

# SD 法による大規模印象評価にもとづくアノテーションを支援する可視化

飯島 緋理<sup>†</sup> 伊藤 貴之<sup>†</sup>

<sup>†</sup>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

E-mail: <sup>†</sup>{iijima.akari, itot}@is.ocha.ac.jp

**あらまし** 画像の印象を推測する機械学習の構築において、学習に使用する訓練データ作成は多数の画像に少人数でタグ付けする作業が主流である。しかしタグ付け作業者が有する個人の印象回答が学習結果に依存する問題がある。本研究では画像から受ける印象の個人差を小さくするため、多人数の印象回答値を参照して印象のタグ付けを半自動化する過程の可視化を提案する。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を用いてファジィ決定木を作成する。決定木によって自動タグ付けされたのち、目視確認及びタグ付け再考が可能になるような可視化システムを提供する。また、類似する画像を近くに配置した画像一覧を可視化することで、クラスタリングされた結果を画像特徴から観察する。決定木と類似画像一覧表示の可視化を連動させることで決定木の可読性が向上した評価結果について述べる。本報告では、1500 枚の女性着衣服画像に 43 人の作業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し、ユーザー操作による本研究の有用性を検証した。

**キーワード** SD 法, 可視化, 決定木, アノテーション

## 1. はじめに

SD 法[1]は概念の意味を測定するためにアメリカの心理学者 Osgood らが提唱した手法である。SD 法を用いた印象評価は学術的にも産業的にも多くの場面で重要な印象評価手段となる。印象評価データの活用例として、機械学習を用いた印象推定があげられる。画像を閲覧して回答する形式の印象評価を例にすると、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できるようになる。

一般的に、アノテーション作業は少人数で実施することが多い。1 人あたりの作業量が多いため、作業者の疲労が蓄積され、回答の信頼性が減少する可能性がある。そのため、近年ではアノテータのタグ付けの過程や結果を可視化し、訓練データの分布理解及びアノテーションの修正、品質保証を図る研究が活発である。Muller ら[2]はアノテータのタグ付け作業中を観察することによって 3 つのアノテーション設計におけるデザインパターンを提示しているが、実際はアノテーション自体には決まったルールが存在しない。その上、アノテータは専門家・非専門家を問わない。したがって、アノテーション設計は自由度が高い分、作業者の特性や作業者にかかる時間・負荷、タグの選定方法を吟味しなければ、訓練データが学習結果の足かせとなる問題が発生しうる。

ここで、アノテーションが個人の主観によって決められる場合を考える。Wu ら[3]は衣服の画像に対してアノテータから収集したキャプション・文章と、製品メタデータや属性ラベルを保持する大規模なデータセ

ット作成をした。しかし、収集するキャプションや文章は個人によって様々な回答が得られるため、全回答をアノテーションとして採用するには学習結果に対する信頼性が低くなる。

また、画像の印象を推測する機械学習の構築を考えた時、使用する訓練データは印象のタグが付いた画像を用いる。ここで、従来のアノテーション設計に則り、訓練データを構成する全ての画像の印象に対応するアノテーションを 1 人の作業者が付与するとする。この場合、作業者が各画像に対して個人的に有する印象をアノテーションとして付与するため、機械学習の挙動が作業者に依存してしまうという問題点がある。

この問題を解決する一手段として、複数の作業者に各画像の印象を回答させて、1 人あたりの作業量を減らしつつ、その回答結果を総合的に判断して各画像にアノテーションを付与する、ということが考えられる。この手段を採用する場合にも、印象推定の適切さが訓練データの品質に大きく影響されることには変わりはない。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が重要な課題となる。

本報告では、複数の作業者間で個人差の出る印象評価に対して、作業者群による過去の印象回答値を参照してこれからタグ付けする画像の印象を推定することにより、印象のタグ付け作業を半自動化する研究に取り組んでいる。その一環として本報告では、SD 法による大規模な印象評価によるアノテーションのデータ構築方法と、ファジィ決定木にもとづく印象推定の過程を可視化する手法を提案する。本手法では以下の 3 つの工程から構成されるものである。

- 画像群に対して大規模な印象評価を実施。

- 印象評価結果から自動的に画像のタグ付けをし、タグ付け結果を可視化。
- タグ付け可視化結果を基に人の手によるアノテーションを支援するシステムを提供。(アノテーションの半自動化)

1 枚の画像あたりの印象タグ付作業を少人数ではなく大人数で実施することで、1 人あたりの作業負荷を軽減しつつ、感性情報の個人差の解消につなげる。これを達成するには、大人数の印象回答値を分析できる可視化システムが肝要となる。可視化では、各画像の印象値からファジィ決定木を生成し、さらに類似画像一覧表示をする。ユーザ操作によって決定木と画像群を連動させることで、決定木の可読性向上と、画像特徴と印象の関係を明らかにする。

アノテーションの半自動化という課題は、機械学習のための訓練データ全般において重要である。第3次 AI ブームの中心的存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。可視化システムにより印象策定の半自動化をサポートし、アノテーション作業の効率化を図る。

