

blog 検索と類似品情報を用いた選定支援システム

山名 健悟[†] 西村 圭亮[‡] 滝沢 敏裕[†] 湯浅 将英[‡] 大山 実^{†‡}

[†]東京電機大学大学院 情報環境学研究科 [‡]東京電機大学 情報環境学部

本システムは、定期的に収集した blog 記事から、ユーザの比較したい対象の情報だけでなく、対象に類似する物の情報を含めて検索し、表示するシステムである。これにより、ユーザは商品を買う際などに購入したい商品と類似品との比較が容易になる。ここでは、特に対象物の属性に着目し、属性と属性値を blog 記事中の文章解析により取り出し利用する。本稿では、選定支援システムの詳細を説明し、さらに blog 記事からの属性の取り出しと類似品情報の利用について述べる。

A Support System for Target Selection

by Blog Retrieval and Similar Targets' Information

Kengo YAMANA[†], Keisuke NISHIMURA[‡], Toshihiro TAKIZAWA[†],

Masahide YUASA[‡], and Minoru OHYAMA^{†‡}

[†]Graduate School of Information Environment, Tokyo Denki University

[‡]School of Information Environment, Tokyo Denki University

We developed a system that helps users make buying decisions. By using this system, users can know not only the information of the interest targets but also that of the similar targets by blog retrieval. The system extracts the attribute information of similar targets from blog articles and shows users the information. Based on the information, the user can easily select the target that he/she wants. In this paper, we explain about details of this system and use of similar targets' information.

1 はじめに

近年、インターネットでの情報発信の手段として、blog が広まっている。そこで、blog に書かれている情報を活用しようという試みが増加しており、blog を対象とした様々なサービスが登場している。たとえば、blog 記事の検索サービスでは blog の更新情報を収集し検索に役立つサービスがある [1]。これは blog の更新が頻繁であるという特徴を活用している。また、blog 記事には商品やサービスなどに関する人々の生の評判情報が反映され

やすいので、それらの評判を検索結果の表示に利用するサービスもある [2]。また、電子掲示板から評判情報を抽出する研究もある [3]。この研究では電子掲示板から抽出した評価表現が肯定的なのか否定的なのかをスコア付けにより判別している。

我々は blog 記事から得られる評判を用いて特に商品の選定を支援するシステムを開発している [4]。このシステムは、商品に対し肯定的な記事と否定的な記事を提示する機能や、評価表現の数を時系列にグラフ表示する機能、

検索対象の類似品を検索する機能を持つ。この際の類似品の選択に関しては、ショッピングサイトから商品の属性(大きさ, 重さなど)を収集し, ユーザに提示することで選択を支援している。この方法では仕様書に掲載されている属性のみを利用して類似品を比較している。

しかし, ユーザが商品を購入する際にはデザインや実際に使ってみたときの操作性を重要視する場合もある。仕様書のみを利用した検索ではこれらに対応できない。

そこで, このような仕様書に書かれていない特徴をblog記事の文章を解析して抽出し, ユーザに提示する。blogには人々の生の意見がより早く書かれやすいので, 有用な情報が得られると考えられる。blogに書かれる情報を類似品の選択に利用することにより, ユーザに提供される情報が増え, さらに効率的な選定支援が期待できる。

本稿ではまず2章で我々の開発しているシステム全体について説明する。そして, 3章で類似品情報を用いた選択支援について説明し, その予備実験について4章で述べる。最後に5章でまとめを述べる。

2 Blog記事を用いた選定支援システム

2.1 システムの概要

本システムは, ユーザがある商品を購入するときに, blog記事を参考にして選択することを効率化するシステムである。たとえば, ユーザがMP3プレーヤを購入したいとする。通常, ユーザはWEB上にあるメーカーの仕様書や, ネット通販のランキング情報などを基に商品を選択する。またはWEB検索やblog検索から当該商品に関する記事を見つけ出すなど, さまざまな情報を自ら集めて選択する必要がある。

本システムは, blog記事内にある評価表現を抽出し, 賛否分類して表示する機能, 評価表現の数を時系列にグラフ表示する機能, 商品の類似品を検索する機能を持つ。ユーザはこれらの機能を用いることで自らblogの記事を探し出す手間を減らし, 商品の選択が容易になる。

2.2 システムの流れ

システム全体の流れを図1に示す。本システムはblog記事を収集し, データベースに保存する。blog記事の収集にはblogの更新情報を集めているサイトで提供されている新着blog記事のRSSを利用する。それを参照しリンク先のblog記事本体のHTMLを保存する。また, 検索結果の表示に利用するため, ニュースサイトのRSSに書かれている情報や, ショッピングサイトに掲載されている商品の情報を収集し, 各々のデータベースに保存している。

