

多数の画像群に対する多人数・多次元印象評価結果の可視化

飯島 紘理[†] 伊藤 貴之[†] Manuela Waldner[‡] Hsiang-Yun Wu[‡] Nicolas Grossmann[‡]

[†]お茶の水女子大学理学部情報科学科 〒112-8610 東京都文京区大塚2丁目1番1号

[‡] Faculty of Informatics TU Wien Karlsplatz 13, 1040 Wien, Austria

E-mail: [†]{g1620501, itot}@is.ocha.ac.jp, [‡]{waldner, wu}@cg.tuwien.ac.at, nicolas.grossmann@tuwien.ac.at

あらまし SD法(Semantic Differential Method)とは概念の意味を測定するためにアメリカの心理学者 Osgoodによって開発された方法である。これは形容詞対にリッカード尺度を用い、ある概念に対する印象を被験者に評価してもらう実験のことを指す。一般的にSD法を採用した印象評価は多数の参加者の回答によって構成されるため、回答者が有する印象の個人差がデータの内容を大きく左右する。本研究ではこの印象評価データを可視化することで、回答群の例外や傾向を分析し、かつ概念×回答者×尺度の三相の改善可能性を検討する。本手法ではまずSD法を採用した印象評価を実施し、続いて各画像の印象値とRGB値に次元削減を適用することで画像の印象分布と色分布を可視化する。次に各尺度の平均値と中央値を重ねて表示することで、画像特徴と回答の関連を観察する。本報告では、300枚の衣服画像に21人の参加者が印象評価を実施したデータを題材として、本手法による可視化の実行例を示す。

キーワード SD法, 可視化

1.はじめに

印象評価は学術的にも産業的にも多くの場面で重要な評価手段となる。学術的には例えば心理実験や感性情報処理などの目的で、産業的には例えば広告戦略や顧客分析などの目的で多用される。

本研究では、回答者に画像を提示して印象を回答させる形式の印象評価を対象とする。ここで、印象評価の回答は個人差があるので、一般的に多数の回答者を集めが必要がある。また、印象評価結果を集計して統計処理するためには、多数の質問（本研究では多数の画像）が必要となる。さらに、画像を見て回答者が受ける印象は多彩なものとなるため、1枚の画像に対しても多数の質問を用意する場合がある。つまり画像に対する印象評価実験では、多人数の回答者が参加し、多数の各々の画像に対して、複数の尺度を回答させることが多い。本報告ではこのような印象評価を多次元印象評価と称する。

印象評価結果を集計可能な形で実施するためには、リッカート尺度などにもとづいたn段階の評価形式で設問を用意することが多い。その形式で設問を用意するための一手段として、本研究では印象評価のための設間にSD法[1]を導入し、対比語への適合度をn段階で回答させる。1枚の画像に対して複数の設問を回答者に提示するものとする。そしてこの回答から得られる集計結果を、多数の画像、多数の回答者、複数の設問を3軸とする多次元データとして扱う。

このような印象評価結果の分析には多角的な視点が必要となる。画像に着目すると例えば、どのような画像集合において回答結果が類似しているか、どのような画像において回答がわかれやすいか、といった点

を分析できる。回答者に着目すると例えば、どの回答者群において回答が類似しているか、どの回答者が他の回答者との回答の差異が大きいか、といった点を分析できる。設問に着目すると例えば、どの設問とどの設問に連動がみられるか、どの設問にて回答がわかれやすいか、といった点を分析できる。

印象評価結果のデータとしての活用例として、機械学習を用いた印象の推定があげられる。印象評価の対象となった画像群を訓練データとして、印象評価結果をそのアノテーションとして付与した形で機械学習を適用することで、テストデータを構成する画像群の印象を推定することができるようになる。このとき、印象推定の適切さは訓練データの品質に大きく影響されると考えられる。そのため訓練データとしての印象評価結果の分布をユーザが理解することは重要である。この点は印象評価結果に限らず機械学習の運用全般において重要なことである。第3次AIブームの中心的存在である機械学習の運用現場では、その挙動に対する説明責任を問われることが多くなっている。訓練データの分布を理解することは、機械学習の説明責任という観点からも重要であると考えられる。

これらの背景を鑑みて我々は、印象評価結果の分布を理解するための可視化手法を提案する。本手法では以下の3つの処理工程から構成されるものである。

- 画像群に対してSD法を採用した印象評価を実施する。
- 各画像の印象値とRGB値に次元削減を適用することで画像の特徴傾向を可視化する。
- 各尺度の平均値と中央値を重ねて表示することで画像特徴と回答の関連を観察する。

