

ロコミマイニングにおける“ロコミ情報”のマイニング手法

○佐野直美 浅倉剛 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よって、個人がモノ・コト・人物・場所等に関する意見や感想、不満等をインターネット上で述べる機会が増えつつある。それらインターネット上に存在する“ロコミ情報”をキーワードを元に収集し、企業が自社製品の評判や風評を把握するための情報を提供するクリッピングサービスも行われている。

我々は、クリッピング調査に更に分析機能を付加し、インターネット上を伝播する“ロコミ情報”から世の中のトレンドや消費者の本音を把握することを目的とする「ロコミマイニング」を提案した[4-7]。本稿では、解析結果を可視化するための、データ解析ツールと組み合わせたマイニング手法について述べる。具体的には、ユーザがキーワードや期間などの条件を設定し“ロコミ情報”を探索することにより、クレイマーの検知、消費者の潜在ニーズ発見、評判伝播の検知などを可能とする。本稿では特にクレイマーの検知について具体的なマイニング手法を述べる。

2. ロコミ情報伝播モデル

我々は“ロコミ情報”がインターネット上を伝播するモデルを図1のように想定している。

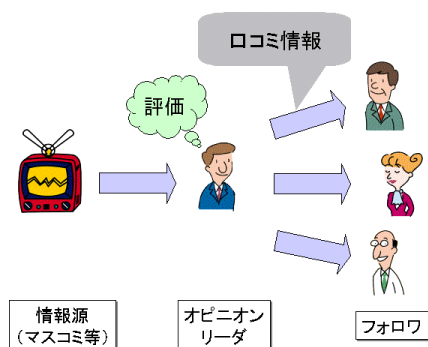


図1. ロコミ情報の伝播モデル

『情報源』となるあるサイトや人から『オピニオンリーダー』が情報を得ると、それに関する『評価』が“ロコミ情報”としてインターネットを介して広まる。“ロコミ情報”を受信した『フォロワー』は、それをまた他に広める場合もあれば、そこで伝播を終了させる場合もある。

ロコミマイニングシステムは、インターネット上において、「どこ／誰から」「どんなメディア／人を介して」「どんな内容」の“ロコミ情報”が「どれくらい」伝播するのかを上記モデルを用いて検証することを目的とする。

3. ロコミマイニングのシステム構成

ロコミマイニングはデータ収集、ロコミ情報認知・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。ロコミマイニングの処理手順とシステム構成を図2に示す。

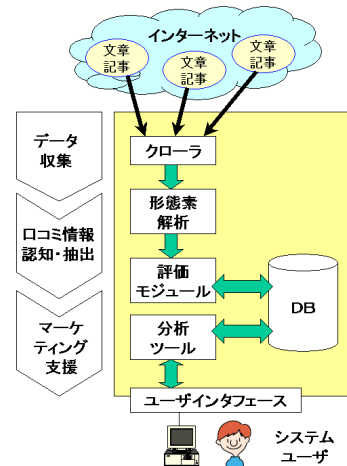


図2. ロコミマイニングのシステム構成

(1) データ収集

インターネット上のBBS、個人HP、ニュースサイト等から文章・記事データを収集することを目標とし、クローラパッケージ「Wget」と「コントローラ」で構成される。ここでは、インターネット上の情報源ごとの収集条件を指定してテキスト情報を収集する。

(2) ロコミ情報認知・抽出

収集したデータが“ロコミ情報”であるか否かの認知を行い、“ロコミ情報”と認知されたデータから、発言内容・発言日時・発言ユーザ・評価等の特徴を抽出しDBに登録することを目標とし、形態素解析パッケージ(Chasen・Cabocha)及び評価モジュールで構成される。ここでは、データの正規化・形態素解析によるキーワードと評価語の分解・評価語から標価値への変換・DBへの登録を行う。

(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“ロコミ情報”の分析を行い、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る

Viral Mining Method for “Word-of-Mouth”
Naomi SANO, Takeshi ASAKURA, Motoyuki SUGANO,
Noriaki SAITO
NTT Information Sharing Platform Laboratories, NTT
Corporation