## 2. 関連研究

印象評価結果を分析し、アノテーションとして印象タグを付与する手法として、決定木を用いることができる。決定木を用いることで、属性が決定されるまでの過程やクラスタリングによってルール化された軌跡を辿ることで、説明性を担保することができる。ファジィ決定木[4]を用いた SD 法による印象分析[5]では、リッカート尺度に基づいた印象値をクラスタリングすることで、確信度を持った属性値を得ることができる。ファジィ決定木は人の感性に合った解釈が可能であり、その過程やルールの説明性が高いことから、印象分析に適した手法として知られている。本研究はファジィ決定木による印象評価の構造化を支援する可視化として位置づけられる。

決定木は必ずしも可読性の高い状態で生成されるとは限らない。専門家のみならず初学者にも決定木から情報を読み取らせるには、可読性の高い視覚表現が有用である。決定木と複数の分析画面を提示することで、決定木の可読性の向上が見込まれる。Muhlbacher ら[6]は、統計学の非専門家でも適切な決定木を効率的に選択できる可視化を提案している。枝切りや丸め処理を施すことで、注目して欲しい箇所を示し、効率的な分析を可能としている。しかし、実際のユーザ評価では、複数ある画面を1つずつ説明することにかかり時間がかかっている。分析画面の提示が10 エリアあるため、口頭での説明時間に2時間を要している。可視

化画面の操作を実施する前の説明が長いと、操作中の作業者の集中力低下につながる。説明時間を少なくするためにも画面数を最小限に留めておく必要がある。本研究では説明資料を準備し、資料数削減のためにも、分析画面の提示は2エリアのみで実装した。

機械学習における畳み込みニューラルネットワーク(CNN)では、学習プロセスを可視化し、デバッグや行き詰まりの阻止、モデルのパフォーマンス向上のための潜在的な方向性の発見として、近年研究が盛んである。Liu ら[7]は、ノードに画像特徴量の表示をしており、対話的操作により詳細を適宜表示させることでユーザのインタラクティブ性を高くしている。クリックによってエッジの強調や拡大・縮小ができる分、画像レイアウトの複雑性が高くなる。

大量画像の階層型データ可視化手法に Image Browser[8]が挙げられる。本手法ではファジィクラスタリングにより画像を分類し、決定木によって階層型データを構築する。このようにして構造化された画像群を閲覧するには Image Browser が有用である。Liu ら[9]は、Image Browser によって区画化された画像群に対して、画像特徴が類似する画像を隣接させるように再配置することによって、より視覚的に画像特徴の分布を表現する手法を提案している。本研究においては画像特徴にもとづいて画像群を配置することで、本研究の目的に合致した分析が可能となる。

## 3. SD 法を用いた印象評価

### 3.1. 尺度の因子選択

本研究では使用する画像群を女性衣服の画像に限定した。ファッション・衣服における SD 法属性について、多くの文献がある。堀尾ら[10]は衣服画像の印象は、活動・嗜好性/力量性/質感・立体感の3因子から構成されていると報告している。質感・立体感の因子は媒体の表現方法の影響が大きく、衣服のみが写っている衣服画像ではなく人物が着衣している衣服画像において構成可能である。したがって本研究では質感・立体感の因子を採用するため、衣服画像を人物が着衣している衣服画像に限定した。一方、村山ら[11]は衣服画像の印象は、自己主張/他者からの推薦/穏健性/審美性の4因子から構成されていると報告している。また、古川ら[12]は衣服画像の印象は、色彩/正統性/年齢・性別/装飾性の4因子から構成されていると報告している。以上の複数の因子から、性差・年齢差を考慮しない因子であり個人差が明らかに大きく出ない因子である色彩/立体感/正統性/穏健性/装飾性の5種類の因子を本研究では採用する。

### 3.2. パイロットテストと尺度決定

3.1 節で選択した因子から列挙した形容詞対を表1に示す。印象評価を実施するにあたり、回答する項目

数を増やしすぎることによって、作業者の疲労につながり、ひいては回答の信頼性が下がる可能性がある。これを回避するために我々は、本研究の印象評価に適用しうる形容詞対を列挙した上で、類似する意味を持つ形容詞対のいずれかを削除する、あるいは本研究で使用する画像に適切な尺度といえない形容詞を削除する、という手順で印象評価のための形容詞対を選出した。この工程のために我々はパイロットテストを実施した。画像には印象評価にも用いる 10 枚の着衣画像を用い、表 1 にある 15 項目の形容詞対に対して、回答者(20 代女性 15 名)から各評価項目に対して 5 段階評価の回答を収集した。この評価結果をもとにして我々は、任意の形容詞対間の距離行列を python の scipy ライブラリを用いてユークリッド距離計算により算出した。

続いて各形容詞対について、他の形容詞対との距離の平均値を求める。この平均値が大きいほど、他の尺度との相関がないことを意味する。この平均値を「相違度」と称して表 1 に示す。本研究では、この相違度が最大となる形容詞対を因子ごとに抽出する。以上の工程によって選出された形容詞対を表 2 に示す。

