類似した感性をもつ人物集合抽出に基づく画像の印象推定の検討

小嶋　万理[†]　駒水　孝裕[††,†]　川西　康友[†††,†]　道満　恵介[††††,†]

井手　一郎[†]　中澤　満[†††††]　Yeongnam　CHAE[†††††]　Björn　STENGER[†††††]

† 名古屋大学大学院情報学研究科　〒464–8601 名古屋市千種区不老町
†† 名古屋大学数理・データ科学教育研究センター　〒464–8601 名古屋市千種区不老町
††† 理化学研究所ガーディアンロボットプロジェクト　〒619–0288 京都府相楽郡精華町光台 2–2–2
†††† 中京大学工学部　〒470–0393 愛知県豊田市貝津町床立 101
††††† 楽天グループ株式会社　楽天技術研究所　〒158–0094 東京都世田谷区玉川 1-14-1 楽天クリムゾンハウス

あらまし　本報告では，人が画像から受ける印象を推定する手法について検討する．画像から受ける印象は人によって異なるが，その傾向について，一定のステレオタイプが存在すると考えられる．本研究では，人の属性に着目し，印象の受け方に大きく関わる属性をデータ駆動的に選択することで，画像に対する印象推定の性能を向上させる手法を検討する．しかし，性別や年代など人物属性は多様であり，それらの組合せも膨大であるため，十分な量のデータは容易に収集できない．この問題に対して，画像クラスタリング及び人物属性クラスタリングによって類似した感性をもつ人物集合を抽出した上で，推定モデルを用いることで，限られたデータでの印象推定を実現する．4,000 枚のカーペットの画像に対して，クラウドソーシングにより 24 個の印象語をアノテーションしたデータセットを用いた実験の結果，人物属性組合せの集合抽出が印象推定の性能向上に有効であることを確認した．
キーワード　印象推定，クラスタリング，人物属性

Towards Estimating Impressions on Images by Grouping Personalities Sharing Similar Sensibilities

Banri KOJIMA[†], Takahiro KOMAMIZU[††,†], Yasutomo KAWANISHI[†††,†], Keisuke DOMAN[††††,†],

Ichiro IDE[†], Mitsuru NAKAZAWA[†††††], Yeongnam CHAE[†††††], and Björn STENGER[†††††]

† Graduate School of Informatics, Nagoya University　Furo-cho, Chikusa, Nagoya, Aichi, 464–8601 Japan
†† Mathematical and Data Science Center, Nagoya University　Furo-cho, Chikusa, Nagoya, Aichi, 464–8601 Japan
††† Guardian Robot Project, RIKEN 2-2-2 Hikaridai, Seika-cho, Souraku-Gun, Kyoto, 619–0288 Japan
†††† School of Engineering, Chukyo University　101 Tokodachi, Kaizu-cho, Toyota, Aichi, 470–0393 Japan
††††† Rakuten Institute of Technology, Rakuten Group, Inc.　Rakuten Crimson House, 1-14-1 Tamagawa, Setagaya-ku, Tokyo

Abstract　In this report, we study a method for estimating impressions that people receive from images. Impressions received from images differ from person to person, but there seems to be a certain stereotype regarding the tendency of these impressions. In this study, we focus on the attributes of people and investigate a method to improve the performance of impression estimation for images grouping attributes that are significantly related to the way impressions are received in a data-driven manner. However, it is not easy to collect a sufficient amount of data because attributes of people have large variety and their combinations are enormous. To solve this problem, a group of people with similar sensitivities are extracted by image clustering and person attribute clustering on the collected data, and then use an estimation model to realize impression estimation with a limited amount of data. Experimental results on a dataset of 4,000 carpet images annotated by 24 impression words indicate that attribute clustering is effective in improving the estimation accuracy.
Key words　Impression estimation, Clustering, Personal attributes

