

テキストマイニングによる ドメイン共通・固有の感性価値の構造化

後藤 杏奈[†] 橋本 翔^{†‡} 長田 典子[†]

[†] 関西学院大学理工学部/感性価値創造インスティテュート 〒669-1330 兵庫県三田市学園上ヶ原 1 番

[‡] 西南学院大学商学部 〒814-8511 福岡市早良区西新 6-2-92

E-mail: [†] {gotoanna.nagata}@kwansei.ac.jp, ^{†‡} s-hashimoto@seinan-gu.ac.jp

あらまし 本研究は、ドメインに共通・固有の感性価値をレビューデータから抽出する手法を提案する。まず階層的クラスタリングを用いて、印象語の出現率をもとに共通の価値を持つ複数ドメインのクラスタに分類する。次に、全てのクラスタに共通している印象語を用いて、LDA でトピック抽出することでクラスタ共通の感性価値の抽出が可能となる。これらを再帰的に繰り返すことにより、ドメイン共通・固有の価値が明確化され、価値の階層構造化を図ることができる。

キーワード 自然言語処理, 感性工学, クラスタリング, トピック抽出

Structuring domain-common and domain-specific emotional values through text mining

Anna GOTO[†] Sho HASHIMOTO^{†‡} Noriko NAGATA[†]

[†] School of Science and Technology / Kwanse Gakuin Institute of Kansei Value Creation, Kwanse Gakuin University 1 Uegahara, Gakuen, Sanda-shi, Hyogo, 669-1330 Japan

[‡] Faculty of Commerce, Seinan Gakuin University, Nishiara, Hukuoka-shi, Hukuoka, 814-8511 Japan

Abstract This study proposes a method for extracting emotional values common and unique to domains from review data. First, hierarchical clustering is used to classify the data into clusters of multiple domains that share common values based on the occurrence rate of impression words. Next, using the impression words common to all the clusters, LDA is used to extract topics, which enables the extraction of sensory values all of the clusters share. By repeating these steps recursively, the common and unique values of the domains can be clarified, and a hierarchical structure of values can be created.

1. はじめに

プロダクトデザインの分野においてユーザのニーズを把握し、プロダクト設計に反映させることは重要である [1]. 特に、ユーザのニーズや好みの多様化に伴い、機能や価格、信頼性といった従来のものづくりにおける価値に加えて、使用感や快適性などの感性的な価値にも注目が集まっている [2]. 経済産業省が取りまとめた「感性価値創造イニシアティブ」[3] では、「生活者の感性に働きかけ、感動や共感を得ることで顕在化する価値」を「感性価値」として定義し、感性を活用したモノづくりを推進している。この感性的な価値を扱う手法として感性工学の分野では、特定のドメインに特化した感性の自動指標化が行われている。

しかし、多種多様なプロダクトが市場に溢れている現在、特定のドメインに特化した指標化は、ドメインごとに一から指標化を行う必要があるため、人的・時間的負荷が高く、汎用性が低い。そのため、異なるドメインでも適用可能な感性の指標化の必要性が高まっている。例えば、家電や衣服などの異なるカテゴリの

ドメインにおいても、デザイン性や質感などが共通の価値として考えられる。このような共通の価値を抽出することで、より広範囲なドメインでの感性価値の適用が可能になる。また、ドメインの価値からドメイン共通の価値を除くことで、ドメインで重視されている固有の感性価値を明らかにできる。こうしたドメイン共通・固有の価値を抽出し、構造的に理解することは、さまざまな価値創造の場面で有用であると考えられる。

そこで本研究では、階層構造を仮定した広範囲なドメインを対象として、ドメインに共通・固有の感性価値の構造化を行う手法を提案する。

2. 関連研究

価値に関する研究は、古くから経済学、心理学、社会学等の分野で行われてきた。近年の代表的なものとして、Schwartz ら[4]の価値の理論的枠組みの研究がある。Schwartz らは個人に関わる価値を 19 に分類し (図 1)、関連の強い価値を近隣に、競合する価値を反対側に配置している。例えば、左上に普遍、その隣に慈悲、

