

商品改善に繋がる比較意見の抽出 A Study of Comparative Opinion Extraction for Remediation

石澤 善雄[†] 細見 格[†]
Yoshio Ishizawa Itaru Hosomi

1 概要

商品の使用者が、機能や性能などに関して競合製品との違いを書いた意見文は、使用者が使っている商品の開発者にとって商品の改善のために有効である。なぜなら、開発者は使用者がどのような機能や性能に不満があるのか、もし不満があるならばどう改善すれば満足するののかのヒントが得られるからである。我々は、このような同種の商品やメーカーが比較評価されて書かれている意見を抽出する方式を提案・実装し、Yahoo! 知恵袋のデータから携帯電話などを対象として比較意見を抽出した。その結果、商品改善に有効な比較意見が得られることを確認した。

2 序論

近年、商品の使用者が、インターネット上のブログなどで使用商品に関する事実や意見を大量に発信するようになってきている。使用者が商品に対して持ったこのような意見には、「携帯電話 B の画面が暗くて見にくい」などの否定的意見 (不満) が含まれており、商品の開発者が使用者の不満を把握できれば、不満を改善するための対策を施した次期商品を開発することができ、次期商品においてユーザ満足度が向上すると考えられる。

しかし、使用者が書く不満には、「携帯電話 A には B があるのに、携帯電話 C に B がないのはおかしい」などの比較を用いて、携帯電話 C には携帯電話 A とは違って B がないので満足していないことを表した比較意見や、「携帯電話 D を 1 週間充電しなくてもずっと使えるようにしてくれたらいいのに」などの、現状の携帯電話 D に不満を持っているのはわかるが、実現できるかわからない使用者要求が書かれている意見も存在する。ユーザが書き込んだ意見を利用して消費者の期待に沿った商品開発をするためには、意見の中でも、ある商品では実現されているが、使用者の商品では実現されていないことが比較して書かれた比較意見を抽出することが重要である。なぜなら、具体的にどこまで改善すれば使用者が満足するのかが比較対象の商品からわかり、かつ他の商品で実現されている機能や性能なので、実現可能な使用者要求だからである。

山崎らは、競合事物との比較から目的事物の優位点や欠点、相異点などを知るために、競合事物を比較した意見文をテキストから抽出し、競合事物対がどのような関係にあるかを判定する手法を提案した [2]。この手法で抽

出する意見文のうち、競合事物対のそれぞれの特徴を表した意見文が、我々が抽出する比較意見文に類似しており、比較文抽出規則によって比較意見文を抽出している。

本論文では、下記の例のような評価の良し悪しを表す肯定評価表現と否定評価表現に関して、比較可能な肯定評価表現と否定評価表現を人手で対応付けて比較評価表現辞書を作成し、比較意見文を抽出する。比較意見文の例と改善された状態を以下に示す (下線は評価表現)：

- 携帯電話 E の液晶画面は 綺麗だが、携帯電話 F は 汚い (改善状態は、携帯電話 F の液晶画面が携帯電話 E と同等の綺麗さになった状態)

以上のように、比較意見文は現商品の不満点と改善状態を与えてくれるので、商品開発に役立てることができる。

また、比較意見文は商品に関してだけでなく、様々な対象について書かれる。我々は高齢者に適した対象についての話題を提供し、高齢者のコミュニケーションの活性化を目指す研究を行っている [1]。その対象と類似した対象に関しても言及された比較意見を利用して話し合うポイントを提供することで、他の高齢者との会話接点を持ち易くなり、コミュニケーションの活性化に繋がると考えている。

3 比較意見文抽出方式

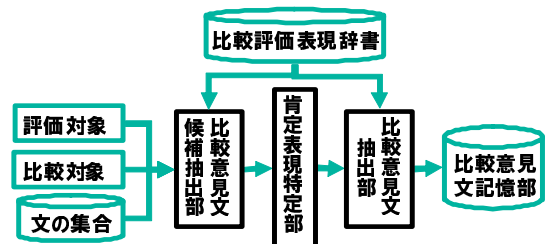


図 1: 提案方式の構成図

提案方式の構成図を図 1 に示す。まず、提案方式は比較意見文候補抽出において、文集合中の評価対象を表す語と比較対象を表す語が含まれる 1 文または連続する 2 文から、評価対象に関する評価表現と比較対象に関する評価表現が含まれている 1 文または連続する 2 文を比較意見文候補として抽出する。ここで、文集合は、掲示板などへの商品使用者による書き込みの集合であり、評価対象と比較対象はそれぞれ評価をしたい事物とそれに競合する事物である。次に、肯定表現特定部において、比較意見文候補に改善された状態が表現されているかを評価表現を使って判定し、改善された状態が表現されている比較意見文候補を抽出するために、評価表現のうち、

[†]日本電気株式会社 情報・ナレッジ研究所, NEC Knowledge Discovery Research Laboratories.

