

ステルスマーケティングに着目した Web レビューの信憑性判断支援

- テキストマイニングを用いて -

日大生産工 (院) ○三上 智行 日大生産工 柴 直樹 齋藤 敏雄

1. はじめに

Web の匿名性の高さに付け込み、企業がいわゆる「サクラ」や「やらせ」行為に走り問題となった事件があった[1]。ユーザレビューにおいてもメーカ側の思惑を孕んだステルスマーケティング目的のサクラによるレビューが潜んでいる可能性がある。

そこで、本研究ではレビュー信憑性の判断支援を目的とし、サクラによるレビューに着目する。サクラレビューの発見手法としてテキストマイニングを使用し、複数のレビュー文書を対象としてキーワードの出現時期を分析することで不自然なレビュー傾向を抜き出す可能性について検討する。

本研究は、ステルスマーケティングの撲滅を目指しているわけではなく、信憑性に対する意識の薄いユーザにレビュー情報を鵜呑みにせず取捨の必要性があるということの提示と、極端なやらせ行為抑制の一因となることを狙う。

2. 関連研究

Web ユーザレビューの評価情報の時系列変化に着目した研究[2]では、キーワードに関する評価情報を用いて時系列解析を行うことで、そのキーワード、あるいは複数のキーワードに関する評価内容の時系列変化を表現するシステムを構築している。時系列に基づき、評判がどのように変化しているかを特徴づけることができる。評判の変化はサクラレビューの目的そのものであり、本研究に大きく関わる。ただし、サクラレビューの発見手法として扱う場合、サクラレビューが短期に集中して投稿される可能性を想定するならば、月ごとの分析のみでは適切ではないため、分析期間に再考が必要と考えられる。

3. 提案手法

メーカや販売企業がステルスマーケティングを行う場合、ブログ運営者に報酬を与え製品を称賛させる方法や、内部社員や関連企業に指示を出し、オンラインストアで自社製品に高評価を与えるなどの方法が想定できる。本研究では後者に着目する。

企業がステルスマーケティングに即効性を求めた場合、レビュー内容を指示したうえで、短い期間に集中して高評価のレビューを投稿する可能性が考えられる。製品に高評価を与えるレビューの局時的な発生や、偏った視点のレビューの集中が見られたとき、やらせの可能性を疑うべきであるが、それが平均評価に影響を及ぼしユーザの目に触れつつも、その不自然な局時性がユーザに伝わるとは限らない。そこで、まずレビュー文章に対して日本語形態素解析器「MeCab」[3]を用いることで、語句単位にレビュー内容を分解する。形態素解析では、例えば、「本を読んだ」というテキストを、図1に示すように「本」、「を」、「読ん（読む）」、「だ」と分割し、同時に各語句の品詞の特定を行う。これにより、文書の内容を直接表現する名詞、形容詞、動詞といった内容語のみを抽出し、データを整理でき、統計分析を行い易くすることができる。

<u>本</u>	<u>を</u>	<u>読ん</u>	<u>だ</u>
名詞	助詞	動詞	助動詞

図1 Mecabによる形態素解析実行例

そして分割された語句をもとに、複数のレビューに含まれる語句とタイムスタンプの総合的な分析を行う。特定の語句の出現頻度と時間軸のグラフを出力することで、時間経過における評価の推移を提示する。なお、出現頻度は、期間中の語句出現回数/期

Credibility Judgment Support of Web Review, Focusing on the Stealth Marketing

- Using Text Mining -

Tomoyuki MIKAMI, Naoki SHIBA and Toshio SAITO

間中のレビュー件数とする．また，特定語句について，一定期間の区分における出現頻度と，全体期間における出現頻度を比較することで，語句の不自然な集中期間を検出し，該当部分を視覚的に提示する．そこで，全体期間中を一定間隔で区切り，第 i 番目の期間における語句 X の出現頻度を X_i ， X_i の標準偏差を S とするとき，全体から見た期間 i における語句 X の集中度を表す値 ZX_i を， X_i のいわゆる Z 値として以下のように定義する．

$$ZX_i = \frac{X_i - \bar{X}_i}{S} \quad (1)$$

ここで， $\bar{X}_i$ は X_i の平均を表す．

4. システム設計

4.1 概要

システム化における運用対象として，オンラインストア“Amazon.co.jp”[4]を想定する．ユーザはAmazon.co.jp で取り扱われる任意の商品を指定する．システムは指定された商品に対し提案手法を適用し，図2 のようにグラフとして提示する．

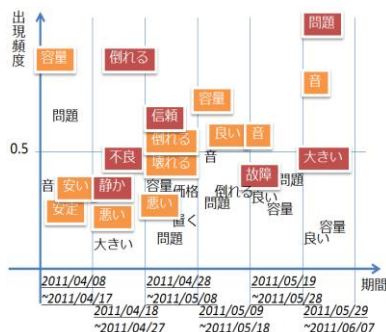


図2 提示グラフのイメージ

また，語句の集中度の高さに応じて色分けを行い，直感的に判別し易くする．これにより，ユーザは語句の不自然な集中を知ることができ，信憑性判断に活かすことができる．

4.2 システム設計における検討項目

レビュー文章は格式張ったものではなく，同じ意味を持つ別の言葉が混在する可能性が非常に高いため，それらの扱いを考慮する必要がある．

ユーザレビューの分析に際し，当然レビューの対象となる何らかの製品・サービスの情報が伴う．テ

キストマイニングにより単語の抽出およびポジティブな・ネガティブな単語の判定を行う場合，レビュー対象製品の特性を視野に入れるべきである．

レビュープロフィールを新たな評価情報として分析に取り入れようとした場合，製品紹介ページに掲載されているレビュー文書だけでなく，レビュープロフィールを取得する必要がある．そのためには，データ収集及び分析の手法に検討を要すると考えられる．

4.3 システムの評価方法

研究結果の評価を行う際，実際にサクラレビューが含まれるサンプルデータを用いて，パラメータのチューニングと提案手法の有効性を確認する方法が考えられるが，実際にステルスマーケティングが行われた確証のあるユーザレビューの取得は簡単とはいえないため，その入手が課題といえるだろう．

5. おわりに

本稿では，サクラレビューの発見手法としてテキストマイニングを使用し，複数のレビュー文書を対象としてキーワードの出現時期を分析することで不自然なレビュー傾向を抜き出す可能性について検討した．本研究により，ユーザレビューの信憑性判断の新たな一要素を提供し，ユーザがユーザレビューをより良く活用できる環境につながることだろう．

参考文献

- [1] “経済産業省主催の県民説明番組への意見投稿呼びかけに関する事実関係と今後の対応（再発防止策）について”，2011，
<http://www.kyuden.co.jp/library/pdf/press/2011/h110711-2.pdf>，最終アクセス日：2011年9月25日
- [2] 打田裕樹ほか，“Web ユーザレビューにおける評価情報の時系列変化の可視化”，知能と情報 Vol.22, No.3, pp.377-389, 2010
- [3] MeCab ホームページ：
<http://mecab.sourceforge.net/>
- [4] Amazon.co.jp： <http://www.amazon.co.jp/>