

レビューテキストに基づく感性の自動指標化による 印象の全体部分関係の構造化

Structuring the Part-whole Relationship of Impression
by Automatic Indexing of Kansei based on Review Texts

小向井美来^{*1}

Miku Komukai

橋本翔^{*1}

Sho Hashimoto

長田典子^{*1}

Noriko Nagata

^{*1}関西学院大学

Kwansei Gakuin University

In recent years, useful information for product design has been extracted from online reviews that describe evaluations or opinions on products. Since the product evaluation on these reviews often consists of partial impressions and the whole impression that synthesize them, the user's impression of the product has a part-whole relationship. If one or only a few parts relate to the whole impression, the part that determines the whole impression is clear. However, in the case of products consisting of many parts, since each impression of multiple parts affects the whole impression, the parts attributed to changing the whole impression are not clear, and this makes it difficult to achieve an arbitrary impression on the design. In this study, we propose a method of structuring the part-whole relationship of impressions based on reviews in order to acquire product knowledge that easily achievable on the product design. The method consists of 4 steps; (1) extracting the main whole impression in the target product domain; (2) extracting the main impression of each part in the target product domain; (3) analyzing the impact of the main impressions of each part on the main whole impressions; and (4) extracting elements related to the main impression of each part. Analysis of the actual data has shown that some main whole impressions of the product can be estimated from the partial impressions.

1. はじめに

機械学習やテキストマイニングなどの分野では、プロダクトに対する感想や意見などが記述された Web 上のレビューから、プロダクトに対する有用な情報が抽出されている。例えば、レビューからユーザが対象に感じた評価極性（肯定的/否定的）を推定する感情分析 (Sentiment Analysis) [Medhat 17][Fang 15] や、その評価極性の要因となる対象の具体的な注目点を抽出する Aspect Extraction [Poria 16] などが存在する。これらは感情に着目し、プロダクトに対する有用な情報を抽出している。

特に、プロダクトデザインに反映させやすい知識を獲得するための分析手法として、心理量と物理量をつなぐ感性の階層構造を仮定し、印象を推定する分析手法が存在する [Hashimoto 19][山田 18][山田 19]。これらの研究では、印象に着目し、感情はプロダクトに対する印象を介して喚起されると仮定することでプロダクトデザインへの反映を容易にしている。

また、評価には階層性がある [讃井 86][讃井 87]。プロダクトの評価を例に挙げると、靴の場合、「色がシンプルで衣服と合わせやすく便利」や「靴底が柔らかく足に負担がかからず便利」といったように、色のシンプルさや靴底の柔らかさから便利という印象が喚起されており、シンプル・柔らかいと便利の間に印象の階層性が存在する。オントロジーやシソーラスなどの分野では、概念の階層構造の 1 つであり、下位概念が上位概念に包含され、下位概念が上位概念の部分にあたる全体部分関係について研究されている。全体部分関係はプロダクトの評価においても存在する。上記の例の場合、色のようなデザインや靴底のような機能が部分にあたり、それらの評価を総合した靴の総合評価がある。そのため、プロダクトに対する印象の評価を分析する際には、部分が全体的な評価に与える影響を考慮す

る必要がある。

そこで、本研究ではプロダクトデザインに反映しやすい知識獲得のため、プロダクトを細分化した部分の評価と、プロダクト全体の評価の関係をレビューに基づき構造化する手法を提案する。具体的には図 1 に示すような 3 つの階層から構成される感性の階層構造を仮定する。感情は「人の事物に対する反応や、その結果生じる態度・行動」、印象は「事物の性質がどのようなものであるかの評価」、形態要素は「対象物そのものや、それを構築する要素」を示す。本研究では中間層の印象層において全体部分関係の構造化を行うことで、部分の評価と全体の評価の関係を明らかにする。