表 1 選択された因子から形容詞対の選出。

因子	形容詞対		相違度
「色彩」	暗い	明るい	5.03
	鈍い	鮮やか	4.78
「立体的」	直線的	曲線的	5.01
	平面的	立体的	4.74
	フィットした	ルーズな	5.42
「正統性」	高級な	質素な	5.84
	洗練された	無粋な	5.57
	フォーマル	カジュアル	6.02
「穏健性」	日常的な	非日常的な	5.08
	普通	異常	4.74
	無個性	個性的	4.61
	自然	不自然	4.86
「装飾性」	くすむ	目立つ	4.47
	地味な	派手な	4.37
	シンプル	ゴージャス	4.77

### 3.3. 画像の前処理

本研究は背景が無地単色である女性の着衣画像<sup>\*1</sup>を対象として印象評価を実施してする。ここで、人物の人相・表情・髪型などが統一されていないと、衣服以外の要因が印象評価に影響する。そこで本研究では、着衣画像から顔を削除する。具体的には、まず OpenCV を用いて顔領域を検出し、検出された顔領域から上を背景色と同じ色で隠す。更に背景削除処理<sup>\*2</sup>を実施することで、全画像の背景色を統一させる。

### 3.4. 印象評価の収集と可視化データの構築

表 2 に示す形容詞対と 3.3 節の画像群を用いて、印象評価を実施した。画像を 1,500 枚、回答者を 43 名、尺度を 5 項目として、5 段階評価で回答を収集した。

43 人をランダムに 3 グループに分けることによって、1 人あたり 500 枚の画像の印象評価を回答している。これは回答者の作業量を減らすためである。したがって、画像 1 枚あたりの回答人数は 14~15 人である。内訳を表 2 に示す。

表 2 印象評価データ。

画像	女性着衣画像 1500 枚	
回答者	43 人(全員 20 代 / 女性 37 人男性 6 人 / 日本国籍 35 人中国国籍 8 人)	
「色彩」尺度 1	暗い	明るい
「立体的」尺度 2	フィットした	ルーズな
「正統性」尺度 3	フォーマル	カジュアル
「穏健性」尺度 4	日常的な	非日常的な
「装飾性」尺度 5	シンプル	ゴージャス
評価段階	5 段階	

回答者には、画像を 1 枚ずつランダムに表示させた。この時、回答者には自身の PC を使用してもらい、ブラウザは Chrome を使用するよう指定した。また、作業中は画面の明るさを変更しないように指示をした。これは画面の明るさが変化することで衣服画像に対する印象が変わることを防ぐためである。回答者には作業時間分の謝金を用意した。

## 4. 可視化手法

### 4.1. ファジィ決定木

本研究ではクラスタリング手法にファジィクラスタリングを採用する。一般的なクラスタリング(いわゆるハードクラスタリング)は、各クラスタへの帰属は 0 か 1 の 2 値で判断されるが、ファジィクラスタリングは[0, 1]の範囲の実数で各クラスタへの帰属を算出するため、曖昧で柔軟なクラスタ表現が可能である。本手法ではクラスタへの帰属値をクラスタリングの確信度の値として扱う。印象評価にはリッカート尺度を用いているため、間隔尺度の水準を満たしていると仮定する。ファジィ決定木の構築手順は以下の通りである。

- クラスタリング数と決定木の深さを決定。
- クラスタリング結果の確信度を算出。
- 確信度をもとに決定木を構築。

本手法による印象推定の手順を図 1 に示す。ここでは 1 枚の画像の 5 段階評価「暗い-明るい」を例とする。現時点での実装ではクラスタ数を 3 に固定し、印象が「暗い」「明るい」「どちらでもない」の計 3 クラスが各クラスタに対応するものとする。また、ファジィ決定木の構築に際して、木の深さの最大値をあらかじめ指定している。我々の実験では、決定木の深さを 6 以上に指定すると決定属性の再帰や繰り返しが発生し頻繁に出現したため、最大深さを 5 にした。

各画像に対して複数の参加者が回答することを前提にして、図 1①に示すように 5 段階評価の各値を選択した人数を集計し、続いて図 1②に示すように各値の人数比を算出する。この値から、Python の fcmeans

<sup>\*1</sup> 使用画像 : <https://www.kaggle.com/dqmonn/zalando-store-crawl/>

<sup>\*2</sup> 背景削除 : <https://ja.clippingmagic.com/>

ライブラリ<sup>\*3</sup>を用いて、図 1③に示すようにファジィクラスタリングの確信度を算出する。そして確信度が一番大きい値をクラスタリング結果として採用することで、印象推定を完了する。



図 1 右の画像が印象項目「暗い-明るい」から「暗い」印象であると策定されるまでの手順。3つのクラス「暗い」「どちらでもない」「明るい」から確信度の高いクラスに策定される。

