

文の構造化による口コミ評価の分析・検索

二本木 智洋 * 住田 一男 **

* 東京工業大学大学院総合理工学研究科 ** (株)東芝 研究開発センター

1. はじめに

他人のコメントや評価を参考にして、様々な商品やサービスを購入することは一般的に行われている。また、消費者のコメントや評価が有用な情報であることは、企業がアンケート調査等を頻繁に行っていることから明らかである。近年ではインターネットを利用して評価を収集する試みも盛んになっている。中でも、口コミ検索サイト[1]と呼ばれるサービス(例えば Power to The People[2]など)では、ユーザーに商品やサービスのコメントや評価(口コミ評価)を自由に登録してもらい、それを別のユーザーが参照することが可能である。しかし、手間がかかるので、ユーザーは口コミ評価の中の全ての自由記述文を読むことはできない。一括して概要を把握できる機能が提供できれば、効率的な意見把握が可能となる。

本研究では、多商品分野に渡る大量の自由記述文から、評価に関する表現を抽出して定量化し、効率的な分析・検索機能を提供する手法を考案する。

2. 評価表現における<対象>と<評価>の抽出

自然言語文から有用な情報や知識を抽出しようとする試みが行われている([3]など)。本研究では、口コミ検索サイトに寄せられた自由記述文から、ある<対象>に関する<評価>(一般的な「対象」「評価」と区別するために以下で用いる)という対をなす構造を用いて評価表現を抽出する。

李らは商品の自由記述アンケートから情報を抽出し、分析を行っている[4]。抽出される情報は単語単体であり、商品の特徴や他商品との相対関係の視覚化に重点が置かれている。また、立石らはインターネット上から人の意見を抽出することを

行っている[5]。ある特定の商品に対する意見を収集・検索するものであり、多商品を対象とする場合にはそのまま適用できるものではない。

評価表現抽出は、図1に示すように3つのステップから構成されている。

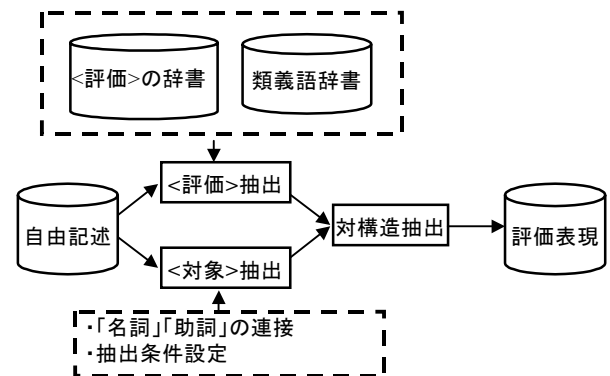


図1 評価表現抽出の流れ

(1)<評価>の抽出

ある商品に対する自由記述文をもとに<評価>になりうる表現の辞書(人手で事前に作成)と類義語辞書に基づき、該当する表現を<評価>として抽出する。(類義語辞書の作成には[6]を利用した)。<評価>の抽出例を図2に示す。

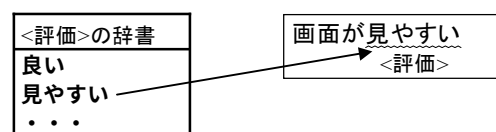


図2 <評価>の抽出例

(2)<対象>の抽出

自由記述文から「名詞」「助詞」の接続を<対象>として抽出する。ただし、語と助詞の種類により<対象>として抽出しない条件を設定した。<対象>の抽出例を図3に示す。

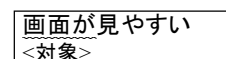


図3 <対象>の抽出例

(3)対構造の抽出

(1)(2)により抽出された<対象>と<評価>を、同一文中で対応付け、対の構造を抽出する。

Analysis and Retrieval of the Word-of-mouth Estimation by Structurizing Sentences

* Tomohiro Nihongi ** Kazuo Sumita

* Tokyo Institute of Technology ** TOSHIBA Corp.

