レビューテキストと画像を用いた 機械学習によるプロダクトの感性指標構築*

鈴木秀通** 飛谷謙介*** 橋本 翔** 山田篤拓** 長田典子**

Modeling Kansei Index of Product by Machine Learning Using Review Text and Image

Hidemichi SUZUKI, Kensuke TOBITANI, Sho HASHIMOTO, Atsuhiko YAMADA and Noriko NAGATA

In the field of Kansei (affective) engineering, the approach is often taken of modeling product's Kansei index to meet user's affective needs. This study work on modeling Kansei index automatically by machine learning using review text and images of products on the web. The proposed method follows: (1) Extraction of the main impressions of target domain and calculation of text impression scores that express the strength of each impression from review text by text mining, (2) creation of the product image data set with the training label made from the distribution of evaluation to products' impression by human and (3) construction of the deep neural network that estimates image impression score of product using the data set. Wristwatches were applied to proposed method as target domain. Then, estimation accuracy of constructed deep neural network was verified. As a result, high correlation coefficient 0.67 was confirmed between image impression scores and text impression scores, and effectiveness of the proposed method was confirmed. In addition, since the result exceeded correlation coefficient 0.51 calculated from estimation result of another deep neural network which has not learned distribution of evaluation, it was shown that learning the distribution was effective for improving estimation accuracy.

Key words: kansei (affective) engineering, kansei index, deep learning, appraisal dictionary, text mining

1. 緒 言

プロダクトデザインにおいて、ユーザのニーズを的確に把握し、実際のデザインに反映させることは重要である。特に近年は、ユーザニーズの多様化に伴い、機能や価格、信頼性といったこれまで重要視されてきたニーズだけでなく、使い心地や快適性などの感性的なニーズや個人ごとに異なる嗜好や欲求（ウォンツ）も着目されている¹⁾。感性的ニーズやウォンツを把握することの有用例として、商品開発の現場においてはプロダクトに対して人が感じる価値や印象の要因を知ることができ、上流工程へのフィードバックが可能になること、オンラインモール等においては個人の感性的なニーズに応える商品を推薦するサービスが提供でき、ユーザの満足度向上に貢献するといったことがあげられる。感性的なニーズを扱う上で信頼性が高く有効な方法論のひとつとして感性工学のアプローチがあげられ、さまざまなドメインに対して適用されている^{2)~4)}。具体的には心理学的実験と統計解析を利用することによって、感性情報の特徴である個人のばらつき、状況依存性、並びに潜在性（因果関係の希薄さ）といった諸課題⁵⁾を解決し、感性指標として顕在化する方法が挙げられる⁶⁾⁷⁾。しかしながらこれらの手法において、精度良く感性指標を構築するためには複数回の実験による心理学的データの取得とその分析が必要である。そのため、人的および時間的な負荷が高いという問題が存在して

いる。

一方で、深層学習を始めとする機械学習手法の発展に伴い、様々な分野において技術の自動化が急速に進んでおり、感性情報処理分野においてもこれらの機械学習手法が適用されつつある⁸⁾⁹⁾。しかしながら、近年の機械学習手法には大量の学習データが必要であり、感性情報を扱う場合は前述の人的および時間的な負荷が高いという問題が研究の推進に対して障害となる。

そこで本論文では、Web上に存在する大量のプロダクトのレビューテキストを統合的に利用して感性指標を構築し、次に感性指標から算出されるプロダクト毎の印象値を教師ラベルとして、プロダクト画像を学習する手法を提案する。最終的にはプロダクト画像から、その感性的印象を推定することを目的としている。提案手法の概念図を図1に示す。図1から提案手法では、はじめに先行研究¹⁰⁾¹¹⁾に基づいたテキストマイニングにより、対象ドメインから受ける主要な印象（印象トピック）をレビューテキストから抽出する。その後、プロダクトから受ける印象トピックの強さを表現する値を各プロダクトに対する複数のレビューテキストからそれぞれ算出し、これらの値をテキスト印象得点とする。次に、上記で得られたデータを用いて学習用データセットを作成する。具体的には、テキスト印象得点から教師ラベルを作成し、レビューテキストと紐づいたプロダクト画像に付与する。これにより、プロダクト画像と、その画像から受ける印象トピックの強さを表現する教師ラベルの組み合わせからなるデータセットを作成する。その際、個人による評価のばらつきを考慮した教師ラベルを作成する。最後に、作成したデータセットを用いて、プロダクト画像の印象推定をクラス分類問題として解く深層学習モデルを、印象トピックご

