



情報検索サービスを利用したWebコンテンツの印象 評価

メタデータ	言語: jpn 出版者: 室蘭工業大学SVBL 公開日: 2010-07-20 キーワード (Ja): キーワード (En): 作成者: 鈴木, 育男, 平尾, 光, 久保, 洋 メールアドレス: 所属:
URL	http://hdl.handle.net/10258/490

情報検索サービスを利用したWebコンテンツの印象評価

鈴木育男¹⁾, 平尾 光(B4)²⁾, 久保 洋²⁾

1) 室蘭工業大学 サテライトベンチャービジネスラボラトリー,

2) 室蘭工業大学 工学部 情報工学科

1. はじめに

インターネット (Web) を情報源とするニュース (情報) は, 普段の生活にとって欠かせないものとなっている. 日々のニュースやブログの記事の閲覧から仮想店舗での Web ショッピングなどの商取引まで, Web の利用は多岐にわたっている. このような Web 上の情報は, 追加・更新されることにより, その情報の量は日々増加している. それにより, Web 上に構築される情報空間が巨大になり, リンクを辿る事による情報獲得には, 多くのコスト (労力, 時間) を必要とするようになってきている. そのため, 現在では Yahoo や Google といった情報検索サービスが Web 上での重要な役割を担っている.

これらの情報検索サービスでは, 検索のための入力文字 (検索クエリ) の選択が, 検索効率に大きな影響を与える. つまり, 検索サービスを使用するユーザが, 適切な検索文字を入力しないと目的の情報が入手できないということを意味している. これでは, 各ユーザが望むような情報を獲得するには, まだまだ多くのコストを必要とする問題の解決には至らないことになる.

本研究の目的は, ユーザごとに適切な情報を呼び出す (推薦する) ための情報提供サービスシステムの研究開発である. 本稿では, ユーザ自身に適した情報を発見するための手段である情報フィルタリングの一手段として, Web 上に存在する情報 (検索結果) を利用した Web コンテンツの印象評価の検証実験を行い, 提案手法の情報フィルタリング能力について報告する.

2. 提案手法

2.1 情報フィルタリング

現在, 我々が情報を検索する際には, 情報検索支援サービスを用いることが一般的である. これらの情報検索サービスでは, 事前に Web 上の各ページを巡回・解析して分類しておくことで, ユーザが問い合わせた検索クエリに対する結果を提示する. ただし, 検索結果を解析順に提示するだけでは, 必要とする情報を獲得するまでに時間コストが大きくなるので, 各検索サービス企業ごとに結果出力順を調整している. このような情報選別に用いられる技術を総称して, 「情報フィルタリング」と呼んでいる.

情報検索サービスでは, コンテンツ内容に基づくフィルタリングが行われており, ページ内容やリンク内容な

どといった Web ページ作成者側の興味に関して解析し分類を行う. Google の検索結果に用いられるページランク技術もこれに分類されると考えられる. また, あるユーザと興味や関心が類似する別のユーザが以前に利用・評価した情報から, 提供すべき情報を判断する手法を協調フィルタリングと呼ぶ.

ユーザ個人に特化した情報提示を考えた場合には, ユーザ自身の情報 (好みや性格など) を把握すると同時に, 情報フィルタリングを通じて情報源の分類もしっかり行うことで適切な情報を提供する確率を上げることが可能となる.

2.2 検索情報による Web コンテンツ評価の提案

我々は, 各ユーザに適した情報を提供するために, ユーザ自身の情報を利用すると同時に, Web 上に存在する情報をフィルタリングして利用することを考えている. それは, あるコンテンツに対してユーザ自身の意見 (主観的データ) と Web 上の他人の意見 (客観的データ) の 2つのデータを利用することである. 一般に, 情報フィルタリングの処理に関しては, 初期段階において情報データ数の少なさから有効な情報が発見できないという問題が発生する. この問題は, 「コールドスタート問題」と呼ばれているが, Web 上に既にある情報を利用することによって, このコールドスタート問題を回避できると考えられる.

新たに, Web 上の情報を収集・解析して情報フィルタリングするには, 多くのコストが要求される. そこで, 情報検索サービス企業が提供する検索結果情報を二次使用する手法を提案する. 具体的には, コンテンツ中の文章から出現頻度の高い単語を抜き出し, その抽出単語と用意した印象評価語 (楽しい・悲しい) とを「AND 検索」によって検索し, その検索結果から Web コンテンツに対する印象評価を決定する. この手法により, Web にアクセスする世間一般のユーザが任意のコンテンツに対してどのような印象を持っているかに関する間接的な統計データとして利用できる.

本稿では, 提案手法による印象評価とユーザによるアンケートによる印象評価との比較を行い, 提供される検索結果を利用した Web コンテンツ (ブログやページ) に対する印象評価の可能性について検証実験を行い, 提案手法の有効性を調査した.

