

口コミマイニングシステム

浅倉剛 佐野直美 菅野元之 齊藤典明

日本電信電話株式会社 NTT 情報流通プラットフォーム研究所

1. はじめに

インターネットの普及に伴い、ユーザから配信される商品や企業に関する評判情報はマーケティング分野で重要視され始めている。例えば、インターネット上に存在する文章を構造化して「対象」と「評価」を含むものを「口コミ」として抽出して活用する事例が報告されている [2]。一方、筆者らはインターネット上を伝播する「口コミ情報」を広く収集、分析することで、世の中のトレンドや消費者の本音を把握することが可能な「口コミマイニング」を提案した [5][6]。本稿では、「口コミマイニングシステム」の機能構成を紹介する。

2. 口コミマイニングの位置付け

企業は消費者ニーズやクレームを把握するため、アンケート調査を実施したり、自社でお客様相談室や口コミ投稿サイトを運営している。しかし、モデレータの介在による検閲やサイレントクレイマー検知の難しさ等の面で、従来の方法では限界がある。そこで、口コミマイニングでは、インターネットからキーワードに基づいた文章収集を行うクリッピング調査の上に分析機能を加えることで、消費者の本音を把握することを目指す。図1に口コミマイニングの位置付けを示す。

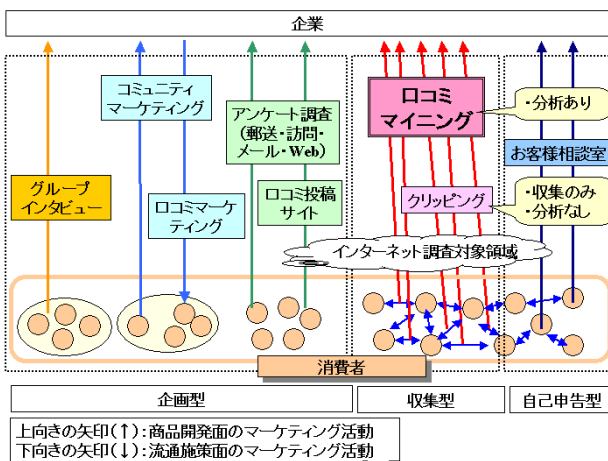


図1. 口コミマイニングの位置付け

3. 口コミマイニングの処理手順

口コミマイニングの処理手順を図2に示す。口コミマイニングはデータ収集、口コミ情報認知・抽出、マーケティング支援の3つの処理手順で実現される。

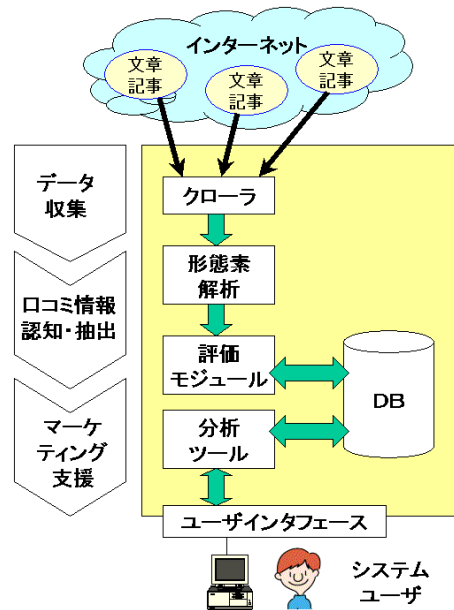


図2. 口コミマイニング処理手順

(1) データ収集

クローラパッケージを用いて、インターネット上のBBS、ニュースサイト、個人Webページ、企業Webページ等から文章・記事データを収集することを目標とする。

(2) 口コミ情報認知・抽出

形態素解析、評価モジュールを用いて、収集されたデータが「口コミ情報」であるか否かの認知を行い、「口コミ情報」と認知されたデータから「発言内容」「発言日時」「発言ユーザ」「評価」等の特長を抽出しDBに登録することを目標とする。

(3) マーケティング支援

システムユーザが分析ツールを用いて、DBに登録された「口コミ情報」の分析を行う。「クレイマーの検知」「消費者の潜在ニーズの検知」「評判伝播の検知」等の情報を得られることを目標とする。