3.1. 印象に基づいたドメインの構造化

3.1.1. 評価対象の抽出

印象に基づいたドメインの構造化を行うにあたり、まずレビューテキストの前処理を行い、評価対象（印象語とその対象）を抽出する。

評価対象（印象語とその対象のペア）を抽出するにあたり、まずGINZAの係り受け分析を用いて、修飾関係のペアを抽出する。次に、得られた修飾関係のペアには印象語以外のペアも多く含まれるため、評価表現辞書[7]を用いて、印象語とその修飾関係にある単語のペアのみを抽出する。また、表記揺れの統一を行うため、抽出した修飾関係のペアをjuman++[8]を用いて正規表現に変換する。最後に出現率の低い評価対象はノイズになるため、得られた評価対象のうち出現率5%以下の評価対象は除く。

3.1.2. 非負値行列因子分解を用いた行列の次元削減

3.1.1で得られた評価対象の出現率を値とする非負値行列 X ($m \times n$)を作成する。 m は製品、 n は評価対象である。出現率は、製品の評価対象の出現回数を製品数で割った値である。この行列を非負値行列因子分解[9]を用いて次元削減を行う。非負値行列因子分解は、特徴空間の低ランク近似に基づく次元削減手法で、データの情報を圧縮し、重要な特徴を抽出できる。

またランク数は、事前に決める必要がある。通常、ランク数 k は観測ベクトルの次元 n やデータの個数 m より小さく設定される。ランク数決定の指標には、cophenetic coefficientという係数を用いる[10]。この値が大きくなり低下し始める最小値をランク数とする。

3.1.3. 階層的クラスタリングによるドメインの構造化

非負値行列因子分解により、行列 X は低次元の行列 W と H に分解される。行列 W は、行が製品 m 、列がランク数 k からなる行列である。この行列 W に対して、階層的クラスタリングを行う。階層的クラスタリングには、最短距離法を用いる。この手法を用いて、レビュー内で出現する評価対象が類似している製品を同じクラスター（サブドメイン）に分類することで、印象に基づいたドメインの構造化を行う。以下クラスターをサブドメインと呼ぶ。

次に、各クラスターのcoefficientの値が高い評価対象を確認する。全ての評価対象のうち、上位30%で全てのサブドメインに共通して出現している評価対象を、ドメイン共通の評価対象とする。

3.2. ドメイン共通の感性価値の抽出手法

本章では、ドメイン共通の印象トピックを抽出する。そのため、3.1.3で得られたドメイン共通の評価対象を

用いて、トピック抽出を行う。

3.2.1. 行列の作成

トピック抽出を行うにあたり、行列を作成する。まず、3.1.1で得られた各製品の評価対象を使用する。これらの評価対象から、3.1.3で得られたドメイン共通の評価対象のみを抽出する。各製品のドメイン共通の評価対象を行とした行列を作成する。

3.2.2. LDAを用いた印象トピックの抽出

3.2.1で作成した行列をBow (Bag Of Words)を用いて、ベクトル化する。このベクトル化した行列を入力として、印象トピックの抽出を行う。ドメイン共通の評価対象のみを入力としているため、得られたトピックはドメイン共通の印象トピックとなる。Bowは、単語の出現回数に基づいたベクトル化手法である。またトピック抽出には、LDA (Latent Dirichlet Allocation)を用いる。LDAはトピック分布にディレクレ事前分布を仮定し、ベイズ推定する手法であり、文書中の単語の「トピック」(話題やカテゴリ)を確率的に求める言語モデルである。文書中の各単語が潜在的なトピックから生成されていると仮定して、そのトピックを文書集合から教師無しで推定することができる。