図 1: 人物属性の組合せの集合抽出の概要

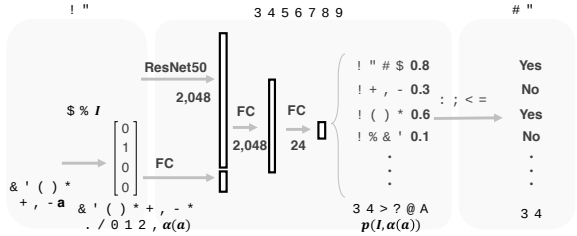

図 2: 印象推定のモデル概要

図 3: アノテーションインタフェース

ルとする．

$$\mathbf{V}(\mathbf{a}) = [\mathbf{V}(I_1, \mathbf{a}), \mathbf{V}(I_2, \mathbf{a}), \cdots, \mathbf{V}(I_{k_i}, \mathbf{a})] \tag{2}$$

ここで I_{k_i} は画像クラスタの数である。最後に，これに基づいて，人物属性の組合せをクラスタリングすることで，類似した感性をもつ人物属性の組合せの集合を抽出する．図 1 の例では，行方向に人物属性の組合せ，列方向に各画像クラスタに対する印象語スコアを画像クラスタ数分並べてある．なお，クラスタリング時に次元の呪いを避けるために，主成分分析により，累積寄与率 90% まで特徴ベクトルを次元削減する．

3.2 印象推定

印象推定部分は先行研究 [4] の推定モデルを参考にした．この推定モデルの概要を図 2 に示す．入力は画像 I と人物属性の組合せ $\mathbf{a}$ である．まず、画像 I から画像特徴を抽出する．以下の実験では ImageNet [5] により事前に学習された ResNet50 を用いた．次に，人物属性の組合せ $\mathbf{a}$ は，それが含まれる全ての人物属性の組み合わせの集合を one-hot ベクトル $\alpha(\mathbf{a})$ として表現した．以上により，2,048 次元の画像特徴と，人物属性の組合せの集合数 k_A 次元のベクトルが得られる．$\alpha(\mathbf{a})$ に対する入力画像 I の印象語スコアベクトル $\mathbf{p}(I, \alpha(\mathbf{a}))$ は，以下のように計算できる．

$$\mathbf{p}(I, \alpha(\mathbf{a})) = f(I, \alpha(\mathbf{a}); \theta) = g(h(e(I), \alpha(\mathbf{a})); \theta) \tag{3}$$

$$h(e(I), \alpha(\mathbf{a})) = \sigma(\mathbf{V}(\mathbf{a}) \cdot \alpha(\mathbf{a})) + e(I). \tag{4}$$

ここで，θ はモデルのパラメータセット，$\mathbf{V}(\mathbf{a})$ は人物属性の組合せの集合の特徴ベクトル，σ は活性化関数を，e は特徴抽出器，g は回帰器をそれぞれ表す関数である．ここで，h は融合層を表す関数である．融合層では，人物属性の組合せの集合を表すベクトルを画像特徴ベクトルと同じ次元のベクトルに埋め込む．この融合層に続いて，2 つの全結合層によって印象語スコアを推定する．モデルの学習には画像特徴と人物属性の組合せの集合の特徴ベクトル $\mathbf{V}(\mathbf{a})$ に加え，その画像に対してアノテーションした人物属性組み合わせの集合のアノテータが受けた印象の平均値 $\mathbf{y}(I, \alpha)$ の 3 つが与えられる．ここで $\mathbf{y}(I, \alpha)$ の各要素は各印象語スコアを表している．$\mathbf{y}(I, \alpha)$ と，損失関数 L は以下で計算される．

$$\mathbf{y}(I, \alpha) = \frac{\sum_{s \in S(I,\alpha)} \mathbf{v}(s)}{\|S(I, \alpha)\|} \tag{5}$$

$$L = \|\mathbf{p}(I, \alpha(\mathbf{a})) - \mathbf{y}(I, \alpha(\mathbf{a}))\|^2 \tag{6}$$