本研究の構成は以下の通りである。2章では関連研究について、3章では提案手法について述べる。そして4章で本手法の分析例と考察について、5章で本研究のまとめと今後の課題について述べる。

2. 関連研究

本研究が採用するSD法[1]は、アメリカの心理学者Osgoodらが提唱した手法である。SD法では形容詞対への適合度を回答させ、その平均値や中央値から印象を判断する。一般的にはリッカート尺度などを回答値として提示するとともに、複数の項目を設問とする場合には間隔尺度の水準を満たしていると仮定する。

続いて、SD法による印象評価データの分析方法について紹介する。三相因子分析[2]はSD法の回答の個人差を考慮したデータ分析手法であり、本研究と同様に概念×回答者×尺度の三相を分析の対象とする。別の分析手法として探索的ポジショニング分析[3]がある。しかしこれらの手法では、ある相の因子負荷の表と、ある相に次元削減を適用した2次元空間(因子空間[4])の両方の提示が必要[5]であり、単一の可視化結果としてデータを観察することができない。それに対して本研究では、探索的ポジショニング分析に色と概念(画像)を組み合わせて、単一の画面でデータを可視化する。

因子分解も印象評価の有用な手段である。尺度のみに注目した因子分解[6]を適用した分析もあれば、その尺度を基本的な因子として価値因子、活動性因子、力量性因子の3因子[7]に分解した分析も知られている。本研究では尺度に対する因子分解はまだ適用していないが、これを適用することも今後可能である。

画像表示には次元削減(PCA)を適用する。表示された画像分布を観察するにあたってユーザインターフェースが重要視される[8]。本研究では可視化システムを実装する際に対話操作機能が充実したライブラリを使用することで、画像の重なりが発生した場合でもユーザ操作で重なりを可能な限り少なくする。

3. 印象評価と可視化

本研究では1章でも述べた通り、画像群に対してSD法を採用した印象評価を実施し、続いて各画像の印象値とRGB値に次元削減を適用することで画像の印象分布と色分布を可視化する。次に各尺度の平均値と中央値を重ねて表示することで画像特徴と回答の関連を観察できる可視化手法を提案する。本章では提案手法の処理手順について述べる。3.1節では印象評価データの形式、3.2節では各画像の印象値とRGB値の次元削減、3.3節では各尺度の平均値と中央値の表示方法、3.4節では回答者の特徴傾向の可視化について説明す

る。

図1 本手法で実装した可視化システム



3.1 印象評価の収集とデータの構築

本研究では多数の画像を用意し、その画像の内容にふさわしい複数の形容詞対を尺度として設定する。そして複数の回答者を集めて形容詞対への適合度を回答させる。以上によって収集された回答群から、「画像」「回答者」「尺度」を3軸とする印象評価データを構築する。なお、以下の説明では回答形式を5段階のリッカート尺度とし、各画像の各尺度において平均値と中央値を算出するものとする。

3.2 各画像の印象値とRGB値の次元削減

続いて本手法では、各尺度の平均値と中央値で構成される多次元ベクトルを各画像に対して設定し、これに主成分分析(PCA)を適用することで各画像を2次元空間に配置する。この次元削減結果により、画像の回答傾向が類似している画像は画面上で近くに配置される。また、各画像のRGB値にも同様に主成分分析を適用して画像の色が類似している画像を近くに配置されるようにする。各画像の回答が類似しているか、RGB値が類似しているかの2通りの画像分布を表示できるようになる。図1にある可視化システムの左上のボタンと選択によって「印象値(平均値)」、「印象値(中央値)」、「RGB値」の3種類の中からユーザが観察したい画像分布を表示させることができる。

3.3 各尺度の平均値と中央値の表示

3.2節に示した処理によって表示させた画像付近に、各尺度の平均値と中央値を表示する。SD法にて採用するリッカート尺度の1から5段階に対して、図2のように緑・赤の2色相を用いた色を割り当てる。このとき、平均値・中央値ともに四捨五入して整数値とした上で色を割り当てる。図3では例として、尺度が5個ある場合に、各尺度の中央値を順に表示している。

評価段階： 1 2 3 4 5

図 2 5 段階評価の色設定

尺度1 尺度2 尺度3 尺度4 尺度5 3.2 1.8 2.1 4.0 4.5



図 3 画像付近に中央値または平均値を各尺度順に表示.