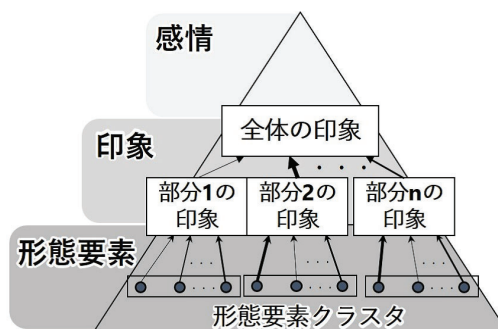


図 1: 感性の階層構造における印象の全体部分関係の構造化

2. 関連研究

プロダクトに対する有用な情報を抽出するため、レビューを用いた感情分析 (Sentiment Analysis) や Aspect Extraction などが行われている。しかし、評価極性の推定や、トピック抽出による評価極性の要因となる対象の抽出などにより得られ

連絡先: 関西学院大学理工学部人間システム工学科

〒669-1337 兵庫県三田市学園 2 丁目 1 番地

E-mail: nagata@kwansei.ac.jp

る情報は、形態要素から直接的に感情につながる要因や評価を推定するものである。そのため、ユーザは対象をどのように評価しているかが明らかでなく、得られた知識をプロダクトデザインに直接反映することは困難である。そこで Hashimoto et.al.[Hashimoto 19] は、感情と形態要素の間に印象層を仮定することで、プロダクトデザインに反映させやすい知識の獲得を可能にした。しかし、Hashimoto et.al. は少数の部分から構成されるプロダクトを対象としており、複数の部分からなるプロダクトの場合、複数の部分が全体の印象に影響するため印象変化に寄与する部分が明解ではなく、デザインに反映させる十分な知識が獲得できない。

また、概念間やことば間にはさまざまな階層構造が存在する[細野 00] が、その中でよく取り上げられる関係の 1 つとして全体部分関係がある。全体部分関係は下位概念が上位概念に包含される階層構造で、下位概念が上位概念の部分にあたる構造である。全体部分関係を明らかにする研究はオントロジーやターミノロジー、シソーラスなどの分野で行われている。オントロジーの分野では、全体部分関係は部分概念、部分概念のクラス制約、部分概念の役割を表すロール概念の 3 つから構成されている[溝口 99][細野 00]。またターミノロジーの分野では、部分概念が包括概念との対比で明確にとりあげられており、本質的部分概念、非本質的部分概念、区別特性を持つ部分概念の 3 つの特徴をあげている[ISO 00]。シソーラスの分野では、身体組織と器官、地理的場所、学問分野、階層的社会的構造の 4 つの明示的な全体部分関係がある。しかし、これらの分野では印象などの評価の全体部分関係を扱っていない。一方、全体部分関係に着目した研究は心理学の分野でも行われている。心理学の分野では印象の形成に関して、いくつかの部分印象から全体印象がどのように形成されるかについて研究されている。浅川ら[浅川 93] は、印象が異なる複数の部分から構成される緑地において、全体イメージがどのように形成されるかを事例的に検討した。方法としては SD 法を用い、緑地利用者に対する現地での質問紙調査及び学生を対象とした室内におけるスライドテストによってイメージ調査を行った。これによって全体の印象の判断には、季節や評価尺度によっても異なるが、多くの場合、強い印象を与える部分印象が深く関与することを明らかにした。このように印象が異なる複数の部分が、全体に包含される場合、対象の評価において各部分の印象が影響する。そのため、各部分印象を統合した全体印象によって対象の評価を判断しなければならない。

3. 提案手法

手法は(1)対象プロダクトドメイン全体の主要な印象の抽出(2)対象プロダクトドメインにおける各部分の主要な印象の抽出(3)各部分の主要な印象が全体の主要な印象に与える影響分析(4)各部分の印象に関連する形態要素の抽出、という 4 つのプロセスから構成されている。(1)(2)(3)は印象層の構造を分析し、(4)は印象層と形態要素層との関係性を明らかにする。