ユーザが検索キーワードを入力すると, 本システムは保存されたblog記事からの検索を実行する。そして, ニュースや商品の情報を利用し, 次節に述べる3種類の方法でユーザに結果を提示する。

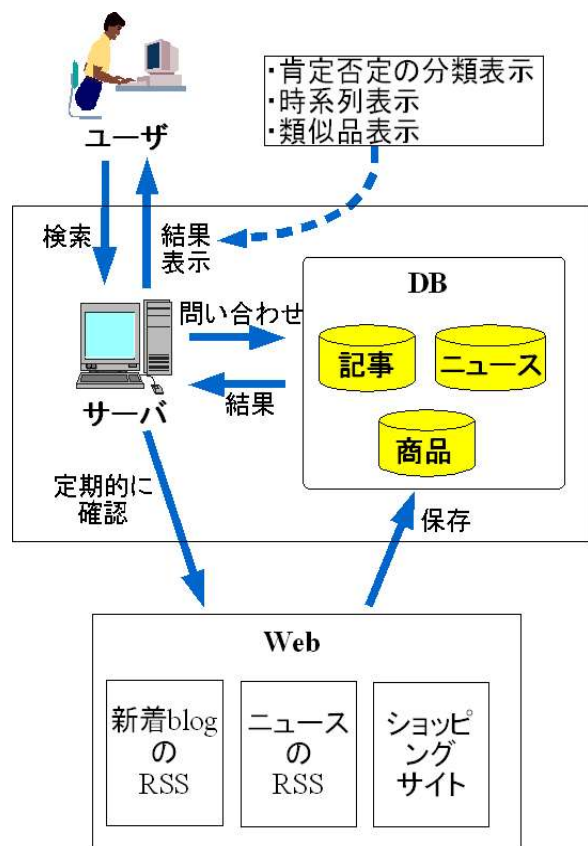


図1 システム全体の流れ

2.3 検索と結果の表示

ユーザが興味対象の記事を検索するときにはキーワードを入力することにより, 保存さ

れた blog 記事からの全文検索をする。検索には全文検索システムの Namazu[5]を利用する。

検索結果は以下の3通りの方法でユーザに提示する。

(1) 肯定的記事と否定的記事に分けた表示

検索結果中の評価表現の数により、記事を肯定的なものとするか否定的なものに分けて表示する。これにより検索対象の長所と短所が明確化でき、ユーザは的確かつ迅速な判断ができる。評価表現の抽出には評価表現用の辞書を作成し利用している。辞書の作成には藤村らの研究[3]の手法を用いた。ここではblog記事から形容詞と形容動詞を抽出して評価表現とした。それをスコア付けすることにより賛否を判別し、辞書を作成している。その辞書を用いてblog記事中から文字列一致により評価表現を抽出する。

「iPod」で検索した場合の結果表示の例を図2に示す。検索された記事中の評価表現の数を肯定/否定ごとに数え、その数によって記事を分類している。ユーザはこの表示結果から見たい記事をクリックすることにより、blog本体の記事を閲覧することができる。

| good | bad |
|--|--|
| <p>1. iPod無敵のトリセツ-ただのマニュアルじゃ満足できないアナタのための iPod超活用術-通販ショピング11e.jp (score:2004) Date:2005/06/30 01:55:37 cache Positive:9, Negative:0</p> <p>2. Apple iPod 20GB (Click Wheel) Mac&PC [M8282J/A]-通販ショピング11e.jp (score:1233) Date:2005/06/12 04:05:33 cache Positive:4, Negative:3</p> <p>3. iPod Style Navi - blog:iPodのライバルは登場するか?~ロンドン (score:1205) Date:2005/07/25 19:58:31 cache Positive:2, Negative:0</p> | <p>1. 「モノ」が好き!LGのiPod shuffleクローン (score:676) Date:2005/06/29 09:27:38 cache Positive:0, Negative:3</p> <p>2. 複写日記 iPodの不調 (score:249) Date:2005/07/03 03:34:48 cache Positive:0, Negative:2</p> <p>3. armrock: brand-new iPod (score:143) Date:2005/06/29 01:08:19 cache Positive:0, Negative:1</p> |



図2 肯定/否定で記事に分けた表示

(2) 時系列表示とニュースとのリンク

検索をした際の評価表現の数を特定の期間

ごとに合計し、その推移をグラフとして表示する。これにより、検索対象の評価の移り変わりを見ることができる。

さらに評価が大きく変化したときの理由を知ることができれば、興味対象の判断に有効である。そのためにblog記事以外にニュースサイトの記事を同時に利用し、評判の時系列表示とニュースの記事を並行して表示する。図3に検索文字列を「PSP」としたときのグラフを示す。上部の折れ線が肯定表現の数を、下部が否定表現の数を表している。同時に、ニュースサイトから抽出したトピックを日付順に並べて表示している。図3からPSPの発売のニュース(ニュース2,3)と同時期に肯定的表現の数が急激に増えているのがわかる。このように、グラフ表示により評判の変動の原因を知ることができ、選定支援に役立てることができる。

