原著論文

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

山田 篤拓, 橋本 翔, 長田 典子

関西学院大学

Automatic Impression Indexing based on Evaluative Expression Dictionary from Review Data

Atsuhiko YAMADA, Sho HASHIMOTO and Noriko NAGATA

Kwansei Gakuin University, 2-1 Gakuen, Sanda-shi, Hyogo 669-1337, Japan

Abstract : To meet user's affective needs, a number of efforts have been addressed to index kansei. However, the conventional method based on subjective evaluation experience take much time and effort to experiment and analyze. To solve this problem, we propose a method that automatically indexes impressions which is the first phase of attempting to index kansei based on evaluative expression dictionaries from the review data. This method consists of three steps. First, we collect evaluation words and categorize them as impression or emotion using word classes and the evaluative expression dictionary. Second, we extract the impression topics using the topic model, which uses only the evaluation words of impressions. Finally, we score each product for each impression topic, using the frequencies of evaluation words and term-scores. The results of an application of the method to the review texts of wristwatches and the subjective evaluation experiment show the validity of our method.

Keywords : Automatic indexing, Evaluative expression dictionary, Review data, Text mining

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、プロダクトに反映させることは重要である。特に近年はユーザのニーズや好みの多様化にともない、さまざまなプロダクトが市場に溢れていることから [1]、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]。この感性的な価値を扱う方法論としては、感性工学のアプローチが最も有用で信頼性が高いとされている [3]。

感性工学の具体的なアプローチには、心理学的データを利用する感性の指標化手法と、身体動作を利用する行動観察手法がある [4]。特に指標化手法は基本的な技術としてデザイン分野をはじめとしたさまざまな分野で適用されている [5-8]。指標化手法においては、指標化の対象とされる感性のモデルとして、物理量と心理量をつなげる階層構造が仮定されている。ここで仮定される感性の階層構造は、先行研究によってさまざまである [9-11] が、これらは評価対象の物理的刺激から、印象やイメージの喚起を経て、「好き」を含む態度・行動・感情が表れる、という過程を多段的に表現している点で共通している。そこで我々はこれらの共通点をもとに、図1に示すような形態要素、印象、感情という3つの階層から構成される感性の階層構造を定義する。各階層の定義を最下層から順に以下に示す。

① 形態要素

対象 (刺激) の物理的なパラメータを表す。

② 印象

対象に関する知識や経験と、対象から新たに得た情報に基づいて、その対象が「自分にとってどうか」を評価した結果を表す [12]。ここでの「自分にとってどうか」とは、対象の性質に関する個人の心理的な反応を示している [11]。

③ 感情

「人が心的過程の中で行うさまざまな情報処理のうち、人、物、出来事、環境についてする評価的な反応」を表す [13]。ここでの「評価的な反応」には評価対象に対する態度や行動も包括される。

印象の定義においては、「長いー短い」といった低次の印象と、「美しいー汚い」といった高次の印象が存在することが考えられるが、低次と高次を分ける明確な基準はなく、研究者により印象内の各単語の位置づけが異なることが多いため、印象内での階層の分割は行わなかった。また感情層の定義は、心理学における感情の定義のうち最も広義 [14] とされている Ortony ら [13] による感情の定義である。感情を印象の上位層と定めた理由は、例えば鈴木ら [12] において「印象によって、私たちは、そのモノに対して「好きだー嫌いだ」「欲しいー欲しくない」などの態度や行動を決めます。」と述べられているように、感情に含まれる態度や行動は印象を通して生じると考えられているからである。このような階層構造を仮定した感性の指標化により、感情の推定結果からその原因となる印象要因を明らかにし、印象要因を介して形態要素を特定することでプロダクトデザインに容易にフィードバックすることが可能となる。

図1 感性の階層構造

上述のような感性の階層構造を仮定して行われる指標化手法は、複数回の主観評価実験による心理学的データの取得とその分析が必要とされる。そのため主観評価実験に基づいた指標化手法は人的・時間的に負荷が高く[15]、専門的な知識が必要とされる。以上の理由から、これらの問題を解決する新たな指標化手法が望まれている。

そこで本研究では、Web上の商品レビューなどのテキストデータを利用した実験をともなわない感性の自動指標化手法として、印象の自動指標化手法を提案する。Web上には商品レビューや審査コメントなど、プロダクトに対する感想や意見などの評価が記述されているテキストデータ（以下、レビュー）が大量に存在している。機械学習やテキストマイニングといった人工知能分野ではレビューから対象に対してユーザが感じた評価極性（肯定的/否定的）を推定する感情分析（Sentiment Analysis）[16, 17]や、その評価極性の要因となる評価対象の具体的な注目点を抽出するAspect Extraction[18]がトピック抽出等の手法を用いて行われている。しかしながら評価極性の推定や注目点のトピックの抽出は、形態要素から直接的に感情に繋がる要因や評価（e.g. シルエットが好き）を推定するものであり、印象要因を介して（e.g. シルエットが上品だから好き）はいない。そのため、得られた知見をプロダクトの設計情報に直接反映するのは困難である。そこで本研究では感性の階層モデルに則った感性の自動指標化の第一段階として形態要素と印象を結び付けることを目的とし、印象の自動指標化手法を提案する。