3.1 全体印象の指標化

全体印象の指標化には Hashimoto et.al.[Hashimoto 19] の手法を用いる。Hashimoto et.al. は、レビュー内の評価語には、評価者の感情を表現する語と対象の持つ印象を表現する語が混在しているが、アプレイザル理論[佐野 11]における、内評価/外評価という評価語のカテゴリ情報に基づいて評価語を分類することで、対象の印象のみを分析することができ、感情表現を除いた客観的な指標化を可能としている。指標化の具

体的な方法としては、はじめに複数の評価表現辞書を用いて評価語の収集・分類を行う。次に、得られた印象語のみを入力としたトピック抽出により、対象プロダクトドメインにおける主要な印象を抽出する。最後に、各プロダクトに対して単語の出現頻度と term-score、単語の分散表現によって得られる印象語間の類似度に基づき、印象トピックごとの得点(以下、印象得点)を算出するというものである。トピック抽出の際、入力とする印象語については、ほとんどのプロダクトに対するレビューで用いられているような印象語と、レビューでの出現頻度が過度に少ない印象語は、対象プロダクトドメインを特徴づける印象語として適していないため、入力から除外する。各レビューの印象得点の具体的な算出方法は、まず、レビュー内で出現する各印象語の当該トピックにおける term-score を合計した値を、そのトピックにおけるレビューの得点として算出する。また、1 つのレビュー内で同じ印象語が複数回記述されている場合には、その印象語の term-score を回数分足し合わせる。さらに、term-score 上位に含まれるが、レビュー内に出現しなかった印象語に関しては重相関係数を計算し、それを印象語の出現頻度の代わりとして用いるというものである。プロダクトの印象得点は、そのプロダクトに対して書かれた各レビューの印象得点の平均をとることで算出する。

3.2 部分印象の指標化

部分印象の指標化には対象プロダクトドメインの、評価における部分にレビューを分離し、各評価部分において Hashimoto et.al.[Hashimoto 19] の手法を用いる。

3.3 全体印象と部分印象の構造化

重回帰分析を用いて全体印象と部分印象を構造化する。目的変数は対象プロダクトドメインに属する各プロダクトの全体印象の得点、説明変数は対象プロダクトドメインに属する各プロダクトの部分印象の得点とする。なお各得点は 3.1, 3.2 で付与されたものを用いる。印象得点は各印象ごとにスケールが大きく異なるため標準化を行う。また、プロダクトによってはレビューで評価されていない部分があり、部分印象の得点に欠損値が存在するため欠損値補完を行う。欠損値補完の方法としては、連鎖方程式に基づく多重代入法(multiple imputation by chained equation; MICE)[buuren 10][野間 17]を用いる。連鎖方程式に基づく多重代入法は、対象となる解析データセットにおいて、複数の変数にまたがって非単調な欠損が生じたときに、全ての利用可能なデータを用いて多重代入法を実行できる方法である。多重代入法における補完値の生成方法は予測平均マッチング(predictive mean matching; pmm)を用いた。

3.4 形態要素の指標化

印象語と名詞の係り受け関係に基づき、各部分印象に関連する形態要素を抽出する。係り受け関係は日本語構文・格・照応解析システム KNP Ver4.19 を用いて推定する。具体的には、各部分印象トピック term-score 上位の印象語に係っている名詞と係られている名詞を抽出する。抽出された名詞が各部分の印象に関連する形態要素である。ただし、名詞には同じ意味や似た意味の単語が多くあるため、それらを統合する必要がある。そのため、まず Word2Vec を用いてレビュー内の各単語を分散表現にする。その後、各名詞間のコサイン類似度を距離に変換したデータを入力としたクラスタリングを行う。なお、-1 から 1 までの値で表現されるコサイン類似度を 1 から減じることで 2 から 0 の距離に変換する。また、クラスタリングはワード法によって階層型クラスタリングし、クラスタ数はクラスタの集中度と孤立度に関する指標によって決定する。これにより、名詞の統合を行う。