ここで，$\mathbf{v}(s)$ はアノテータ s の印象ベクトル，$S(I, A_j)$ は画像 I のアノテータの中で人物属性の組合せの集合 A_j に含まれる人物属性組合せをもつアノテータの集合である．事前学習した ResNet50 のパラメータを固定したまま，パラメータ θ と特徴ベクトル $\mathbf{V}(\mathbf{a})$ を最適化する．出力は 24 の各印象語に対する印象語スコアである．

表 1: 各人物属性の詳細

人物属性	詳細
性別	男性，女性
年代	10・20，30，40，50，60
会員ランク	Regular・Silver，Gold，Platinum，Diamond
居住地	北海道・東北，関東，中部，近畿，中国，四国，九州，その他
年収	～400 万円，400～600 万円，600～800 万円，800～1,000 万円，1,000～1,200 万円，1,200 万円～
婚姻状況	独身，既婚
子供の有無	なし，あり
車の保有状況	保有していない，保有している
職業	研究・開発・技術者，専業主婦（主夫），営業・販売，自営業，専門職（医師・看護師・弁護士など），学生，公務員・団体職員，財務・経理，事務職，パート・アルバイト，教職員，管理職，企画・マーケティング，総務・人事，会社経営・役員，契約社員・派遣社員，広報・広告・デザイン，自由業（フリーランス），働いていない，その他

4. データセット

本手法の学習と評価のために，電子商取引（EC）サイトから収集した商品画像とクラウドソーシングサービスによって収集した画像に対する印象の回答のアノテーションデータの組からなるデータセットを構築した．

4.1 画像の収集方法

画像は EC サイト上でカーペットのカテゴリで販売されている 1 万件以上の商品画像をクローリングした．この際に画像中にテキストが含まれているとアノテーション時に影響を与える可能性があるため，テキスト入りの画像や，重複している画像を目視で削除した．また，人物が映り込んでいる商品画像については肖像権を考慮し，これらも目視で削除した．最終的にカーペットのカテゴリの画像 4,000 枚を選択した．

4.2 印象語の定義

使用した印象語は先行研究 [4] と同じものである．先行研究では，まず EC サイト上で使用される検索キーワードを収集した．次に検索キーワードを形態素解析し，「ヨーロッパスタイル」などの「スタイル」を含む形容詞や単語を抽出した．そして，検索キーワードの使用頻度を考慮し，合計 24 の印象語を選択した．具体的には，おしゃれな，可愛い，小さい，大きい，安い，高級な，和風，シンプルな，長い，短い，北欧風，エレガントな，かっこいい，欧風，透明な，カフェ風，軽い，重い，涼しい，暖かい，アジアン，モダン，アンティーク，ナチュラルである．

4.3 画像に対する印象のアノテーション

クラウドソーシングサービスである楽天超ミニバイト [6] を利用してアノテーションした．アノテータには図 3 のように画像 1 枚と印象語 1 語，それに対して「Yes」と，「No」のボタンが表示される．アノテータは画像に対してその印象語の印象を受ければ「Yes」，受けなければ「No」と回答する．アノテータはこの質問を各画像に対して 24 の印象語について連続して回答する．その際，印象語は無作為な順番で表示される．アノテータの属性として，性別，年代，EC サイトの会員ランク，居住地，婚姻状況，子供の有無，職業，年収，車の保有状況を収集した．各人物属性の詳細は表 1 に示す．画像 1 枚あたり 92 人のアノテータが 24 語の印象のアノテーションした．それにより合計 368,000 件のアノテーションデータを得た．そこからアノテータの属性の中で未回答などが含まれるデータは削除し，最終的に 273,163 件のアノテーションデータを用いて実験した．

表 2: 印象語に対する各人物属性の平均及び累積特徴量重要度

人物属性	平均特徴量重要度	累積特徴量重要度
(画像クラスタ番号)	0.8191	0.8191
職業	0.0650	0.8841
年代	0.0321	0.9162
性別	0.0290	0.9452
居住地	0.0184	0.9636
年収	0.0154	0.9790
会員ランク	0.0080	0.9870
婚姻状況	0.0053	0.9923
子供の有無	0.0040	0.9963
車の保有状況	0.0037	1.0000