* 原稿受付 令和元年5月12日

掲載決定 令和元年8月21日

** 関西学院大学 理工学部/感性価値創造研究センター
(兵庫県三田市学園2-1)

*** 正 会 員 関西学院大学 理工学部/感性価値創造研究センター

とに構築する。以上により、感性指標構築に要する大量の学習データを自動で作成することが可能となり、加えて、大規模画像データを用いた機械学習による感性指標の精緻化と、その結果としてのプロダクト画像からの印象推定が実現される。

2. 関連研究

感性工学の分野では、主観評価実験に基づいて取得した心理学的データを利用して、対象物の感性指標を構築する手法が一般的に用いられる。ここで感性指標とは、感性のモデルを「感情—印象—物理量」の3層からなる階層構造として表現し⁷⁾、各層における主要因子を抽出し、主要因子間の対応関係を層内および層間で定義したものである。類似研究として人工知能分野における感情分析 (sentiment analysis) などがあるが、それらは感情と物理量を直接結びつけるモデルであり、仮定する枠組みが異なっている。印象層を介することでヒト (感情や価値) とモノ (物理要因) の対応関係における感性的な価値形成の根拠 (因果関係) が明らかになり、プロダクトデザインへのフィードバックが容易になる。ただし本論文では、以降、感性指標として「印象—物理量」の下位2層の関係のみを議論の対象とする。

主観評価実験に基づいた従来手法⁶⁾⁷⁾の手順を以下に述べる。はじめに、対象ドメインのプロダクトから受ける印象を表す評価語の収集・選定を行う。評価語の収集については、先行研究からの抜粋や、対象から受ける印象を自由記述で回答する実験を実施する。また評価語の選定については、収集した評価語が対象ドメインのプロダクトから受ける印象を表す単語として適切であるかの判定を目的とした、適合度実験を実施する。その後、評価語間の意味的な距離の測定を目的とした対比較実験を実施する。その結果をもとに評価語空間の構造化を目的とした多次元尺度構成、およびクラスタ分析を行うことで、代表性・網羅性の高い評価語セットを作成する。次に、実験刺激の収集・選定を行う。刺激 (例えば対象ドメインにおける画像群) に対しても評価語セットの作成と同様に、収集・選定、刺激空間の構造化を目的とした実験・分析を行うことで、代表性・網羅性の高い刺激セットを作成する。最後に、対象ドメインのプロダクトの印象に寄与する主要な要素 (主要因子) の抽出および、各刺激がもつ主要因子に対する重みの算出を行う。具体的には、まず Semantic Differential (SD) 法による主観評価実験を行い、上記の手順で収集した評価語セットを用いて刺激セットの印象の強さを評価する。その後、得られた評価データをもとに多変量解析 (因子分析等) を行う。この分析によって対象ドメインにおける主要因子を抽出し、各刺激がもつ主要因子に対する重みを算出する。これにより、その主要因子が意味する性質 (印象) を各刺激がどの程度有しているかを定量化することが可能となる。

以上のように、主観評価実験に基づいた従来手法を用いた場合、ユーザの感性的なニーズをプロダクトデザインに対し正確に把握することが可能である。しかしながら、感性指標構築の各段階において実験・分析が必要であるため、人的および時間的な負荷が高いといった問題が存在している。これらの問題を解決するために本研究では、深層学習モデルによりプロダクトの印象を推定することで感性指標を自動構築する。また、学習用データセットについても、Web上に存在するプロダクトの画像とレビューテキストの収集、および先行研究¹⁰⁾¹¹⁾に基づいたテキストマイニングにより自動で作成する。

一方で近年、機械学習により言語と画像という2つのモダリティを扱う研究分野が発展しており、画像キャプション生成や Visual Question Answering (VQA)¹²⁾などのタスクを扱った研究が報告されている。画像キャプション生成は、画像に説明文を自動で付与するタスクである。代表的な研究¹³⁾では、Convolutional Neural Network (CNN)¹⁴⁾を用いて画像の特徴を抽出し、抽出した情報を Long Short-term Memory (LSTM)¹⁵⁾に入力することで画像の適切な説明文を生成している。VQAは、ある画像に関する質問分に対して正しく回答するタスクである。VQAを扱った主要な研究¹²⁾では、CNNによって画像から抽出した情報と LSTMによって質問文から抽出した情報をニューラルネットワークで統合させることで、質問文に対する適切な応答を出力している。しかしながらこのような研究の多くは、画像内に存在している物体や状況の説明を対象としており、画像に対する感性的な情報である印象を扱う研究は少ない。審美的な解析を目的として作られた画像データベースである AVA (Aesthetic Visual Analysis)¹⁶⁾においても、画像から喚起される人の感情と、画像が持つ感性的印象とが区別されておらず、ユーザのニーズを把握したりプロダクトデザインにフィードバックするなど感性工学的に有効なモデルを構築することは困難である。