4. 実データの分析

4.1 対象プロダクトドメイン

本実験の対象プロダクトドメインはバイクのレビューとした。その理由は、形態要素が多く、デザイン面や機能面など評価される部分が多いためである。レビューは価格.com[価格.com]にて2020年1月15日までに書かれたものを用いた。分析対象としたバイクは70種で、対象とするレビューの数は1,655件となった。形態素解析の結果、総単語数は234,929語、総語彙数は21,908語となった。また、評価部分はデザイン、エンジン性能、走行性能、乗り心地、取り回し、燃費、価格の7つとした。

4.2 全体印象の指標化

収集された評価語は2145語で、そのうち印象語は1644語となった。これらの印象語を入力としたトピック抽出を100回行い、perplexityからモデルを決定した。その結果、「便利」「シンプル」の2つの頑健な全体印象トピックが抽出された。ただし、今回はHDP-LDAを100回実行し、他の99モデルの20%以上に同様の解釈ができるトピックがある場合、頑健性の高いトピックと判断した。同様の解釈ができるかを判断する基準としてはterm-score上位10語の印象語中5語以上が共通することと定めた。

4.3 部分印象の指標化

各評価部分のレビューから収集された評価語数と、印象語数を表1に示す。部分別にこれらの印象語を入力としたトピック抽出を100回行い、perplexityからモデルを決定した。その結果、デザインでは「高い」「大きい」「オーソドックス」、エンジン性能では「スムーズ」「粘り」、走行性能では「安定」、乗り心地では「快適」「柔らかい」、取り回しでは「大きい」「低い」「安定」「操作性」、燃費では「悪い」「延びる」「想像以上」、価格では「得」「妥当」の全17の頑健な部分印象トピックが抽出された。また、頑健性の判断は4.2と同様の基準で行った。

表 1: 各評価部分の評価語数と印象語数

評価部分	デザイン	エンジン性能	走行性能	乗り心地	取り回し	燃費	価格
評価語数 (語)	572	697	658	564	398	309	348
印象語数 (語)	433	530	496	419	293	252	267

4.4 全体印象と部分印象の構造化

全体印象トピック「便利」を目的変数、4.3の全17の部分印象トピックを説明変数とした重回帰分析を行った。全体印象トピック「便利」のterm-score上位5語および重回帰分析の標準解において、全体印象トピック「便利」に対して絶対値の大きい係数を持つ部分印象トピックとそれらのterm-score上位5語を表2に示す。ここで、(-)は各評価語の否定的表現であることを示す。例えば「悪い(-)」は、「悪い」という単語について「悪くない」や「悪くはない」などのような使われ方をしている場合に、(-)を付与している。この結果から、バイクの便利さは乗り心地が強く影響することが示唆された。また、走行性能と取り回しの2つの評価部分において「安定」という印象トピックが抽出されているが、バイクの便利さは走行性能の「安定」に強く影響されることが示唆された。ただし、変数の数に比べてデータサイズが小さいため、今回報告する数値は多重共線性による影響により不安定であると考えられる。今後はデータの拡充方法を考える必要がある。

4.5 形態要素の指標化

4.3の全17の部分印象トピックに関連する形態要素を係り受け関係に基づき抽出した。その中で4.4において全体印象ト

表 2: 全体印象に影響度の高い部分印象

	トピック名	係数	term-score 上位 5 語
全体印象トピック	便利	-	静かだ、便利だ、改善、太い、妥当だ
部分印象トピック	乗り心地の快適	1.034	大きい、快適だ、強い、悪い(-)、きつい
	乗り心地の柔らかい	0.946	高い、柔らかい、高速、良い(-)、同じだ
	デザインの高い	0.572	高い、安っぽい、小さい、シャープだ、逆だ
	走行性能の安定	0.567	安定、高速だ、走れる、高速、安定感

