

印象評価にもとづくアノテーション作業の半自動化を支援する可視化システム

飯島緋理¹⁾(非会員) 伊藤貴之¹⁾(正会員)

1) お茶の水女子大学

A Visualization System for Assistance of Semi-Automated Annotation Tasks based on Impression Evaluations

Akari Iijima¹⁾ Takayuki ITOH¹⁾

1) Ochanomizu University

{ijijima.akari, itot}@is.ocha.ac.jp

概要

画像の印象を推測する機械学習の構築において一般的に、学習に使用する訓練データ作成は多数の画像にアノテーションを付与する作業が必要である。しかし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低減し、かつ作業者の負担を低減するために、多人数の印象回答値を機械学習した結果にそって画像の印象のアノテーション付与を半自動化するシステムと、その過程を可視化する手法を提案する。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生成する。このファジィ決定木によって画像を自動分類したのち、その結果と過程を可視化することで、ユーザによる画像再分類を支援する。決定木の表示と類似画像の一覧表示を連動させることで、決定木の可読性を向上させ、アノテーション付与の傾向理解を促す。本論文では、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すとともに、ユーザ実験による本研究の有用性を検証する。

Abstract

The construction of training datasets for machine learning to estimate the image impression of generally requires annotating a large number of images. Here, it is problematic in that the training results depend on the individual impression responses of the operators. In this study, we propose a semi-automatic system for annotating impressions of images based on the results of machine learning of the impression responses of many operators, and a method for visualizing the annotation process in order to reduce the individual differences in the impressions of images and the burden on the operator. First, the system evaluates the impressions using the SD method, and then generates a fuzzy decision tree using the impression values of each image. The fuzzy decision tree automatically classifies images, and then visualizes the results and the process of classification to assist the user in reclassifying images. By linking the display of the decision tree with a similarity-based image browser, we improve the readability of the decision tree and support understanding of annotation trends. In this paper, we present an example of visualization using this method on data from 43 workers' impression evaluations of 1,500 images of women's clothing, and verify the usefulness of this research through user experiments.

1 はじめに

コンテンツの印象をアノテーションとしてコンテンツに付与する場面は多い。例えばオンラインショッピングサービスにおいて、商品の印象をアノテーションにして、それを手がかりに商品の検索や推薦ができるような機能を設けることがある。また近年では、深層学習を用いてコンテンツの印象を推定する研究が活発に発表されている [1, 2] が、このような技術ではコンテンツの印象をアノテーションとして付与した訓練データを要するのが一般的である。そこで我々は、上述のような用途を想定して、画像コンテンツの印象をアノテーションとして付与する作業を支援する研究に取り組んでいる。具体的には、アノテーションとなる複数の形容詞対を SD 法 [3] にもとづいて選択し、各コンテンツの印象を的確に表現する形容詞をアノテーションとして付与する、という作業の支援を目標とする。

ここで、画像の印象を推測する深層学習の構築を例にする。このような深層学習では、印象に関するアノテーションが付与された大量の画像を含む訓練データを用意する必要がある。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 人の作業者が付与とする。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、深層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方法があげられる。しかし、これら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変わりはない。そのため、訓練データのアノテーション分布をユーザが理解することが重要な課題となる。その解決の一手段として、訓練データの分布を可視化する研究 [4, 5] が報告されており、アノテーションの修正支援などにも活用が可能である。

これらの問題に対して我々は、旧来型の機械学習を用いて比較的少ない枚数の印象回答値を学習し、この機械学習モデルを用いて画像を自動分類し、その結果を対話操作によって修正することで、比較的少ない手間で画像のアノテーションを実現する手法を開発している。その一環として本論文では、SD 法による印象評価結果と

ファジィ決定木にもとづく画像の自動分類過程と、そしてその画像分類結果を可視化する手法を提案する。提案手法では、ユーザ操作によって決定木と画像表示を連動させて可視化することで、決定木による自動分類の傾向理解を促進し、対話操作による再分類を支援する。この再分類結果に沿って画像にアノテーションを付与することで、アノテーション付与の作業負荷の低減を目指す。

本手法が有用な状況の例として、同一データに対してアノテーションのやり直しが必要になった状況が考えられる。例えば「明るい⇔暗い」という形容詞対に関するアノテーションを付与する際に、

- 全ての画像に排他的択一で「明るい」「暗い」のいずれかを付与する。
- 全ての画像に排他的択一で「明るい」「暗い」「明るくも暗くもない」のいずれかを付与する。
- 全ての画像に排他的択一で「とても明るい」「やや明るい」「やや暗い」「とても暗い」のいずれかを付与する。
- 特に明るい画像にのみ「明るい」を付与し、特に暗い画像にのみ「暗い」を付与し、それ以外の画像にはアノテーションを付与しない。