そこで本研究では、感性的印象をレビューテキストからテキストマイニングにより抽出し、抽出した印象の強さを表す教師ラベルをプロダクト画像に付与したデータセットを用いて深層学習を行う。これにより、レビューテキストから抽出した情報に基づいてプロダクト画像の印象を推定することが可能となる。

3. 感性指標の自動構築手法

本章では、提案手法について具体的に述べる。3.1節では感性指標 (印象層) の構築手法、すなわち先行研究¹⁰⁾¹¹⁾に基づいたテキストマイニングにより、Web上から収集したレビューテキストから印象トピックを抽出し、その後テキスト印象得点を算出するまでの手順を述べる。3.2節では、プロダクト画像の印象を推定する深層学習モデルの構築に用いる学習用データセットを作成する手法について述べる。3.3節では、感性指標 (物理層) を構築するための深層学習モデルの構造および学習の手順を述べる。

3.1 レビューテキストを用いたプロダクトの感性指標 (印象層) の構築

先行研究¹⁰⁾¹¹⁾で著者らが提案した手法は、テキストマイニングにおいて、日本語アプレイザル評価表現辞書¹⁷⁾を利用して感情語と印象語を分離して収集することによって、感性モデルの感情層と印象層に対応させる方法である。以下に概要を記述する。

まず、Web上から収集したレビューテキストに対して、評価語の収集・分類を行う。評価語の収集については、まず各レビューテキストを単語ごとに区切ることを目的とした形態素解析を行う。その後、感性的な表現をもつ単語 (評価語) を、複数の評価表現辞書¹⁸⁾の情報および品詞情報に基づいて収集する。上記の手法で収集した評価語群には、人の感情に関わる評価語 (感情語) と、プロダクトの印象に関する評価語 (印象語) が含まれる¹⁹⁾。しかしながら、感情語 (例えば、「欲しい」や「うれしい」) にはプロダクトデザインに関する特徴が含まれていないため、プロダクトに対する人の感情が推定できても、そ

Fig. 1 Workflow of Impression Estimation of Product Image

の推定結果からプロダクトに対するニーズを正確に把握することは困難であると考えられる。一方で、印象語（例えば、「小さい」や「高級感がある」）はプロダクトの印象に関する単語であるため、プロダクトに対するニーズを把握するための単語として有効であると考えられる。そこで、前章で述べた感性モデルの3層の階層構造に則り、収集した評価語を感情語と印象語に分類し、本研究では印象語のみを感性指標構築の対象として使用する。評価語の分類は、アプレイザル理論¹⁷⁾における内評価／外評価の概念に則って行う。内評価とは、評価対象に対する評価者の感情、もしくは感情を表す行為を示す表現であり、外評価とは、評価対象の特徴を示す表現である。日本語評価アプレイザル辞書には評価表現に対する内評価／外評価の属性が収録されているので、内評価の属性が付与された言葉を感情語、外評価の属性が付与された言葉を印象語に分類する。

次に、上記で収集した印象語を用いて印象トピックの抽出を行う。ここでは、文章を入力として文章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のひとつである Hierarchical Dirichlet Process LDA (HDP-LDA)²⁰⁾ に、上記の手順で収集した印象語のみで構成された文章を入力する。その結果抽出されたトピックを印象トピックとする。

最後に、プロダクトの各印象トピックのテキスト印象得点を、レビューテキストごとの印象語の出現頻度と各印象トピックにおける各印象語の重要度を用いて算出する。その際、各印象語の重要度として、文章内における単語の重要度の尺度である TF-IDF²¹⁾ をトピックに応用した term-score²²⁾ を用いる。term-score の値が大きいほど各印象トピックにおける重要度が高いといえる。先行研究¹⁰⁾¹¹⁾ では、各プロダクトに対する複数のレビューテキストごとに、印象トピックの得点を上記の手法で算出し、それらの平均値をテキスト印象得点としている。一方で本研究では、個人による評価のばらつきを考慮する。そのため、複数のテキスト印象得点の平均値ではなく、すべてのテキスト印象得点を教師ラベルとして使用する。

3.2 学習用データセットの作成

3.2.1 プロダクトデータの収集・選定条件

データセットの作成に際して、まず、対象ドメインにおけるプロダクトデータとして Web 上に存在しているレビューテキ

ストおよび画像を収集する。その際、3.1 節で述べたテキストマイニング手法の要件を満たすために、印象語を含むレビューテキストの数が 10 件以上のプロダクトデータを対象とする。

次に、収集したプロダクトデータの選定を行う。Web 上から収集したプロダクト画像の中には、プロダクト以外の物体や背景が映っているものや、プロダクトが見切れているものが存在する。これらの画像は、モデルの推定精度の低下につながると考えられる。そこで、プロダクト全体が写っており、かつ背景が白で他の物体が写っていない画像を選定し、データセットに利用する（選定されたプロダクト画像を、以下では単に画像と表記する）。