ピック「便利」に影響力の高かった4つの部分印象トピックの結果を表3に示す。また、対象のレビューに含まれる名詞は17,885語だった。ここからレビューの40%以下にしか含まれていない単語と固有名詞を除去した結果、名詞のクラスタリングに用いる単語は270語であった。これらをクラスタリングした結果、44の形態要素クラスタが得られた。表3に関係する17の形態要素クラスタを表4に示す。乗り心地の快適にはエンジン音や排気音などの音、シートやライダーの姿勢などのポジションが関連することが明らかとなった。

表 3: 部分印象トピックに関連する形態要素クラスタ

部分印象トピック	term-score 上位の語	関連する形態要素クラスタ
乗り心地の快適	大きい	音 (7), ボジション (3), サスペンション (3), スクリーン (2)
	悪い (-)	ボジション (2)
	きつい	自分 (2)
乗り心地の柔らかい	高い	停車 (9), 足付き (5), ボジション (5), イメージ (3), 噂 (2)
	柔らかい	サスペンション (13), ボジション (3), タンデム (3)
	良い (-)	ボジション (2), タンデム (2)
	同じだ	ボジション (2)
走行性能の安定	安定	コーナー (6), 慣らし (4), 峠 (3), 速度 (2), サスペンション (2), 重量 (2)
	高速だ	慣らし (6), 車体ジャンル (5), 速度 (5), 峠 (2)
	走れる	峠 (4), 車体ジャンル (2), コーナー (2), 重量 (2), イメージ (2)
デザインの高い	高い	停車 (3)
	安っぽい	イメージ (2)
	シャープだ	イメージ (5), デザイン (2)

表 4: 名詞クラスタリング結果

形態要素クラスタ	名詞クラスタリング結果
音	エンジン音, 排気音, 振動, 音, 辺り
ボジション	前傾, 体, ボジション, 長時間, 姿勢, 腰, 腕, 長距離, ライディング, シート, 尻
サスペンション	サスペンション, セッティング, サス, 前後, 調整, ギャップ, 設定, リア, 段差, 路面, フロント
スクリーン	効果, スクリーン, 風圧, カウル, 風
自分	自分, 僕, 私, 方, 人
停車	停車, 位置, ステップ, 信号待ち, ハンドル, ミラ
足付き	足付き, 身長, シート高, 足つき, 足, 月, 両足, 膝, 女性
イメージ	感じ, イメージ, 印象, 感覚, 感
噂	話, 他, 車両, 笑, 所, 何, ネット, 事, 点
タンデム	レベル, タンデム, 事 (-), 乗り心地, 問題 (-)
コーナー	コーナリング, カーブ, コーナー, コーナーリング, ハンドリング
慣らし	慣らし運転, 距離, km, 通常, 走行, 慣らし, 走行距離
峠	峠道, ワインディング, 操作, 走り, 峠, 運転, 道, 乗り
速度	速度, 巡航, 高速道路, 限界, 追い越し, km/h, スピード, 流れ
重量	重量, 車重, 軽量, 車体, バランス, コンパクト, 取り回し, お陰, 重心
車体ジャンル	バイク, マシン, 理由, オススメ, 基本, 大型バイク, リッターバイク, タイプ, スクーター, 選択, 初心, 中, バイク (-)
デザイン	デザイン, 個人, 目, 色, 語り, エンジン性能, ヘッドライト

5. おわりに

本研究では、プロダクトの評価の全体部分関係をレビューに基づき構造化する手法を提案した。手法は4つのプロセスからなっており、まず、主要な全体印象の抽出を行い、次に主要な部分印象の抽出を行う。その後、抽出された印象を重回帰分析により構造化することで、全体印象に影響力のある部分印象を