続いて、図 1 に示した手順で算出される確信度をもとに決定木を生成し、対話的に平行座標プロットを表示する Python モジュール<sup>\*4</sup>を適用して決定木を表示する。図 2 に決定木の構造のイメージ図を示す。

ここでは、1つの枝は1画像を表しており、木構造を上位階層から下位階層にたどることで、各画像がクラスタリングされる工程を表現する。決定ノードは、分類条件と条件判定を受けた画像数(図 3①)を表し、各画像に対して範囲[0,1]で算出される確信度の詳細(図 3②)が表示される。

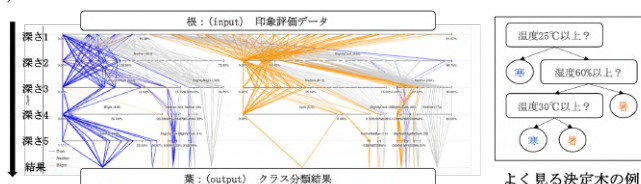


図 2 決定木の構造図。(左)本手法の決定木。(右)よく見る決定木の例。

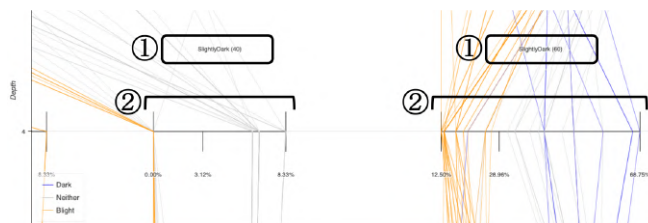


図 3 決定ノードの詳細。①分類条件 (条件判定を受けた画像数)。②確信度の詳細。

分類条件には評価時のリッカート尺度 5 段階を使用している。木構造の葉ノードは、クラスタリング結果となる各クラスを表現している。色はクラスごとに異

なり、色自体には意味を持たない。この可視化結果から、各画像がクラスタリングされる過程を観察できるだけでなく、その確信度も推察可能である。概して、浅いノードでクラスが決定される画像は確信度が高い傾向にあり、逆に深いノードでクラスが決定される画像は確信度が低い傾向にある。

## 4.2. 類似画像群の一覧表示

4.1 節でクラスタリング結果による画像の木構造を得ることから、これを画像群に適用する。本手法では、決定木のクラスタリング結果ごと及び深さごとに枠を生成し、木構造の葉ノードに対応する領域に画像群を表示する。さらに各枠に所属する画像群を一覧表示するため、画像の RGB 値を多次元ベクトルとし、これに主成分分析(PCA)を適用して各画像を 2 次元空間に配置する。これにより、色分布が類似する画像を近くに配置する。詳細を図 4 に示す。

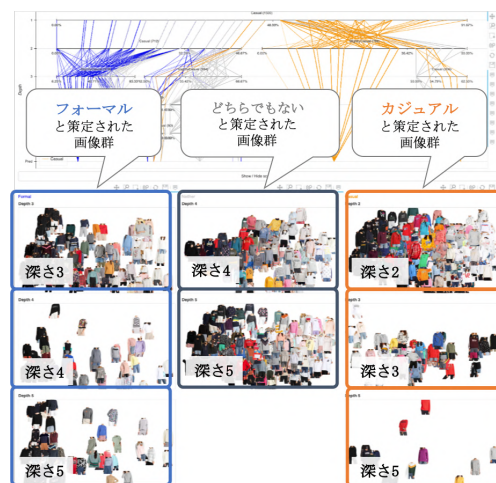


図 4 類似画像群の一覧表示の詳細。

## 4.3. ユーザインターフェース

本手法を操作するためのユーザインターフェースを図 5 に示す。このユーザインターフェースによってアニメーションの半自動化を目指す。決定木により画像(本研究では 1500 枚)がすでに分類済みで、この分類可視化結果をもとに人の手で根拠のある再分類ができる。クラスの再分類は画像中央にある点描画をダブルクリックすることでできる。クラスの再分類は図 6 にあるように、点描画の色が変わることによって確認できる。この点描画は決定木と画像群一覧表示の間にあるボタンによって表示・非表示の操作ができる。画像特徴をよく観察したい時にユーザはこのボタンを活用する。

操作後は、「Save」の青いボタンをクリックすると、再分類結果のダウンロードが可能であり、JSON 形式で再分類結果データを取得できる。ウィンドウ上部にあるタブを切り替えることで尺度ごとの可視化結果を観察することができる。平行移動、拡大縮小、再描画、ホバー機能やフィルタリング機能は、ウィンドウ右側

<sup>\*3</sup> fcmeans : <https://pypi.org/project/fuzzy-c-means/>

<sup>\*4</sup> iparacoord : <https://github.com/intellygenta/InteractiveParallelCoordinates/>