また感情分析のアプローチには、ラベル付きデータから学習する手法と評価表現辞書を用いる手法がある[19]。ラベル付きデータから学習する手法では、感情を表すラベルを総合評価（星）等から取得し、機械学習を行い分析する。しかし印象に関するラベル付きデータ等はなく、実験なしでは入手困難である。一方、評価表現辞書を用いる手法はあらかじめ評価や感情に関する表現を登録しておき、それらを用いて分析する。このアプローチであれば印象の分析に対しても適用可能であり、さらに多分野、多言語への展開など拡張性も高い。

本論文では、評価表現辞書に基づく感情分析のアプローチを感性の階層モデルに適用した印象の自動指標化手法を提案する。本論文の構成は以下の通りである。2章では、従来の主観評価実験に基づいた感性の指標化手法の流れを述べる。3章では、本手法の詳細を説明する。4章では本手法を実在するレビューデータに適用した結果を示す。5章では、主観評価実験を実施し、4章で得られた結果と実験によって得られた人の評価との比較により、手法の有効性を検証し考察する。

2. 先行研究

感性工学において一般的に用いられるさまざまな主観評価実験に基づいた感性の指標化手法[20-23]をまとめたフローを図2に示す。主観評価実験に基づいた指標化には、以下の3つの段階が必要である。

- ① 対象の製品分野に存在しているプロダクトの印象を記述する評価語の収集・選定を行う。収集では、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してもらう自由記述実験を実施する。また選定では、収集された評価語が対象製品分野のプロダクトを評価する単語としてふさわしいか判定する適合度実験を実施する。その後、評価語間の距離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MDS、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。
- ② 刺激（対象製品分野に存在しているプロダクト群）に対しても評価語セットの作成と同様に、刺激の収集・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。
- ③ 上記で得られた評価語セットを用いて刺激セットを評価するSD（Semantic Differential）法による評価実験を実施し、得られた評価データをもとに因子分析を行う。この分析によって対象製品分野における主要因子を抽出し、各刺激に対して因子得点を算出する。これによって、その因子の意味する性質（印象）を各プロダクトがどの程度有しているかがわかる。

ただし研究や対象製品分野によっては複数の実験・分析の一部のみを実施することもある。

以上のように主観評価実験に基づいた手法では、指標化の各段階において複数の実験・分析が必要となり、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。

3. 評価表現辞書に基づいた印象の自動指標化手法

3.1 概要

本研究の評価表現辞書に基づいた印象の自動指標化手法のフローを図3に示す。印象の指標化を行う対象製品分野のプロダクトに対する大規模なレビューデータを用いて、自動で指標化を行う。提案手法は以下の3つの段階から構成され、それぞれ図2に示した従来の指標化手法の3段階と対比させている。

- ① 複数の評価表現辞書と形態素解析後の一部の品詞を用いて、評価語の収集・分類を行う。分類は1章で定義した感性の階層構造に基づいて行われ、収集した評価語を感情の評価語（以下、感情語）と印象の評価語（以下、印象語）に分類する。その後、得られた印象語のみを入力としたトピック抽出を行う。これによって得られるトピックを「印象トピック」とし、主観評価における主要因子として扱う。なお感情語については今後の研究課題とし本研究では用いない。

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

図2 主観評価実験に基づいた従来手法

図3 評価表現辞書に基づいた提案手法

- ② 刺激の収集・選定と刺激空間の構造化に対応した操作は行わず、大量のレビューデータをすべて用いる。
- ③ 各レビューの印象語の出現頻度と各トピックにおける各印象語の重要度を用いて印象トピックごとのプロダクトの得点化を行う。これによって得られる得点を「印象得点」とし、主観評価における因子得点、つまり各プロダクトが有する印象の度合いとして扱う。

このように本手法は人による実験や分析を行わないため、人的・時間的負荷の少ない印象の指標化処理が可能である。次節からは手法の詳細について述べる。

3.2 前処理

大規模なレビューデータから印象の自動指標化を行うにあたり、はじめに各レビューデータを単語ごとに区切るために形態素解析を行う。また、あらかじめ指標化に必要な品詞の除去や表記ゆれの統一を行うことで、のちのプロセスにおける精度向上を図る。

(1) 形態素解析 形態素解析は図4に示す例のように文章を意味のある単語に区切り、品詞を判別する処理である。形態素解析を行うことで単語間に区切りのない日本語の文章を、解析に適した単語ごとに区切られたデータに変換する。形態素解析には形態素解析エンジン MeCab 0.996 [24] と辞書 mecab-ipadic-2.7.0 [注1] を用いる。使用する品詞は名詞、動詞、副詞、形容詞、接頭詞の5種類で、助詞、助動詞、感動詞、接続詞、フィラー、その他に分類される単語はすべて除去する。また各単語を原型に変換し、名詞のナイ形容詞語幹は直前の単語と連結する。接頭詞は次に続く単語が名詞の場合連結し、それ以外の場合は除去する。

私	は	昨日	時計	を	買い	まし	た
名詞 (代名詞)	助詞 (係助詞)	名詞 (副詞可能)	名詞 (一般)	助詞 (格助詞)	動詞 (自立)	助動詞 -	助動詞 -