少なくとも比較対象の評価表現が肯定表現である比較意見文候補を抽出する。最後に、比較意見文抽出部において、評価対象と比較対象に対する評価表現が比較可能な評価表現であるか、及びその評価に差があるかを、比較評価表現辞書及び評価の良し悪しの程度を数値化した評価極性を使って調べる。比較対象の方が評価対象より高評価を得ている比較意見文を抽出するために、比較意見文候補の評価表現の評価極性の差が閾値以上である比較意見文を選択する。

山崎ら [2] の方式では相互比較可能な評価表現を含まない文も抽出されてしまうが、提案方式では比較可能な評価表現の組を辞書に登録することで、山崎らの問題を解決する。

4 評価実験と抽出結果

提案方式を実装し、携帯電話、製造メーカ、キャリアに関する評価対象と比較対象を用いて、Yahoo!知恵袋のデータから比較意見文抽出を実施した。

Yahoo!知恵袋データの「> インターネット、PC と家電 > 携帯電話、モバイル」カテゴリーに含まれる質問と、その質問に対するベストアンサーを紐付けて 1 組とし、質問とベストアンサーの組 125,590 組を比較意見文抽出に用いた。データには 1,070,596 文が含まれ、1 組当たり約 8.5 文から構成されている。

また、比較され得る評価表現の対応付けを行うために、既存の評価表現辞書を用いた [4]。使用した評価表現辞書には、肯定評価表現 2569 表現、否定評価表現 2132 表現、中立評価表現 2880 表現が含まれている。肯定評価表現、否定評価表現、中立評価表現に対して、それぞれ評価極性として、+1、-1、0 が割り当てられている。このうち、評価表現辞書の肯定評価表現と否定評価表現を人手で対応付けた比較表現辞書を用意して、抽出された文中の 2 つの評価表現に対して肯定評価表現と否定評価表現の対応が付いていれば評価極性差が閾値以上であると判定し、比較意見文の抽出を行った。人手で対応付けた肯定評価表現と否定評価表現の対は 7,879 組であった。

Yahoo!知恵袋のデータを対象に、比較意見文の抽出を行った結果の例を示す (□ は評価対象または比較対象、下線は評価表現)：

- 携帯電話の型番
 - (a) [携帯電話 S] だとメモリースティックしか対応してないので、今度の機種変でも S 社じゃないとメモリースティックは 使えないし。[携帯電話 H] なら MiniSD なので、次の機種変でも S 社以外の機種なら継続して MiniSD 使えるし
- 携帯電話の製造メーカ
 - (a) [H 社製品] は携帯電話 H で PC サイトビューアーが 使える から PC と同じように検索を掛けると沢山の情報が見られます。[P 社製品] は

携帯電話 P で PC サイトビューアーが 使えない 為に、検索を掛けても情報が出てきません

● 携帯電話のキャリア

- (a) 赤外線通信ですけど、D 社 → [A 社] は 出来る んですが A 社 → [D 社] が 出来ない んです

評価対象と比較対象が、携帯電話の型番の場合 1 件、携帯電話の製造メーカの場合 2 件、携帯電話のキャリアの場合 10 件、の計 13 件が抽出され、抽出された比較意見文から、改善の方向性がわかる比較意見文が抽出できることが確認できた。

抽出数が少ない理由を調査するために、評価対象と比較対象が含まれており、それらに対応する評価表現が存在する意見文をデータのサンプリングによって調べた結果、約 67,800 文 (6.3%) 含まれると考えられ、サンプリングで得られた比較意見文から評価表現辞書に登録されている評価表現の不足と中立評価表現の多さが原因と推測される。そのため、評価表現辞書に含まれる評価表現の充実と、中立評価表現も含めて評価表現の対応付けをすることができれば、再現率を高めることができると考えている。

5 まとめ

商品の改善に繋がる比較意見文を抽出する方式を提案・実装し、Yahoo!知恵袋のデータから、携帯電話、携帯電話の製造メーカ、携帯電話のキャリアを評価対象および比較対象として、比較意見文の抽出を試みた。その結果、107 万文から計 13 件が抽出され、商品改善に繋がる比較意見文の抽出を確認できた。抽出数の少なさに関してデータを調査し、評価表現辞書への評価表現の追加、中立評価表現も含めた肯定評価表現と否定評価表現の対応付けが課題であると確認した。

6 謝辞

本研究は、総務省の委託研究「ライフサポート型ロボット技術に関する研究開発」プロジェクトの成果である。

参考文献

- [1] Yoshio Ishizawa *et al.*, A method for conversational topic recommendation to appropriate user group, ICDIM, 270-275 (2011).
- [2] 山崎義隆、乾健太郎、松本裕治、競合事物間における比較関係認識、情報処理学会研究報告. 自然言語処理研究会報告 Vol.2011 No.5, pp.1-7 (2011).
- [3] 国立情報学研究所、「Yahoo!知恵袋データ」の提供について、
http://www.nii.ac.jp/cscenter/idr/yahoo/chiebkr2/Y_chiebukuro.html (Jun. 2009).
- [4] 土田正明、水口弘紀、久寿居大、対象-属性-評価の 3 項関係同定による評判情報抽出、言語処理学会第 13 回年次大会論文集、pp.412-415、2007.