4. 口コミマイニングシステム

口コミマイニングシステムの構成を図3に示す。

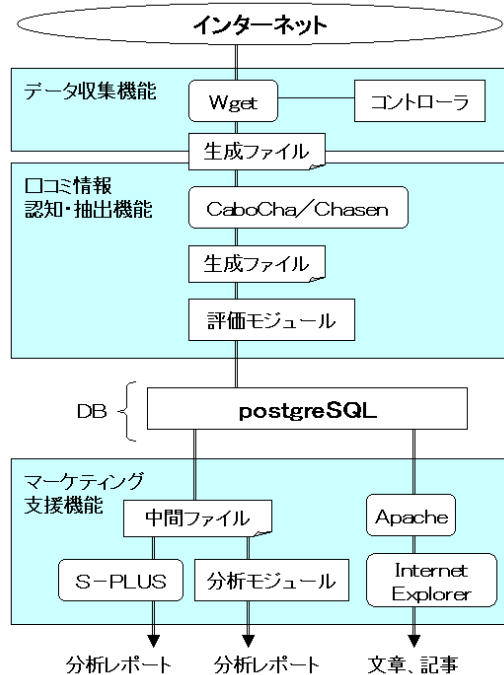


図3. 口コミマイニングシステム構成図

(1) データ収集機能

データ収集機能はクローラパッケージ「Wget」と「コントローラ」から構成される。ここではインターネット上の情報源ごとの収集条件(クローラ起動条件)を指定して文章・記事データを収集する。

(2) 口コミ情報認知・抽出機能

口コミ情報認知・抽出機能は形態素解析パッケージ「CaboCha」「Chasen」及び「評価モジュール」から構成される。ここでは以下の機能を実現する。

- ・情報源毎の、メタデータを抽出後、収集ファイルのデータフォーマット/文章フォーマットの解析を行い、データを正規化する。
- ・正規化されたデータを、形態素解析パッケージ「CaboCha」「Chasen」に入力することにより、キーワードと評価語に分解し、評価語を評価値に変換する。

- ・口コミ情報をデータベース「PostgreSQL」に登録する。

(3) マーケティング支援機能

マーケティング支援機能は分析ツール「S-PLUS」、「分析モジュール」、Webサーバ「Apache」、Webブラウザ「Internet Explorer」から構成される。ここでは「S-PLUS」及び「分析モジュール」を用いてデータベースに登録された「口コミ情報」の条件絞込みや伝播分析を実現す

る。また、「Apache」及び「Internet Explorer」で元の文章・記事の参照を実現する。

5. 実現例

口コミマイニングの出力画面例を図4に示す。

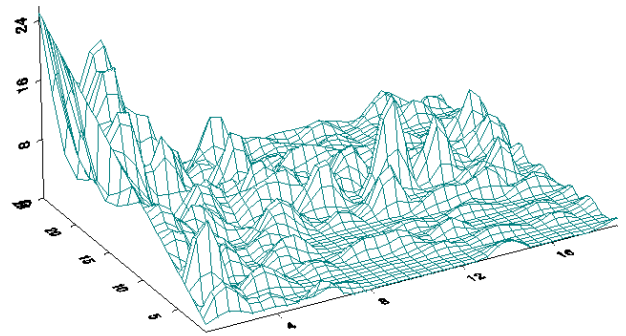


図4. 出力画面例

これは、インターネット上の口コミ解析手法を報告した[6]際に収集したデータを分析したものである。ある入力キーワードを含んでいる文章・記事データをインターネットの各種サイトから収集し、これら文章・記事の中に多く記述されている単語の出現回数を時系列で表示している。

6. おわりに

本稿では、口コミマイニングシステムの機能構成を示した。今後は口コミマイニングの有効性を検証していく。

参考文献

- [1] 立石ほか, "インターネットからの評判情報検索", 情報処理学会研究会報告, NL-144-11, pp.75-92, 2001.
- [2] 二本木ほか, "文の構造化による口コミ評価の分析・検索", インタラクシオン 2002 論文集, pp.175-176, 2002.
- [3] 立石ほか, "Web上の自動意見分析-情報抽出とテキストマイニングの融合-", 情報処理学会第64回全国大会, pp.3-19-3-20, 2002.
- [4] 齊藤ほか, "話題の自動抽出による電子メールの情報組織化手法" 情報処理学会論文誌, Vol39, No.10, pp.2907-2913, Oct.1998.
- [5] 菅野ほか, "口コミ情報解析による情報遷移の把握に関する一考察", FIT2002 M-39, 第四分冊 P.111-112, 2002.
- [6] 齊藤ほか, "インターネット上の口コミ解析手法の提案", 情報処理学会研究会報告, GN 45-12, pp.65-70, 2002.