

# インターネットから評判情報を抽出する口コミ分析技術 ユビdeコミミハサンダー™

Analysis of Word-of-Mouth Information from Consumer-Generated Media on Internet

長野 伸一      溝口 祐美子      稲葉 真純

■ NAGANO Shinichi      ■ MIZOGUCHI Yumiko      ■ INABA Masumi

近年、商品のレビューをインターネット上に公開したり、他の人が書いたレビューを読んで商品を購入する消費者が増加し、インターネット上の口コミ情報の自動的な収集や分析が重要になっている。

東芝は、インターネット上のブログから商品の評判情報を抽出する口コミ分析技術“ユビdeコミミハサンダー™”を開発した。このユビdeコミミハサンダー™によって、指定した商品に関する消費者の評価意見及びその商品と比較評価されている話題商品を抽出し、単語の意味を文脈に沿って正確に判定するため、従来の検索技術と比べて消費者の評価表現を高い精度で抽出できる。

Active utilization of word-of-mouth information appearing on the Internet has been increasingly employed as a business approach in recent years. Toshiba has developed a word-of-mouth analysis engine for analyzing information from consumer-generated media (CGM) on the Internet. The main feature of this technology is its application of ontology to a text analysis procedure in addition to conventional natural language processing. This makes it possible to extract evaluation expressions related to consumer products from blog articles with high accuracy. In the process of applying the engine to our business activities, we have also developed support tools for ontology development and analysis tools for market research.

## 1 まえがき

近年のブログやソーシャルネットワークなどのCGM (Consumer Generated Media: 消費者生成メディア) のWebサービスの爆発的な普及もあって、商品のレビューをインターネット上に公開したり、他の人が書いたレビューを読んで商品を購入する消費者が増加し、インターネット上の“口コミ情報”が注目されている。この“レビューを書く→読む”のサイクルが、口コミ情報の質及び量を向上させ、消費者の購買行動に大きな影響を与えており、新しいメディアとしての価値が注目されている<sup>(1)</sup>。このように、商品を製造、販売する企業にとって、インターネット上の口コミ情報は無視できない存在になってきている。マーケティングリサーチや商品企画などのビジネスへの積極的な活用に向けて、口コミ情報の自動的な収集や分析に関する研究開発が盛んに行われている<sup>(2), (3)</sup>。

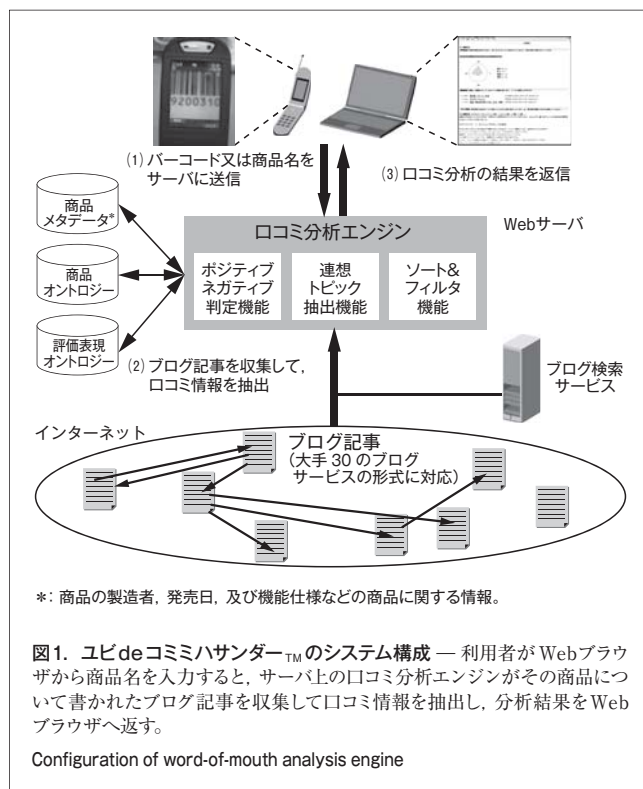
東芝は、消費者の口コミ情報への関心の高さに着目し、インターネット上のブログから評判情報を抽出する口コミ分析技術“ユビdeコミミハサンダー™”を開発した<sup>(4)</sup>。ここでは、ユビdeコミミハサンダー™の概要と特長について述べる。

## 2 ユビdeコミミハサンダー™の概要

ユビdeコミミハサンダー™は、インターネット上のブログから商品に関する評判情報を抽出する口コミ分析技術である。パソコン上のWebブラウザや携帯電話の専用クライアントアプリ