といったアノテーションのバリエーションが考えられる。本手法では同一データに対して 1 回だけ印象回答（本研究ではリックカート尺度による 5 段階評価）の収集を実施した上で、ファジィクラスタリングのクラスタ数を修正しながらファジィ決定木を再生成することで、既存の印象回答結果をそのまま活用しつつ、異なるバリエーションのアノテーションを自動生成することが可能になる。

また別の有用な状況の例として、コンテンツが動的に増大する状況への対応が考えられる。例えばオンラインショッピングサービスにおいて新商品が追加された時、深層学習の訓練データにデータオーグメンテーションを適用した時などにも、本手法にて既に構築されたファジィ決定木を活用することで、少ない手間でのデータ追加が可能になる。

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 [6] を用いた SD 法による印象分析 [7] 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。

決定木を効果的に可視化し、他の画面とあわせて提示することで、決定木の可読性の向上が見込まれる。Muhlbacher ら [8] 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。しかしこの手法では、分析画面の提示が 10 種類もあるため、使用方法の説明に 2 時間を要している。それに比べて本手法は 2 種類の提示画面のみで決定木の理解を促進している。

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。このような画像群を閲覧する Image Browser の研究も多数発表されている。一例として Liu ら [4] は、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。本研究においても画像特徴にもとづいて画像群を配置した Image Browser を用いる。Image Browser をアノテーション修正支援に応用した研究もある [5] が、本手法は画像自動分類の過程も同時に可視化している点が独自性がある。

印象評価結果をアノテーションとして付与したコンテンツデータの用途の典型的な例として、コンテンツの印象を推定する機械学習のための訓練データとしての実用が考えられる。印象を推定する機械学習の例として、McCurrie ら [1] は人間の表情がもたらす印象を推定する機械学習を提案している。また Gltrk ら [2] は YouTube のビデオを題材として、音声のみ、動画のみ、テキストのみ、あるいはこれらの組み合わせを訓練データとして印象を推定する機械学習を提案し、訓練データの違いがもたらす性能の違いを比較している。

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 [9] 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 [10] 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 [11] 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。

SD 法を用いた評価は衣服画像以外にも多種報告されている。例として、杉浦ら [12] は都市公園のイメージ分析に、杉原ら [13] は音楽の感性分析に、池田ら [14] は欧文書体の印象分析に SD 法を活用している。これらの題材の中でも女性衣服は、前述 [9, 10, 11] のとおり因子分析の研究が既に充実しており、尺度の選択が容易であると判断した。さらに女性衣服はオープンな画像データベースが充実しており、研究での利用が容易である。これらを理由に本研究では、女性衣服画像を題材に選んだ。

3.2 パイロットテストと尺度決定

3.1 節で選択した因子の各々について、先行研究 [9, 10, 11] 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、回答する項目数が多すぎると、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は回答の項目数を最小限に抑えることが望ましい。一方で、同一の因子に属する形容詞対には、回答が類似すると予想されるものが多数あることから、同一の因子に属する複数の形容詞対について回答を求めることは効率的ではないと考えた。そこで我々は、本研究の印象評価に適用しうる形容詞対の中から、他の形容詞対と回答が類似すると予想される形

容詞対を除外し、各因子について1つの形容詞対のみを実験対象とする、という方針をとることにした。

以上の方針に沿って、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、リッカート尺度による5段階評価の回答を収集した。あらかじめ筆頭著者がパイロットテストの回答を試したところ、10分未満で全て終了したことから、負荷の軽い作業であると判断して、回答者には謝金などは課せず、回答者の作業時間も計測しなかった。

この評価結果をもとにして我々は、表1に示す形容詞対の全てのペア間について距離を算出し、距離行列を生成した。具体的な手順は以下の通りである。

1. i 番目の画像の j 番目の形容詞対について、15名の5段階評価の平均値 a_{ij} を求める。これを各画像の各形容詞対に適用する。
2. 10枚の画像の各々における a_{ij} の値を j 番目の形容詞対における10次元ベクタとする。これを各形容詞対に適用する。
3. j 番目の形容詞対と k 番目の形容詞対のユークリッド距離 $dist_{jk} = \sqrt{\sum_{i=1}^{10} (a_{ij} - a_{ik})^2}$ を求める。これを形容詞対の各ペアに適用する。

続いて各形容詞対について、他の形容詞対とのユークリッド距離の総和を求める。 j 番目の形容詞対についての総和は $sum_j = \sum_{i=1, i \neq j}^{15} dist_{ij}$ となる。この値が大きいくほど、他の尺度との相関がないことを意味する。この値の平均を「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、表1にて*印で示された5種類の形容詞対を選出した。