3.2.2 教師ラベルの作成

提案手法では、画像の印象を推定するタスクを、クラス分類問題として解く深層学習モデルを構築する。そのため、学習データとして教師ラベルを画像に付与する必要がある。作成手順を以下に述べる。

はじめに、レビューテキストの選定を行う。前項で述べた条件により収集したレビューテキストには必ず印象語が含まれるが、その数はレビューテキストごとに異なる。先行研究^{10,11)} におけるテキスト印象得点の算出手法の性質上、含まれる印象語の数が少ないほどテキスト印象得点は過少に算出される傾向を示す。そのため、各プロダクトに対するレビューテキストの内、含まれる印象語が多い上位 10 件のレビューテキストを印象トピックごとに選定する。その後、選定したレビューテキストからテキスト印象得点をそれぞれ算出し、教師ラベルの作成に使用する。

次に、テキスト印象得点によるクラスターリングを行う。テキスト印象得点は連続量であるため、教師ラベルを作成するためには離散量に変換する必要がある。そこで、まず k-means 法を用いて印象トピックごとに 7 つのクラスターに分類する。次に、テキスト印象得点が小さいクラスターから順に 1 から 7 までのクラスラベルを割り当て、各レビューテキストに付与する。付与されたクラスラベルの値は、レビューテキストが表す印象の強さを 1 から 7 の 7 段階で表す。クラスター数を 7 とした理由は、後述する主観評価実験と、印象の強さを評価する段階を統一するためである。

Table 1 Detail of CNN architecture. The input and output sizes are denoted as $height \times width \times channels$

Layer	Input size	Output size	kernel	Stride
Conv1	128 × 128 × 3	128 × 128 × 64	7 × 7	2
Max pool1	128 × 128 × 64	64 × 64 × 64	2 × 2	2
Conv2	64 × 64 × 64	64 × 64 × 256	5 × 5	2
Max pool2	64 × 64 × 256	32 × 32 × 256	3 × 3	2
Conv3	32 × 32 × 256	32 × 32 × 512	3 × 3	1
Conv4	32 × 32 × 512	32 × 32 × 512	3 × 3	1
Conv5	32 × 32 × 512	32 × 32 × 512	3 × 3	1
Max pool3	32 × 32 × 512	16 × 16 × 512	2 × 2	2
FC1	1 × 1 × 4096	1 × 1 × 4096	-	-
FC2	1 × 1 × 4096	1 × 1 × 4096	-	-
FC3	1 × 1 × 4096	1 × 1 × 7	-	-

最後に、クラスラベルの値を one-hot ベクトルで表現し、これらを教師ラベルとする。

以上の手順により、印象トピックごとに 10 個の教師ラベルが付与された画像から成る学習用データセットを作成する。これにより、個人による評価のばらつきを考慮した学習が可能となる。

3.3 深層学習モデルを用いた感性指標（物理層）の構築

提案手法では、画像の印象を推定するモデルとして、CNN を使用する。使用する CNN の構造は、CNN-M²³⁾ に基づいている。CNN の詳細を表 1 に示す。このうち Conv1, Conv2, Conv3, Conv4, Conv5 は畳み込み層であり、Max Pool1, Max Pool2, Max Pool3 は Max Pooling 層である。また、FC1, FC2, FC3 は全結合層であり、7 クラス（7 段階の印象の強さ）の分類問題を解くため出力層のユニット数は 7 とする。

CNN の学習における最適化手法として momentum 法を使用し、重みの加算の割合は 0.9、バッチサイズは 256 とする。また、過学習を抑制するために weight decay 法を適用し、正則化のパラメータは 0.0005 とする。各層での重みの初期値には He の初期値²⁴⁾を適用し、学習率の初期値は 0.001 とする。活性化関数には Rectified Linear Unit (ReLU) 関数を使用し、最初のふたつの畳み込み層において Batch Normalization²⁵⁾を適用する。また、出力層では softmax 関数を使用する。使用した softmax 関数を式 (1) に示す。

$$p_i = \exp(y_i) / \sum_{k=1}^n \exp(y_k) \quad i = 1, \dots, n, n = 7 \quad (1)$$

ここで、 y_i は、クラスラベル i に対応する前層の各ユニットとその結合重みを乗じた値であり、 y_k は、前層の値に対して、結合重みを乗じて総和した値である。この活性化関数を出力層で使用するによって、クラスラベル i に対する尤度 p_i を算出する。この算出した尤度 p_i と画像の教師ラベルを用いて式 (2) に示す誤差 E を算出し、これを最小化するように重みを更新することによってネットワークの重みを学習する。

$$E = - \sum_{i=1}^n t_i \ln(p_i) \quad n = 7 \quad (2)$$

式 (2) は交差エントロピー誤差関数であり、 t_i は教師ラベルの i 番目の要素の値である。誤差 E が減少しない状態が 5 エポック続いた場合、学習率を 10 分の 1 にする処理を 2 回まで