3.2.3. 価格.comの評価項目と比較

価格.comの評価項目には、製品ごとに口コミの満足度を表す評価項目が設定されている。例として、自動車の評価項目を図3に示す。この価格.comの対象ドメインの評価項目と、3.2.2で得られたドメイン共通の印象トピックの比較を行う。具体的には価格.comの対象ドメインの共通の評価項目と、本手法の分析結果であるドメイン共通の印象トピックが一致しているか、または共通項目として価格.comに足りていない項目がないか考察を行う。

満足度	★★★★★ 4
エクステリア	★★★★★ 4
インテリア	★★★★☆ 3
エンジン性能	★★★★★ 4
走行性能	★★★★★ 5
乗り心地	★★★★★ 4
燃費	★★★★☆ 3
価格	★★★★☆ 3

図4. 価格.comの評価項目

3.3. 再帰的適用による感性価値の構造化手法

本研究は、様々なドメインを対象に感性価値を構造化することが目的である。3.1では印象に基づいてドメインの構造化を行い、3.2ではドメイン共通の印象

トピックを抽出した。本章では、3.1のドメインの構造化により、同じサブドメインに分類された製品に対し、3.1、3.2の手法を同様に適用する。この手法を各層（ n 層・ $n+1$ 層・ $n+2$ 層...）に再帰的に適用する。

ただし、3.1、3.2の手法を各層に適用する際に、前処理を加える必要がある。例として、 n 層の価値を抽出する際、第1層から第 $n-1$ 層で得られたドメイン共通の評価対象を除き、3.1、3.2の手法を適用する。 $n+1$ 層でも同様に、第1層から第 n 層で得られた共通の評価対象を除いて、手法を適用する。このように前層までのドメイン共通の評価対象を除いて分析することで、ドメイン固有の価値を明確化できる。この一連の手法を各層に対して再帰的に繰り返すことで、ドメイン共通・固有の価値が段階的に明確化され、感性価値の階層構造化を図ることができる。これにより、どの層においても感性価値を抽出でき、構造化を行えることを示し、本手法の有用性を検証する。

4. 実データ分析

4.1. 対象ドメイン

本実験では、日本標準産業分類（総務省、令和5年6月改定）[11]より、製造業の下位分類項目である「機械器具」を対象ドメインとする。実際には価格.com[12]より、対象ドメインに含まれる「自動車・バイク・自転車・掃除機・洗濯機・扇風機・冷蔵庫・テレビ」の8製品のレビューテキストを収集した。レビューテキストは2023年9月11日までに書かれたものを使用している。収集された8製品の製品数を表1に示す。

表1. 対象ドメインの製品数

	自動車	バイク	自転車	掃除機	洗濯機	扇風機	冷蔵庫	TV
商品数	668	696	314	949	877	308	964	369

4.2. 印象に基づいたドメインの構造化

4.2.1. 評価対象の抽出

ドメインのレビューテキストに対し前処理を施し、評価対象の抽出を行った結果。評価対象の種類数224,527語が得られた。このうち、出現率5%以下の評価対象を除去し、1,050語の評価対象を分析に用いる。

4.2.2. 非負値行列因子分解を用いた行列の次元削減

非負値行列因子分解への入力行列は、行が8製品、列が評価対象1,050語、値が評価対象の出現率である。本実験でランク数は2に決定した。非負値行列因子分解により、入力行列 X （8行×1051列）を W （8行×2列）、 H （2行×1051列）の2つの低次元行列に分解で

きた。

4.2.3. 階層的クラスタリングによるドメインの構造化

非負値行列因子分解で得られた行列 W （8行×2列）を用いて、階層的クラスタリングを行った。クラスタリングの結果、クラスター1は「自動車・バイク・自転車」、クラスター2は「掃除機・洗濯機・テレビ・扇風機・冷蔵庫」となり、図5のように印象に基づいて構造化された。これは乗り物と家電の2つのサブドメインに分類できたと解釈できる。

次に、各クラスターの coefficient の値が高い評価対象を確認した。全ての評価対象のうち、上位30%で全てのサブドメインに共通して出現している評価対象として103語が得られた。これらの評価対象が、機械器具ドメイン共通の評価対象となる。