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を Fashion-MNIST *1 から抽出して印象評価に用いる。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、まず顔領域を検出し、顔領域から上を背景色と同じ色で隠

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
「装飾性」	自然⇔不自然	4.86
	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

す。更に背景削除処理を実施することで、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1にて*印で示した5種類の形容詞対と、3.3節の画像群を用いて、印象評価を実施した。Fashion-MNIST から無作為抽出した1,500枚の画像を用いて、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、1人あたり500枚の画像の印象評価を割り当てることで、回答者の作業量を抑えた。つまり画像1枚あたりの回答人数は14~15人となった。

回答者には画像を1枚ずつランダムに提示した。登校が制限されているコロナ禍での実施だったため、回答者には自宅にて自身の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。回答者には2時間分の所属大学規定の謝金を支払い、2時間以内で500枚の画像の印象評価を実施するように指示した。作業途中の休憩については特に指示しなかった。また、2時間を超過する場合には申告するように指示したが、全員が2時間以内で作業を終えた。このように、3.2節にて示したパイロットテストにより形容詞対を5種類に厳選したことで、回答者の作業時間を2時間以内に抑えることができた。

*1 <https://www.kaggle.com/dqmonn/zalando-store-crawl/>

4 ファジィ決定木による画像分類過程の可視化

4.1 ファジィ決定木

本研究では井上ら [15] の手法に類似した手法でファジィ決定木を生成する。ここで本研究では、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。

1. クラス数と決定木の深さを指定する。
2. ファジィクラスタリングを順次適用する。
3. 各クラスの確信度にもとづいて決定木を構築する。

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「暗いー明るい」を例とする。現時点での実装ではクラス数を 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。この値を 5 次元ベクトルとみなしてファジィクラスタリング [16] を適用し、図 1③に示すように確信度を算出する。そして各画像について、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。以上のクラスタリング処理を形容詞対の各々について実施する。

以上の処理によって形容詞対ごとにファジィクラスタリング結果を導出したら、この結果を組み合わせることで決定木を生成し、生成結果を表示する。決定木の生成



図 1 右の画像が印象項目「暗いー明るい」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。

には井上らの手法 [15] と同様に、期待獲得情報量最大化原理にもとづいた処理手順を採用する。図 2 に決定木の構造のイメージ図を示す。ここでは、1 本の折れ線は 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、各画像が分類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 3(1)) 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 3(2)) が表示される。例えば「明るい」という決定ノードに対して、確信度が 0 に近ければ「明るくない」という確率が高く、1 に近ければ「明るい」という確率が高い。0.5 に近ければ「どちらともいえない」と解釈される。

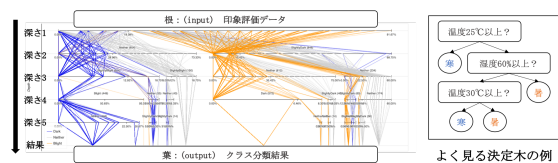


図 2 決定木の構造図。(左) 本手法の決定木。(右) よく見る決定木の例。

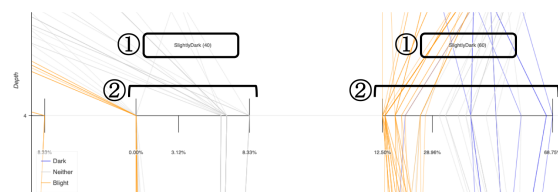


図 3 決定ノードの詳細。(1) 分類条件 (条件判定を受けた画像数)。(2) 確信度の詳細。

分類条件には評価時の 5 段階リッカート尺度を使用している。木構造の葉ノードは各クラスを表しており、クラスごとに固有の色で表示される。この可視化結果から、各画像が木構造に沿って分類される過程を観察できるだけでなく、その確信度を読み取れる。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 0 または 1 に近い傾向にあり、逆に深いノードでクラスが決定される画像は確信度が 0.5 に近い傾向にある。以下、前者を「断言的」と称し、後者を「逡巡的」と称する。図 4 は形容詞対「フォーマル-カジュアル」について生成された決定木の例である。図 4(2)(3) は「フォーマルでもカジュアルでもない」というクラスへの決定ノードであり、図 4(2) はこれへの確信度の範囲 [0.06, 0.42] であり、図 4(3) は確信度の範囲が [0.42, 0.83] である。図 4(3) よりも断言的である図 4(2) を通過した画像は浅いノードでクラスが決定し、図 4(2) よりも逡巡的である図 4(3) を通過した画像は深いノードでクラスが決定しているのがわかる。なお図 4(2) を通過している画像の多くは、確信度が [0.00, 0.32] である図 4(1) を通過している。つまり断言的なノードを 2 つ続けて通過した画像が浅いノードでクラスを決定している傾向にあることがわかる。同様に図 4(4) も、確信度が [0.00, 0.35] である断言的なノードであり、ここを通過した画像は浅いノードでクラスを決定していることがわかる。