図4 MeCabを用いた形態素解析の例

(2) 表記ゆれの統一 商品レビューや審査コメントは多数のユーザが記述するため、表記ゆれが多く発生する状況にある。表記ゆれの影響により指標化の精度低下が懸念されるため、表記ゆれ問題の解決策として形態素解析後に抽出されたすべての単語に対して表記ゆれの統一を行う。表記ゆれの統一には日本語形態素解析システム JUMAN [25] の意味情報の1つである「代表表記」を用いる。代表表記は、同じ単語の表記集合から、代表的な表記（原則として新聞記事での高頻度表記）を選んだものである。具体的な処理は、MeCabにより分割された各単語を JUMAN に転送し解析する。JUMAN による解析の結果、単語が1つの単語として認識されて、かつ意味情報として代表表記が含まれる場合にはその単語を代表表記に統一し、2つ以上に分割された場合には、正確な表記ゆれ統一が行っていないと判断し、表記ゆれ統一を行わない。

3.3 評価語の収集・分類

全体の流れとしては、まず前処理後のレビューデータから複数の評価表現辞書と単語の品詞情報を用いて評価語を収集し、次に収集した評価語を感情語と印象語に分類する。以下に具体的な手法について述べる。

(1) 評価語の収集 形態素解析後のいくつかの品詞と、評価に関する複数の辞書を用いて、レビュー内の単語から次の条件に当てはまる単語を評価語として収集する。各条件は、①品詞が形容詞自立、②品詞が名詞の形容詞語幹、③評価表現辞書 [26, 27] に含まれる、④日本語評価極性辞書（用言編、名詞編）[27, 28] 中の「評価」の属性が付与されている単語に含まれる、である。

(2) 評価語の分類 上記の手法で収集された評価語には、1章の階層構造の最上位層である感情と、中位層の印象にあたる評価語が混在している。そこで先行研究 [29] で定義されている評価表現のカテゴリを用いて評価語を分類する。評価表現には〈内評価〉と〈外評価〉という2つのカテゴリ

が存在する。内評価とは、「楽しむ」「憤慨する」など評価対象に対する評価者の感情、もしくは「笑う」「泣く」など感情を表す行為を示す表現である。外評価とは、評価対象の特徴を示す表現である。上記の定義より、内評価が感情語、外評価が印象語に相当すると考えられる。そこで、評価表現に対して内評価/外評価の属性が付与されている日本語評価アプレイザル辞書[30]を用いて、収集した評価語のうち内評価の属性のみが付与されている評価語を感情語、それ以外の評価語を印象語として分類する。なお、この辞書内で「満足する」など内評価と外評価の両方に含まれる評価語は、印象語とした。

3.4 印象トピックの抽出

前節で得られた印象語のみで構成される文書を入力として Hierarchical Dirichlet Process LDA (HDP-LDA) [31] を適用し、印象トピックを抽出する。HDP-LDAは、文書を入力とし、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 Latent Dirichlet Allocation (LDA) [32] をノンパラメトリックベイズモデルに拡張したものである。LDAは、初期のパラメータとして出力のトピック数を指定する必要があるが、HDP-LDAは学習により適切なトピック数を自動的に決めることができる利点がある。

前処理として、同じプロダクトに対するレビューをすべて連結し、各プロダクトの文書を作成する。このとき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は、各対象を特徴づけるのに適していないため、入力の対象から除外する。またレビューでの出現頻度が過度に少ない印象語についても、対象を特徴づける共通の印象語として適していないため入力の対象から除外する。ここでは50%以上のプロダクトに対するレビューに含まれている印象語、および0.1%以下にしか含まれていない印象語を除去した。事後分布の近似手法としては周辺化ギブスサンプリング法(collapsed Gibbs sampling)を用いて、サンプリング回数は500回とする。HDP-LDAにおける各ハイパーパラメータは、 $\alpha=2.0$, $\beta=0.3$, $\gamma=0.01$ と定めた。

3.5 印象得点の算出

各プロダクトの各印象トピックに対する印象得点を、レビューごとの印象語の出現頻度と各トピックにおける各印象語の重要度を用いて算出する。

各トピックにおける各印象語の重要度として term-score [32] を用いる。term-scoreとはLDAにおける、あるトピック内の単語の重要度の尺度の1つであり、ある文書内の単語の重要度の尺度として用いられる Term Frequency Inverse Document Frequency (TF-IDF) [33] をトピックに応用したものである。TF-IDFは、それぞれの単語の文書内の出現頻度を表すTF値と、それぞれの単語がいくつの文書内で共通して用いられているかを表すIDF値との積をとり、これを単語の重みとして用いる。つまりその値が大きいほど各文書において重要度の高い単語といえる。これを応用した term-score は、

データ		得点付与 (商品 p , トピック k , レビュー r)			
レビュー r , 商品 p の 評価語の出現頻度		ImpWords	freq _r	term-score _k	freq _r × term-score _k
ImpWords		華奢な	2	0.35	0.70
		可愛い	1	0.29	0.29
		小さい	0	0.38	0.00
		...	...	...	...
		合計値			0.99

プロダクト p の印象得点				
	トピック1	トピック2	...	トピック k ...
レビュー1	0.38	0.02	...	0.96
...	...	...	...	...
レビュー r	0.01	0.32	...	0.99
...	...	...	...	...
レビュー n_p	0.53	0.12	...	1.04
平均値	Score _{p,1}	Score _{p,2}	...	Score _{p,k} ...

