

ユーザー評価を用いた併用商品情報抽出システムの開発

岩淵 莉央† 中島 陽子‡ 本間 宏利‡ 増山 繁*

† 釧路高専 情報工学科

‡ 釧路高専 創造工学科

* 豊橋技術科学大学

honma@kushiro-ct.ac.jp

1 はじめに

近年、インターネットの普及により、気軽にインターネットショッピングで商品を購入する機会が増加している。ショッピングサイトは商品を実際に利用したユーザーのレビューを参照できるようになっており、商品を購入する際にはある程度の指標にすることが可能である。また、ショッピングサイトの機能としておすすめ商品の機能はあるが、ある商品と相性の良い商品を勧める機能を備えているショッピングサイトはまだ存在しない。

そこで、本研究では、任意の商品を入力とし、商品レビューを用いて入力商品と併用している商品(併用商品)の情報を抽出し、併用商品をユーザーに提示するシステムの開発を目的とした。

2 用語の定義

本研究の併用商品は化粧品に限定し、ユーザー評価は、化粧品紹介サイトの@cosme¹で各ユーザーが記述した商品レビューを用いる。レビューの内容から、併用表現、商品情報、評価表現の3つを定義し、併用商品の抽出を行った。商品レビューの例を以下に示す。

シャネルのCCクリーム50と一緒に使用しているからなのか、毛穴もないように見えるくらい綺麗に出来上がります!

「一緒に使用」のように、他の商品を併用していると判断できる表現を併用表現、「綺麗」といった商品情報に対する意見や評判、および、評価を表す表現を評価表現と定義する。さらに、商品カテゴリや「シャネル」などのブランド名、「CCクリーム50」といった商品名をまとめて商品情報と定義する。

¹<http://www.cosme.net>

@cosmeに登録されている商品情報から、商品名を発見できるように、以下の形式で商品情報リストを生成する。本研究では、実験のために併用表現は12種類用意し、商品は約15000種、ブランドを約3000社、カテゴリを123個取得した。

商品情報リスト

商品(商品名, URL, ブランド名, カテゴリ名),
ブランド(ブランド名),
カテゴリ(商品カテゴリ名)

3 実験

本研究は、OSにLinuxを用い、メインメモリ3000MBの仮想マシン上で動作させた。また、言語解析ツールのmecabとcabocha、HTMLを解析するBeautifulSoupを使用している。本項では、実験を行うために利用したツールについて説明する。

3.1 BeautifulSoup

BeautifulSoup²はPythonライブラリで、HTMLやXMLの要素を取得することができる。本研究では、商品情報やレビューなどを@cosmeから取得する際に用いた。

3.2 言語解析ツール

本研究ではレビューの内容を解析するため、形態素解析ツールのMeCab³と構文解析ツールのCaboCha⁴を用いた。

²<http://kondou.com/BS4>

³<http://taku910.github.io/mecab/>

⁴<http://taku910.github.io/cabocha/>

3.3 商品レビューの取得

本研究では、任意の商品名とその商品のカテゴリ、ブランドが入力として与えられた場合、商品レビューを用いて併用商品の商品名、ブランド、カテゴリを出力する。

併用商品を抽出するため、まずは入力した商品のレビューを取得する。レビューは Python ライブラリの BeautifulSoup を用いることでウェブページから取得する。商品情報を入力すると、商品情報リストにあるレビューの URL を抽出し、商品レビューをリストにして取得する。

3.4 併用商品の抽出

商品情報リストを用いて、レビューの内容から商品名、ブランド名、カテゴリ名、併用表現を探索する。

商品名と併用表現を発見した時、その商品名を併用商品とする。さらに、商品名ではなくブランド名とカテゴリ名を発見した場合、商品情報リストからそれらの情報を持った商品を抽出し併用商品とする。

3.5 併用商品の好評率

商品の相性が良ければ、レビューはその商品を褒める内容になる。そこで、評価表現を肯定と否定の二極に分類し、レビュー内に肯定の評価表現が多ければ、併用商品は相性が良いと判断する。評価表現の極性は東北大学の乾らが作成した、日本語評価極性辞書 [1][2] で定める。

さらに、全レビューで相性の判定を行った後、一つの併用商品に対してどれだけのレビューが相性が良いと判断したのかを示す割合を好評率と定義し、好評率が高いほど、併用商品は入力した商品と相性が良くなる。

3.6 実験結果

ランコム (ブランド名) のタンミラクコンパクト (商品名) を入力した実行結果の中で、カテゴリ名がリキッドファンデーションの併用商品を表 1 に示す。この商品のレビューは 620 件あり、そこから 192 種類の併用商品を抽出することができた。

取得した商品レビューを評価対象として手作業による併用商品抽出を行い、抽出した商品の精度を求めた結果は 54.84% となった。

表 1: 実行結果

リキッドファンデーション		
商品名	好評率	レビュー数
タンミラク リキッド	75%	8
コントロールカラー	50%	4
EAU DE TEINT	50%	4
	⋮	
パリュール アクア	0%	1

4 おわりに

本研究では、商品レビューから言語解析を用いて併用商品の抽出を行い、レビュー内の評価表現から併用商品の好評率を求めた。

実験の結果、併用商品抽出の精度があまり高くないことがわかった。考えられる原因として、例えばレビューを書いたユーザーがリキッドファンデーション (カテゴリ名) を、リキッドやリキッドファンデのように省略したため抽出できなかったこと、用意した併用表現が少なかったことが挙げられる。よって、これらの問題を改善することで、精度を向上させることが可能である。

また、今回は化粧品に限定して抽出を行ったが、洋服やコンピュータ部品なども商品同士の相性を持っているため、本システムを適用できるようにしていきたいと考えている。

謝辞

本研究は豊橋技術科学大学高専連携教育研究プロジェクトの助成 (2015-2016) を受けて行われました。

参考文献

- [1] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一, “意見抽出のための評価表現の収集”, 自然言語処理, Vol.12, No.3, pp203-222, 2005.
- [2] 東山晶彦, 乾健太郎, 松本裕治, “述語の選択選好性に着目した名詞評価極性の獲得”, 言語処理学会第 14 回年次大会論文集, pp.584-587, 2008.