3.4 回答者の特徴傾向の可視化

回答者の特徴傾向を理解するために各回答者の回答した総数を積み上げ棒グラフを用いて回答結果（5段階評価）の内訳を可視化する。図 1 の左下にその例を示す。横軸は回答者を回答番号順に並べ、縦軸の回答の総数を、未回答を黒色、1 から 5 段階的回答を図 2 にある色設定で表示する。また、各画像の回答結果の詳細を見たいとき、その画像を選択すると図 1 の右にあるように回答の内訳を表示する。

4. 実施例

我々は本手法を Python の可視化ライブラリ Bokeh を拡張することにより実装した。Bokeh は平行移動や拡大縮小といった視点処理、クリック操作等によるオブジェクト選択機能、画像として可視化結果保存、といった対話操作機能が充実したライブラリである。

本事例では使用する画像を衣服画像に限定した。印象評価データとして表 1 に示す 2 種類のデータを用意した。尺度の選択には古川らの文献[9]を参考にした。本事例ではまずデータ 1 を可視化して結果を考察し、その検討結果にもとづいてデータ 1 の尺度の一部を変更し、画像の枚数と回答者の人数を増やす形でデータ 2 を作成した。

我々は以下の 3 つの観点から印象評価結果を観察した。

- 回答または RGB 値が類似する画像の特徴傾向
- 尺度が適正かどうか
- 各回答者の回答傾向

4.1 節ではデータ 1 について、4.2 節ではデータ 2 について、印象評価データの可視化結果と考察を示す。

表 1 印象評価データの内訳

	データ 1	データ 2
画像	137 枚	300 枚
回答者	女性 14 名	女性 15 名 男性 6 名
尺度 1	Formal – Casual	Formal – Casual
尺度 2	Cool – Cute	Masculine – Feminine
尺度 3	Cold – Warm	Modern – Classic
尺度 4	Dull – Vivid	Dull – Vivid
尺度 5	Dark – Bright	Simple – Gorgeous
評価段階	5	5

4.1 データ 1 の可視化と考察

印象評価データ 1 を可視化した結果を図 4～7 に示す。図 4 は全体の可視化結果で、画像分布は回答が類似する画像が近くに配置されている。図 2 の色設定で表示された実数値が平均値を表示している。緑色の文字の割合が多いことから、5 段階で 1 または 2 の回答結果が多いことがわかる。



図 4 画像付近に平均値を表示させた可視化結果.

図 5 (上) は画像を非表示にして一つの尺度の中央値のみを表示した可視化である。この可視化結果から、回答が類似する画像が画面上で近くに配置されていることがわかる。なお、図 5(上) は尺度 Formal – Casual、図 5 (下) は Cold – Warm の可視化結果である。ここからわかるることは Formal – Casual は 5 段階評価の 1 や 5 が多いのに対して、Cold – Warm は評価段階 2, 3, 4 の回答が多いことである。この結果から、前者の方が明快な印象をつけやすく、後者は明快な印象をつけにくいということがわかる。

我々の実装では、図 4 に示した画像群の中から特定の画像に関する回答結果の詳細を見たいとき、その画像を選択すると図 6 にあるように回答の内訳を表示する。この可視化結果から、図 6(左) にある 2 枚の画像は Dark – Bright に関して 5 段階評価の 4 または 5 であり、つまり両者ともに明るい印象を有していたことがわかる。背景画像から衣服の色まで異なる衣服画像である

にもかかわらず、どちらも明るい印象という回答結果であることから、形容詞対 Dark – Bright に対する解釈が回答者間で異なる可能性が示唆される。

回答者の 5 段階評価の傾向を可視化した結果を図 7 に示す。5 段階評価の色設定は図 2 に示した通りで、黒色は未回答データがあることを指す。図 7 では回答者が左右に並んでおり、縦軸が回答の総数を表している。この可視化から、回答者 5, 7 は評価段階の中央である 3 を多く選択する傾向が強く、これは控えめに印象評価をしたことがわかる。逆に回答者 11, 13 は評価段階の 1 と 5 を多く選択する傾向が強く、これは積極的に印象評価をしたことがわかる。回答者 1 は未回答が多いことから、印象評価の遂行を催促する必要があることがわかる。