図5 本手法を操作するためのユーザインタフェース。

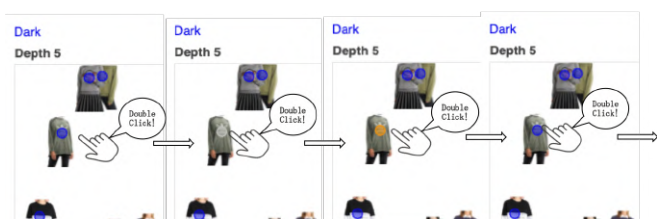


図6 ダブルクリックで人の手による再分類方法。

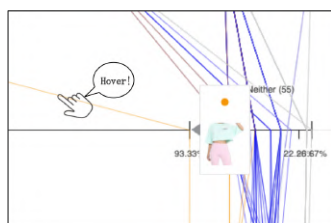


図7 ホバー機能の詳細。

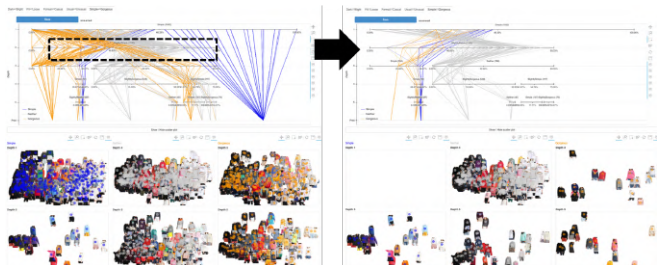


図8 フィルタリング機能の詳細。(左)フィルタリング前。(右)フィルタリング後。

のアイコンを選択することで使用でき、これらはどのプロットエリアでも操作可能である。ホバー機能は、図7にあるように、決定木の枝部分と、画像の中心にある点描画にカーソルを乗せることで、各画像が拡大されたものと初期分類結果のクラス色を点で表示している。フィルタリング機能は、例えば図8(右)で深さ2の決定ノードをドラッグアンドドロップで選択すると、図8(左)のように、選択された枠の中の決定ノードを通過する枝と、その枝に該当する画像が画像群一覧表示と連動してフィルタリング表示される。これを活用することによって、ユーザは気になる決定ノードや枝と画像特徴を見比べたり、分析したいクラスのみを絞ってクラス分類の傾向を観察することができる。

## 5. 実行例

本事例では表2に示す印象評価データを使用し、可視化している。我々は本手法を Python 3.9.5 の可視化ライブラリ Bokeh 2.4.0<sup>\*5</sup> 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可視化ライブラリである。また、4.1 節にあるように、ファジィ決定木を構築する際の分類クラス数は3つで、深さは最大5までに固定している。

本章では、決定木のみ可視化結果、類似画像一覧表示のみの可視化結果、決定木と類似画像一覧表示のフィルタリング機能を用いた可視化結果を示す。

### 5.1. 決定木から得られる知見

図9から図13の全体を通して、各尺度のクラスタリング分類結果について、3つのクラスの間には大きな偏りの存在は確認されなかった。画像群は全ての尺度について、3クラスをほとんど網羅していることがわかる。今回の実行例では、表2に示す印象評価を実施する前に、少人数の参加者によるパイロットテストを実施して、尺度が適切であるか、尺度間に強い相関がないか、といった点をあらかじめ見積もっており、これが効果的であったものと考えられる。しかし、図12「ゴージャス-シンプル」と図13「非日常的な-日常的な」の決定木に注目すると形が似ていることがわかる。パイロットテストによって印象因子ごとに項目を決定したが、この2つの尺度に関しては衣服においてやや同義であることが観察できた。

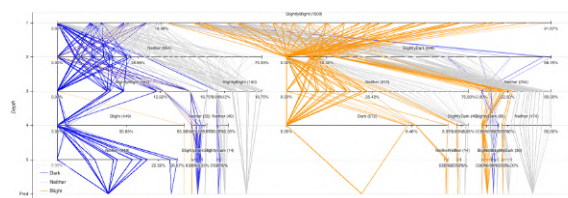


図9 尺度「暗い-明るい」のファジィ決定木。

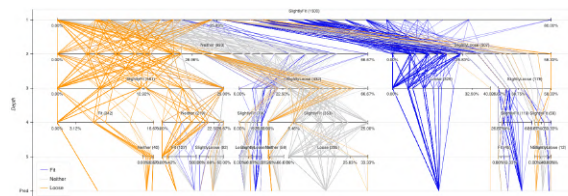


図10 尺度「フィット-ルーズ」のファジィ決定木。



図11 尺度「フォーマル-カジュアル」のファジィ決定木。

<sup>\*5</sup> Bokeh : <https://bokeh.org/>

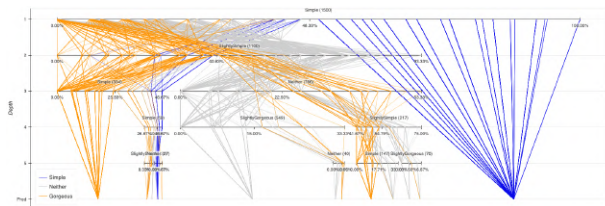


図 12 尺度「シンプル-ゴージャス」のファジィ決定木.