図5 印象得点算出手順

トピック内における単語の重みをつけることができ、値が大きい単語ほど各トピックにおいて重要度が高いといえる。トピック k での単語 v の term-score は、あるトピックの語の出現確率を β として次式で表される。

$$\text{term-score}_{k,v} = \hat{\beta}_{k,v} \log \left(\frac{\hat{\beta}_{k,v}}{(\prod_{j=1}^K \hat{\beta}_{k,j})^{\frac{1}{K}}} \right)$$

この式において、出現確率を表す β がTF値に対応し、残りの部分がIDF値に対応している。

印象得点の算出手順を図5に示す。あるプロダクトのトピックごとの得点は、そのプロダクトに対して記述されたレビューごとに各印象トピックの得点を付け、それらの平均をとることで求める。各レビューの各トピックについての得点化は、印象語の出現頻度と、各印象語に付与された term-score により算出する。具体的には、各印象語のそのトピックにおける term-score を合計した値を、そのレビューの当該トピックの得点とする。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は、その都度その印象語の term-score を足し合わせる。以上の処理をまとめると、あるプロダクト p のトピック k の得点 Score_{p,k} は、そのプロダクトのレビュー数を n_p 、印象語全体を ImpWords、レビュー r 内の印象語 w の出現頻度を freq_{r,w} として次式で表される。

$$\text{Score}_{p,k} = \frac{1}{n_p} \sum_{r=1}^{n_p} \left(\sum_{w=1}^{\text{ImpWords}} \text{term-score}_{k,w} \cdot \text{freq}_{r,w} \right)$$

4. 自動指標化実験

本章では、大規模なレビューデータに対して本研究の提案手法を適用した結果を示す。

4.1 対象製品分野とデータ概要

本実験での対象製品分野は腕時計とした。その理由として、検証実験によって結果の妥当性を判断する上では、視覚的印象が主であり、かつ多様な印象が持たれる製品分野が適していることが挙げられる。触覚的(素材)・嗅覚的・味覚的印象が優位となる製品分野は実物が必要となり、また聴覚的

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

印象が優位となる製品分野は刺激提示に時間を要するため、検証実験に適しているとはいえない。腕時計はデザインが豊富であることから視覚的印象に多様性が存在する点で好ましく、また腕時計のデザインとしての構成要素が主に文字盤とベルトに限定されるため、刺激の統制の面でも適している。そこで本研究では対象製品分野として腕時計を採用した。

腕時計のレビューデータとしては、楽天株式会社が運営するオンラインモール「楽天市場」の公開データセット[注2]である「みんなのレビュー・口コミ情報」の中から、2010年から2012年の3年間で記述された楽天市場の腕時計の商品レビューを用いた。

商品レビューにおいて、レビュー数が極端に少ないプロダクトは、プロダクトに対する評価の信頼性が低いと考えられるため、本実験における対象のプロダクトは、10件以上のレビューが書かれたプロダクトとし、10件に満たないプロダクトおよびそのレビューは対象から除外した。またレビューのタイトル情報にも対象を評価する記述があると考え、レビュー内容とタイトルを連結して1つのレビューとした。楽天市場の商品レビューは、その店で商品を購入してなくてもレビューを書くことができるが、今回は購入フラグが1（購入あり）のレビューのみを対象とした。その結果、腕時計の全商品数は4,557点で、対象とするレビューの数は203,086件となった。形態素解析の結果、総単語数は4,514,325語、総語彙数は19,689語となった。

4.2 結果

得られた腕時計のレビューに対して、前章で提案した手法を適用した結果を示す。

(1) 評価語の収集・分類 収集された評価語は3,097語であり、そのうち感情語が493語、印象語が2,604語となった。評価語の収集・分類結果の一部を表1に示す。

(2) 印象トピックの抽出 印象トピックの抽出にあたり、3.4節で述べた前処理を適用し出現頻度の条件を満たさない印象語の除去を行った結果、印象語を含むレビューが10件以上ある商品の数は4,557点中3,644点となった。それらのデータ内の印象語を入力としたHDP-LDAの結果、トピック数は15となった。各トピックのterm-score上位4語の印象語と、それらの印象語から人が解釈した結果を表2に示す。この結果から得られたトピックは以下の4種類に分類される、

表1 評価語の収集・分類結果（一部抜粋）

収集	分類
助かる、簡単だ、惚れる、軽い、心配、手頃だ、カジュアルだ、細い、一番、ビックリ、落ち着く、結構、多い、無料、カッコイイ、慣れる、愛用、ショック、お買い得、頼む、不安だ、安心、最高、小さい、安っぽい、嬉しい、残念だ	感情語
	印象語