5. 実　　験

5.1 利用する人物属性

人物属性全てを使用すると組合せの数が指数関数的に多くなってしまうため，印象に関連する人物属性を選択する．この選択のために LightGBM [7] を用いた印象推定により各属性の特徴量重要度を調べた．9 つの人物属性に加えて，画像クラスタリングの結果も印象推定に寄与しているかを調査するために画像クラスタ番号も加えた．このときの画像クラスタリングは k-means 法で $k_I = 100$ とした．その理由はこの後に述べるクラスタ数の決定において，適切なクラスタ数 k_I が分からないため，大きな値を設定した．特徴量重要度は各印象語の推定における各人物属性の重要度を計算し，最後に各人物属性の平均値を計算した．その結果を集計したものが表 2 である．

画像クラスタ番号が最も特徴量重要度が高いことから，印象推定において画像クラスタリングが有用であることがわかった．人物属性においては累積特徴量重要度 98 %で平均値が 1 桁下

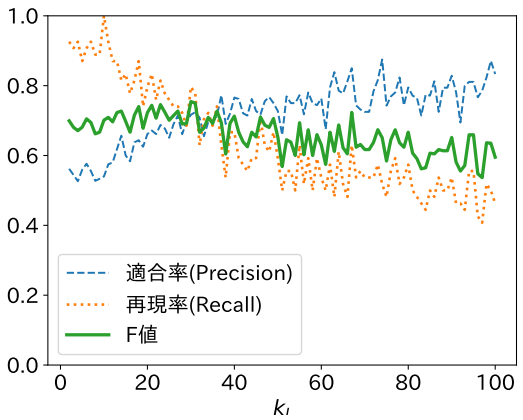

図 4: クラスタ数 k_I の変化による F 値

がった．この結果から印象推定における人物属性は性別，年代，居住地，職業，年収の 5 つを用いることにした．

5.2 人物属性の組合せの集合抽出におけるクラスタリング

5.2.1 画像クラスタリング

まず画像クラスタリングにおけるクラスタリング手法として k-means 法を採用した．このとき，最適なクラスタ数 k_I を決める必要がある．ここでいう最適なクラスタ数 k_I というのは人が実際に似ていると感じる画像が同じクラスタに属していることが多い k_I のことを指す．クラスタ数 k_I を決めるにあたり，似ている画像がクラスタリング時に同じクラスタに属するかを確認するために画像対を作成した．作成手順は以下の通りである．

（1） 画像 4,000 枚を小さな k_I でクラスタリングをする．本実験では $k_I = 10$ でクラスタリングした．

（2） 各クラスタから画像を 20 枚ずつ選択し，画像対を 10 対作成する．これを全てのクラスタに対して行う．

（3） 画像数が多いため，似ていない画像対の割合が多い，そこで，似ている画像対を増やすために似ていない画像対をいくつか削除し，似ている画像対を追加する．最終的に似ている／似ていない対が 50 対ずつとなるようにする．

本研究を知らない第三者に，作成した 100 対に対して似ている／似ていないのラベル付けをさせた．似ているとラベル付けされた画像対は 54 対，似ていないとラベル付けされた画像対は 46 対となった．画像 4,000 枚を様々な k_I でクラスタリングしたときに，ラベル付け結果との一致度合を確認した．クラスタ数 k_I を決める基準として F 値を算出した．その結果を図 4 に示す．F 値が最も高い 3 つの k_I をここでは画像クラスタリングにおける適切なクラスタ数の候補とした．更に，似ているとラベル付けされた画像対が同じクラスタに属する割合，または似ていないとラベル付けした画像対に異なるクラスタに属する割合のどちらかが特に高く，F 値も高い k_I もそれぞれ 1 つずつ選択し，k_I の候補に追加した．これにより 4,000 枚の画像クラスタリングにおいて適切であると予想されるクラスタ数 k_I の候補は{22, 23, 30, 31, 73}となった．