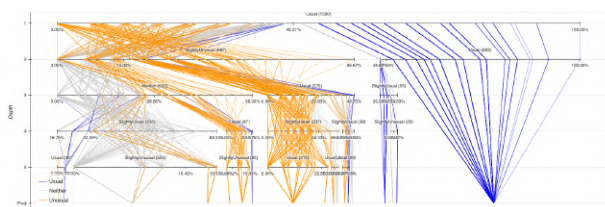


図 13 尺度「日常的な-非日常的な」のファジィ決定木.

また、決定木を観察することで、印象を回答しやすい画像群を発見することも容易である。例えば図 11「カジュアル」、図 12「シンプル」と図 13「日常的な」に関しては、決定木の深さが浅いところでクラスが策定されていることから、画像群の中でも印象の確信度が高い尺度であることが推察される。今回の画像群は衣服の中でもトップスが多く、さらに無地の衣服やトレーナーが多く含まれており、これらが印象を回答しやすい衣服の例であったと考えられる。

## 5.2. 類似画像群の一覧表示から得られる知見



図 14 尺度「暗い-明るい」の類似画像群一覧表示。(左)「暗い」と深さ 5 で策定された画像群。(中央)「どちらでもない」と深さ 3, 4, 5 で策定された画像群。(右)「明るい」と深さ 4, 5 で策定された画像群。

図 14 は尺度「暗い-明るい」に対する類似画像群の一覧表示例である。全体を観察してわかるように、「明るい」印象は色が鮮やかな衣服が多く、「暗い」印象は黒に近い衣服が多い。「どちらでもない」印象には青や緑の衣服、あるいは複数の色を組み合わせ合わせた衣服が多いことが観察できる。

## 5.3. 決定木と類似画像群の一覧表示のフィルタリング機能から得られる知見

図 15(左)は尺度「フォーマル-カジュアル」の可視化結果である。決定木から対となるはずの「フォーマル」と「カジュアル」の混合分類が見られたので、その部

分をフィルタリングすると、図 15(右)にあるようなフィルタリングされた画像群を観察できる。図 15(右)中の赤矢印で示している衣服で、同じ深さで同じ色/形/デザインが似ている衣服が存在した。2 つの画像群に所属する衣服画像の違いはモデルのポーズと髪の毛が少し写っているかないかである。衣服の印象に限定するために顔削除や背景削除して印象評価を実施したが、着衣画像であればある程度、モデルの髪形や体格が影響する可能性が示唆される。



図 15 尺度「フォーマル-カジュアル」の混合分類が見られる箇所をフィルタリングした詳細。(左)決定木の可視化。(右)フィルタリング後の画像群一覧表示の詳細。

尺度「フォーマル-カジュアル」に関わらず、他の尺度でも決定木において混合分類が見られるが、それらはどれも深さが深いところでクラス策定されている。このことから混合分類が見られる画像群はクラス分類の確信度が低いものであり、ファジィクラスタリングの最終的なクラス分類の判断基準は正しく働いていることがわかる。

## 6. 評価実験

本章では本手法を用いた可視化の有用性と、システムの操作性と有効性を示すための評価実験について述べる。評価者は 20 代女性 9 名で、所属グループとしては以下の 3 グループに分けている。

- ML グループ：機械学習の専門家 3 人。
- VIS グループ：可視化の専門家 3 人。
- NONE グループ：機械学習/可視化の非専門家 3 人。

本報告では機械学習の専門家 3 人のグループを「ML グループ」、可視化の専門家 3 人のグループを「VIS グループ」、機械学習/可視化の非専門家 3 人のグループを「NONE グループ」と定義する。

訓練データのアノテーションを題材にしているため、機械学習の専門家を混じえている。また、決定木の理解への難しさから可視化の専門家と非専門家とで比較対象とした。本システムを使用し、評価者にはタグ付け可視化結果をもとに人の手による再アノテーションをしてもらう。

評価手順としてまず、評価者に説明資料を読んでもらってから本システムを用いたタグ付けを開始しても



らった．本システム※6はリモートサーバに置かれているため，評価者には URL からアクセスして作業を開始してもらう．ブラウザは Chrome で固定し，画面の明るさを作業中に変更しないように評価者に指示をした．その際に，画面録画による操作ログを残してもらい，さらに分類結果をダウンロードしてもらった．最後に可視化や操作に関する評価項目に答えてもらい，評価者には作業時間分の謝金を用意した．評価項目は以下の4つであり，これら全てに関してコメントや感想を自由記述として回答してもらった．これらの手順には特に制限時間は設けていない．

- 決定木について．
- 類似画像一覧について．
- 画面操作について．
- タグ付け作業について．

## 6.1. 画面録画による操作ログによる評価結果

画面録画による操作ログから観察できたことを述べる．どのグループも，1 尺度にかかる作業は約 10～20 分で，5 尺度あるため大体その 5 倍の約 50～100 分ほど作業に時間がかかっていた．作業時間平均の詳細を表 3 に示す．

