

商品レビューからの代替商品情報の抽出

堀川 泰輝[†] 赤石 美奈[‡]法政大学情報科学部^{†‡}

1. はじめに

近年、楽天市場やAmazon等のECサイトが普及している。多種多様な商品情報を簡単に閲覧できる反面、情報の膨大さゆえに、ユーザは好みに合う商品を見つけることは容易ではない。

また、ユーザの購入目的を満たす商品は多数存在し、それらの形や色、本来の機能や用途は多岐に渡る。このため、ある商品は一つの購入目的に着目すれば、他の商品で代替が可能な場合がある。しかし、ユーザ自身がそれらの商品同士の代替関係をすべて網羅することは難しく、既存のシステムにおいても、それらの関係に基づいた商品を提示できるものとはなっていない。

そこで本研究では、代替商品情報を商品レビューから自動抽出することを目的とする。商品レビューには「〇〇の代わりに商品を購入した」等の記載があり、企業側が発信するだけでは得られない詳細な利用者目線の意見を抽出することを可能とした。

2. 関連研究

本研究と関連する研究として、本田ら[1]の研究がある。閲覧商品の組み合わせによりユーザの利用目的を判定し、ECサイトにおける推薦商品に代わる代替商品の提示を提案している。目的の判定には、商品レビューに含まれる単語の出現頻度の χ^2 乗値を用いている。代替商品提示の研究はなされているが、商品レビューに明示されている代替商品情報を、構文解析を用いて抽出を行う研究はあまりなされていない。

3. 提案手法

構文解析を用いた抽出ルールを商品レビューに適用して、商品間に存在すると考えられる代替関係を自動抽出するシステムを提案する。商品レビューから抽出された語と、レビュー元の商品名との対のデータを代替商品情報とする。本章において、提案手法の流れを説明する。

Extraction of substitute relations among products from reviews with syntactic analysis

^{†‡}Taiki Horikawa and Mina Akaishi

^{†‡}Faculty of Computer and Information Sciences, Hosei University

3.1. 解析対象とするデータ

本研究で使用するレビューデータは、楽天市場の「みんなのレビュー・口コミ情報」である。レビューデータには、投稿者に関する情報、商品に関する情報、商品に対する評価情報等が登録されている。2012年の1年分の約2,600万件のレビューデータを解析対象とし、本研究では、商品名とレビュー本文を取得する。

3.2. 代替商品情報の抽出ルール

商品レビューに適用する抽出ルールについて説明する。まず、「代わりに」という語彙パターンを含むレビュー文を収集する。次に、収集したレビュー文に対して係り受け解析を行い、「代わりに」に係る語を抽出する。本節において、各処理の詳細について説明する。

3.2.1. 解析対象文の絞り込み

本研究では、レビュー本文に「代わりに」を含む文のみを対象とし、構文解析を行った。人手により商品レビューを分析した結果、代替品情報を含み、かつ構文解析をしたときに特徴を捉えやすいため、この語彙パターンを用いることとした。これより、解析対象としたレビューデータの内、約5万件のレビューを取得した。

表1に、「代わりに」を含み、かつ代替商品情報を含む商品情報の例を示す。

表1 代替品情報を含む商品情報例

	商品名	レビュー文
例1	アロマ ディフューザー	陶器の下部隙間からもれる灯りが常夜灯の <u>代わり</u> になり、本来の照明を消せるので便利です。
例2	ボディ ローション	私はあまり塗りすぎると肌が赤くなるので、香水の <u>代わり</u> に香り付けするくらいで楽しんでいきます。

3.2.2. 代替品情報の抽出

本研究では、構文解析器 CaboCha を用いて係り受け解析を行う。3.2.1 節で示したレビュー文に係り受け解析を行い、「代わりに」に係る文節の付属語や活用を除いた、自立語の原形を抽出する。

表1のレビュー文を読むことで、例1では「アロマディフューザー」と「常夜灯」、例2では「ボディローション」と「香水」がそれぞれ代替関係にあることがわかる。例1では、文節「常夜灯の」が「代わりに」に係り、例2では、文節「香水の」が「代わりに」に係るため、「常夜灯の」と「香水の」をそれぞれ自立語の原形に直すことで「常夜灯」と「香水」を抽出することが可能となっている。

提案手法での抽出ルールでは、「代わりに」に係る語のみを抽出対象としており、代替商品情報を含むかどうかに関わらず、抽出結果は係り受け解析の結果のみに依存する。また、抽出結果が、「暑い」や「涼しい」等の形容詞、「壊れる」や「食べる」等の動詞の場合があり、これらはモノの名前として適切ではない。そのため、抽出する品詞は名詞に限る。

4. 評価実験

抽出結果と商品が、代替関係として妥当であるかを調べるための評価実験を行った。実験には大学生9名に協力してもらった。抽出した代替商品情報から、無作為に40件のデータを選び出し、それらを被験者に提示し、アンケート調査を行った。被験者には、それぞれ40件のデータに対し、代替することが可能であるか、また、代替が可能であると判断できる場合に、その代替関係は意外であるかについて答えてもらった。

5. 実験結果

提示した代替商品情報40件の内、代替することが可能という評価を得た商品の割合の平均、代替が可能だと判断できたもののうち、意外だという評価を得た商品の割合の平均、被験者ごとの評価の平均からみた標準偏差（標準偏差1）、商品ごとの評価の平均からみた標準偏差（標準偏差2）を表2に示す。

全データ数40件の内、83%の割合で代替することが可能という評価を得られた。商品レビューからの代替商品情報抽出は既存研究がない。そのため、他の手法との比較をすることは出来ないが、実験結果から提案手法は、商品レビューから代替商品情報を自動抽出するための一つの方法として、その有効性を示すことができたのではないかと考える。また、代替可能と判断されたもののうち、

39%の割合で代替関係として意外だという評価を得られたことから、既知である代替商品の知識以外の、新たな知識をユーザに提示できる可能性を示すことができた。

表2 代替商品情報に対するアンケート調査結果

	平均	標準偏差1	標準偏差2
代替が可能	0.83	0.06	0.24
代替関係として意外	0.39	0.12	0.25

6. まとめ

考案した抽出ルールを商品レビューに適用し、商品間に存在する代替関係を自動抽出する手法を提案した。

実験結果から、抽出した代替商品情報のうち、83%の割合で代替が可能であるという評価を得ることができた。また、代替が可能だと判断された83%のうち、39%の割合で、代替関係として意外であるという評価を得られたため、既知である代替関係以外の、新たな知識をユーザに提示できる可能性を示すことができた。

今後の課題としては2点挙げられる。1つ目は、抽出ルールの改善である。代替商品情報を含むが、「代わりに」に係る文節がないため抽出を行うことができないパターンが見受けられた。そのため、他の文法上の特徴を見つけ出し、抽出ルールの改善を行う必要がある。2つ目は、他の語彙パターンにおける代替商品情報の抽出である。本研究ではすべてのレビューの中で、「代わりに」を含むレビューに対してのみ解析を行った。しかし、「代わりに」を含むレビューだけが代替商品情報を含むとは限らないため、他の語彙パターンにおける抽出方法も検討する必要がある。

謝辞

本研究には、楽天株式会社提供の公開データを利用して頂きましたことに、感謝致します。

文献

- [1] 本田達也, 北山大輔, 角谷和俊, “オンラインショッピングサイトにおけるレビューを用いた商品対の目的判定による商品推薦” 情報処理学会関西支部 支部大会