行い、3 回目で学習を終了させる。

以上により、画像に対する 1 から 7 のクラスラベル（プロダクトの印象の強さ）の尤度をそれぞれ推定する。また本研究では、推定精度の検証のために、推定結果として出力された確率分布から式 (3) を用いてプロダクトごとの期待値 EX を算出し、テキスト印象得点との相関係数を算出する。式 (3) において c_i は、7 段階の印象の強さを示している ($c_1 = 1, c_2 = 2, \dots, c_7 = 7$)。CNN によって推定した、画像の印象トピックの強さを表現するこれらの値を、画像印象得点とする。

$$EX = \sum_{i=1}^n c_i p_i \quad n = 7 \quad (3)$$

4. 実験 1：レビューテキストを用いた感性指標（印象層）の構築と学習用データセットの作成

提案手法の有効性を確認するための検証実験を行う。提案手法はプロダクトの画像とそれを評価するレビューテキストが存在する任意のドメイン（例えばファッション、コスメティック、インテリア、自動車等）に対して適用可能である。本研究では対象ドメインを腕時計とする。腕時計とした理由は、第 1 に意匠性が高く多様な印象を喚起するので、多種類の印象トピックが抽出される可能性が高いと期待されたためである。第 2 に、一方でデザインを構成する要素が比較的少ないため、印象とそれを喚起するデザイン要素との関係が解釈しやすく、その結果として、本研究の有用性・妥当性を定量的に示しやすいと考えたためである。第 3 に実際に多くの良質な画像が収集できたためである。

4.1 プロダクトデータの収集・選定

プロダクトデータは、楽天株式会社が運営するオンラインモールである「楽天市場」から、楽天ウェブサービス API を用いて収集した。その際、価格が 10 万円以下であることを 3.2.1 項で述べた収集条件に追加した。その結果、2811 点のデータを収集した。収集した画像枚数は 2811 枚、レビューテキスト数は 252228 件となった。また、レビューテキストから日本語アブレイザル評価表現辞書に基づき収集した印象語は 3880 語となった。その後、データセットに利用する画像として 1936 枚を選定した。選定された画像および選定されなかった画像の一部を図 2 に示す。

4.2 教師ラベルおよびデータセットの作成

まず教師ラベルを作成するために、収集した印象語を入力とした HDP-LDA により 9 個の印象トピックを抽出した。各印象トピックを表 2 に示す。このうち、印象トピック 1（上品だ、可愛い、小さい）、および印象トピック 5（重い、高級、質感が良い）は視覚的印象に関するトピックであると考えられる。また、印象トピック 3, 6, 8 は腕時計特有の機能に関する印象であり、印象トピック 2, 7, 9 はプロダクトの使い心地に関する印象、印象トピック 4 は触覚的印象であると考えられる。本研究では、画像の印象を推定する CNN を構築するため、視覚的印象に関するトピック 1 と 5 のテキスト印象得点から教師ラベルを作成した。その後、教師ラベルを対応する画像に付与してデータセットを作成した。最終的に、1 枚の画像には、2 つの印象トピックごとに印象語を多く含む上位 10 件のレビューテキストのテキスト印象得点から作成した、20 個の教師ラベルが付与された。

(a) Product images selected for use in the data set

(b) Product images not selected

Fig. 2 Example of Product Images Collected from the Web

Table 2 Extracted impression topics

Impression topics	Impression words
1	sophisticated, pretty, precious, tiny
2	hard to see, dressed, affordable, easy
3	accurate, thin, useful, unnecessary, easy
4	hard, smart, casual
5	heavy, high grade, durable, nicely textured
6	cool, light, hard to see, childish
7	childish, affordable, breakable, enough
8	undisturbing, forgettable, affordable, loose
9	luxurious, fulfilling, overjoyed, unfriendly

4.3 テキスト印象得点の妥当性検証

テキスト印象得点の妥当性の検証を目的として、画像の各印象トピックに対する評価を7段階で回答する主観評価実験を実施し、1936枚の画像の印象に対する評価得点を収集した。実験は、クラウドソーシングサービスである「クラウドワークス」を利用して実施した。実験では、各実験参加者は画像を1枚ずつ見て、各評価項目（印象トピック1と5）に対して「非常に当てはまらない」、「当てはまらない」、「やや当てはまらない」、「どちらでもない」、「やや当てはまる」、「当てはまる」、「非常に当てはまる」の7段階で評価した。その際、画像の提示順序は実験参加者ごとにランダムとし、1枚の画像を20人が評価するようにした。最終的に、20人分の評価結果から、7段階の評価を1から7の値に変換した場合の平均値を算出し、提示した画像の印象に対する人による評価得点とした。その後、これらの評価得点とテキスト印象得点との相関係数を算出した。