表2 印象トピック抽出結果

	解釈結果	term-score 上位4語
1	重厚な高級感	高級だ、重い、重厚だ、質感
2	丁寧で迅速	丁寧だ、迅速だ、素敵だ、最高
3	値段相応	安っぽい、相応、おしゃれ、堅い
4	軽くて見易い	軽い、見易い、薄い、簡単だ
5	暗くて見づらい	ライト、見難い、暗い、カッコイイ
6	子供用にタフ	子供、タフだ、丈夫だ、迅速だ
7	上品で素敵	上品だ、素敵だ、小さい、華奢
8	おしゃれでカラフル	見易い、子供、違う、カラフルだ
9	アンティーク	アンティーク、ダーク、ナチュラルだ、馴染む
10	軽くて子供に合う	軽い、子供、邪魔、小さい
11	面白くて変わった	面白い、おしゃれだ、変わる、珍しい
12	目覚ましに便利	目覚まし、スッキリ、気付く、便利だ
13	高度で本格的	高度、本格、ハイだ、落ちる
14	(製品仕様)	ワイドだ、無垢だ、平均、潤滑
15	無骨でタフ	タフだ、無骨だ、剛健、正確だ

①視覚的印象(トピック1, 4, 6, 7, 9, 11)、②触覚的印象(トピック5, 8, 10)、③機能・サービスに対する印象(トピック2, 3, 12, 13, 15)、④製品仕様(トピック14)。いずれのトピックも腕時計に関する印象語の集合として解釈可能であったことから、妥当な結果が得られたと考えられる。ただし、製品仕様のトピック14については、ワイドサイズのベルトや「ステンレス無垢」等、本来は製品仕様を指す分野特有の一般名詞として判断されるはずの単語が、表記ゆれ統一の段階で名詞の形容詞語幹として誤って抽出されたものであった。

(3) 印象得点の算出 前項で得られたすべての腕時計に対して全15トピックの印象得点を算出した。これらの結果の妥当性を検証するために次章で性能評価を行う。

5. 印象得点算出の性能評価

算出された印象得点と主観評価実験で得られた評価得点との一貫性を確認するために、腕時計の画像から受ける印象を回答してもらう主観評価実験を実施し、相関係数分析および有意差検定の2つの分析により結果の比較を行った。

5.1 主観評価実験

(1) 実験参加者 実験には大学生、または大学院生20名(男性11名、女性9名、平均年齢21.6($SD=1.17$)歳)が参加した。

(2) 実験刺激 実験刺激としては、腕時計画像から受ける視覚的印象の評価の一致度を検証するため、視覚的印象に関すると判断した5つのトピック(1, 4, 6, 7, 11)の印象得点が上位の腕時計画像を10枚ずつ、計50枚を用いた(表3参照)。トピック9(アンティーク、ダーク、ナチュラルだ、馴染む)も視覚的印象に関するトピックと考えられるが、印象得点が上位の商品画像を複数確認したところ、図6のように時計の雰囲気に合わせてディスプレイ上に、腕時計が並べられている商品画像が多くみられ、全体のデザインの把握や刺激の統一において問題があると考えられるため用いなかった。

表3 印象得点算出の性能評価の実験刺激として用いた、視覚的印象のトピックにおける印象得点が上位の商品画像

トピック	term-score 上位4語	印象得点が上位の商品画像
1	高級だ、重い、 重厚だ、質感	
4	軽い、見易い、 薄い、簡単だ	
6	子供、タフだ、 丈夫だ、迅速だ	
7	上品だ、素敵だ、 小さい、華奢	
11	面白い、おしゃれだ、 変わる、珍しい

図6 トピック9の印象得点上位の商品画像

元のデータセットには商品情報の画像リンクが切れているものが多いため、商品画像は、商品名や品番の情報をもとにWeb上から画像を収集した。Web上に画像がないものや刺激の統一が困難なもの、評価されている商品と同じか不明確なものは刺激として用いず、刺激として使用できる印象得点上位10点の商品画像をトピックごとに収集して用いた。

刺激画像は、統一した背景上に腕時計の画像を張り付けたものを作成した。画像サイズは400×400ピクセルに統一し、腕時計本体の大きさは縦の画素数が320ピクセルになるように拡大・縮小を行った。腕時計の画像はデザインがわかりやすいよう文字盤が正面にある画像を用いた。背景は一般に画像評価実験の背景で使用する18%グレーを用いた。

(3) 評価項目 実験では表4に示す5つの評価項目を用いた。各評価項目は、視覚的印象に関するトピックのterm-score上位4語を1つにまとめたものであった。トピック1の「質感」、トピック6の「子供」、トピック7の「華奢」、トピック11の「変わる」は評価の回答が困難であると考え、レビューでの用いられ方に従って、以下の通り実験用に単語を変換した。まず「質感」はレビュー中で「質感がなかなかで、とて

表4 実験で用いた評価項目（カッコ内は修正前）

トピック1	高級だ、重い、重厚だ、質感が良い（質感）
トピック4	軽い、見易い、薄い、簡単だ
トピック6	子供向けな（子供）、タフだ、丈夫だ、迅速だ
トピック7	上品だ、素敵だ、小さい、華奢な（華奢）
トピック11	面白い、おしゃれだ、変わった（変わる）、珍しい

も気に入ってます」、**「質感色合いは期待通りでした」**、**「質感に大満足です」**のように質感の良さを述べる際に用いられているので、「質感が良い」とした。また「子供」は「子供も満足しています」、「子供の少し遅れた誕生日プレゼントにしました」、「子供っぽい感じでした」のように、腕時計が子供をターゲットにしていることを述べる際に用いられているので「子供向けな」とした。「華奢」は「予想より華奢で自然な存在感です」、「はめてみると華奢な感じでオシャレな雰囲気でした」、「でも華奢な感じのデザイン・色共に、まさに求めているもので◎」のように名詞の修飾語として用いられているので「華奢な」とした。「変わる」も同様に「変わった時計」、「本当デザインが変わってて、カッコイイです!!」、「変わったデザイン」のように名詞の修飾語として用いられているので、「変わった」とした。