ケーションが、Webサーバ側の口コミ分析エンジンと連携して動作する。ユビdeコミミハサンダー™のシステム構成を図1に示す。

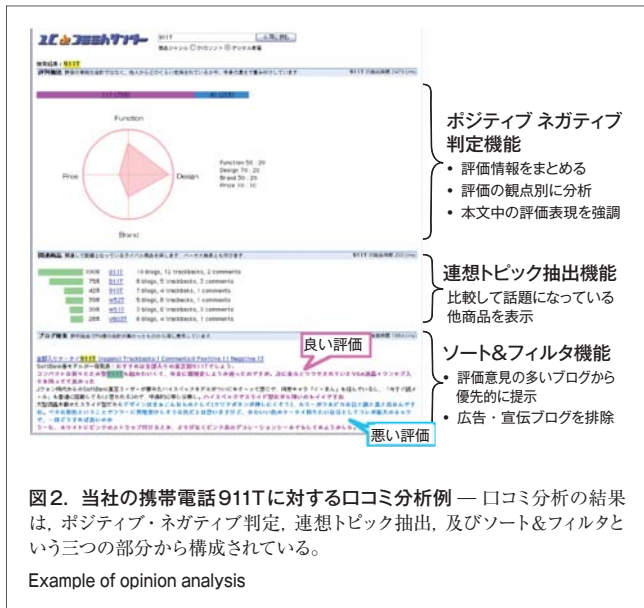
利用者がWebブラウザから商品名を入力すると、サーバ上



\*: 商品の製造者、発売日、及び機能仕様などの商品に関する情報。

の口コミ分析エンジンが、その商品について書かれたブログ記事をインターネット上から収集して口コミ情報を抽出し、分析結果をWebブラウザへ返す。携帯電話の場合、カメラで商品のバーコードをスキャンし、サーバ側でバーコードから商品名へ変換し、ブログの収集を行う。

ユビdeコミミハンダー™はポジティブ・ネガティブ判定、連想トピック抽出、及びソート&フィルタという三つの機能から構成される。当社製の携帯電話911Tに対する分析例を図2に示し、各機能の概要を以下に述べる。



- ポジティブ・ネガティブ判定機能** ブログ記事の本文から、指定商品に関する評価意見を含む文を抽出して、良い評価を表す文か、悪い評価を表す文かを判定し、評価意見の重要度スコアを算出する。図2では、良い評価と悪い評価のスコアの比が3対1となっている。また、レーダーチャートは四つの観点（機能 (Function)、デザイン (Design)、ブランド (Brand)、価格 (Price)) のそれぞれについて、良い評価と悪い評価のスコアの合計値を表示している。これにより、評判の全体傾向を把握できる。
- 連想トピック抽出機能** ブログ記事には、複数の商品を比較して意見を述べているものが多く存在する。この機能は、指定した商品と比較されている別商品のうち、出現頻度が高いものを抽出する。図2では、六つの比較商品が抽出されている。
- ソート&フィルタ機能** ブログ記事を、評価意見の重要度が高い順に並び替える機能、及び販売業者による営利目的や広告などのブログ記事を除外する機能を提供する。また、(1)で良い評価、悪い評価と判定された文に色を付けて、強調して出力する。