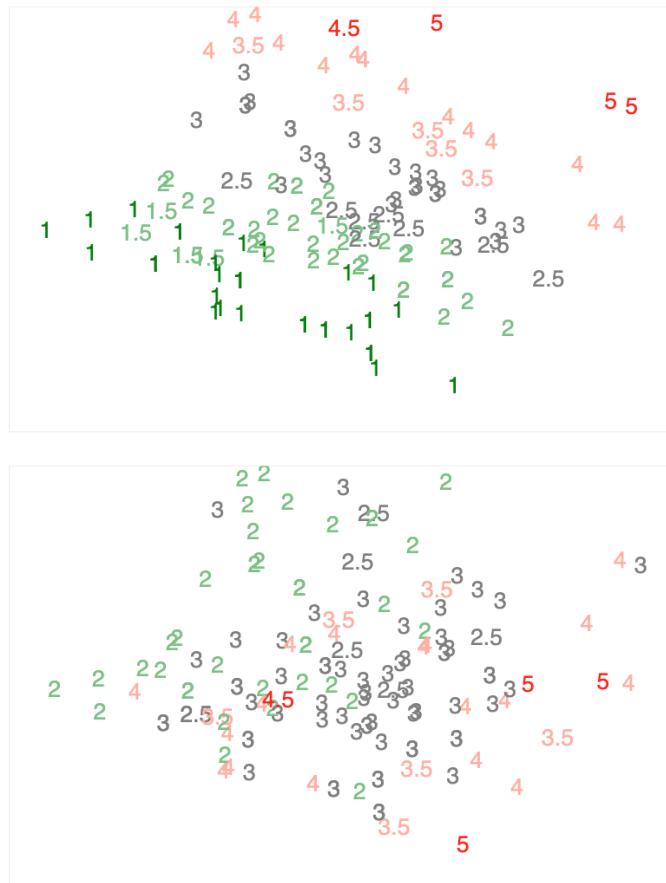


図 5 各画像の（上）Formal – Casual
(下) Cold – Warm の中央値。

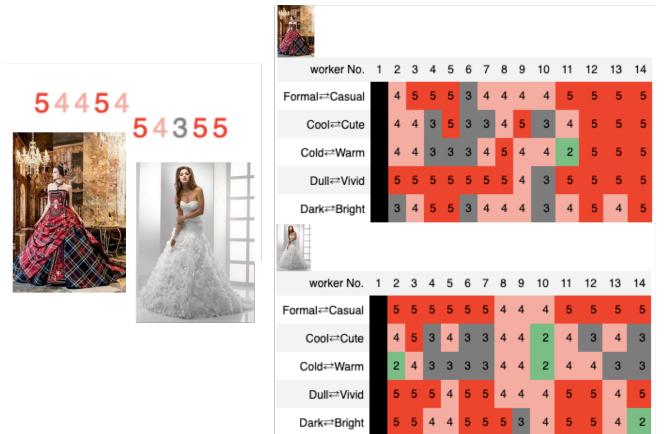


図 6 (左) 選択した画像。
(右) 選択した画像に対する回答結果の詳細。

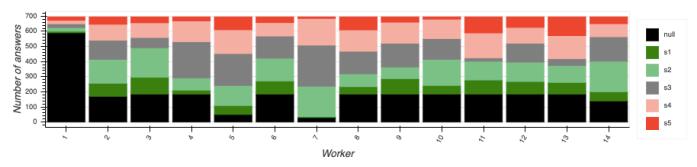


図 7 (データ 1) 回答者の 5 段階評価の回答の傾向。

4.2 データ 2 の可視化と考察

データ 1 の可視化を通して我々は以下の問題点を導いた。

- 尺度の再考の必要がある。データ 1 の可視化を通して、回答者によって複数の解釈が可能な形容詞対や、英語でのニュアンスを理解しにくい形容詞対があることが示唆された。例えば Cold – Warm という形容詞対の場合、服装自体が寒そうなのか暖かそうなのか、雰囲気が冷たいのか温かいのか、という 2 つの解釈があり得ることがわかった。Dark – Bright という形容詞対についても同様である。
- 回答者が 20 代の女性のみであることから、このデータのみから議論を一般化するのは望ましくないと考えた。

以上の観点から、問題のある尺度を変更しながら画像の枚数を増やして「データ 2」を作成した。そして女性だけでなく男性の回答者を増やし、さらに日本人だけではなく海外国籍の回答者も増やして回答を収集した。なお本原稿の提出時点ではまだ回答を収集中であり、未回答の設問が多く残っている点に注意されたい。