(4) 実験課題 実験課題は、腕時計の画像を見て5つの評価項目に回答してもらうことであった。回答は「非常に当てはまる」、「当てはまる」、「やや当てはまる」、「どちらでもない」、「やや当てはまらない」、「当てはまらない」、「非常に当てはまらない」の7段階で行った。

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

図7 実験画面例

図7に実験画面の例を示す。実験試行が始まると、ディスプレイ上に腕時計の画像と評価入力欄が呈示される。このように、視覚的印象に関する5つのトピックそれぞれに対して、10枚の画像を呈示した。試行回数は全50回であった。なお刺激画像は同じトピックから抽出された商品が連続しないようランダムに呈示し、評価項目は実験参加者ごとにランダムな順とした。

(5) 実験装置および実験環境 実験刺激は、24.1インチの液晶ディスプレイ（EIZO社、ColorEdge CG246）に呈示した。ディスプレイ解像度は1920×1080ピクセルであった。またColorNavigator6を用いてディスプレイ輝度のキャリブレーションを行った。実験参加者は1名ずつ椅子に腰かけ、実験を行った。また実験参加者の課題に対する妨げになることを防ぎ、周辺光の影響による刺激の見え方の変化をなくすため、実験は防音の暗室で行った。

5.2 結果と考察

(1) 分析1：得点の相関 印象得点算出結果の有効性を確認するために、各腕時計に付与された印象得点と、主観評価実験（以下、実験と略する）で得られた評価得点との相関を確認した。実験で得られた各腕時計の得点は、7件法の「非常に当てはまらない」を1、「非常に当てはまる」を7として得点化し、実験参加者全体の平均値とする。また各トピックで算出された印象得点は、0から1の範囲に正規化した。

算出された印象得点と実験で得られた評価得点との相関係数を表5に示す。トピック6（子供向けな、タフだ、丈夫だ、迅速だ）と、トピック7（上品だ、素敵だ、小さい、華奢な）は、相関係数0.7以上の強い正の相関がみられた。またトピック1（高級だ、重い、重厚だ、質感が良い）とトピック11（面白い、おしゃれだ、変わった、珍しい）は中程度の相関がみられた。トピック4（軽い、見易い、薄い、簡単だ）に関しては相関がみられなかった。

トピック6, 7に関しては強い相関がみられ、提案手法の有効性が示された。一方、他の3つのトピックに関しては強い相関がみられなかった。この結果の原因を探るため、トピックごとに算出された印象得点を横軸、実験で得られた評価得点を縦軸として各腕時計をプロットした。トピックごとのプロット結果を図8に示す。プロット結果より、印象得点が上位の腕時計は、

表5 本手法により算出された印象得点と実験で得られた評価得点の平均値との相関係数

トピック	1	4	6	7	11
相関係数	0.622	-0.081	0.743	0.787	0.576

図8 各腕時計画像の実験で得られた評価得点(縦軸)と算出された印象得点(横軸)

相関がみられなかったトピック4を除いて、対応する評価項目においても比較的高い評価を受けていることがわかる。しかし実験で得られた評価得点が中位(3:やや当てはまらない~5:やや当てはまる)の腕時計と下位(1:非常に当てはまらない~2:当てはまらない)の腕時計の印象得点はともに0.0~0.3の範囲に集まっており、評価得点に差があるにもかかわらず印象得点の分布には差がみられない。このことから相関がでなかった原因は主に、人の評価得点が高くない、つまり印象の強度が低い腕時計に対して、適切な印象得点が付与されていないことであるとわかった。印象の強度が低いプロダクトは、レビュー内でもその印象に対する記述が少ない傾向にあり、今後そういったレビューに対しても情報を補完する必要がある。

(2) 分析2: 平均値の有意差検定 前項において、印象得点が上位の腕時計は対応する評価項目において比較的高い評価を得ていることがわかった。本項では、分散分析を用いて各評価項目に対応したトピックの印象得点が上位の腕時計の平均値が、それ以外の各トピックの腕時計の平均値に比べて統計的に高い評価を得ているのかを検証する。

まず評価項目とトピックとの交互作用を調べるため、実験参加者とトピックと評価項目と腕時計を要因とした四要因分散分析を行った。対象としたのは、実験に参加した20名の実験参加者で、分析は、実験で使用した5つの評価項目について行った。トピックと評価項目との交互作用がみられた場合、各評価項目においてトピックの単純主効果が有意であるか確認する。その結果有意であれば多重比較を行い、評価項目ごとに実験で得られた評価得点が高いトピックを確認する。

分散分析の結果を表6に示す。評価項目、実験参加者、腕時計の単純主効果は0.1%水準で有意であった。またトピック

と評価項目との交互作用は0.1%水準で有意となった。トピックと評価項目との交互作用が確認されたため、各評価項目におけるトピックの単純主効果の検定を行った。その結果を表7に示す。すべての評価項目においてトピックの単純主効果が0.1%水準で確認された。最後に、多重比較の結果を踏まえた各評価項目のトピックごとの評価得点の平均値のグラフを図9に示す。この多重比較の目的は、各評価項目に対応したトピックの腕時計の平均値とそれ以外の各トピックの腕時計との平均値の差を見ることであるため、棒グラフでは対応するトピックと有意であったトピックにのみ有意確率を表示し