また、本手法の適用先としてオンラインモールの商品推薦サービスを想定した場合、より多くのプロダクトに対して網羅的に尤もらしい印象得点を付与することよりも、人による評価得点が低い商品に対して高い印象得点を付与しないことが重要である。そこで、評価得点とテキスト印象得点との相関係数を算出するとともに、テキスト印象得点が上位10%以上の194

Table 3 Correlation coefficient of text impression score and evaluated score of wristwatch images by human

Number of impression topic	Correlation coefficient
1	0.63
5	0.45

(a) Topic1

(b) Topic5

Fig. 3 Plot of Text Impression Score and Evaluated Score by Human for Topic1 and Topic5

個のプロダクトの内、人による評価得点が4以上となるプロダクトをポジティブサンプル、人による評価得点が4以下となるプロダクトをネガティブサンプルとして適合率を算出し、テキスト印象得点の妥当性を検証するための指標とした。

4.3.1 結果

印象トピック1と5について、算出された相関係数を表3に示す。また適合率を算出した結果、印象トピック1では適合率が0.93であり、印象トピック5では0.86であった。加えて、テキスト印象得点を横軸、人による評価得点を縦軸として1936個のプロダクトをプロットした結果を印象トピックごとに図3に示す。

Table 4 Scale of strength of correlation

Correlation coefficient	Strength of correlation
0.0 ~ 0.2	There is almost no correlation
0.2 ~ 0.3	There is a weak correlation
0.3 ~ 0.4	There is any correlation
0.4 ~ 0.5	There is a moderate correlation
0.5 ~ 0.7	There is a fairly high correlation
0.7 ~ 1.0	There is a strong correlation

4.3.2 考 察

相関係数の値から相関関係の強さを示す尺度²⁶⁾を表4に示す。本研究では、この尺度を基に相関関係の強さを測る。表3から、トピック1（上品だ、可愛らしい、小さい）のテキスト印象得点と人による評価得点との間に、相関係数0.63のかなり高い相関関係が確認された。また、トピック5（重い、高級、質感が良い）のテキスト印象得点と人による評価得点との間に、相関係数0.45の中程度の相関関係が確認された。さらに、算出した適合率については、本研究で想定している適用先において期待される精度としては十分と考えられる。以上より、テキスト印象得点の妥当性が示された。

トピック5において高い相関関係が確認されなかった原因は、トピック5に含まれている印象語である「重い」にあると考えられる。「重い」という印象には、視覚だけでなく触覚も関わっており、レビューの筆者は実物のプロダクトに対して両方の感覚に基づいて評価を行っていると考えられる。一方で、主観評価実験では対象が画像であるため、評価者は触覚的印象を考慮することが困難である。このような、プロダクトの印象を評価する際の状況の差が、トピック5におけるテキスト印象得点と人による評価得点との相関関係に大きく影響したと考えられる。

5. 実験2：プロダクト画像を用いたCNNによる感性指標（物理層）の構築と印象推定

続いて前章で作成した学習用データセットを用いて、CNNによる感性指標（物理層）の構築および印象推定実験を行う。本研究では、画像の印象に対する評価得点との相関がより高かった、印象トピック1（上品だ、可愛らしい、小さい）のテキスト印象得点を用いて作成したデータセットをCNNの構築に使用した。

画像の印象を推定するモデルを構築するために、以下の手順で11-fold cross validationを行った。まず、1936枚の画像を1760枚の訓練データと176枚をテストデータに分割した。次に、訓練データを用いてCNNを学習させた。その際、画像には印象トピック1の教師ラベルがそれぞれ10個付与されているため、学習は17600枚分の画像を用いて行われた。最後に、構築したCNNによりテストデータの画像印象得点を推定した。以上の手順を、全ての画像がテストデータとして画像印象得点を推定されるように繰り返した。

また、本研究で扱っている人間の主観的評価には個人によるばらつきが生じる。このばらつきには個人の感性のような個人ごとの特徴の違いが含まれているため、機械学習における教師ラベルとして使用する際、このばらつきを測定誤差として対象ごとに平均化して扱う場合と、系統的な誤差として平均化せずに扱う場合とでは、学習に異なる影響が生じると考えられる。

Table 5 Difference of correlation coefficient by difference of training label

	Training label	Correlation coefficient
Proposed method	Each text impression score	0.67
Comparative method	Average of text impression score	0.51