### 3 ユビdeコミミハンダー™の特長

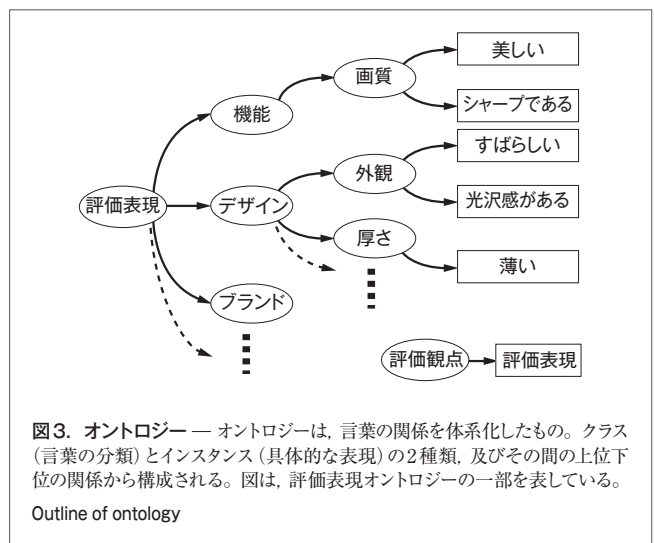
#### 3.1 文脈による意味の違いを判別

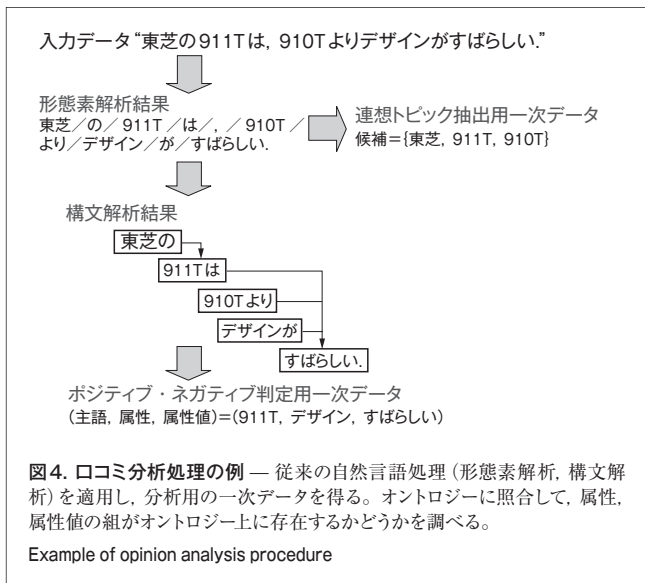
一般に、商品に対する評価表現、例えば“高い”という単語をキーワード検索すれば、口コミ情報を含むブログを簡単に発見できると思われるかもしれない。しかし、単語の意味は、使用される対象や文脈によって異なる。“高い”という単語が、デザインや性能に関して用いられる場合は良い評価を意味するが、価格に関して用いられる場合は一般に悪い評価を表す。また、“価格は決して高くはない”というように否定文で用いられる場合は、悪い評価ではなく、むしろ良い評価を意味することもある。このように、単に評価表現を機械的に検索して抽出するのではなく、その表現が出現する文における文脈を見て単語の意味を判定することが必要である。

この問題を解決するために、ユビdeコミミハンダー™では、従来の自然言語処理技術に加えて、オントロジーを導入している。オントロジーは辞書の一種で、ある分野で使われる語彙(ごい)、及び語彙間の関係を体系化したものである<sup>(5)</sup>。オントロジーを参照することで、言葉と言葉の関係性や同義の言葉を調べることができ、文脈に応じた意味の違いを正確に判別して、評価表現を高い精度で抽出することができる。更に、ブログ記事を評価の観点別に良い評価と悪い評価に分類することができる。

評価表現に関するオントロジーの例を図3に示す。だ円は評価表現の分類(評価観点)を表し、長方形は具体的な評判表現を表す。これは、携帯電話端末に関する良い評価を意味する表現を体系化したものの一部で、例えば“すばらしい”という表現は、デザインに関して使われる表現として定義されている。

ブログの文章を自然言語処理技術で解析すると、主語(商品名)、属性(評価観点)、属性値(評価表現)が抽出される。属性と属性値をオントロジーと照合することによって、対象商





品に対する評価が良い表現か悪い表現かを判定する。判定処理の例を図4に示す。ここでは, 入力文は携帯電話端末のデザインに関して良い評価をしていると判定している。

### 3.2 ブログ読者の感性に近い判定

一般のブログサービスには, トラックバックと呼ばれるブログ記事間に参照リンクを張る仕組みが備わっていて, このリンク機能により, 同じ商品に関心を持つ“同好の士”が集まりやすく, ブログ利用者間のコミュニティ形成の基本となっている。

従来のWeb検索エンジンは, ブログ記事中の単語の出現頻度やブログ記事に対する被リンク数に基づいて, ブログ記事の重要さを算出している。したがって, 商品に関する評価意見を含んだブログ記事も, 評価意見を含まないブログ記事も同じ程度に重要であると判定されてしまう。

ユビdeコミミハンダー™では, ブログ記事に書かれている評価意見の内容とブログ記事間のリンク関係の両方を利用して, ブログ記事の重要度を算出している。これにより, ブログ読者の印象に近い判定ができる。例えば被リンク数が同じであっても, 評価意見を含まないブログ記事は重要度を低く, 評価意見を含むブログ記事は重要性を高くしている。また, 良い評価を述べているブログ記事が多いなかで, 悪い評価を述べているブログ記事は, 著者の強い意志を尊重して重要度を高くしている<sup>(注1)</sup>。