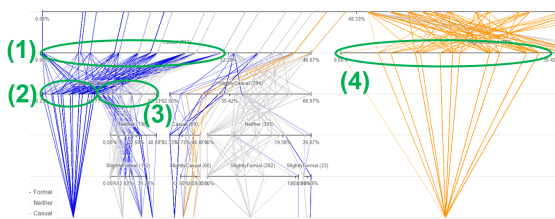


図 4 決定木の深さと確信度の関係。浅いノードでクラスが決定される画像は確信度が 0 または 1 に近い傾向にある。

4.2 分類された画像群の一覧表示

4.1 節で示す手法で生成された木構造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。具体的には、第 1 主成分を 2 次元

空間の水平方向に、第 2 主成分を垂直方向に割り当てることで、各画像の 2 次元空間中での位置を特定する。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 5 に示す。

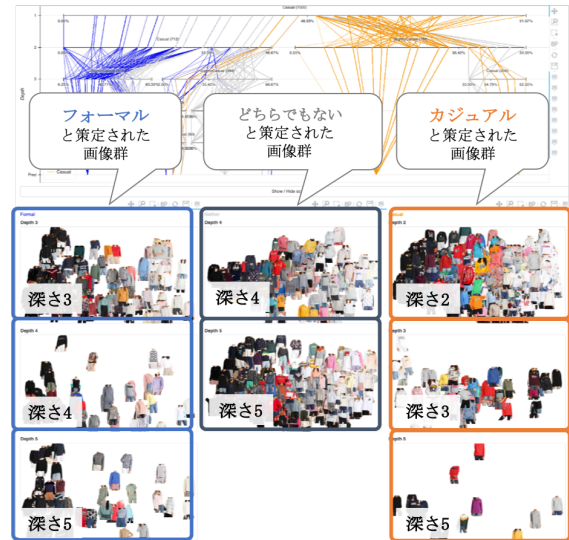


図 5 類似画像群の一覧表示。

4.3 ユーザインターフェース

我々が開発したユーザインターフェースでは、図 5 に示すように、画面の上半分にファジィ決定木を、画像の下半分に画像群の一覧を表示する。本研究では、このユーザインターフェース上でアノテーションの半自動化を支援する。

決定木により分類された画像群の一覧表示結果をもとに、ユーザは対話操作によって分類結果を修正できる。特定の画像を指定して再分類する際には、一覧表示部に表示された画像の中央部をダブルクリックする。図 6 に示すように、点描画の色が変わることによってクラスの再分類結果を確認できる。また、決定木と画像群一覧表示の間にあるボタンによってこの点描画の表示・非表示を切り替えられる。

その他の操作機能として、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。また、ウィンドウ右側のアイコン選択により、平行移動、拡大縮小、再描画、ホバー機能、フィルタリング機能の操作が可能である。ホバー機能を用いて図 7 に示すように、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、各画像の拡大表示を施し、さらにその初期分類結果のクラスを点の色

で表示する。フィルタリング機能を用いて、決定木の決定ノードをドラッグアンドドロップで選択すると、そ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。例えば図8(左)のように、ドラッグアンドドロップ操作によって矩形領域を指定すると、その領域内部にある決定ノードを通過する枝と、その枝に対応する画像のみが、図8(右)に示すように表示される。

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。また、分類修正操作が終了したら「Save」ボタンをクリックすることで、再分類結果をJSON形式で保存できる。