ことを目標とし、市販のデータ解析ツール(S-PLUS)、分析モジュール、Web サーバ(Apache)、Web ブラウザ(IE)で構成される。ここでは、システムユーザが条件設定画面において、キーワード・期間・対象とする情報源など任意の条件を入力することにより“ロコミ情報”を絞り込み、伝播分析を行い、必要に応じて元の文章・記事の参照も行う。

4. “ロコミ情報”マイニング手法

3項で示したロコミマイニングシステムを利用して、以下の情報を探索することが可能である。

- 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

システムユーザがクレイマーを探索する場合の画面遷移例を図3に示す。

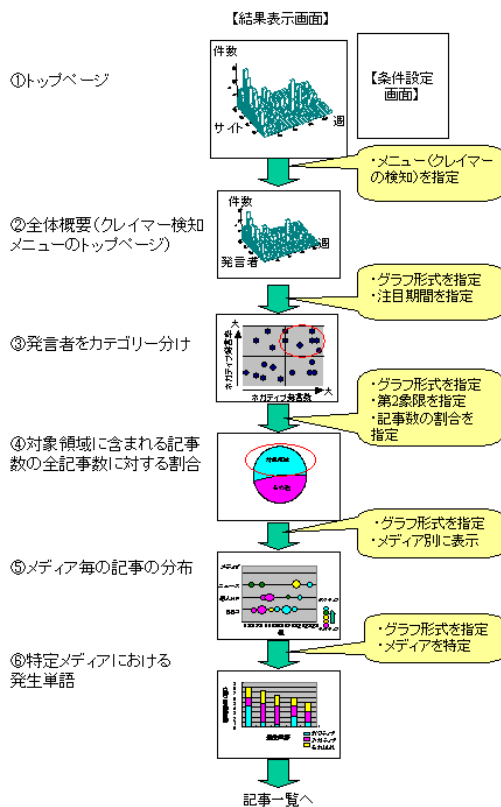


図3. クレイマー探索を行う場合の画面遷移例

図3の各画面の説明を以下に示すと共に、クレイマーを探索する流れを説明する。なお、すべての「結果表示画面」は「条件設定画面」と対になって表示され、「条件設定画面」においてメニューや絞り込み条件を設定し『グラフ表示』ボタンを押下することにより、次のグラフが「結果表示画面」に表示される。

- ① トップページ; ロコミマイニングシステムの起動の際に表示される画面。Web サイト毎の発言件数を時系列に示した 3D 棒グラフで、本画面により“ロ

コミ情報”の全体像を把握する。

- ② クレイマー探索のためのトップページ; 画面①から「クレイマー探索」メニューを指定、必要に応じてキーワードを設定した場合のトップページ。発言者毎の発言状況を把握する。
- ③ 発言者のカテゴリ分け; 画面②に示された発言者を、ネガティブ発言の件数・割合で分類する。件数、割合共に大きい、すなわち第 2 象限に属する「クレイマー」と思われる発言者を把握する。
- ④ 対象記事の割合; 画面③において第 2 象限に属する発言者全体の発言件数が、全発言件数に占める割合を示す。これにより、クレイマーの発言が無視できる程度のものであるかどうか確認する。
- ⑤ メディア毎の記事の分布; クレイマーの発言がどのメディア(ニュースサイト/個人HP/BBS等)で多く行われているかを把握する。
- ⑥ 特定メディアにおける発生単語; 画面⑤で特定した着目すべきメディアにおいて、クレイマーが話題にしている単語を把握する。

以上の一連の操作を行い、最終的に画面⑥に関連する記事の一覧を確認することにより、ある『キーワード』について『誰が』『どんな』ネガティブ発言(＝クレーム)を発しているかを把握することができる。例えばシステムユーザが企業の場合であれば、自社の悪評が広まる前に防止策を講じることが可能となる。

5. おわりに

本稿では、ロコミマイニングにおいてシステムユーザが“ロコミ情報”を探索する際の画面遷移例を示した。今後は更にシステムユーザが的確に所望する結果に辿り着くことが可能なマイニング手法の検討を進めていく。

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, "文の構造化によるロコミ評価の分析・検索", インタラクション 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] 斎藤ほか, "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, "ロ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] 斎藤ほか, "インターネット上のロ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.
- [7] 浅倉ほか, "ロコミマイニングシステム", インタラクション 2003.