## 4 オントロジー整備支援ツール

3.1節で述べたように, オントロジーはユビdeコミミハンダー™の主要な要素の一つであり, オントロジーの整備状況が

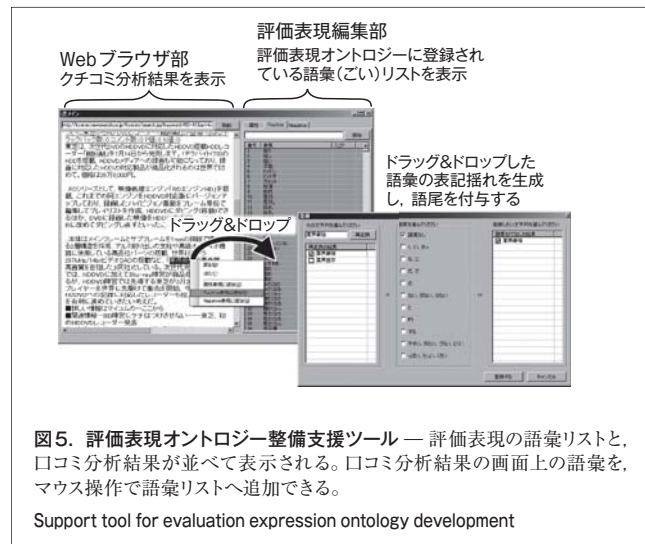
(注1) Web広告研究会の調査によれば, ブログでは7割が良いことを述べる傾向が見られる<sup>(6)</sup>。

口コミ分析の精度に大きく影響する。現在, 整備作業に掛かるコストを低減し, 品質の高いオントロジーを構築する技術や方法論の整備を進めている。ここでは, その取組みの一つである, 一般技術者向けの整備支援ツールについて述べる。

評価表現オントロジー整備支援ツールは, オントロジーについて深く精通していない一般の技術者によるオントロジー整備作業を支援するもので, 評価表現の編集機能と, 口コミ分析結果のレビュー機能とを備えている。単なる辞書の編集ツールではなく, 編集作業と分析結果のレビュー作業とをインタラクティブに行えることが特長である。

ツールの画面を図5に示す。図5の評価表現編集部は, 属性(評価観点)と属性値(良い評価表現及び悪い評価表現)をリスト表示する。評価表現オントロジーの階層構造はシステム内で固定しており, 整備作業者は階層構造を意識することなく編集できる。

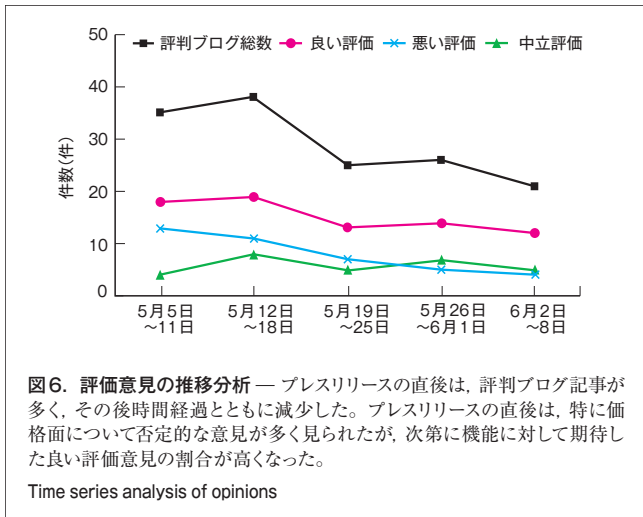
図5のWebブラウザ部は口コミ分析の実行画面で, 図2に示した分析結果情報が表示される。良い評価及び悪い評価と判定された文は色付きで表示されるため, 整備作業者は, 評価表現の登録漏れや登録誤りを容易に発見できる。また, Webブラウザ部上でマウスを使って表現を選択し, 評価表現編集部へドラッグ&ドロップして表現を追加できる。このとき, 語彙の表記揺れや語尾の候補を提示したウィンドウが開くため, 効率よく評価表現を整備することができる。



## 5 マーケティング分析への応用

### 5.1 評価意見の推移分析

ある商品に関するブログコミュニティにおける評判を, ユビdeコミミハンダー™を利用して定期的に分析することによって, 評価意見の推移を調べることができる。ある商品について, その商品の販売企業によるプレスリリース直後の1か



