

Blog 記事を用いた選定支援システム

山名 健悟[†] 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[‡]

1 はじめに

インターネットの発達にともなって、多くの人がウェブサイトを開設し情報を発信するようになった。その中でも blog と呼ばれる日記型のサイトを作るサービスが広まってきている。blog サービスはブラウザに表示されたフォームに文章を入力するだけで、HTML ページを更新できるという特徴を持つ。このように更新方法が簡単なので blog サイトは通常の web サイトよりも更新頻度が増える。この blog に書かれている情報を活用しようという試みが近年増加しており、blog を対象とした検索サービスが登場している[1][2][3]。これらでは独自に収集した blog 記事から興味のある記事を検索することができる。また、blog 記事には人々の意見がリアルタイムに反映されることが多い。したがって、blog 記事からは人々の生の評判情報が取得できると考えられる。

興味対象の評判をわかりやすく知るためには、対象の長所と短所を知ることや時系列における評判の移り変わりを知ることが重要である。ユーザの興味対象に関する評判情報を提供することによって、ユーザの選定を補助することができる。

そこで、本システムでは blog サイトから興味対象に関する記事を抽出する。抽出した記事から対象の長所と短所を明確化することや評判情報の時間的推移を知ることができる。これらを提供することによってユーザの興味対象の選定を支援できる。

2 本システムの機能

以下の機能を持つシステムを開発している。

2.1 blog 記事の収集

本システムではまず blog 記事を収集し、キャッシュとして保存する。blog の更新情報を集めているサイトで新着 blog 記事の URL が提供されているので、そこからリンク先の記事 HTML を収集する。また、有名 blog サービスのサイトで提供されている新着 blog の URL から記事を収集する。図 1 にその流れを示す。blog サイトが更新されると、blog 記事収集サイトに更新が通知される。blog 記事収集サイトはその情報を基に新着記事の URL リストを作成する。次に、本システムがそのリストを取得する。本システムはリストに従って新着記事の HTML を取得し、保存する。

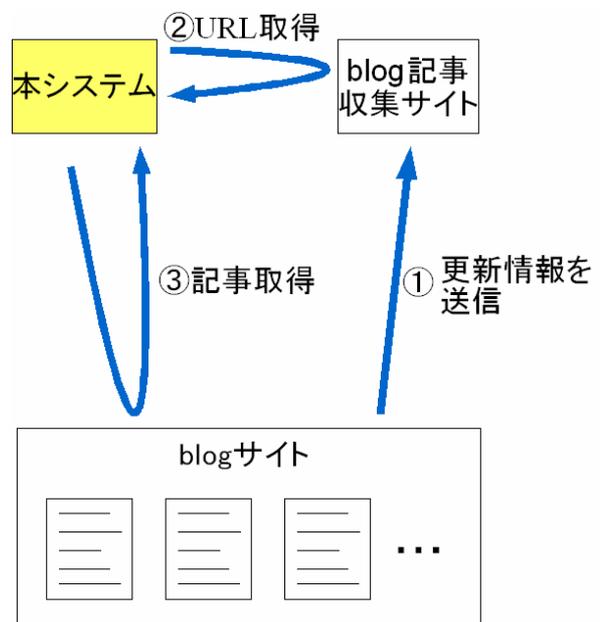


図 1 blog 記事保存の流れ

2.2 検索

ユーザが興味対象の記事を検索するときには保存された記事からの全文検索を行う。検索には全文検索システムの Namazu[4]を利用する。

A Supporting System to Select a Target by Blog Articles

Kengo YAMANA[†] Toshihiro TAKIZAWA[‡]

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

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

[‡]School of Information Environment, Tokyo Denki University

2.3 評価表現の抽出

検索に加えて評価表現を抽出する。検索された記事の中にある評価表現を抽出し、肯定否定で分ける。肯定表現と否定表現を別々に数えることで、その記事が検索対象に関して肯定的なのか否定的なのか調べる。現在は評価表現の辞書を手動で作成し、それを用いて文字列一致により評価表現を抽出している。

2.4 検索結果の表示

検索結果は以下の 3 通りの方法でユーザに提示する。

2.4.1 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、ユーザは的確な判断ができる。「PSP」で検索した場合の結果表示の例を図 2 に示す。

good	bad
<p>1. はてなダイアリー - 超兄記。(w) Date: Sun, 12 Dec 2004 21:45:25 肯定表現は17個, 否定表現は3個あります。</p>	<p>1. ブログdeショッピング: PSPがようやく戻ってきましたよ Date: Mon, 27 Dec 2004 15:45:59 肯定表現は10個, 否定表現は2個あります。</p>
<p>2. インターネットの本屋さん別館 PSP 予約受付情報 Date: Sat, 25 Dec 2004 18:13:14 肯定表現は3個, 否定表現は10個あります。</p>	<p>2. ブログdeショッピング: PSP持っていない人 手に入るぞ Date: Mon, 13 Dec 2004 10:55:30 肯定表現は10個, 否定表現は1個あります。</p>
<p>3. 戦国風丁髷【せんごくふうちゃんまげ】: NDS vs PSP Date: Wed, 05 Jan 2005 23:54:41 肯定表現は2個, 否定表現は10個あります。</p>	<p>3. ブログdeショッピング: PSPまだこないな～ Date: Tue, 21 Dec 2004 19:51:02 肯定表現は10個, 否定表現は1個あります。</p>

肯定的な記事

否定的な記事

図 2 肯定/否定で記事を分けて表示

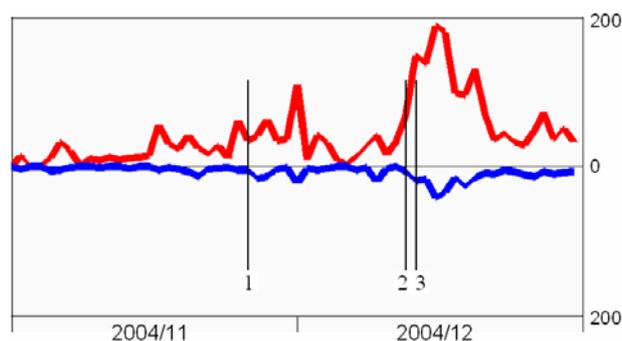
2.4.2 時系列表示とニュースとのリンク

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とで、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。

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。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。



News: 1 2004/11/25 PSPでオークション
2 2004/12/11 いよいよ明日発売
3 2004/12/12 PSP発売

図 3 グラフ表示のイメージ

2.4.3 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とで、ユーザは類似品との比較が容易にできる。類似品情報の検索にはショッピングのサイトや掲示板のスレッドのトピック名やタイトルを用いる。

3 まとめ

本論文では、ユーザの興味対象に関する評判情報を効率的に収集でき、対象を選定する際の判断材料を増やすことができるシステムを提案した。

現在までに本システムには約 45 万件の blog 記事が集まっている。今後は実際に使用することによりシステムを評価していく。

参考文献

- [1] 未来検索 livedoor <http://sf.livedoor.com/>
- [2] Bulkfeeds <http://bulkfeeds.net/>
- [3] blogWatcher <http://blogwatcher.pi.titech.ac.jp/>
- [4] 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.