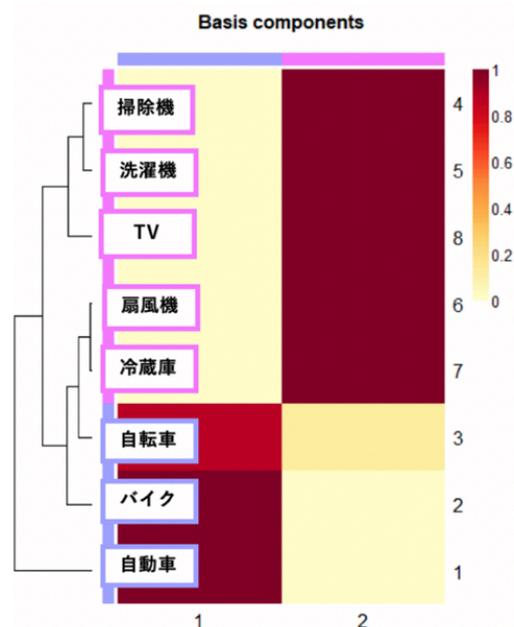


図5. 印象に基づいたドメインの構造化

4.3. ドメイン共通の感性価値の抽出

4.3.1. 行列の作成

4.2.1で得られた機械器具ドメインの評価対象から、4.2.3で得られた機械器具ドメイン共通の評価対象のみを抽出し、行列を作成した。

4.3.2. LDAを用いた印象トピックの抽出

4.3.1の行列をBowでベクトル化し、LDAでトピック抽出をした結果、トピック数は5個となった。各トピックの重要度の高い上位5語と、それらの印象語から人が解釈した結果を表2に示す。これらの5つのトピックがドメイン共通の価値となる。また、3章4章の結果から、図6のようにドメインを構造化することができ、ドメイン共通の価値を抽出できた。

表 2. トピック抽出の結果

Topic	解釈結果	上位5語
1	性能・パフォーマンス	"コストが良い" "性能が良い" "必要だ十分" "パフォーマンスが良い" "パフォーマンスが高い"
2	音	"良い音" "音が良い" "反応が良い" "良い感じ" "事多い"
3	背の高さ (サイズ)	"背が高い" "良い感じ" "背が低い" "高い有る" "余裕有る"
4	価格の安さ	"コストが良い" "値段安い" "シンプルだデザイン" "価格安い" "使い勝手が良い"
5	デザイン	"良い買い物" "シンプルだデザイン" "買い物出来る" "良い感じ" "デザインが良い"

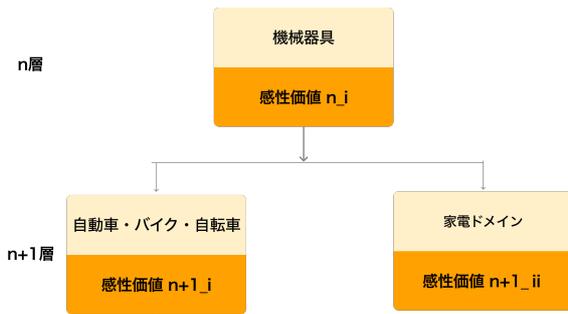


図 6. 機械器具ドメインの感性価値の構造化

4.3.3. 価格.com の評価項目と比較

価格.com の評価項目で、全てに共通している項目は「デザイン」のみである (表 3)。それに対し、本手法の分析結果では「デザイン・性能/パフォーマンスの良さ・価格の安さ・サイズ感・音の良さ」の5つが共通の項目として得られた。本手法は、口コミのデータを元に分析を行なっているため、製品を使用している消費者視点では、これら5つの項目が機械器具ドメインで共通して重要視されていることが分かった。このように、本手法では、製品の評価項目の自動設計にも役立てられることが示唆された。