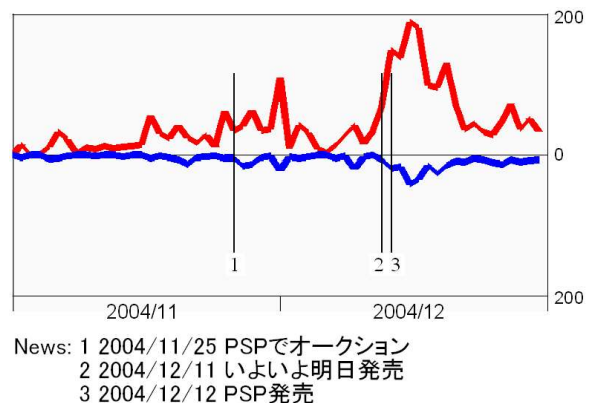


図3 blogの評価表現の数とニュースの表示

(3) 検索対象の類似品との比較

検索対象の類似品がある場合は、それについても同様に表示する。これにより、ユーザは類似品との比較が容易にできる。

類似品検索をするためには、対象の商品の属性情報が必要となる。たとえば、MP3プレーヤでは属性情報として各MP3プレーヤの名称、製造元、値段、PCとの接続方法、タイプ(HDD/半導体メモリ)、連続再生時間、その他の機能、再生フォーマット、電源の種類、重量をショッピングサイトから取得する。本システムは、これらの情報を組み合わせて類似品を検索することができる。

| 商品名 | メーカー | 値段 | 接続 | タイプ | 連続再生時間 | その他機能 | フォーマット | 電源 | 重さ |
|----------------------|-------|-------|-----------------|---------------|--------|-----------|------------------|----|------|
| iPod MA079J/A (20GB) | APPLE | 29400 | IEEE1394/USB2.0 | HDDタイプ (20GB) | 15時間 | ストレージデバイス | MP3,AAC,WAV,AIFF | 内蔵 | 166g |

この製品の類似品を検索します。

メーカー
 値段 ±0 円以内
 接続
 タイプ
 連続再生時間 ±0 時間以内

その他機能
 フォーマット
 電源
 重さ ±0 g以内

類似品検索

図4 類似品検索画面

図4に iPod の類似品を検索する際の画面を示す。ここでチェックを入れた属性が同じまたは指定した範囲内である商品を iPod の類似品一覧として表示する。それらの類似品についても blog 検索ができるので、ユーザは検索対象との比較が容易にできる。

3 類似品情報を用いた選択支援

3.1 概要

前述のように、類似品比較のための商品の仕様の取得にはショッピングサイトを利用している。しかし、これらの仕様以外の特徴を類似品選択に利用したい場合がある。たとえばデザインや操作性などである。これらは、実際に利用した人の意見から取得できる情報である。これらの情報は人々の生の意見が書かれやすい blog 記事から取得する。このように商品の仕様として表しにくい情報も利用することにより、より効率的な類似品の選択ができる。

ここではユーザがある商品を検索したときに blog 記事から類似品の情報を得て、類似品選択に利用することを述べる。

3.2 類似品情報利用までの流れ

図5に類似品情報を取得し、類似品選択に利用するまでの流れを示す。

まず、ユーザが検索キーワードとして MP3 プレーヤである「商品A」と入力したとする。商品AがMP3プレーヤに属するので(図5①)、システムはMP3プレーヤに属する商品名全てを blog 検索し、記事を収集する(図5②,③)。次に、集めた記事から「デザインが良い」や「画面がきれい」などの商品に関する情報を取得する。ここで「デザイン」や

「画面」などの評価の対象を属性とし、「良い」や「きれい」という評価自体を属性値とする。最後に、これら取得した属性と属性値含めた商品Aの類似品一覧を図4に追加する形でユーザに提示する(図5④)。