以下、データ 2 を可視化した結果を示す。ここで図 8, 9 では、RGB 値の距離が近い画像が画面上で近くに配置している。図 8 にある通り、色分布が類似してみえる画像ほど近くに配置されていることがわかる。図 9 (上) は尺度 Modern – Classic の回答の中央値を可視

化結果である。図 9 (下) は赤丸で囲まれた画像の拡大図と回答の中央値を示した。主成分分析によって配置された 4 つの画像の距離が近いことから、これらの衣服画像は RGB 値が類似している画像である。左の 2 つの衣服画像に関してはトップス・ボトムスとともに同じ色合いと形をしていることから回答も類似していることが示唆される。しかし、右の 2 つの衣服画像はどちらも白いドレスであるにもかかわらず、回答の中央値に大きな差異がある。このことから、衣服の色が同色でも、形が少しでも異なれば印象値も変わることが示唆される。あるいは、今回使用した衣服画像のデータでは画像ごとに異なる人物がモデルとなっているため、そのモデルの顔や髪の毛、肌の色が影響した可能性もあげられる。



図 8 RGB 値が類似する画像が近くに配置された画像分布。



図 9 (上) 各画像の Modern – Classic の中央値。
(下) 赤丸で囲まれた画像の拡大図と各中央値。

次に、回答者の 5 段階評価の傾向を可視化した結果を図 10 に示す。色設定は 4.1 節と同じである。黒色の未回答が全体的に多いため、データ 2 は 1 つの画像に対して回答が少ないことがわかる。そのため現時点での可視化結果からは、データ 1 に比べて各個人の回答内容が可視化結果に大きく影響する可能性がある。

以下、図 11~13 では回答が類似する画像が画面上で近くに配置されている。まず、日本人と海外の被験者の違いを考察できる可視化結果を図 11, 12 に示す。この可視化結果から、特に尺度の Dull – Vivid と Modern – Classic の回答傾向に違いをはっきりと観察できた。日本人の被験者は 5 段階評価のうち 2, 3, 4 を回答する傾向が比較的強く、海外の被験者は 1 または 5 の回答が目立つ。日本人は他国に比べて、控えめな印象評価をしていることがわかる。また、海外と日本での印象の解釈が特に Dull – Vivid と Modern – Classic において異なっている可能性もあり、その影響で海外の回答者に比べて日本人には白黒つけにくい尺度であるという可能性もある。

続いて、男性と女性の被験者の違いを考察できる可視化結果を図 13 に示す。図 13(上) は各画像の Formal – Casual に対する女性の回答の中央値である。図 13(下) は同じく男性の回答の中央値である。5 段階評価のうち Formal である 1 に近いほど中央値は緑色で表示されており、その逆に Casual である 5 に近いほど中央値は赤色に表示される。ここで図 13 にある赤丸で囲まれた画像は男性・女性の回答結果ともに同じ画像である。しかし、同じ画像にもかかわらず女性の回答結果では Casual の印象を持ち、男性の回答結果では Formal の印象を持つ。このことから Formal – Casual の尺度は男女で捉え方が異なることがわかる。しかし図 10 から考察した通り、データ 2 では各画像に対して回答が少なく、各個人が持つ印象が大きく可視化結果に反映される可能性がある。あるいは回答者が 5 段階評価の形容詞対にして誤った回答をした可能性も考えられる。