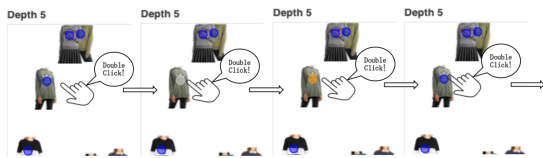


図6 ダブルクリック操作による画像の再分類。

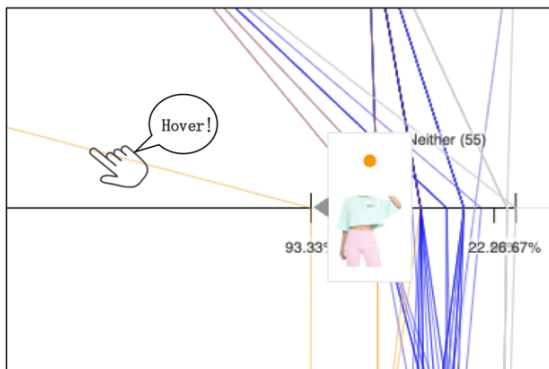


図7 ホバー機能。

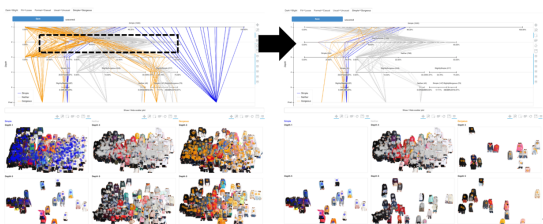


図8 フィルタリング機能。

5 実行例

3章で説明した印象評価データを可視化した例を示す。我々はPythonの可視化ライブラリBokeh 2.4.0*2を拡張することにより本手法を実装した。また4.1節にあるように、ファジィ決定木の分類クラス数は3で、深さは最大5とした。本章では、決定木のみでの可視化結果、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9~13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。

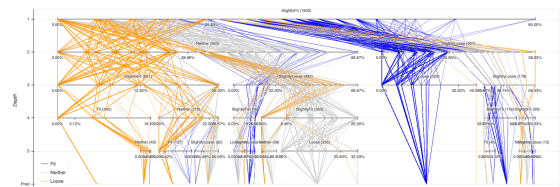


図10 尺度「フィット-ルーズ」のファジィ決定木。

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

*2 <https://bokeh.org/>

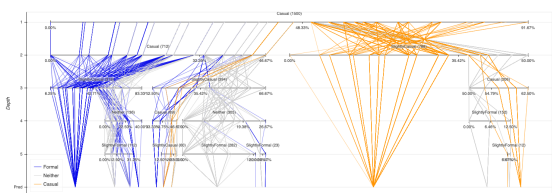


図 11 尺度「フォーマル-カジュアル」のファジィ決定木.

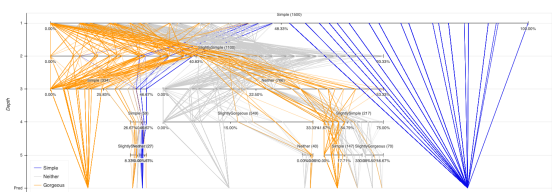


図 12 尺度「シンプル-ゴージャス」のファジィ決定木.

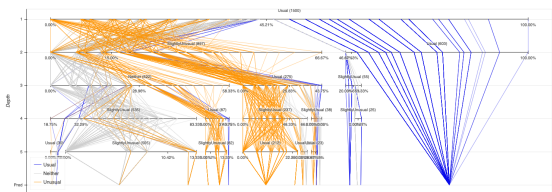


図 13 尺度「日常的な-非日常的な」のファジィ決定木.

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、断言的な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 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。「フォーマル」と「カジュアル」が混合するクラスが生じる結果となったため、フィルタリング機能を用いてそれに該当する画像群を図 15(右) のように表示した。ここで図 15(右) 中の赤矢印で示している衣服は、同じ色・形を有しデザインも似ていた。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写真り方である。衣服の印象のみを評価するために顔削除



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)「明るい」と深さ4, 5で策定された画像群。(右)「どちらでもない」と深さ3, 4, 5で策定された画像群。

や背景削除して印象評価を実施したが、それでもモデルの髪形や体格が評価に影響する可能性が示唆される。

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が逡巡的なものであり、ファジィクラスタリングの最終的なクラス分類の判断が正しく働いていることがわかる。

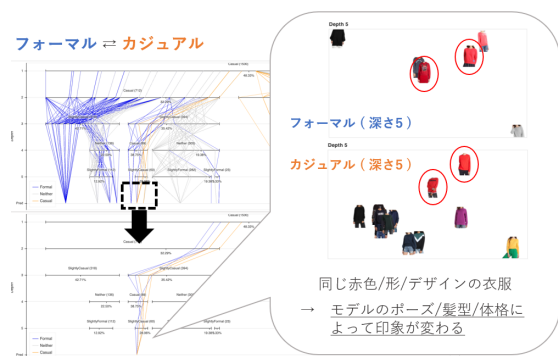


図 15 尺度「フォーマル-カジュアル」の混合分類が見られる箇所をフィルタリングした詳細。(左) 決定木の可視化。(右) フィルタリング後の画像群一覧表示の詳細。

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代の情報科学系の学生9名で、以下の3グループで構成される。本研究では訓練データを題材にしていることから、機械学習に従事する評価者を招聘

