

# 商品レビューにおける主観的表現と有用性に関する検討

木浪貴博† 小林亜樹†

† 工学院大学工学部情報通信工学科

## 1 はじめに

EC サイトにおける商品レビューはユーザの商品購入時の判断材料に活用されており、商品レビューの有用性に関する研究も進められている。例えば、山澤ら [1] は、ユーザの商品購入検討時の意思決定に寄与する文を有用なレビュー文と定義し、文単位での分類を形態素情報を用いた SVM によって試み、結果を人手による有用性に関するタグ付け処理と比較することで有用性判定ができると結論づけている。また、落合ら [2] は、商品比較を目的として、レビュー文章からの有益な潜在的トピック抽出手法を提案している。一方、原田ら [3] はオンライン書評における主観、客観表現の割合を分析し、オンライン書評が新聞書評より主観的だという結論を導いている。

商品レビューが多くの利用者の手になることから、全般に主観的であるだけでなく、時に投稿者の個人的感想に偏った内容となることはしばしば目にするものである。そこで本研究では、レビューの主観客観視点と有用性との間の関係性について分析を試みる。具体的には、形態素レベルでの主観、客観表現の存在とレビュー文章の有用性との間の関係性を調査する。

## 2 調査手法

### 2.1 概要

商品レビューには Amazon.com のデータを用いた。当該サイトでは商品レビューの提示順序として有用度の高い順を設定できる。そこで、レビュー毎の有用度をこの

順序を用いて決定する。具体的には、提示順序の順位情報として取り扱う。

商品単位で有用性の高いレビュー群（高有用性群）と低いレビュー群（低有用性群）とについて、それぞれ形態素レベルで語の出現について調査を行う。調査対象となる語は別途選定しておくものとする。

### 2.2 主観語、客観語

形態素レベルで主観、客観表現を抽出するため、主観語、客観語を辞書として定義する。主観、客観の意味合いについては新潮国語辞典 [4] を参考にし語の選定を行った。具体的には、分析対象となるレビューとは別のレビューデータから主観的表現につながる主観語、客観的表現につながる客観語それぞれを筆者が選定した。選定した語の例は次の通りである。

**主観語** 思う, 感じる, 面白い, 良い など

**客観語** 観点, 検討, 評価, 好評 など

これらをそれぞれ主観語辞書、客観語辞書と呼ぶ。主観語は、164 語、客観語は 7 語を選定した\*1。

### 2.3 手順

複数の商品について、商品毎にレビュー文データを取得する。高有用性群、低有用性群に属するそれぞれのレビュー毎に MeCab を用いて形態素に分割し Bag-of-words として取り扱う。各レビューについて辞書中に出現する語の頻度や位置などを計測し、群毎の平均値などで分析を行う。

An investigation of subjective representation and usefulness in customer reviews text

†Takahiro Kinami, †Aki Kobayashi

†Department of Information and Communications Engineering, Faculty of Engineering, Kogakuin University.

\*1 ただし、語尾の活用を含む。

### 3 分析

#### 3.1 対象データ

Amazon.com の商品レビューより、レビュー件数 300 件以上の商品をカテゴリの異なる 4 商品選定した。それぞれの商品レビューについて、有用性の高い順に提示された順位情報を用いて、高有用性群、低有用性群となるレビューを収集した。ここで、それぞれの群のレビュー件数は各 50 件である。

以下、商品名は a~d のシンボルで表現するが、具体的には、

- a. I-O DATA USE 接続ビデオキャプチャー GV-USB2
- b. SHARP スタンダートタービンヘッド太ブサイクロンクリーナー ページュ系
- c. 【Amazon.co.jp 限定】ZEPEAL ゼピールサーキュレーターホワイト DKS-20W
- d. 白夜行 (集英社文庫)

である。

#### 3.2 結果と考察

各商品各群のレビュー集合毎における主観語数、客観語数、全語数、それぞれの相対頻度の表を表 1 に示す。

全商品に渡って、低有用性群の全語数が高有用性群に比して少ない。商品レビューにあっても文章量と有用性との間に関係があることが伺える。客観語はレビューではあまり用いられず、辞書選定時点で語数が極めて限られていた。オンラインレビューが新聞などのそれに比して主観的であるという報告から類推される結果である。すなわち、主観語を用いて記述される傾向が強いことが

表 1 出現頻度

商品	主観語	客観語	全語	主観相対	客観相対
商品 a (高)	4.48	.20	183.9	.0323	.0023
商品 a (低)	1.52	.14	58.9	.0324	.0033
商品 b (高)	4.34	.20	149.9	.0371	.0012
商品 b (低)	2.12	.14	63.6	.0395	.0006
商品 c (高)	3.62	.20	155.6	.0304	.0010
商品 c (低)	1.84	.02	64.6	.0345	.0001
商品 d (高)	4.76	.10	242.2	.0216	.0010
商品 d (低)	3.16	.12	143.4	.0258	.0013

確認された。これは、選定した主観語辞書、客観語辞書の妥当性を示唆するものといえる。

その上で、主観語の相対出現頻度を見ると、いずれの商品にあっても、低有用性群のほうが主観語頻度が高くなっている。商品 a についてはほとんど差がなく、事例数もわずかであることから統計的な議論はできないが、レビューの有用性と主観語の出現との間の関係の可能性を示唆する。仮説としては、主観語の出現が少ないレビューほど有用である、となる。ただし、個々のレビュー毎の分散が大きいため、そのまま有用性判定指標とはできない。

なお、有用性の特に高いとされるレビューでは、主観語、客観語ともに出現頻度が他のレビューと比べて低い傾向が見られた。これは、文章量が相対的に多い中で、評価を下すような表現自体が少ないことを示唆しており、結果的に、データの事象などの事実を記述しているものと考えている。

主観語客観語の枠組みだけでなく、評価を下す文、あるいは表現の出現と事実との比率などといった他の指標による分析を今後の課題としたい。

### 4 おわりに

レビュー文中の主観客観語の出現と有用性の関係について調査した。調査の大規模化や他の指標の分析などは今後の課題である。

### 謝辞

本研究の一部は科学技術研究費 (基盤 (A)) No. 26242013 による。

### 参考文献

- [1] 山澤 美由紀, 吉村宏樹, 増市博, “Amazon レビュー文の有用性判別実験”, 情報処理学会研究報告, pp.15-20, 2006.
- [2] 落合恵理香, 小林一郎, “商品の評価を対象としたレビュー文章の分析”, 言語処理学会 第 18 回年次大会 発表論文集, pp.1176-1179, 2012.
- [3] 原田隆史, 吉村紗和子, “オンライン書評と従来の書評との違いの分析”, 情報知識学会誌, Vol.20, No.2, pp.65- 72, 2010.
- [4] 久松潜一, 山田俊雄, 築島祐, 小林芳規, “新装改訂 新潮国語辞典 一現代語・古語一”, pp.487, pp.936, 1992.