表 3: 人物属性の組合せの集合数による印象推定精度

比較設定	人物属性の組合せの集合数	精度（%）↑
A	1	72.4
B	4,704	68.1
C	6	71.2

5.2.2 属性組合せクラスタリング

属性組合せのクラスタリングには k-means 法を用いた．このときのクラスタ数 k_A の候補は{2, 3, 4, 5, 10, 20, 30, 40, 50}とした．このクラスタ数を選んだ理由としては，人物属性の組合せの集合の適切な数を決める手法をまだ検討中であるためである．そのため暫定的に決めたクラスタ数 k_A による結果に基づいて，クラスタリングをし，この後の印象推定により精度を確認し，適切なクラスタ数の検討をする予定である．

5.3 印象推定

5.3.1 評価方法

印象推定の実験では 4,000 枚の画像を無作為に 2,880 枚，320 枚，800 枚に分割し，それぞれを訓練セット，検証セット，テストセットとした．訓練セットのみ画像に対して 20 倍のデータ拡張をした．画像に対してした処理は，無作為に左右反転，±5 度の範囲で回転，0.9 から 1.1 倍の範囲で拡大縮小である．データ拡張によって生成された画像に対応するアノテーションは拡張前の元画像と同じとした．評価指標としては出力された各印象語のスコアをしきい値で 2 値化したときの「Yes」または「No」と，テストセットのアノテーションとの一致割合を精度として採用した．この 2 値化の際のしきい値は，訓練セットと検証セットのアノテーション全体の平均値に設定した．この評価では，5 回の交差検証をした．人物属性の組合せの集合抽出の 2 回のクラスタリング時における様々なクラスタ数 k_I，k_A の組合せによって得られた属性組合せを用いて印象推定をし，精度を確認した．

5.3.2 比較設定

人物属性の組合せの集合抽出の有効性を確認するために，集合を作る場合と作らない場合及び先行研究 [4] と同じ設定の 3 通りの条件で実験した．

- 比較設定 A：人物属性の組合せの集合を 1 つのみで印象推定した場合である．これはすべての人物属性を 1 つの人物属性の組合せの集合であるとして印象推定をすることに相当する．
- 比較設定 B：人物属性の組合せの集合を抽出せずに印象推定した場合である．これは各人物属性を 1 つの人物属性の組合せの集合であるとして印象推定をすることに相当する．
- 比較設定 C：ステレオタイプ的に人物属性の組合せの集合を決めた先行研究 [4] と同じ人物属性の組合せを用いて印象推定した場合である．先行研究の人物属性の組合せは使用属性が性別（男性／女性）と年代（ヤング／ミドル／シニア）のみで組合せ数は 6 である．

5.3.3 結　　果

比較設定 A，B，C を用いたときの印象推定の精度を表 3 に示す．2 段階のクラスタリングにおけるいくつかのクラスタ数

表 4: 画像クラスタ数及び人物属性の組合せの集合数を変えて人物属性の組合せの集合を抽出した結果を用いた印象推定精度

精度（%）↑		人物属性の組合せの集合数（k_A）								
		2	3	4	5	10	20	30	40	50
画像クラスタ数（k_I）	22	70.3	69.6	69.4	68.6	68.4	67.9	67.2	67.5	67.6
	23	70.1	69.5	69.5	68.9	67.7	67.2	67.6	67.6	67.3
	30	71.3	69.9	70.0	68.7	67.7	67.5	67.4	67.6	67.9
	31	**71.8**	69.9	69.8	69.2	67.7	67.4	67.8	67.5	67.5
	73	71.6	71.4	70.7	70.9	67.8	68.1	68.3	67.9	68.2