した。また、決定木や画像一覧表示の視認能力を評価するために、可視化に従事する評価者を招聘した。

- ML グループ：機械学習の研究に従事する 3 人。
- VIS グループ：可視化の研究に従事する 3 人。
- NONE グループ：機械学習/可視化の研究に従事しない 3 人。

本実験では提案システムを使用してもらい、評価者には可視化結果をもとに人手による画像再分類してもらった。なお本実験では、コロナウイルス感染防止に配慮し、各自の PC での作業画面録画と自由記述によるコメント提出による質的評価を実施した。

評価者にはまず説明資料を読んでもらってから、サーバに構築された本システム^{*3} にアクセスして作業を開始してもらった。ウェブブラウザには Chrome を指定し、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。以上の作業が終わった後に、可視化や操作に関する評価項目に答えてもらった。評価項目は決定木、類似画像一覧、画面操作、アノテーション作業の 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。以上の手順には制限時間は設けていない。

6.1 画面録画からの考察

画面録画からの考察を述べる。どのグループも、1 尺度にかかる作業は平均 10~20 分^{*4}であった。5 尺度の各々の作業時間に加えて、作業開始前の操作練習時間や、作業終了後の評価コメント記入時間を含めて、評価者には所属大学規定の 3 時間分の謝金を支払った。

ML グループの評価者は、機械学習や予測モデルの理解があるからか、決定条件が逡巡的な画像を中心に再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ながら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、決定木を可視化する意義が伺えた。

VIS グループの評価者は、「どちらでもない」に属する画像から再分類を開始し、可視化結果全体を何度も眺めながら再分類する傾向があった。つまり、VIS グループ

の評価者は情報可視化の「まず全体を概観してから局所に注目する」という基本操作に沿って画像を再分類していた。このような操作は、ML グループと NONE グループには見られないものだった。

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。このことから、機械学習の原理や可視化の操作に精通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は一定の訓練を要することがわかった。

6.2 再分類結果による評価結果

Fashion-MNIST から無作為抽出した 1,500 枚の画像を提案システムにより自動分類したのち、作業者によって再分類された画像数を表 2 に示す。なお表 2 にて、ML1~ML3 は機械学習を専攻する作業者 3 人、VIS1~VIS3 は可視化を専攻する作業者 3 人、NONE1~NONE3 はいずれにも該当しない作業者 3 人である。同一画像を複数回にわたって再分類した際には 1 枚として加算されている。形容詞対ごとに画像数の平均を求めると、「日常的な-非日常的な」が最大 (70.89 枚) であり、「フィット-ルーズ」が最小 (39.56 枚) であった。グループごとに画像数の平均を求めると、ML が 57.93 枚、VIS が 30.47 枚、NONE が 77.33 枚となった。いずれにしても、平均 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。つまり、この作業は全ての画像を人手で分類する方法と比較して手間が小さいものであり、決定木による自動分類結果を出発点にした作業が分類作業の負荷軽減に寄与していることがわかる。

6.3 操作・システムに関する評価コメント

評価者による自由記述コメントの内容を紹介する。

決定木に関しては、「理解に時間がかかった」「可読性の向上を期待したい」という声が NONE グループのみからあげられた。その一方で ML グループと VIS グループからは、フィルタリング機能と類似画像一覧を併用しながら決定木を観察することで、画像分類過程を容易に理解できたとの声が多かった。このことから、本研究が想定するユーザ（機械学習または可視化の経験があるユーザ）には類似画像群一覧表示とフィルタリング機能が可読性の向上に有効であることが示された。

類似画像一覧に関しては、尺度「暗い-明るい」以外はうまくいかなかったという意見があった。この解決策として色以外の画像特徴量の次元削減による画像配置を

^{*3} <https://user-test-imp.sakura.ne.jp/lab/evaluation/en/decition.html/>

^{*4} 作業時間平均の詳細は文献 [17] の表 3 に掲載されている。

検討したい。また、特に尺度「ゴージャス-シンプル」では、生地感やデザイン・ロゴといった細部に注目するために、画像拡大操作の反復を要したという意見があった。解決策として、ホバー機能による詳細情報表示の再設計など、新機能の追加や改良の余地があると考えている。

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ったのを除くと、総じて全グループから、豊富な対話操作機能が便利だった、またダブルクリック操作での再分類も便利だったとの声が多くあった。

分類作業の精度に関しては、「自動分類結果を出発点とした再分類操作により精度が向上すると感じた」「決定木や画像一覧の表示をみながら作業したおかげで再分類のための基準にブレが生じなくなり精度の向上につながるので、決定木と画像一覧の表示は有意義であった」などの意見があった。