表 3 本システムで再分類作業にかかった時間の平均．

	ML	VIS	NONE
「色彩」暗い-明るい	15m11s	15m14s	15m00s
「立体的」フィット-ルーズ	14m14s	17m52s	17m03s
「正統性」フォーマル-カジュアル	14m55s	15m04s	12m37s
「穏健性」日常的な-非日常的な	14m32s	16m18s	15m53s
「装飾性」シンプル-ゴージャス	11m52s	19m12s	16m55s

ML グループは機械学習や予測モデルの理解があるからか，決定条件の確信度が低いものを中心にクラス分類する傾向があった．また，フィルタリング機能を多用しており，決定木による分類過程と画像との関係性を注意深く観察している様子が見られた．

VIS グループは基準を決めてから再分類する傾向があり，これは ML グループと NONE グループと大きく異なる動きであった．例えば，図 16(上)の赤矢印の流れは，ML グループと NONE グループの再分類作業の流れの傾向を表している．これら 2 つのグループは左上から右下へと流れるように再分類作業していたが，VS グループは図 16(下)の赤矢印の流れのように，「どちらでもない」から再分類を開始し，全体の画像分布を何度も眺めて，自分の基準をブレないように心がけている様子が観察できた．全体像から分析に入るといふ，普段から分析慣れしている可視化の専門家らしい操作ログ結果が見れた．このことから，本システムの全体像の見やすさによって，個人の主観が首尾一貫するような再分類作業を可能としている．

NONE グループは決定木をあまり使用しておらず，操作は平行移動と拡大縮小のみの作業が目立った．

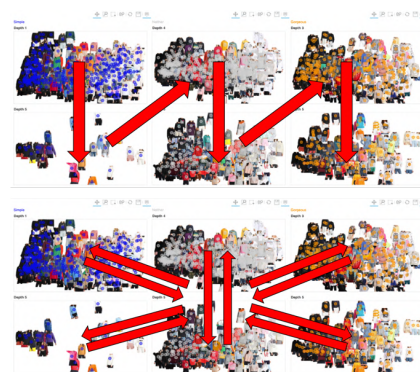


図 16 画面録画の操作ログから観察できる再分類作業の流れの傾向．(上)ML グループと NONE グループ．(下)VIS グループ．

このことから，操作や分析慣れしていないユーザにとって，決定木の理解や，インタラクティブな操作による分析が難しかったことが観察できる．

## 6.2. 再分類結果による評価結果

表 4 分類済み画像から再分類された画像数の平均．

	ML	VIS	NONE
「色彩」暗い-明るい	55.98	26.65	47.99
「立体的」フィット-ルーズ	27.65	<b>21.99</b>	67.98
「正統性」フォーマル-カジュアル	56.65	26.66	93.31
「穏健性」日常的な-非日常的な	80.97	37.65	<b>93.99</b>
「装飾性」シンプル-ゴージャス	68.29	38.31	83.32

評価者には作業後，再分類結果をダウンロードしてもらった．表 4 は，1500 枚の自動分類された画像が作業によって再分類された画像数の平均である．再分類された画像は最大で平均 93 枚であり，約 100 枚以内の再分類操作で作業が完結していることがわかる．決定木による自動分類の結果が再分類の支援に優位に働いていることが分析できる．

## 6.3. 操作・システムに関する評価結果

作業後，評価者に自由記述のコメントやアンケートに回答してもらった．

まず，決定木に関しては，理解に時間がかかったため，可読性の向上を期待する声が NONE グループには多かった．その一方で ML グループと VIS グループは，フィルタリング機能を多用し，類似画像一覧と並行して観察することで理解がしやすかったとの声があった．決定木のみを使用であれば，専門家・非専門家にかかわらず可読性は低いことは予想していたため，類似画像群一覧表示とフィルタリング機能による補助が可読性の向上に有効に働いていることが示された．

次に，類似画像一覧に関しては，尺度「暗い-明るい」以外ではうまくいかなかったという意見があったため，RGB 値以外でも次元削減する必要性を感じた．また，尺度「ゴージャス-シンプル」では，生地感やデザイン・ロゴといった衣服画像の細部まで注目する必要があったため，何度も拡大画像を必要とした．尺度ごとに

※6 本システム： <https://user-test-imp.sakura.ne.jp/labo/evaluation/en/decition.html/>

よって類似画像一覧の役目に一貫性がないことは、評価者には操作に慣れるまでの時間を必要とさせてしまう。ホバー機能による詳細情報表示の再設計や、画像群の特徴の見方を変更できるような新機能追加や改良の余地があると感じた。

画面操作に関しては、NONE グループでは、操作が少し難しいと感じたという意見があがった。普段から分析慣れしていないユーザにとっては、システムで作業する以前に、操作の理解に時間がかかったことがわかる。しかし、総じて全グループから、インタラクティブな操作が豊富で、ダブルクリック操作でのクラス変更が便利だったとの声が多くあった。

最後に、タグ付け作業に関しては、タグ付けの精度が上がると感じた、可視化により分類基準のブレはなかった、分類する判別要素が人によって違うことを実感した、などの評価があった。印象のアノテーションを支援するという意味において、この可視化システムは全体的に効果があったと思われる。