3. 検証実験

3.1 実験手順

提案手法に対する検証実験は、以下の手順で実行した。

1). サンプル文章の収集

今回の検証実験において、サンプルとなる Web コンテンツの文章は、文字数が 250~500 程度の「喜・怒・哀・楽・特に意味なし」と思われる文章を用意した。「喜」に関するサンプル文章を図 1 に示す。

全国のサッカー担当記者の投票による 2006 年度フットボラー・ザ・イヤー(年間最優秀選手)に 7 日、J1・浦和の日本代表 DF 田中マルクス闘莉王(25)が選ばれた。浦和からの選出は初めて。闘莉王は有効票 205 票のうち 77 票を集めた。2 位は中村俊介(スコットランド・セルティック)の 34 票で、3 位はワシントン(浦和)の 32 票だった。

図 1: 「喜」に関するサンプル文章

2). アンケート調査による印象評価

サンプル文章を被験者に読んでもらい、それに対する印象を 5 段階の SD 法によりアンケート調査した。アンケート項目は、「楽しい⇔悲しい」のような感情モデル対と「丸っぽい⇔四角っぽい」のような印象対である。

3). 検索情報による印象評価

前処理として、サンプル文章中の単語について出現頻度を調査する。その後、上位 3 つの単語と感情・印象語との AND 検索を実施し、過半数を取った評価対から文章全体の印象を評価した。印象評価指数 (印象度) は、次式で算出した。

$$\frac{\text{AND検索でのヒット数}}{\text{抽出単語のヒット数}} \times 100 \quad (3.1)$$

4). アンケート結果と提案手法結果との比較

最後に、両調査の印象評価を比較する。ただし、アンケート調査においては、「どちらとも思えない」が 100% のとき、両評価対が同率であった場合は、印象決定不能として比較対象から除外した。

3.2 検証結果

前小節の手順で検証実験を行った。表 1 に、先に示した「喜」のサンプル文章で出現頻度が 1 位だった「票」という単語の検索情報による印象評価の結果を示す。

表 1: 「票」に関する印象評価の結果

評価項目 1	印象度 1	印象度 2	評価項目 2
楽しい	1.62	4.74	悲しい
怒り	5.92	6.07	恐れ
期待	0.87	4.18	驚き
嫌悪	8.80	4.04	受容
大人っぽい	1.33	1.08	子供っぽい
丸っぽい	2.48	8.03	四角っぽい

「楽しい⇔悲しい」の評価項目では、「悲しい」の印象度が高いことから「票」を含むこの文章は「悲しい」の印象を与えると分類される。今回用意した印象評価項目対は、11 対である。各印象評価項目対に対する調査を用意した 5 つのサンプル文章について実施し、アンケート調査からの印象評価と比較した結果を表 2 に示す。

表 2: 検索情報を利用した印象評価の結果

	比較可能個数	一致数	不一致数
サンプル 1	8	6	2
サンプル 2	9	7	2
サンプル 3	6	1	5
サンプル 4	4	2	2
サンプル 5	7	4	3
合計	34	20	14

この結果から、実際の印象評価 (アンケート調査) と検索情報を利用した印象評価との一致率は、59% であった。これは、決して高い結果ではない。このような結果になった理由として、検索対象の単語の抽出に出現頻度のみを用いたこと、検索単語の係り受けを考慮に入れていないことなどが考えられる。今後は、これらの課題を改善し、一致率を上げていく必要がある。

4. おわりに

本研究では、ユーザに適した情報の提供を実現するために、Web 上の情報検索結果を利用した情報フィルタリングの手法を提案した。今後は、検証実験で得られた課題の解決に向けた手法の改善を行っていくとともに、検索サービス企業が提供する Web API を利用してシステムの自動化などに取り組んでいく予定である。

◇参考文献◇

- [1] 村上晴美: Web ブラウジング履歴からの自己の興味の理解, 日本心理学会第 65 回大会発表論文集, p. 9, (2001).
- [2] 三浦典之, 山名早人: ユーザの検索履歴を用いた情報検索システムの提案, 第 65 回情報処理学会全国大会, pp. 3-115~3-116, (2003).
- [3] 鈴木泰裕, 高村大也, 奥村学: Weblog を対象とした評価表現抽出, 人工知能学会, セマンティックウェブとオントロジー研究会, SIG-SW0-A401-02, (2004).
- [4] 小原恭介, 山田剛一, 絹川博之, 中川裕志: Blogger の嗜好を利用した協調フィルタリングによる Web 情報推薦システム, 第 19 回人工知能学会全国大会, 2C2-02, (2005).
- [5] 大鹿広憲, 佐藤学, 山名早人: Google を活用した英作文支援システムの構築, DEWS2005, (2005).