

# 商品に対するレビュー情報を用いた商品特性の抽出

—ゴルフポータルサイトを対象として—

高橋 凌太† 生田目 崇‡ 大竹 恒平†

東海大学情報通信学部経営システム工学科† 中央大学理工学部経営システム工学科‡

## 1. はじめに

EC (Electronic Commerce) サイトの普及により、商品の購入や商品に関する情報取得が容易にできるようになった [1]. 中でも、商品の評価に点数を設けたレーティングサービスや、文章による商品評価 (レビュー) は、主要なコンテンツの一つである. 商品に関する文章や数値的評価は、商品の購入を検討している顧客にとって、貴重な情報源になる. 特にレビューには、実際購入した商品に対する感想や商品の使用感など、重要な情報が含まれており、購買意思決定に大きな影響を与えている [2].

本研究は、あるゴルフポータルサイトを対象にゴルフ商品に対して投稿されたレビューから商品特性を抽出することを目的とする. 具体的には、投稿されたユーザーレビューに対して自然言語処理を行い、商品評価のレビューに含まれる特徴的な表現 (単語) を特定する. これらの単語を用いて、商品に対する注目すべき項目を明らかにする.

## 2. データ概要

本研究では、ゴルフポータルサイトより提供を受けた商品レビューデータを使用する. なお、レビューデータに含まれる商品レビューは、サイトの会員により投稿されたものである.

データ期間は、2015年8月5日～2018年7月30日であり、総レビュー数は98,265件であった. 本研究では、商品特性を抽出するにあたり、ある程度のレビュー数を有する商品を対象とする必要があると判断し、最低10件以上のレビュー数を有する201商品を対象とした. なお、総レビュー数は6,231件であり、商品は7商品カテゴリ (アイアン, ウェッジ, ウッド, シューズ, パター, ボール, ユーティリティ) のいずれかに属するものであった.

## 3. レビューを用いた商品特徴の特定

本研究では、レビューデータに対して自然言語処理解析を行うことで、商品特性の抽出を試みる. はじめに、各商品に対して行われたレビュー全てを1つの文書としてまとめた. 次に、各文書に対して形態素解析を行い、品詞ごとに分解した. ここでは、形態素解析のための日本語辞書である MeCab を用いた. 形態素解析の結果を用い、本研究では、名詞、動詞、形容詞、形容動詞、固有名詞、地名、組織名の7つの品詞を抽出した. 次に、各商品の特徴を表すことを期待した単語 (特徴語) の特定を試みる. 具体的には、TF-IDF 法により、特徴語を特定する. TF-IDF 法とは、単語の出現頻度 (TF 値) 及び単語の逆文書頻度 (IDF 値) により、文書に含まれる単語の重要度を算出する手法である. TF-IDF 値は、次の式(1)～(3)で求める.

$$TF-IDF_{i,j} = tf_{i,j} \times idf_i \quad (1)$$

$$tf_{i,j} = \frac{n_{i,j}}{\sum_S n_{S,j}} \quad (2)$$

$$idf_i = \log \frac{|D|}{|\{d: d \in t_i\}|} \quad (3)$$

ここで、 $n_{i,j}$  は、文章  $j$  の単語  $i$  に関する出現頻度である.  $\sum_S n_{S,j}$  は、文章  $j$  内のすべての単語の出現頻度、 $|D|$  は総文書数である.  $|\{d: d \in t_i\}|$  は、単語  $i$  を含む文書数である. 本研究では、各文書において TF-IDF 値の上位 20 単語を特徴語として定めた. 次に、それぞれの商品に関する特徴語について、その単語単体での意味を筆者が判断し、以下の5つのゴルフに関連するタグ付けを行った (表 1).