今回の実験は「機械学習に従事する評価者」「可視化に従事する評価者」という制約を満たす少人数の参加者で実施したことから、数量的な評価を実施しても有意な結果は導きにくいと考え、自由記述コメントによる評価を実施した。より大規模な評価を実施することになった際には、System Usability Scale (SUS) や NASA Task Load Index (NASA-TLX) などのスタンダードな基準にもとづく数量的な評価を実施したい。

6.4 ベースライン手法との比較実験

提案手法が採用したファジィ決定木による画像分類結果の有効性を検証するために、k-means 法をそのまま適用した実装（以下ベースライン手法と称する）による分類結果との比較実験を実施した。前節までの実験参加者とは異なる 1 名による実験結果を表 3 に示す。ここで「枚数」は再分類した画像の枚数であり、「所要時間」は作業の開始から終了までの時間である。

この結果から、提案手法のほうが再分類の枚数がやや少ない傾向にあり、ベースライン手法と比べても作業量の低減に貢献していることがわかる。それに対して、所要時間は 5 尺度のいずれにおいても提案手法のほうが大幅に小さくなっている。その理由として、ファジィ決定木による分類結果は確信度をともなった結果であることから、例えば「深い階層で分類された画像群だけを注意深く眺める」「浅い階層で分類された画像群は短時間で眺める」といった考えに立って作業を高速化できること

が考えられる。

7 まとめと今後の課題

本論文では、複数の作業員間で個人差の出る印象評価に対して、作業員群による過去の印象回答値を参照してファジィ決定木を生成し、それをを用いて画像を自動分類することにより、印象のアノテーション作業を半自動化するとともに、その過程を可視化することでアノテーションの修正を支援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が効果的であることがわかった。

今後の課題として、まずユーザインタフェースの拡張について検討したい。決定木の可視化においては、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。また、ラベリング作業を支援するユーザインタフェースに関する先行研究 [18, 19] を参考にして、参加者のユーザビリティを高めたい。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携を深め、両者の関係を理解しやすいデザイン設計を考えたい。

画像群の一覧表示にも再考の余地がある。画像の RGB 値に単純に PCA を適用して次元削減した場合に、明度や色相に応じたグラデーションを描くように画像群が配置される。6.3 節にて紹介したコメントからもわかるように、この配置は「暗い-明るい」といった尺度には効果的であるが、他の尺度には必ずしも効果的には限らない。そこで画像の RGB 値を直接次元削減する以外に、各種の画像特徴値を算出してそれを次元削減するという形での画面配置も検討したい。さらに同時に、PCA 以外の次元削減手法として、t-Stochastic Neighbor Embedding (tSNE) や Manifold Approximation and Projection (UMAP) などの適用も検討したい。

本研究では 5 段階評価によって画像群への印象を回答させてきたが、このような絶対評価は参加者にとって必ずしも簡単ではない。そこで相対評価結果を用いたクラ

スタリング手法 [20, 21] の適用なども検討したい。

今回適用したモデルや画像群に限らず、他のモデルや画像群を用いての実験にも取り組みたい。具体的には、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、決定木以外のモデルも採用して比較対象実験を実施したい。加えて、女性の衣服画像だけでなく、異なる画像群での印象評価も実施したい。

さらに、本研究に関する追加評価実験として、評価者たちが画像を再分類した結果を第三者に提示し、適切なアノテーションが効率的に実現されているかを評価したい。

以上の機能拡張や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での実証を進め、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(課題番号 20K11917) の助成に関するものです。ユーザー実験にご協力していただいた皆様に感謝いたします。

