

オントロジーを用いた Web からの評判情報抽出サービス

Ontology-based Reputation Extraction Service from the Web

川村 隆浩 長野 伸一 稲葉 真純 溝口 祐美子 長 健太
 Takahiro Kawamura Shinich Nagano Masumi Inaba Yumiko Mizoguchi Cho Kenta

(株) 東芝 研究開発センター

Corporate Research & Development Center, Toshiba Corp.

In this paper, we introduce a web-based service that extracts reputations of a product from the internet. If a user inputs the product name, the service first collects articles reviewing the product from weblogs, bbs, and so on. Also, it analyzes their contents using metadata and ontologies with conventional NLP techniques. Then, it indicates the reputations (positive or negative) from the overall and several pre-defined aspects. Further, other related products that are the subject of much discussion in the articles. This paper illustrates two technical points regarding use of metadata and ontologies, and summarizes an evaluation in a case that we applied it to a market research for a vehicle.

1. はじめに

近年、ネット上のクチコミに注目が集まっている。理由の1つは、商品のレビューをユーザの視点でブログなどに公開する人が増え、商品購入の際にそれらレビューを参考にする消費者が増加していることが挙げられる。現に、ネットユーザの購買行動にブログの書き込みが大きく関与していることを示すデータもある [1]。一方で、製品を製造または販売する企業がクチコミをマーケティングに活用し、製品の改良や広告戦略の参考にしている例も多い。更に、クチコミを収集・検索するサイトへは新たなターゲット型・検索連動型広告の掲載先として熱い視線が注がれている。そのため、クチコミの自動的な収集、検索に関して現在、盛んに研究が進められている [2]。

本論では、これらクチコミ情報の発信元が主にブログなどの半構造化データであることに着目し、従来の構文解析などの自然言語処理に構造化された情報から得られるセマンティクス(メタデータとオントロジー)を組み合わせたことを特徴とするクチコミ情報抽出技術を紹介する。

2. ユビ de コミミハサンダーとは

ユビ de コミミハサンダーとは、商品名を入力するとインターネットから商品のメタデータを取得し、関連するブログを収集し、その商品に関するクチコミ情報をユーザに提示するサービスである。抽出結果は、評判の良し悪しを棒グラフとレーダーチャートで提示するポジティブ・ネガティブ(PN)判定、関連する商品で話題となっているものを提示する関連トピック抽出、および有用と思われる順にソートされ、意見箇所がハイライトされたブログ本文で構成される。システム構成図と出力例を図 1, 2 に示す。

本システムの技術的な特徴は以下の 2 点である。1 つ目は、ブログエントリ(ブログ内の個々の記事)間の相関関係を利用し、各エントリの重要さを重み付けし、評価をバイアスしている点である。従来の研究では、互いに独立した文書の集まり(新聞など)を対象としている場合が多く、ユーザが Web 上のリンクを辿りながらさまざまな意見に出くわす状況を想定していない。しかし、多くのトラックバックによって賛意が付けられているブログと、1 つもトラックバックがついていないも

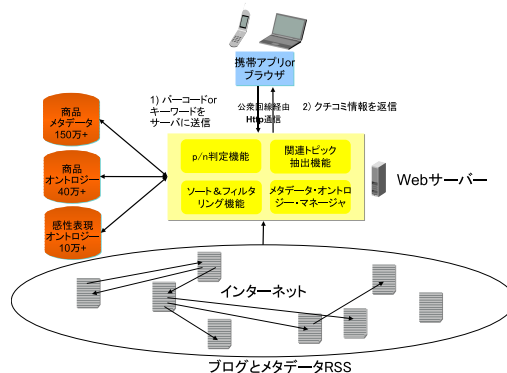


図 1: システム構成

のとではユーザの心理に与える影響は異なるだろう。そこで、ブログのメタデータ RSS(RDF Site Summary) からエントリ単位でのトラックバックやコメントの数、著者毎の平均的なトラックバックやコメントの数、当該著者が他の関連商品についても意見を述べているかどうか、エントリとトラックバック、コメントとの時間的な開き、などを基にルールを設定し、各意見に重み付けを行っている。

2 つ目の特徴は、構文解析の過程でオントロジーを活用している点にある。図 3 に商品オントロジーと感性表現オントロジーの構成を示す。構文解析時には商品オントロジー内の対象クラスまたはインスタンスのプロパティ名や値を参照し、係り受け関係を抽出する。これにより、例えば DVD タイトルに関する評判を調べる場合には、タイトル名に加えて監督や出演者に関して述べている箇所も効率よく抽出できる。そして、商品クラスに関連付けられた感性表現オントロジー内の属性クラス(インスタンス)、および表現クラス(インスタンス)を参照することで抽出された箇所が、何について(属性)肯定しているのか否定しているのか(表現)を判定する。ここでは、属性と表現が対応づけられているため、属性によって意味が逆転する表現(例えば価格が高いと性能が高い)にも対応できる。