Table 6 Correlation coefficient of text/image impression score and evaluated score by human

Method of getting text/image impression score	Correlation coefficient
Text minig	0.63
CNN	0.81

そこで、個人による評価のばらつきを考慮したデータセットを学習することが深層学習モデルの印象推定精度に及ぼす影響を調べるために、評価のばらつきを考慮しないデータセット、すなわち評価を平均したデータセットを作成して学習し、結果を比較した。作成の手順としては、まず各プロダクトに対する10件のレビューテキストから算出したテキスト印象得点の平均値を求めた。次に、k-means法を用いて各プロダクトを平均値により7つのクラスタに分類し、平均値が小さいクラスタから順に1から7までのクラスラベルを割り当てた。最後に、クラスラベルの値をone-hotベクトルで表現したものを、教師ラベルとして対応する画像に付与して、データセットを作成した。作成したデータセットを用いて提案手法と同様にCNNを構築したものを比較手法として、提案手法と推定精度を比較した。

5.1 結 果

印象トピック1における、提案手法および比較手法により推定した画像印象得点と、テキスト印象得点との相関係数を手法ごとに表5に示す。また、画像印象得点を横軸、テキスト印象得点を縦軸として各プロダクトをプロットした結果を手法ごとに図4に示す。

次に、提案手法により推定した画像印象得点と人による評価得点との相関係数を表6に示す。また、画像印象得点の推定における適合率を算出した結果、適合率は1.0であった。加えて、画像印象得点を横軸、人による評価得点を縦軸として各プロダクトをプロットした結果を図5に示す。最後に、CNNモデルの構築に使用した訓練画像の枚数とモデルの推定性能との関係を表すために、本研究で使用した訓練画像の枚数（1760枚）を100%として、学習に使用した画像枚数の割合を横軸、推定性能（推定した画像印象得点と人による評価得点との相関係数）を縦軸として各深層学習モデルの推定性能をプロットした結果を図6に示す。

5.2 考 察

表5から、提案手法により推定した画像印象得点とテキスト印象得点との間に、相関係数0.67のかなり高い相関関係が確認された。この結果から、提案手法の有効性が示された。また本結果は、比較手法により推定した画像印象得点とテキスト印象得点との相関係数0.51を大きく上回ったため、評価のばらつきを学習することが推定精度の向上に有効であることが示された。

また表6から、提案手法により推定した画像印象得点と人による評価得点との間に、相関係数0.81の強い相関関係が確認された。本結果は、テキスト印象得点と人による評価得点と

Fig. 4 Plot of Image Impression Score and Text Impression Score for Each Method

Fig. 5 Plot of Image Impression Score and Evaluated Score by Human for Topic1

Fig. 6 Transition of the Correlation Coefficient by the Number of Images

の相関係数である 0.63 を上回る。これにより、テキストマイニングにより作成した学習用データセットの精度を、そのデータセットを学習した CNN の推定精度が上回ったことが示された。この原因を探るために、同じプロダクトに対するテキスト印象得点と画像印象得点の違いを、図 3-(a) と図 5 のプロット点の分布を比較することで確認した。その結果、テキスト印象得点の分散は大きく、人が 4 から 7 の高い評価得点を付けたプロダクトに対して、1 から 7 の間でテキスト印象得点が算出されていることが判明した。一方で、画像印象得点の分散は比較的小さく、同様のプロダクトに対して 2 から 4 の間で画像印象得点が推定されていることが判明した。この結果から、本研究で構築した CNN は、推定精度がデータセット内の外れ値に影響されないような、頑健な推定を行っており、このことが人による評価得点との強い相関関係の要因となっていると考えられる。加えて、算出した適合率は 1.0 であり、人による評価得点が低いプロダクトに対して高い画像印象得点を推定する可能性が少ないことが確認された。以上より、本手法を用いて推定した画像印象得点の妥当性が示された。

最後に図 6 から、深層学習に使用する画像枚数が本研究で使用する枚数の 40% (704 枚) 以上であれば、今回の実験結果と同程度の推定性能をもつ深層学習モデルを構築可能であることが示された。

6. 結 言

本研究では、Web 上に存在するプロダクトのレビューテキストと画像を収集し、テキスト印象得点から作成した教師ラベルを用いて学習用データセットを作成した。その際、個人による評価のばらつきを考慮した教師ラベルを作成した。さらに、作成したデータセットを用いた深層学習により CNN を構築し、プロダクト画像から印象を推定することで、プロダクトの感性指標を自動構築する手法を提案した。