表 3. 価格.com の 8 ドメインの評価項目

自動車	バイク	自転車	掃除機	洗濯機	扇風機	冷蔵庫	TV
エクステリア インテリア	デザイン	デザイン	デザイン	デザイン	デザイン	デザイン	デザイン
走行性能	走行性能	走行性能	使いやすさ	使いやすさ	使いやすさ	使いやすさ	操作性
乗り心地	乗り心地	乗り心地	サイズ	サイズ	サイズ	サイズ	サイズ
価格	価格	操作性	静音性	静音性	静音性	静音性	音質
燃費	燃費	使用パーツ	吸引力・ パワー	洗浄力	風量	機能	機能
エンジン 性能	エンジン 性能	メンテナンス 性能	手入れの しやすさ	機能・ メニュー	微風		画質
	取り回し	軽さ	取り回し				応答性能

4.4. 再帰的適用による感性価値の構造化

本章では、4.2 のドメインの構造化により、同じサブドメイン n に分類された乗り物サブドメイン (自動車・バイク・自転車) に 3 章の手法を適用し、さらに構造化することで n+1 層・n+2 層の感性価値を抽出した。

まず、n+1 層の感性価値を抽出した結果を述べる。前処理として、乗り物サブドメインの評価対象から、4.2.3 で得られた機械器具ドメイン共通の評価対象 (表 3) を除去した結果、総評価対象数は 171,787 語となった。この総評価対象を用いて、3 章の手法を同様に適用した結果、サブドメイン 1 は「自動車・バイク」、クラス 2 は「自転車」となった。さらに、n+1 層の乗り物ドメインに共通の価値として「心地よさ・車体の軽さ・ブレーキの効き」が得られた。

次に、n+2 層の感性価値を抽出した結果を述べる。前述の n+1 層の構造化により、分類された「自動車・バイク」サブドメインに対して 3 章の手法を適用した。分析データには、「Suv (車)・Sedan (車)・Compact (車)・Honda (バイク)・Suzuki (バイク)」の 5 製品のレビューテキストを用いた。前処理として、4.2.3 で得られた機械器具ドメイン共通の評価対象と乗り物ドメイン共通の評価対象を除去した。その結果、総評価対象数は 126,275 語となった。この総評価対象を用いて、3 章の手法を適用した結果、サブドメイン 1 は「Suv・Sedan・Compact」、サブドメイン 2 は「Honda・Suzuki」となった。さらに、n+2 層の「Suv・Sedan・Compact・Honda・Suzuki」に共通の価値として「燃費の良さ・加速感・重心/車高の高さ・エンジン性能」が得られた。

このように、3, 4 章の手法を各層 (n-1 層・n 層・n+1 層) に適用していくことで、各層の感性価値を段階的に抽出できることが示唆された (図 7)。

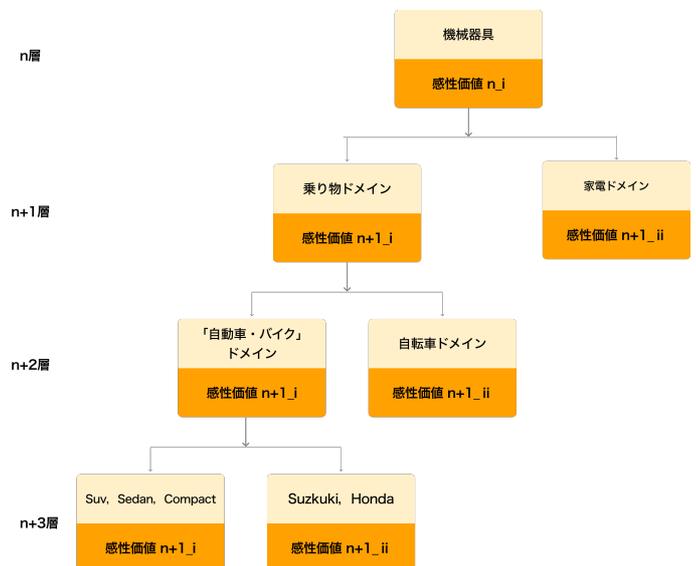


図 7. 再帰的適用による感性価値の構造化

5. 考察

4.3, 4.4 の分析結果を踏まえて、構造化された感性価値について述べる。第 n 層では、機械器具ドメイン共通の感性価値が得られた。また第 n+1 層では、機械