表6 四要因分散分析結果

	自由度	F値	有意確率
トピック*評価項目	16	202.870	.000
実験参加者	19	7.191	.000
評価項目	4	80.114	.000
トピック	0	.	.
腕時計	45	4.146	.000
誤差	4413		

表7 トピックの単純主効果検定結果

	自由度	F値	有意確率
トピック1	4	191.549	.000
トピック4	4	51.207	.000
トピック6	4	35.801	.000
トピック7	4	185.109	.000
トピック11	4	81.516	.000
誤差	995

図9 各評価項目におけるトピックごとの実験で得られた評価得点の平均値

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

ている。トピック4を除いたすべてのトピックの平均値は、対応する評価項目において他のトピックの平均値と比べて0.1%水準で有意に高いことがわかった。トピック4の評価項目においては、トピック6の平均値がもっとも高かった。

トピック4を除いて、対応する評価項目の平均が他のトピックと比べて有意に高かったことから、印象得点が上位の腕時計に関しては実験で得られた人の評価得点と一致しており、印象の強度が高いプロダクトに対する自動指標化が適切に行われることが確認された。トピック4のみ平均が高くなかったのは、トピック4が視覚的印象のみならず触覚的印象にも関係するトピックであったことが原因と考えられる。例えば表3のトピック4の1番の腕時計は、トピック4の評価項目において平均3.70点というあまり高くない評価を受けているが、レビューを見るとほとんどのレビューにおいて「軽い」との評価を受けている。レビューにて「軽い」評価が多く見られた理由として、この腕時計の文字盤の素材がステンレス製に比べて軽いチタンであり、実物を手にしたときに、見た目に反して軽かったとの印象を受けたことがレビューに反映されたものと推測される。このような金属部分の素材の印象のように、見た目だけで判断が難しいものについては、画像のみでの評価と、算出された印象得点の結果が一致しなかったのではないかと考えられる。実物を用いてレビューを記述した状況を再現することによって、トピック4を含む視覚的印象以外のすべてのトピックについても検証可能であると考えられる。

6. お わ り に

本研究では、プロダクトに対する大規模レビューデータを用いた評価表現辞書に基づく感性指標化の試みの第1段階として、印象の自動指標化手法を提案した。提案手法では、評価表現辞書に付与された内評価/外評価のカテゴリ情報を利用して、印象語の収集・分類を行った。これにより印象の評価語のみのトピック抽出を行うことができ、対象製品分野の主要因子である印象トピックを抽出できた。また各レビューに含まれる評価語の出現頻度と各トピックの評価語の重要度尺度を用いて、プロダクトから受ける印象を数値化した印象得点の算出を行った。この手法をネット通販サイトの腕時計の商品レビューに適用し、プロダクトの印象得点を算出した結果、視覚的印象にかかわるトピックについて、人の評価得点と正の相関を示すことが確認された。また、印象の強度が高いプロダクトに対する指標化の妥当性が示された。

今後は人の評価において印象の強度が低いプロダクトに対する印象得点の算出手法を改善する等、精度の向上を図る。また形態要素と印象、および印象と感情についても大規模レビューデータから対応付けを行う手法を検討し、感性の自動指標化手法の確立を図る。さらに多分野、多言語へ展開し、手法の有用性を検証していきたい。

謝 辞

本研究の一部は国立研究開発法人科学技術振興機構 (JST) の研究成果展開事業「センター・オブ・イノベーション (COI) プログラム」の支援によって行われました。

また本研究の自動指標化実験においては、楽天株式会社と国立情報学研究所が提供する「みんなのレビュー・口コミ情報」を使用させていただきました。

注

[注1] <https://sourceforge.net/projects/mecab/files/mecab-ipadic/2.7.0-20070801/>

[注2] https://rit.rakuten.co.jp/data_release_ja/

参 考 文 献

- [1] 徳丸正孝：感性データマイニング，精密工学会誌，82(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，52(12)，pp.694-700，粉体工学会誌，2015.
- [3] Chen, C.H., Khoo, L.P., Chen, K., Pang, J.H., and Huang, Y.: Consumer-oriented product form creation via Kansei engineering, Proceedings of the International Symposium for Emotion and Sensibility – Emotion Research in Practice, pp.184-191, 2008.
- [4] 都甲潔，坂口光一：感性の科学－心理と技術の融合，朝倉書店，pp.32-37，2006.
- [5] Nagamachi, M.: Kansei engineering: a new ergonomic consumer-oriented technology for product development, International Journal of Industrial Ergonomics, 15(1), pp.3-11, 1995.
- [6] 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，日本感性工学会論文誌，15(7)，pp.677-685，2015.
- [7] 橋本翔，田中一品，片平建史，長田典子：刺激と独立な個人の傾向を考慮した新たな三相データの分析法，行動計量学，45(1)，pp.27-38，2018.
- [8] 朴成桓，釜池光夫，長尾徹：乗用車フロントマスクの表情印象と形態要素の関係，デザイン学研究，55(6)，pp.21-28，2009.
- [9] 棟近雅彦，三輪高志：感性品質の調査に用いる評価用語選定の指針，品質，30(4)，pp.96-108，2000.
- [10] 小西秀吾，野中秀俊，栗原正仁：感性語の階層構造を利用した配色における個性の抽出と活用，知能と情報(日本知能情報ファジィ学会誌)，20(1)，p.141-149，2008.
- [11] 片平建史，武藤和仁，橋本翔，飛谷謙介，長田典子：SD法を用いた感性の測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－，日本感性工学会論