k_I，k_A の組合せによる人物属性の組合せの集合抽出を用いて印象推定した結果を表 4 に示す．

結果としては人物属性の組合せの集合数が多くなるにつれて精度が低くなった．提案手法においては画像クラスタ数を大きくすると精度が高くなることが確認できた．

まず比較設定の結果について考察する．すべての人物属性の組合せを別々に扱った比較設定 B の場合に精度が最も低くなった．これは集合 1 つあたり 1 つの人物属性の組合せしか含まれていないため，データ不足が理由だと考えられる．先行研究 [4] と同様の 6 つの属性集合（比較設定 C）による推定結果は比較設定 B よりも +3.1% となった．このことから人物属性の組合せの 1 集合あたりのデータ数を増やすことが推定において有効であると考えられる．人物属性の組合せの集合を 1 つとみなして推定した比較設定 A の精度が最も高いという結果になった．これは比較設定の中で最もデータ数が多いためであると考えられる．

次に提案手法による結果について考察する．人物属性の組合せの集合抽出を用いた印象推定の結果を見ると人物属性の組合せの集合数 k_A が大きくなるにつれて画像クラスタ数にかかわらず精度が低くなっていくことがわかった．これは 1 つの人物属性の組合せの集合あたりのデータ数が少なくなっていくためと考えられる．また，人物属性の組合せの集合を過剰に細かく抽出してしまっていることも考えられる．画像クラスタ数の変化における精度を見ると，画像クラスタ数を大きくするにつれて精度が高くなっていくことがわかる．ここで選んだ画像クラスタ数の中では $k_I = 31$ が最も精度が高かった．しかし、表 3 の比較設定 A よりも精度が下回ってしまった。本実験は k_I の最大値を 100 と設定したが，より大きいクラスタ数が適切な可能性があるので，今後は更に大きなクラスタ数で実験をしてより精度の高い k_I 探す必要がある．

6. おわりに

本報告では，人物属性に着目して，人が画像から受ける印象を推定する手法について検討した．従来の手法では類似した感性をもつであろう人物属性の組合せの集合を暫定的に決めていた．そこで提案手法ではクラウドソーシングにより収集したデータに対して 2 段階のクラスタリングを適用し，人物属性の組合せの集合をデータ駆動的に抽出した．そして画像特徴量と人物属性の組合せの集合を表す one-hot ベクトルを入力とする推定モデルを用いて印象語スコアを出力した．結果としては人物属性の組合せの集合を用いて印象推定をすることで精度が向上することを確認した．今後は，人物属性の組合せの集合の抽出時のクラスタリングにおけるよりよいパラメータ等の検討をする予定である．

文　献

[1] Hidemichi Suzuki, Atsuhiro Yamada, Kensuke Tobitani, Sho Hashimoto, and Noriko Nagata, “An automatic modeling method of kansei evaluation from product data using a CNN model expressing the relationship between impressions and physical features,” in Proceedings of the 21st International Conference on Human-Computer Interaction, pp. 86–94, July 2019.

[2] Linfu Li, Hiroshi Kubo, and Takashi Uozumi, “KANSEI search system of individual preference for food selection toward health support,” KANSEI Engineering International, vol. 6, no. 1, pp. 31–38, June 2006.

[3] Natsuki Sunda, Kensuke Tobitani, Iori Tani, Yusuke Tani, Noriko Nagata, and Nobufumi Morita, “Impression estimation model for clothing patterns using neural style features,” in Proceedings of the 22nd International Conference on Human-Computer Interaction, pp. 689–697, July 2020.

[4] 中本麻友, 川西康友, 出口大輔, 井手一郎, 村瀬洋, 中澤満, Chae Yeongnam, Stenger Björn, “顧客の属性を考慮した商品画像の印象推定法の検討,” 2021 年電子情報通信学会総合大会, D-12-5, March 2021.

[5] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei, “ImageNet large scale visual recognition challenge,” International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, December 2015.

[6] 楽天グループ(株),“楽天超ミニバイト,”https://minijob.rakuten.co.jp/biz/（2022/8/22 アクセス）

[7] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu, “LightGBM: A highly efficient gradient boosting decision tree,” Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 3146–3154, December 2017.