## 7. まとめと今後の課題

本報告では、複数の作業間で個人差の出る印象評価に対して、作業群による過去の印象回答値を参照してこれからタグ付けする画像の印象を推定することにより、印象のタグ付け作業を半自動化する手法を提案した。これにより、印象のタグ付けを目的とした本研究の新しいアノテーション作成手順を適用することで、画像から受ける印象の個人差を解消しつつ、印象タグのアノテーション作業の効率化を実現できた。また、本システムを用いて実際にアノテーション作業を評価者にしてもらい、操作ログと再分類結果やアンケートによって、本手法の有用性を示した。評価結果から、決定木と類似画像一覧を連動して活用することで、決定木の可読性の向上を確認できた。

今後の課題としては、まず印象評価の回答者の偏りを改善する。表 2 にあるとおり、回答者はほぼ 20 代女性が占めているため、男性の回答者を増やしたり、幅広い年代の回答者を集めて大規模な実験を再度実施したい。また、現在の実装では決定木モデルにファジィクラスタリングを採用しているが、異なるモデルを採用して比較対象実験を実施したい。加えて、異なる画像群での印象評価も実施したい。次に、現段階では、決定木のクラス数と深さを固定して可視化しているため、パラメータを可変にして再実験したい。これにより、ユーザ操作によって自由な表現を可能とする。さらに、初学者でもわかるように、類似画像一覧表示と決定木の連動関係をより操作しやすく、理解しやすいデザイン設計を考えたい。最後に、本研究の評価測定を追加実施したい。本研究では、6 章にもあるように、評価者は実際に本システムを使用してアノテーション

作業をした。今後はその評価者たちが再分類した結果を他者に見せ、一般的にタグ付けが適切かどうかをさらに評価してもらう。以上の工程により、本手法において印象の個人差の解消につながっているかを測りたい。そして、これらの機能を実装し、評価実験を完了させた後に、訓練データ作成の現場でのアノテーション作業を支援し、印象推定における学習結果の向上に貢献したい。

## 謝 辞

本研究の一部は、日本学術振興会科学研究費補助金の助成に関するものです。

ユーザーテストにご協力いただいた皆様に感謝いたします。

## 参 考 文 献

- [1] Charles E. Osgood, "The nature and measurement of meaning", *Psychological bulletin*, 197, 1952.
- [2] Michael Muller, Christine T. Wolf, Josh Andres, Michael Desmond, Narendra Nath Joshi, Zahra Ashktorab, Aabhas Sharma, Kristina Brimijoin, Qian Pan, Evelyn Duesterwald and others, "Designing Ground Truth and the Social Life of Labels", *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1-16, 2021.
- [3] Hui Wu, Yupeng Gao, Xiaoxiao Guo, Ziad Al-Halah, Steven Rennie, Kristen Grauman, Rogerio Feris, "Fashion IQ: A new dataset towards retrieving images by natural language feedback", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11307-11317, 2021.
- [4] James C. Bezdek, Robert Ehrlich, William Full, "FCM: The fuzzy c-means clustering algorithm", *Computers & geosciences*, 191-203, 1984.
- [5] OMatsuo, Takeshi, et al., "Sensitivity Information Analysis of Running Shoes Using Fuzzy Decision Tree and Visualization of Analytical Results" *SCIS & ISIS SCIS & ISIS 2010. Japan Society for Fuzzy Theory and Intelligent Informatics*, 2010.
- [6] Thomas Muhlbacher, Lorenz Linhardt, Torsten Moller, Harald Piringer, "Treepod: Sensitivity-aware selection of pareto-optimal decision trees", *IEEE transactions on visualization and computer graphics*, 174-183, 2017.
- [7] Mengchen Liu, Jiaxin Shi, Zhen Li, Chongxuan Li, Jun Zhu, Shixia Liu, "Towards better analysis of deep convolutional neural networks", *IEEE transactions on visualization and computer graphics*, pp. 91-100, 2016.
- [8] Bederson, Benjamin B, "PhotoMesa: a zoomable image browser using quantum treemaps and bubblemaps", *In Proceedings of the 14th annual ACM symposium on User interface software and technology*, 71-80, 2001.
- [9] Xiaotong Liu, Yifan Hu, Stephen North, Han-Wei Shen, "CorrelatedMultiples: Spatially Coherent Small Multiples With Constrained Multi-Dimensional Scaling", *Computer Graphics Forum*, 7-18, 2018.
- [10] 堀尾華子, 加藤雪枝, "衣服の印象評価における媒体間の比較", *日本家政学会誌*, 53(7), 693-701, 2002.
- [11] 村山和弘, 油屋直子, "学生の衣服の色彩嗜好と色彩感情に関する分析", *尚絅学院大学紀要*, 55, 157-164, 2008.
- [12] 古川貴雄, 三浦爾子, "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", *共立女子大学・共立女子短期大学総合文化研究所紀要*, 23, 35-48, 2017.