- 文誌, 17(4), pp.453-463, 2018.
- [12] 鈴木邁, 大澤光: 感性工学と情報科学, 森北出版, pp.58-63, 2002.
- [13] Ortony, A., Clore, G.L., and Collins, A.: The Cognitive Structure of Emotions, Cambridge University Press, 1988.
- [14] 大平英樹: 感情心理学・入門, 有斐閣アルマ, p.5, 2010.
- [15] 秋山修一, 鈴木謙介, Gehrmann Andreas, 永井義満, 石津昌平: MTS法を用いた名詞による感性画像検索システムの構築, 日本感性工学会論文誌, 8(4), pp.1171-1178, 2009.
- [16] Medhat, W., Hassan, A., and Korashy, H: Sentiment analysis algorithms and applications: A survey, Ain Shams Engineering Journal, 5(4), pp.1093-1113, 2014.
- [17] Fang, X., and Zhan, J.: Sentiment analysis using product review data, Journal of Big Data, 2(1), p.5, 2015.
- [18] Poria, S., Cambria, E., Gelbukh, A.: Aspect extraction for opinion mining with a deep convolutional neural network, Knowledge-Based Systems, 108, pp.42-49, 2016.
- [19] 野中尚輝, 中山浩太郎, 松尾豊: オンラインレビューから抽出した消費者の感情に寄与する素性を用いた自動車販売予測, 情報処理学会論文誌データベース (TOD), 10(3), pp.16-25, 2017.
- [20] 井口征士, 猪田克美, 小林重順, 田辺新一, 長田典子, 中村敏枝: 感性情報処理, 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.
- [21] 長町三生, 石原茂和, 西野達男, 松原行宏, 土屋敏夫, 神田太樹, 井上勝雄: 商品開発と感性, 海文堂出版, pp.13-24, 2005.
- [22] 柳澤秀吉, 村上存, 大富浩一, 穂坂倫佳: 感性の多様性を考慮した感性品質の定量化手法 (製品音の設計における感性品質の定量化への適用), 日本機械学会論文集 C編, 74(746), pp.2607-2616, 2008.
- [23] 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, 映像情報メディア学会誌, 71(11), pp.259-268, 2017.
- [24] 工藤拓: 形態素解析器 MeCab, <http://taku910.github.io/mecab> (2018.04.25閲覧).
- [25] 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.7.01, <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN> (2018.04.25閲覧).
- [26] 小林のぞみ, 乾健太郎, 松本裕治: 意見情報の抽出/構造化のタスク仕様に関する考察, 情報処理学会研究報告 NL171-18, pp.111-118, 2006.
- [27] 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), pp.203-222, 2005.
- [28] 東山昌彦, 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.
- [29] 佐野大樹: 『日本語アブ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, pp.115-118, 2011.
- [30] 佐野大樹: 日本語アブレイザル評価表現辞書 (JAppraisal辞書)～態度評価編～ version 1, 言語資源協会, 2011.
- [31] Teh, Y.W., Jordan, M.I., Beal, M.J., and Blei, D.M.: Sharing clusters among related groups: Hierarchical Dirichlet Processes, Advances in Neural Information Processing Systems, 17, pp.1385-1392, 2005.
- [32] Blei, D.M., and Lafferty, J.D.: Topic models, In: Text mining: classification, clustering, and applications, p.34, 2009.
- [33] Salton, G., and Buckley, C.: Term-weighting approaches in automatic text retrieval, Information Processing & Management, 24(5), pp.513-523, 1988.

山田 篤拓 (非会員)

2017年 関西学院大学理工学部人間システム工学科卒業。現在、同大学院理工学研究科修士課程在学中。専門は感性情報学、特にテキストマイニングを用いた感性の自動指標化に関する研究に従事。

橋本 翔 (非会員)

2007年 大阪大学人間科学部卒業。2015年 大阪大学大学院人間科学研究科博士後期課程修了。博士 (人間科学)。2015年より関西学院大学大学院理工学研究科 感性価値創造研究センター 博士研究員。2017年より同大学理工学部研究特別任期制助教で現在に至る。日本行動計量学会, 日本計算機統計学会, 日本分類学会などの各会員。

長田 典子 (正会員)

1983年 京都大学理学部数学系卒業。同年三菱電機(株)入社。1996年 大阪大学大学院基礎工学研究科博士後期課程修了。2003年より関西学院大学理工学部情報科学科助教授, 2007年教授。2009年米国バドュー大学客員研究員。2013年 感性価値創造研究センター長。2015年 革新的イノベーション創出プログラム「感性とデジタル製造を直結し, 生活者の創造性を拡張するファブ地球社会創造拠点」サテライトリーダー。博士 (工学)。専門は感性工学, メディア工学等。