

評判情報検索を用いた商品評価の分析

安江 和洋 都築 直希 間瀬 智史 濱川 礼

中京大学 情報科学部 情報科学科

1 はじめに

携帯電話を使い、ネットを通じて商品に対しての評価・相場を分析しユーザが手軽に入手できるシステムを開発した。2006年3月末に総務省より発表されたBlog利用者数の資料[1]によると現在900万人を突破し、今後も増え続けるといわれている。既存システム[4]として、PC用に検索対象をまとめるシステムは存在するが、外出先での利用を考慮されていない。そこで外出先で一番利用される携帯電話向けにシステムを開発した。

2 研究アプローチ

本研究では以下の点に着目した。

検索対象 より多くの情報を引き出すにはより多くの検索対象となるサイト情報を持っている検索エンジンが必要である。本研究ではGoogleの検索エンジンを利用[2]することにより、サイト・ブログ両方を対象とした検索が可能である。

情報取得 現状では検索の結果を上位から回覧することによって情報を得ていた。しかし、外出先から情報を得たい場合はスピードを求められることが多く、上位から回覧することは時間が掛かる。そこで本文を抽出・評判を解析することによって必要な情報のみを閲覧することができるようにする。

連想検索 閲覧中のページをキーとして検索をすることによって関連性の強い類似ページや類似商品を検索することが可能。また特徴語抽出を用いる事によって閲覧中のページの特徴も表示することができる。

以上より本システムでは検索結果の中から本文を抽出し、日本語処理によって必要な評判情報を取得、さらにユーザが指定したページを元に新しく検索語を調べ再検索するシステムを開発した。

3 提案手法

3.1 本文抽出

Googleの検索エンジンを利用して得られたURLのウェブページからソースを参照・取得し、情報の抽出を行い易いように文章の処理をする。正規表現により、ソース内部に多数存在するタグとマッチングを行っていき、本文の中から不要だと思われる部分の削除、重要だと思われる部分を強調する処理を行う。

3.2 評判情報抽出

サイトの情報を素早く入手するには指標となる物が必要である。そこでサイト内の評判を文章解析・日本語処理によって評価を抜き取りユーザに示すことによって指標とする。評価用辞書として良い・悪い評価になる言葉とその品詞名を記録した辞書を用意した。評価の重み付けとして動詞、形容詞、名詞…の順に優先を低くし、

評価

$$= (\text{良い評価になる単語}) \frac{(\text{動詞の重み} \times \text{出現数}) + (\text{他品詞の重み} \times \text{出現数}) \dots}{\text{良い評価用語の出現数}} \\ - (\text{悪い評価になる単語}) \frac{(\text{動詞の重み} \times \text{出現数}) + (\text{他品詞の重み} \times \text{出現数}) \dots}{\text{悪い評価用語の出現数}}$$

とする。

3.3 特徴語抽出

想検索を実装するために特徴語抽出を行う。そのためには様々な手法があるが、今回、「ある文書の特徴語は、その文書中での出現頻度が高い」ということを第一に考えた。また、それとは別に、文書の中で作成者が最も述べたい部分を示す場合、どのように示すかということも考えた。そこで、以下2つの手法により示すのが一般的であろうと考えた。

- Webページのタイトルに含める
- 強調表示を行う

上記2つの手法がなされている部分は、特徴語である可能性が極めて高い。さらに、強調表示されている部分がすべて名詞の場合、複合名詞として抽出した。これらを加味して特徴語の重み付けを行った。

以上の提案手法を用いてシステム開発するに当

「Analysis of commodity evaluation using reputation information retrieval」
Department of information science, Chukyo-University
Kazuhiro YASUE, Naoki TSUZUKI, Satoshi MAZE, and Rei HAMAKAWA

たり、各手法でより効率的に開発・動作させるために開発言語が違ふ。そこで各機能を CGI とサーバレットにまとめ、サーバで動作し連携が取れるようにした。

4 システム概要

本システムは、提案した手法を用いて評判情報検索システムを構築する。本システムの構成図を図1に示し、流れを説明する。

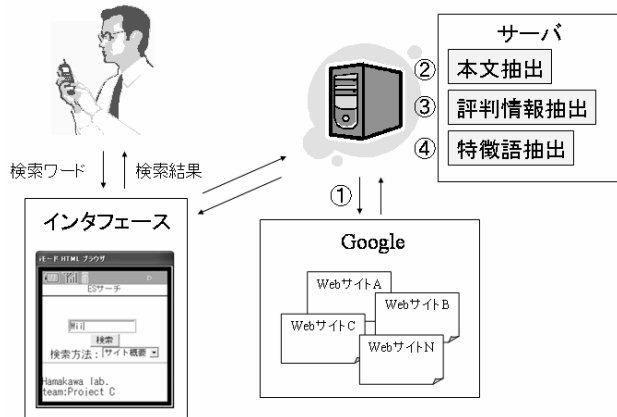


図1 システムの概要図

- i ユーザより入力された検索ワードを Google の検索エンジンを使用して検索をする。
- ii Google より得られた検索結果の本文を URL から参照し本文抽出を行い、不要なタグなどを取り除く。
- iii 得られた文章より日本語処理を行い、評判がどのようになっているかを評価する。さらに検索語より製品の正式名称・新品値段・公式サイトをユーザに表示させる。
- iv ユーザより指定したページをキーとして再検索を行うように指示された場合、文章内より特徴語を見つけ出し、再検索を行う。候補が複数ある場合はユーザが指定できるようにする。

5 評価実験と今後の課題

5.1 評価実験と結果

実際にシステムを稼動し評価実験を行った。評価実験に用いたキーワードは「Wii」[3]である。その結果を図2に示す。

ページ閲覧を繰り返すたびに多くの評価情報を取得する。その結果を表1に示す。さらにユーザの評価とシステムの評価の違いを検証した(表2)その結果より検索結果やその評判抽出等はユーザが思った印象に近い結果が出ているが、逆に要約や特徴語といった項目は評価が悪い。原因として特徴語を抽出する際に本来であれば不要な語まで対象に含めてしまっている点がある。

5.2 今後の課題

先ほどの結果より本文中に辞書とマッチング

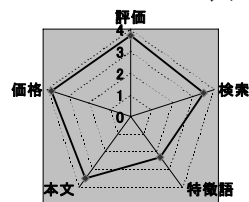
する評価用語が記入されていれば正しく評価情報を取得することができた。しかし、評価に関係の無い部分が対象になってしまいノイズとなるので、それを改善することによりさらなる精度向上を図っていく必要がある。



図2 システム実行

閲覧件数	評価
50件	良い: 5 悪い: 0 評価無し: 45
200件	良い: 21 悪い: 4 評価無し: 175

表1 検索結果



各項目を5段階にて評価

※100回検証した値の平均値

表2 各機能の評価

6 まとめ

本論文では提案手法を用いることによってネット上にある情報を手軽に入手できるシステムの開発方法を論じた。本システムを実際に運用することにより外出先などで携帯電話だけの環境でもより素早く欲しい情報を入手することが図2で実証できた。今後は現段階で把握している問題点を分析していきシステムの改善を図る参考文献

- [1] ブログ及びSNSの登録者数(総務省)
http://www.soumu.go.jp/s-news/2006/060413_2.html
 - [2] "Google SOAP Search API"
<http://code.google.com/apis.html>
 - [3] Nintendo Wii
<http://www.nintendo.co.jp/wii/>
- 関連技術
[4] 東京工業大学奥村研究室: BlogWatcher