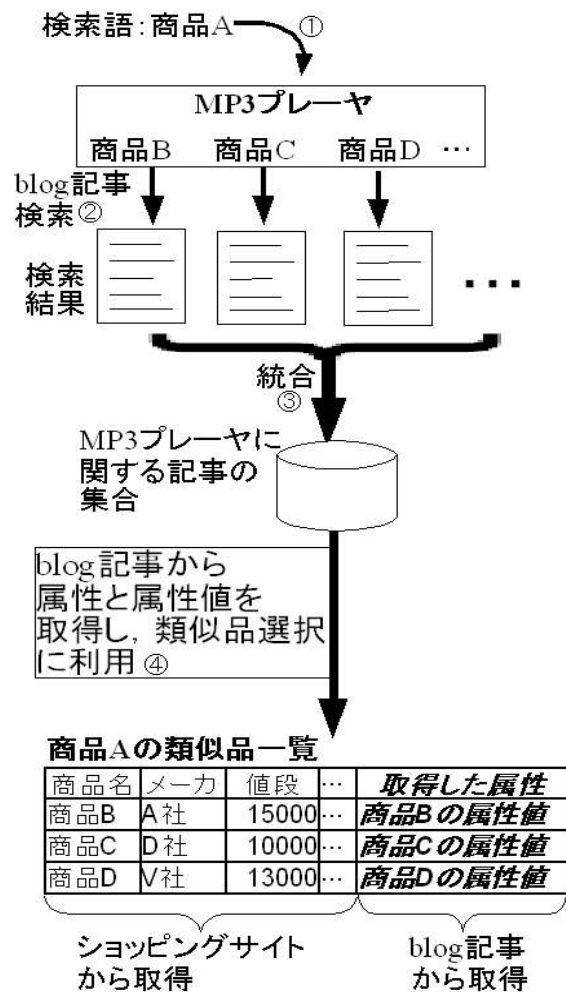


図5 類似品情報を利用するまでの流れ

4 予備実験

blog 記事から属性を抽出するための予備実験として、ノートPCに関する電子掲示板の記事から属性の抽出を行った。

4.1 準備

実験で用いるデータセットは多くの評判情報を含む掲示板として kakaku.com の口コミ掲示板を利用した。記事はノートPCについての掲示板から 46,773 文を収集した。これらの文を CaboCha[7]により係り受け解析し、非評価文除去のため品詞によるフィルタリングを行った。品詞は評価表現によく使用される形容詞と名詞-形容動詞互換、直前の単語の賛否を反転させる助動詞 特殊・ナイと助動詞 特殊の 4 種以外を取り除いた。この処理により 30,385 文を得た。

4.2 属性の抽出方法

前処理によって得られた 30,385 文の中から名詞-形容詞の組み合わせを抽出し、その中で出現頻度の高い名詞を属性とした。また、属性に選ぶ名詞は一般名詞とサ変接続の名詞とした。その結果を表 1 に示す。この表から「ファン」や「キーボード」など仕様書には記述されない属性についても、記事から取得できることがわかった。一方、「気」や「自分」など属性とは関係の無い名詞も属性として取得してしまう問題点がある。これらは属性値情報などを用いて除去することを考えている。

表 1 ノートPCについての属性

| 属性 | 頻度 | 属性 | 頻度 |
|-------|-----|-------|----|
| 液晶 | 146 | 他 | 46 |
| ファン | 109 | 運 | 45 |
| 画面 | 108 | 性能 | 43 |
| キーボード | 98 | ノート | 38 |
| 音 | 87 | 値段 | 37 |
| 気 | 79 | バッテリー | 34 |
| 機種 | 73 | キータッチ | 33 |
| 自分 | 64 | 文字 | 32 |
| 調子 | 52 | 起動 | 32 |
| パソコン | 51 | 購入 | 31 |
| 本体 | 50 | メーカー | 30 |
| 価格 | 49 | 最初 | 30 |
| メモリ | 46 | デザイン | 30 |

5 まとめ

本稿では、我々が開発している商品の選定支援システムについて説明し、特に類似品との比較をする機能について述べた。予備実験により、記事の係り受け解析により仕様書には含まれない属性が取得できることがわかった。これらの情報を本システムに利用することにより、効率的な商品の選定支援を目指す。

今後は実際に blog 記事から属性を抽出し、本稿で提案した方式の評価を行う予定である。

参考文献

- [1]Myblog japan <http://www.myblog.jp/>
- [2]blogWatcher <http://blogwatcher.pi.titech.ac.jp/>
- [3]藤村 滋, 豊田 正史, 喜連川 優, “電子掲示板からの評価表現および評判情報の抽出”, 人工知能学会第 18 回全国大会, 3F1-03 (2004)
- [4]山名 健悟, 滝沢 敏裕, 湯浅 将英, 大山 実, “Blog 記事を用いた選定支援システム”, 情報処理学会第 67 回全国大会, 2U-4 (2005)
- [5]馬場 肇, “Namazu システムの構築と活用”, ソフトバンクパブリッシング (2001)
- [6]西村 圭亮, 湯浅 将英, 大山 実, “評判情報利用によるネットオークションの商品選定支援システムの提案”, 第 4 回情報科学技術フォーラム, D-031 (2005)
- [7]工藤 拓, 松本 裕治, “チャンキングの段階適用による係り受け解析”, 情報処理学会論文誌 43, 6, pp.1834-1842 (2002)