表 1 単語の意味を表すタグ一覧

タグ種別	タグの概要
性能	商品の性能に関する特徴語
デザイン	商品の外観やメーカーに関する特徴語
動作	使用感、利用時の動作に関する特徴語
価格	価格に関する特徴語
アイテム	関連する他のゴルフアイテムに関する特徴語

Extraction of product characteristics using review information -Targeting on the golf portal site-

†Ryota Takahashi, Kohei Otake

School of Information and Telecommunication Engineering, Tokai University

‡Takashi Namatame

Faculty of Science and Engineering, Chuo University

表2に、アイアンカテゴリに属する商品A及びBについて、各タグに分類されたTF-IDF値が上位の特徴語を示す。

表2 アイアンカテゴリの特徴語

タグ種別	アイアンA	アイアンB
性能	球筋, トップ, 上がる, 早い	なし
デザイン	モデル, ヘッド	後継, モデル
動作	打てる, 諦める, 比べる, 振り回す, 振る	引っ掛ける, 全く, 評判
価格	安い	買い換える, 買える, お買い得
アイテム	ボール, ドライバー	なし

表2を用いて、アイアンカテゴリの2つの商品特性を比較する。デザインタグに着目すると、アイアンAは「ヘッド」という単語が抽出された。即ち、アイアンAのヘッド部分に関連するレビューが多いことが推察される。他方で、アイアンBでは「後継」という単語が抽出された。これは、過去に発売された同モデルの後継との比較に関連するレビューが多いことが推測される。性能タグについては、アイアンAのみ「球筋」や「トップ」といった、商品評価に関する特徴語が抽出された。価格タグに関しては、アイアンAでは、「安い」という価格に関連する単語のみが抽出されたが、アイアンBでは「買い替える」、「お買い得」というより具体的な購買行動に関連した単語が抽出された。

以上の結果より、アイアンAは性能や動作といった商品特性が、アイアンBは価格特性が、それぞれの商品に対するレビュー中の特徴語から抽出された。

表3は、シューズカテゴリに属する商品C及び商品Dについて、各タグに分類されたTF-IDF値が上位の特徴語を示す。なお、アイテムタグの特徴語はどちらの商品でも抽出されなかった。

表3 シューズカテゴリの特徴語

タグ種別	シューズC	シューズD
性能	最新	通気, 風通し, 軽量
デザイン	ミズノ, 品質, スパイク	メッシュ, モデル, 夏場
動作	硬い, 慣れる, 足首	涼しい, ベスト, 欲しい
価格	コスパ, 高い	なし

表3を用いて、シューズカテゴリの二つの商品特性を比較する。性能タグに注目すると、シューズCは、「最新」という単語が抽出された。一方、シューズDでは、「通気」、「風通し」、「軽量」という性能をより具体的に表現する単語が抽出されており、シューズDを評価したレビューには、重量感と通気性に言及しているものが多いことが推察される。さらにシューズDのデザインタグに注目すると、「メッシュ」、「夏場」という特徴語が、動作タグでは、「涼しい」といった特徴語が抽出されており、タグ同士の関連性から、「夏場でも使い心地の良い、明瞭・通気性に特徴を持ったシューズ」といった、商品特性が伺える。

一方で、シューズCのデザインタグに注目すると、「ミズノ」、「品質」といった特徴語が抽出されており、メーカーや商品の品質に関連するレビューが多いことが推測される。また、シューズCの価格タグでは、「高い」という単語が抽出されており、価格に対して必ずしも肯定的なレビューばかりではないことが推察される。

以上の結果より、シューズCでは価格特性が、シューズDでは性能、デザイン、動作の特性に関連した商品特性が、それぞれの商品に対するレビュー中の特徴語から抽出された。

#### 4. まとめおよび今後の課題

本研究では、ゴルフポータルサイトの会員により投稿された、ゴルフ商品についてのレビューデータを用いた商品特性の抽出を試みた。アイアンとシューズの2種類の商品に対して投稿されたレビュー文を用い、TF-IDF法により特徴語を定め、特徴語を用いた商品特性の比較を行った。その結果、各商品が有する商品特性の一端が示された。今後は、特徴語の解釈を目的とした感情分析を行い、商品特性の詳細について明らかにする予定である。

#### 謝辞

本研究では JSPS 科研費 19K01945 および 17K13809 の助成を受けたものです。

#### 参考文献

- [1] 経済産業省, 電子商取引に関する市場調査 <https://www.meti.go.jp/press/2019/05/20190516002/20190516002.html> (2020/01/09 閲覧)
- [2] 江川 雄太, 一藤 裕, 今野 将, “ECサイトにおけるユーザレビューが購買行動に与える影響に関する研究について,” 情報処理学会第73回全国大会予稿集, pp. 591-592 (2011).