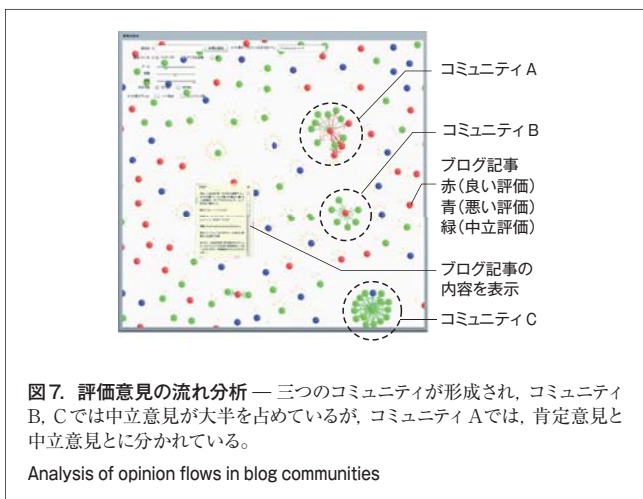
月間、週ごとに分析を行い、その結果を図6にまとめた。

ユビdeコミミハサンダー<sup>TM</sup>が、評価意見を含むと判定したブログ記事の数、及び評価別(良い、悪い、中立)のブログ記事の数の推移をグラフ化している。プレスリリースの直後は、評判ブログ記事が多く、その後、時間経過とともに減少している。分析結果を詳細に見てみると、プレスリリース直後は特に価格面について否定的な意見が多く見られたが、次第に機能に対して期待した良い評価意見の割合が高くなった。

このように、評価意見の件数の推移と、実際にブログ記事に書かれた消費者の生の声とを調べることで、例えば商品のプロモーション効果を測定することができる。

## 5.2 評価意見の流れ分析

ブログコミュニティでは、ブログ記事間の参照リンクが、ブログ利用者間のコミュニティ形成の基本であり、リンクを通して口コミ情報が広まっている。したがって、ブログ記事間の評価意見の伝搬を分析することによって、口コミで広まりつつある商品や、ブログコミュニティで影響力の強いブログ利用者(俗に言うオピニオンリーダー)を発見することができる。



ブログ記事間の参照関係を図7に視角化した。円はブログ記事を表し、ユビdeコミミハサンダー<sup>TM</sup>の分析による評価別(肯定、否定、中立)に色を変えている。図7では、点線の円が示すように、三つのコミュニティが形成されている。コミュニティB、Cでは中立意見が大半を占めているが、コミュニティAでは、肯定意見と中立意見とに分かれていることがわかる。また、いずれのコミュニティも、それぞれの中心に位置するブログ記事が他のブログ記事に影響を与えていることから、重要度の高い意見を含んでいて特に読む価値があると言える。

## 6 あとがき

ここでは、ユビdeコミミハサンダー<sup>TM</sup>の概要と技術的な特長について述べた。また、オントロジー整備作業を支援するツールを示すとともに、口コミ分析技術のマーケティング分析への応用事例について述べた。

ユビdeコミミハサンダー<sup>TM</sup>の 口コミ分析技術は、自動車メーカー向け 口コミ分析業務、及びKDDI(株)のポータルサイトau one ラボのWebサービスにおいて、事業適用された。

今後は、口コミ分析の精度向上に努めるとともに、オントロジーの構築や整備にかかわる作業を効率化するための技術や方法論の整備を進めていく。

## 文献

- 日経BP. “掲示板や口コミ・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\\_kuchikomii](http://www.nikkeibp.co.jp/netmarketing/databox/nmdb/070126_kuchikomii)>. (参照2008-06-15).
- 大塚裕子, ほか. 意見分析エンジン. コロナ社, 2007, 220p.
- Facca, F., et al. Mining interesting knowledge from weblogs: a survey. *Data and Knowledge Engineering*. 53, 3, 2005, p.225 - 241.
- Kawamura, T., et al. Mobile service for reputation extraction from weblogs - public experiment and evaluation. *Proc. AAAI-07*. 2007, p.1365 - 1370.
- 溝口理一郎. オントロジー工学. オーム社, 2005, 280p.
- Web広告研究会. “ブログ書き込み調査” <<http://www.wab.ne.jp/pdf/2005112801.pdf>>. (参照2008-06-15).



長野 伸一 NAGANO Shinichi, Ph.D.

研究開発センター 知識メディアラボラトリー研究主務, 工博。セマンティックWeb及びソフトウェアエージェントの研究・開発に従事。IEEE, 情報処理学会, 電子情報通信学会会員。Knowledge Media Lab.



溝口 祐美子 MIZOGUCHI Yumiko

研究開発センター 知識メディアラボラトリー研究主務。セマンティックWebの研究・開発に従事。Knowledge Media Lab.



稲葉 真純 INABA Masumi

研究開発センター 知識メディアラボラトリー。セマンティックWebの研究・開発に従事。Knowledge Media Lab.