* tomohiro.nihongi@ip.titech.ac.jp

** kazuo.sumita@toshiba.co.jp

3. 抽出実験

ある携帯電話 1 に対する自由記述から＜評価＞の辞書を事前に作成した。この辞書を用いて別の携帯電話 2 に対する自由記述から＜評価＞を抽出した結果を表 1 に示す。

	携帯電話 1	携帯電話 2
＜評価＞の辞書(人手)	294	175
＜評価＞(1)[全抽出数]	294 [512]	77 [217]
再現率 R_1	1.00	0.44
適合率 P_1	0.57	0.35
＜評価＞(2)[全抽出数]	294 [620]	122 [356]
再現率 R_2	1.00	0.70
適合率 P_2	0.47	0.34

＜評価＞(1) : ＜評価＞の辞書を使用

＜評価＞(2) : ＜評価＞の辞書と類義語辞書を使用

$$\text{再現率 } R = \frac{\text{＜評価＞}}{\text{＜評価＞の辞書}} \quad \text{適合率 } P = \frac{\text{＜評価＞}}{\text{全抽出数}}$$

表 1 ＜評価＞の抽出結果

携帯電話 1 の＜評価＞の辞書に類義語辞書を組み合わせることによって、携帯電話 2 においても 0.7 の再現率が得られている。

一方、携帯電話 1、2 のそれぞれの自由記述から＜対象＞を抽出した結果を表 2 に示す。

	携帯電話 1	携帯電話 2
人手による＜対象＞	284	152
＜対象＞[全抽出数]	220 [1120]	110 [762]
再現率 R	0.78	0.72
適合率 P	0.20	0.14

$$\text{再現率 } R = \frac{\text{＜対象＞}}{\text{人手による＜対象＞}} \quad \text{適合率 } P = \frac{\text{＜対象＞}}{\text{全抽出数}}$$

表 2 ＜対象＞の抽出結果

機種に関わらず 0.7 以上の再現率が得られている。＜対象＞＜評価＞構造の抽出例を表 3 に示す。

自由記述	＜対象＞-＜評価＞
使い勝手はとっても良いので、是非お勧めです。	使い勝手-良い

表 3 ＜対象＞＜評価＞構造の抽出例

4. 試作システム

抽出された＜対象＞＜評価＞構造の有用性を評価するために、商品検索を行うシステムを試作した。ある商品分野の中で、自分が商品に求める特性を自由記述で入力すると、それに一致すると思われる商品を一致度順に出力する。試作システムの構成は図 4 のようになっている。また、動作画

面の例を図 5 に示す。

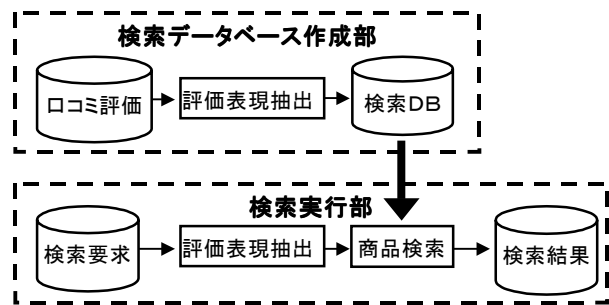


図 4 試作システム

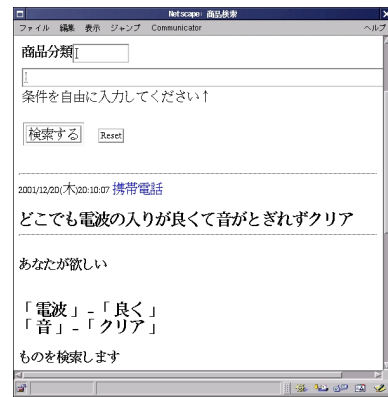


図 5 動作画面の例

5. まとめ

本研究では、＜対象＞＜評価＞の構造を利用して、自由記述から評価表現を抽出した。＜評価＞の辞書と品詞情報を用いることにより、＜対象＞と＜評価＞の抽出は商品の種類に関わらず 0.7 の再現率がみられた。また、得られる情報を利用した商品検索システムを試作した。今後、システム全体としての有効性を評価する。

6. 参考文献

- [1]日経ネットビジネス, No.66, ”口コミ検索サイト ユーザー評価を基に情報抽出”, pp.132-135, 2001
- [2]Power to The People , <http://www.ptp.co.jp>
- [3]IREX 実行委員会：IREX ワークショップ予稿集, 1999
- [4]李ほか：確率的コンプレキシティを用いたルール学習による自由記述アンケート分析, 言語処理学会第 7 回年次大会発表論文集, pp.379-382, 2001
- [5]立石ほか：インターネットからの評判情報検索, 自然言語処理, No.144-011, pp.75-82, 2001
- [6]国立国語研究所：分類語彙表(フロッピー版), 1993