明らかにし、最後に各部分印象に関係する形態要素を明らかにする。これら4つのプロセスによりレビューから自動的に全体部分関係の構造化が達成され、その結果を解釈することで、プロダクトデザインに反映しやすい知識を獲得できる。バイクのレビューを対象とした実験の結果から、いくつかの全体印象は部分印象から推定可能であることが示され、本手法の有効性が示された。今後の課題は、データを拡充する方法の検討と対象プロダクトドメインの評価における部分を自動で抽出することである。

参考文献

- [Hashimoto 19] Hashimoto, S., Yamada, A., Nagata, N: A Quantification Method of Composite Impression of Products by Externalized Evaluation Words of the Appraisal Dictionary with Review Text Data, *International Journal of Affective Engineering*, 2019, vol 18, no. 2, p. 59-65.
- [山田 18] 山田篤拓, 橋本翔, 長田典子: レビューデータを用いた評価表現辞書に基づく印象の自動指標化, *日本感性工学学会論文誌*, 2018, vol 17, no. 5, p. 567-576.
- [山田 19] 山田瑠奈, 橋本翔, 山田篤拓, 長田典子: Twitter データを用いたアプレイザル評価表現辞書に基づく印象の自動指標化, *人工知能学会全国大会論文集 一般社団法人 人工知能学会*, 2019, p. 3Rin239-3Rin239.
- [讃井 86] 讃井純一郎, 乾正雄: レポートリー・グリッド発展手法による住環境評価構造の抽出: 認知心理学に基づく住環境評価に関する研究 (1), *日本建築学会計画系論文報告集*, 1986, vol 367, p. 15-22.
- [讃井 87] 讃井純一郎, 乾正雄: 個人差および階層性を考慮した住環境評価構造のモデル化: 認知心理学に基づく住環境評価に関する研究 (2), *日本建築学会計画系論文報告集*, 1987, vol 374, p. 54-60.
- [ISO 00] ISO704: Terminology work-Principles and methods, *International Organization for Standardization*, 2000.
- [溝口 99] 溝口理一郎, 池田満, 来村徳信 and others: オンントロジー工学基礎論: 意味リンク, クラス, 関係, ロールのオンントロジー的意味論 (「オンントロジーの基礎と応用」), *人工知能学会誌*, 1999, vol 14, no. 6, p. 1019-1032.
- [細野 00] 細野公男: 概念間の関係に関する一考案: 全体-部分関係を中心に, *情報知識学会誌*, 2000, vol 10, no. 1, p. 28-39.
- [浅川 93] 浅川昭一郎, 渡辺大介, 首藤健一: 多面性を有する緑地のイメージ構成に関する事例研究, *造園雑誌*, 1993, vol 57, no. 5, p. 307-312.
- [佐野 11] 佐野大樹: 日本語アプレイザル評価表現辞書 (JAppraisal 辞書) ~能動評価編~version 1, *言語資源協会*, 2011.
- [buuren 10] Buuren, S van., Groothuis-Oudshoorn, Karin: mice: Multivariate imputation by chained equations in R, *Journal of statistical software*, 2010, p. 1-68.
- [野間 17] 野間久史:連鎖方程式による多重代入法, *応用統計学*, 2017, vol 46, no. 2, p. 67-86.
- [価格.com] 価格.com 自動車・バイク バイク 新車・中古車. 価格.com. <https://kakaku.com/bike/>, (参照 2020-01-15).
- [Medhat 17] Medhat, W., Hassan, A., Korashy, H: Sentiment analysis algorithms and applications: A survey, *Ain Shams Engineering Journal*, 2014, Vol 5, no. 4, p. 1093-1113.
- [Fang 15] Fang, X., Zhan, J: Sentiment analysis using product review data, *Journal of Big Data*, 2015, Vol 2, no. 1, p. 5.
- [Poria 16] Poria, S., Cambria, E., Gelbukh, A: Aspect extraction for opinion mining with a deep convolutional neural network, *Knowledge-Based Systems*, 2016, vol 108, p. 42-49.