という単語と，3つ目のブログの特徴語である”コンテスト”という単語の上位概念である”催し”という単語で関連付けられていた．つまり，シソーラスを用いて推薦が行われた結果である．

4 おわりに

本稿では，ユーザの嗜好にあったブログを推薦するブログ推薦システム，DB Empfehlung を提案した．本システムは，ユーザの興味とブログの特徴からユーザの嗜好に合ったブログを推薦するシステムである．ユーザの興味は，ユーザにより示される仕組みとなっており，ブログの特徴はブログを読む全ユーザにより特徴付けられる．また，シソーラスを用いることで，ユーザのドラッグで示される単語だけではなく，ドラッグされた単語の上位概念を取得し，その単語をユーザの興味，およびブログを特徴付ける単語とした．これにより，ドラッグされた単語だけで，類似度を算出するのではなく，類似度に上位概念を考慮することで幅広い推薦が行うことが可能となる．今後，上位概念の重み付けにより，推薦対象がどのように変化するのか検討する必要がある．

本システムには，推薦の精度において，課題がある．本システムでは，ブログを特徴付けるのはユーザであることから，推薦対象となるブログは，予め，他のユーザから特徴付けられている必要がある．これにより，システム稼働時には，的確にブログを推薦することはできない．これを回避するために，ブログがユーザにより特徴付けられる現在の仕組みと，内容に基づくフィルタリングの組み合わせを検討する．

参考文献

1) Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl. Item-Based Collaborative Filtering Recommendation Algorithms. In Proc. the 10th International World Wide Web Conference (WWW10). Hong Kong, pp.285-295, May 1-5(2001).
2) Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, John Riedl. GruopLens: An Open Architecture for Collaborative Filtering of Netnews. Proceedings of the ACM, pp.175-186, 1994.
3) 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ァイリング技術，人工知能学会論文誌，Vol. 19, No. 3
4) 清田 陽司，黒橋 禎夫，中村 順一，工藤 拓．単語の共起情報による重みづけを用いたテキスト検索システム，IREX ワークショップ予稿集, IREX 実行委員会, pp. 53-56, KKR ホテル, 東京都, September, 1999.
5) 池原 悟，村上 仁一，本本 泰博，技術資料単語意味属性を使用したベクトル空間法，言語処理学会論文誌，Vol.10, Number 2, pp.111-128 (2003-4).