器具ドメイン共通の評価対象を除いて分析したことで、乗り物の特徴づける印象トピックが得られたことが分かる。これは、「自動車・バイク・自転車」に共通の価値であると共に、乗り物ドメインに固有の価値とも言える。また、n+2層でも同様に、乗り物ドメインに共通の評価対象を除いて分析したことで、「自動車・バイク」を特徴づける印象トピックが得られ、「自動車・バイク」ドメインに固有の価値が得られた。つまり本手法では、共通の価値を抽出できると共に、下の層に対しては共通の価値を除くことで、固有の価値を抽出できることが示唆された。この手法を各層に適用していくことで、ドメイン共通・固有の価値が段階的に抽出され、感性価値の階層構造化を図ることが可能となった。

6. まとめ

本研究では、レビューテキストを用いてドメイン共通・固有の感性価値を構造化することを目的として、1. 印象に基づいたドメインの構造化, 2. ドメイン共通の印象トピックの抽出, 3. 再帰的適用による感性価値の構造化の3つのプロセスから成る手法を提案した。本手法を各層に再帰的に適用することで、ドメイン共通・固有の感性価値を段階的に抽出でき、階層構造化を行えることが示された。本手法を広範囲のドメインに適用し、感性価値の辞書/コーパスとして利用することで、製品設計や製品評価をはじめとするさまざまな価値創造に寄与することが期待される。

文 献

- [1] 徳丸正孝, “感性データマイニング—製品の印象分析—,” 精密工学会誌, vol. 82, no. 1, pp. 43-46, 2016.
- [2] 豊田直晃, 浅野浩志, 北原路郎, 中田悟, 浅井巖, 高尾泰正, 島田泰拓, 羽多野重信, “粉体層せん断力測定による化粧品用粉体の使用感に関する客観的評価,” 粉体工学会誌, vol. 52, no. 12, pp. 694-700, 2015
- [3] 経済産業省, “感性価値創造イニシアティブ,” 財団法人経済産業調査会, 2007
- [4] S. H. Schwartz, et al., “Refining the theory of basic individual values,” *Journal of Personality and Social Psychology*, vol.103, no.4, pp.663-688, 2012.
- [5] A. Yamada, S. Hashimoto, and N. Nagata, “A text mining approach for automatic modeling of Kansei evaluation from review texts,” *Proc. the 7th International Conference on Kansei Engineering and Emotion Research 2018 (KEER2018)*, Springer, vol.739, pp.319-328, 2018.
- [6] S. Hashimoto, A. Yamada, and N. Nagata, “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*, vol.18, no.2, pp.59-65, 2019.
- [7] 佐野大樹, “日本語アブレイザル評価表現辞書,” の構築〜評価の多様性を捉えるための言語資源の開発〜, 言語処理学会第17回年次大会発表論文集, pp. 115-118, 2011.
- [8] 京都大学黒橋・河原研究室, “日本語形態素解析システム JUMAN++ Ver.1.02.” <https://nlp.ist.i.kyoto-u.ac.jp/?JUMAN%2B%2B> (2024/02/04 閲覧).
- [9] D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” *Advances in neural information processing systems*, vol. 13, 2000.
- [10] A. Takeuchi, T. Saito, and H. Yadohisa, “Asymmetric agglomerative hierarchical clustering algorithms and their evaluations,” *Journal of Classification*, vol. 24, 31 no. 1, pp. 123-143, 2007.
- [11] 日本標準産業分類 (総務省, 令和5年6月改定)
- [12] I. Kakaku.com, “価格.com.” <https://kakaku.com/>.