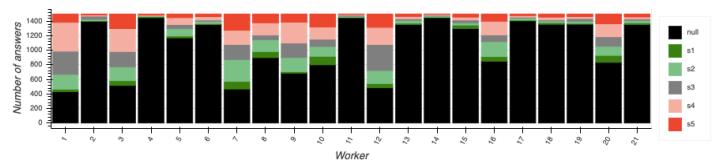


図 10 (データ 2)回答者の 5 段階評価の回答の傾向。

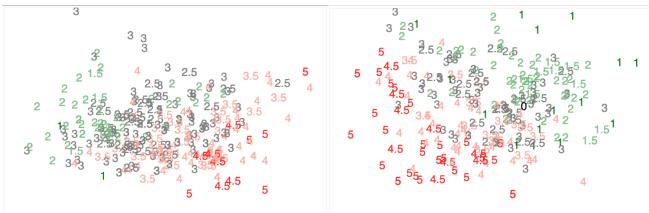


図 11 Dull – Vivid の各画像に対する中央値
(左) 日本人と (右) 海外の被験者による回答

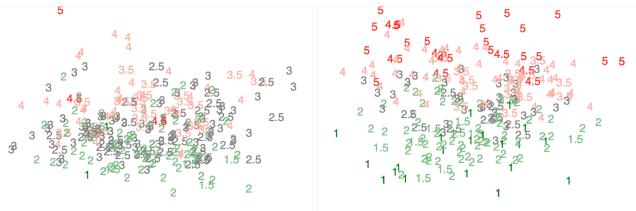


図 12 Modern – Classic の各画像に対する中央値
(左) 日本人と (右) 海外の被験者による回答

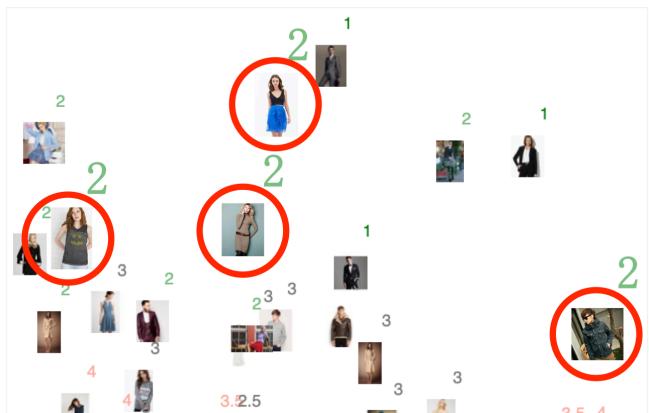


図 13 各画像の Formal – Casual に対する (上) 女性
(下) 男性の回答の中央値.

5.まとめと今後の課題

本研究では画像群に対して SD 法を採用した印象評価を実施し、続いて各画像の印象値と RGB 値に次元削減を適用することで画像の特徴分布を可視化した。本

研究のための実装した可視化手法では、次元削減による可視化結果に対して、各尺度の平均値と中央値を重ねて表示することで画像特徴と回答の関連を観察できるようにした。また、詳細を知りたい対象の画像を選択すると回答の内訳と回答者の回答傾向を観察できるようにした。この可視化手法により、画像、回答者、尺度の 3 つの軸から印象評価データを分析することが可能になると考えられる。

今後の課題は以下の通りである。

- 画像群に対する印象評価データの拡張
- スケーラブルな可視化手法への移行
- 機械学習の訓練データとして用いるためのアノテーションの付与形式の策定

本研究で可視化したデータは画像 300 枚という小規模なデータであったが、機械学習の訓練データとして用いるのであればもっと大量の画像が必要になる。また、データの信頼性を高めるためには偏りのない回答者群による回答が必要である。これらの点に留意しながら、さらに拡張したデータを作成し、さらに大規模な印象評価を進めたい。また、現状の可視化手法ではデータの形式が固定であるためスケーラブルな可視化手法ではない。そこで可視化手法の拡張も必要となる。最後に、画像の印象評価結果をどのような形式でアノテーションとして画像に付与するかを策定したい。これは画像の印象評価結果を機械学習の訓練データとして用いるときに必要となる。

謝 辞

データセットを提供していただいた株式会社 ABEJA 様に感謝いたします。

参 考 文 献

- [1] C. E. Osgood, "The nature and measurement of meaning", *Psychol. Bull.*, 1952.
- [2] 代, "3 相因子分析法による椅子の形態イメージの分析", J-STAGE, 心理学評論, 1982.
- [3] 豊田, "探索的ポジショニング分析", J-STAGE, 心理学研究, 2001.
- [4] M. M. Bradley, P. J. Lang, "Measuring emotion: The self-assessment making and the semantic differential", Pergamon, 2009.
- [5] Y. Kanno, T. Minetoki, T. Bogaki, K. Toko, "Visualization of Flavor of Sake by Sensory Evaluation and Statistical Method", Sensors and Materials, 2018.
- [6] J. Stoklasa, T. Talášek, J. Stoklasová, "Semantic differential for the twenty-first century: scale relevance and uncertainty entering the semantic space", Qual Quat, 2019.
- [7] C. E. Osgood, "Semantic differential technique in the comparative study of cultures", Am. Anthropol, 1964.
- [8] B. Moghaddam, Q. Tiam, C. Shen, T.S. Huang,

“Visualization & Layout for Image Libraries”, CIVR,
2002.

- [9] 古川, 三浦, “ラグジュアリーファッションに注目
したトレンドの定量分析”, Creative Commons,
2017.