オントロジーの他の使い方としては、ブログをフィルタリングする際に商品クラスの上位クラスや下位クラスを検索ワードに追加することで、同名の別ジャンルの商品と区別することがで

連絡先: 〒 212-8582 川崎市幸区小向東芝町 1, e-mail: takahiron.kawamura@toshiba.co.jp

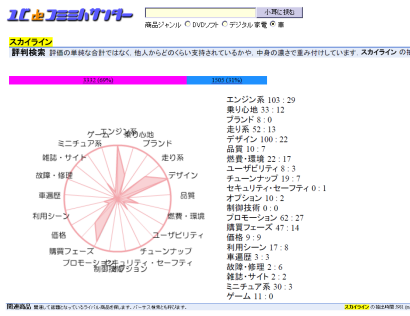


図 2: 出力例

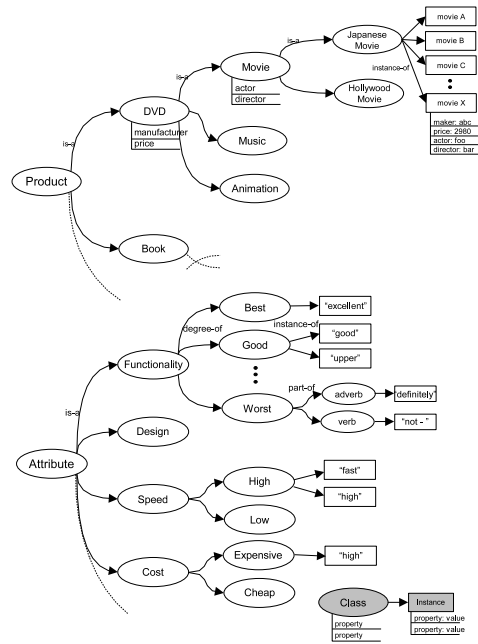


図 3: 商品オントロジーと感性表現オントロジー

表 1: 実験方法

評価対象範囲	センテンス毎	スレッド毎
PN 判定機能の出力	PN 有無の判定	バイアスを加えた PN スコアと比率
人手による作業内容	人手による PN 有無の判定 (各センテンスに対する正解)	スレッドの与える全体印象 (7 段階)
評価内容	PN 判定機能の再現率と適合率	バイアス付き PN スコアがどれだけ人の印象と近いかを検証

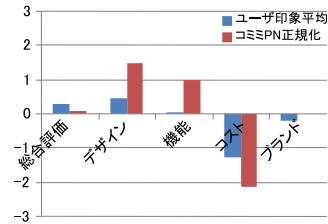


図 4: 読後の印象値とバイアス出力の比較

きる．例えば、東京タワーという作品とタワーそのものについて触れたブログを区別する際に使うことができる．

3. 評価実験

実験は、ある乗用車に関する 5 スレッド (全 1,548 センテンス) の書き込みを対象に、センテンス毎の PN 判定の結果を人手による PN 判断と比較する精度評価と、各スレッドを被験者 18 名に読んでもらい、その結果受けた印象値の平均をシステムの出力するバイアスされた結果と比較する印象評価の 2 種類を行った (表 1 参照)．

まず、精度評価に関しては再現率は 87.80%、適合率は 90.72% という結果が得られた．ここで、再現率は人が P/N と判断したセンテンスの内、いくつ拾うことができたかを表し、適合率は PN 判定機能が PN 有り と判断したセンテンスの内、いくつが実際に PN を含むかを表している．但し、精度はオントロジーの整備状況によって大きく左右される．そのため、今回実験した自動車分野以外では結果が変わってくるのが予想される．

次に、印象評価に関してあるスレッドにおける製品への印象値の比較を図 4 に示す．PN 判定機能の出力は -3 から 3 の間に正規化してある．また、印象値は図にある 5 つの観点 (総合、デザイン、機能、コスト、ブランド) 毎に比較した．5 スレッド × 5 観点の平均では、ユーザの印象値とシステムの出力との間に絶対値で 0.96 ポイントの開きがあることが分かつ

た．全体的にメタデータによるバイアスルールの効きが強く、システムの PN 判定結果が PN 両面においてユーザの印象値よりも大きく出ている．これは、今回の被験者 18 名の中に車を所有していない者が 12 名含まれており、当該分野に詳しくないことが PN いずれの面においても印象を弱めたためと思われる．

4. あとがき

本論では、メタデータとオントロジーを情報検索・抽出に活用する方法を提案し、簡単な実験を通して有効性の一端を示した．今後は更なる評価を進め、精度面と性能面の改良を図っていきたい．

参考文献

[1] 日経 BP, 掲示板やクチコミ・サイトを参考にする消費者は約 4 割, http://www.nikkeibp.co.jp/netmarketing/databox/nmdb/070126_kuchikomi/, 2007.

[2] 稲葉 真純, 長野 伸一, 長 健太, 溝口 祐美子, 川村 隆浩, CGM 分析技術の現状と課題-メタデータ, オントロジーの応用可能性について-, 人工知能学会セマンティック Web とオントロジー研究会, SIG-SWO-A603-06, 2007.