参考文献

- [1] M. McCurrie, F. Beletti, L. Parzianello, A. Westendorp, S. Anthony, W. Scheirer, “Predicting First Impressions with Deep Learning”, 12th IEEE International Conference on Automatic Face & Gesture Recognition, 2017.
- [2] Y. Gltrk, U. Gl, X. Bar, H. J. Escalante, I. Guyon, S. Escalera, M. A. J. van Gerven, “Multimodal First Impression Analysis with Deep Residual Networks”, IEEE Transactions on Affective Computing, 9, 316-329, 2018.
- [3] C. E. Osgood, “The Nature and Measurement of Meaning”, Psychological bulletin, 197, 1952.
- [4] X. Liu, Y. Hu, S. North, H.-W. Shen, “Correlated-Multiples: Spatially Coherent Small Multiples With Constrained Multi-Dimensional Scaling”, Computer Graphics Forum, 7-18, 2018.
- [5] S. Xiang, et al., Interactive Correction of Mislabeled Training Data, IEEE Conference on Visual Analytics Science and Technology (VAST), 57-68, 2019.
- [6] J. C. Bezdek, R. Ehrlich, W. Full, “FCM: The Fuzzy C-means Clustering Algorithm”, Computers and geosciences, 191-203, 1984.
- [7] T. Matsuo, et al., “Sensitivity Information Analysis of Running Shoes Using Fuzzy Decision Tree and Visualization of Analytical Results”, SCIS and ISIS 2010. Japan Society for Fuzzy Theory and Intelligent Informatics, 2010.
- [8] T. Muhlbacher, L. Linhardt, T. Moller, H. Piringer, “Treepod: Sensitivity-aware Selection of Pareto-optimal Decision Trees”, IEEE transactions on visualization and computer graphics, 174-183, 2017.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 杉浦芳夫, 加藤近之, “SD 法による都市公園のイメージ分析”, 総合都市研究 46, 53-79, 1992.
- [13] 杉原太郎, 森本一成, 黒川隆夫, “SD 法を通してみた音楽に対する感性の基本特性”, 映像情報メディア学会技術報告, 25.48, 2001.
- [14] 池田マイケル, “SD 法を用いた本文用欧文書体の印象分析: Gill Sans, Futura, Frutiger, Caslon, Didot, Palatino を対象として”, デザイン学研究, 54(5), 11-18, 2008.
- [15] 井上博行, 王冠, 西村美津恵, “決定木を用いた花束の感性ルール抽出と花束提示システムへの応用”, 知能と情報(日本知能情報ファジィ学会誌), 24(5), 1023-1034, 2012.
- [16] J. C. Bezdek, R. Ehrlich, W. Full, “FCM: The Fuzzy c-means Clustering Algorithm”, Computers & Geosciences, 10(2-3), 191-203, 1984.
- [17] 飯島緋理, 伊藤貴之, “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”, 第 14 回データ工学と情報マネジメントに関するフォーラム (DEIM), F33-2, 2022.
- [18] C.-M. Chang, C.-H. Lee, T. Igarashi, “Spatial Labeling: Leveraging Spatial Layout for Improving Label Quality in Non-Expert Image Annotation”, 2021 CHI Conference on Human Factors in Computer Systems, 306, 2021.
- [19] C.-M. Chang, Y. He, X. Yang, H. Xie, T. Igarashi, “DualLabel: Secondary Labels for Challenging Image Annotation”, Graphics Interface (GI2022), 2022.
- [20] Z. Cui, N. Charoenphakdee, I. Sato, M. Sugiyama, “Classification from Triplet Comparison Data”, Neural Computation, 32(3), 659681, 2020.
- [21] Z. Cui, I. Sato, “Active Classification With Uncertainty Comparison Queries”, Neural Computation, 34(3), 781803, 2022.

飯島緋理



2020 年お茶の水女子大学理学部情報科学科卒業。2022 年同大学大学院人間文化創成科学研究科理学専攻博士前期課程修了。情報可視化に関する研究に従事。

伊藤貴之



1992 年早稲田大学大学院理工学研究科電気工学専攻修士課程修了，日本アイ・ビー・エム (株) 東京基礎研究所 研究員。1997 年博士 (工学)。2005 年お茶の水女子大学 理学部情報科学科助教授。2011 年同大学教授。2017 年 まで同大学シミュレーション科学教育研究センター長兼 任。2019 年から同大学文理融合 AI・データサイエンス センター長兼任。2014 年より 2016 年まで芸術科学会会 長。情報可視化，マルチメディア，インタラクション， コンピュータグラフィックスなどの研究に従事。

表2 分類済み画像から各作業者が再分類した画像数.

	ML1	ML2	ML3	VIS1	VIS2	VIS3	NONE1	NONE2	NONE3
「色彩」 暗い-明るい	50	47	71	19	27	34	67	56	21
「立体的」 フィット-ルーズ	35	34	14	22	20	27	107	90	7
「正統性」 フォーマル-カジュアル	77	54	39	38	12	30	187	71	22
「穏健性」 日常的な-非日常的な	33	131	79	77	6	30	203	71	8
「装飾性」 シンプル-ゴージャス	41	80	84	65	14	36	136	93	21

表3 ベースライン手法との比較実験.

	枚数 (提案手法)	枚数 (ベースライン)	所要時間 (提案手法)	所要時間 (ベースライン)
「色彩」 暗い-明るい	32	40	7分51秒	9分13秒
「立体的」 フィット-ルーズ	24	21	8分37秒	11分08秒
「正統性」 フォーマル-カジュアル	51	60	10分10秒	13分20秒
「穏健性」 日常的な-非日常的な	56	68	10分43秒	13分09秒
「装飾性」 シンプル-ゴージャス	29	35	9分20秒	12分51秒