また、本手法を対象ドメインを腕時計として適用し、その有効性を検証した。その結果、テキスト印象得点と本手法により推定した画像印象得点との間にかかなり高い相関関係が確認され、本手法の有効性が示された。さらに、構築した CNN の印象推定精度が、評価の平均を学習した CNN の印象推定精度を上回ったことから、評価のばらつきを学習することが印象推定精度の向上に有効であることを示した。

今後は、感性指標（感情層）の構築に拡張し、プロダクト画

像から喚起される人の感情や価値感の推定においても本手法が有効であることを検証する。また人の感性的評価に対する本手法のさらなる精度向上を目指して、機械学習におけるテキストと画像の最適な組み合わせ方法について検討を進める予定である。

参考文献

- 1) 豊田直晃, 浅野浩志, 北原路郎, 中田悟, 浅井徹, 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, 粉体工学誌, **52**, 12, (2015) 694.
- 2) M. Nagamachi: Kansei engineering: a new ergonomic consumer-oriented technology for product development, International Journal of industrial ergonomics, **15**, 1, (1995) 3.
- 3) 饗庭絵里子, 高松直也, 沼田晃佑, 柳田修太, 鈴木征一郎, 佐藤暢, 長田典子, 高田勝啓: 年代による感性空間の違い, 日本感性工学論文誌, **15**, 7, (2015) 677.
- 4) 黄峻, 川澄未来子: タイ人向け自動車フロントグリルデザイン提案のための感性の分析, 日本感性工学論文誌, **18**, 1, (2019) 127.
- 5) 井口征士, 猪田克美, 小林重順, 田辺新一, 長田典子, 中村敏枝: 感情情報処理, 電子情報通信学会編ヒューマンコミュニケーション工学シリーズ, オーム社, (1994) 1.
- 6) 飛谷謙介, 松本達也, 谿雄祐, 藤井宏樹, 長田典子: 素肌の質感表現における印象と物理特性の関係性のモデル化, 映像情報メディア学会誌, **71**, 11, (2017) 259.
- 7) 片平建史, 武藤和仁, 橋本翔, 飛谷謙介, 飛谷謙介, 長田典子: SD法を用いた感性の推定における評価の階層性-EPA構造の評価因子の多義性に注目して-, 日本感性工学論文誌, **17**, 4, (2018) 453.
- 8) 三由裕也, 萩原将文: 感情推定を利用した感性的な画像説明文自動生成システム, 日本感性工学論文誌, (2019) TJSKE-D.
- 9) 中村充志, 瀧澤生, 星泰成, 網島秀樹, 陳キュウ: 画像の感性を反映させたフォントの自動生成手法, 日本感性工学論文誌, **17**, 5, (2018) 523.
- 10) 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, 日本感性工学論文誌, **17**, 5, (2018) 567.
- 11) 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, **18**, 2, (2019) 59.
- 12) S. Antol, A. Agrawal, J. Lu, M. Mitchell, D. Batra, C. L. Zitnick, and D. Parikh: Vqa: Visual question answering, Proceedings of the IEEE international conference on computer vision, (2015) 2425.
- 13) K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio: Show, attend and tell: Neural image caption generation with visual attention, International conference on machine learning, (2015) 2048.
- 14) A. Krizhevsky, I. Sutskever, and G. E. Hinton: Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems, (2012) 109.
- 15) S. Hochreiter, and J. Schmidhuber: Long short-term memory, Neural computation, **9**, 8, (1997) 1735.
- 16) N. Murray, L. Marchesotti, F. Perronnin: AVA: A large-scale database for aesthetic visual analysis, 2012 IEEE Conference on Computer Vision and Pattern Recognition, (2012) 2408.
- 17) 佐野大樹: 日本語アプレイザル評価表現辞書-態度評価編-JAppraisal辞書 ver1, 言語資源協会, (2011) 1.
- 18) 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, **12**, 3, (2005) 203.
- 19) 佐野大樹: 『日本語アプレイザル評価表現辞書(態度表現編)』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, (2011) 115.
- 20) Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei: Sharing clusters among related groups: Hierarchical Dirichlet processes, Advances in neural information processing systems, (2005) 1385.
- 21) D. M. Blei, and J. D. Lafferty: Topic models, Text Mining, (2009) 101.
- 22) G. Salton, and C. Buckley: Term-weighting approaches in automatic text retrieval, Information processing & management, **24**, 5, (1988) 513.
- 23) K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman: Return of the devil in the details: Delving deep into convolutional nets, arXiv preprint arXiv:1405.3531, (2014).
- 24) K He, X. Zhang, S. Ren and J. Sun: Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, arXiv preprint arXiv:1502.01852, (2015).
- 25) S. Ioffe, and C. Szegedy: Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167, (2015).
- 26) 岩永雅也, 大塚雄作, 高橋一男: 社会調査の基礎, 放送大学教育振興会, (2003) 1.