

マルチタスク CNN によるテクスチャ柄の階層的感性推定モデル

山下 ののみ[†] 飛谷 謙介^{††} 都賀 美有紀[†] 長田 典子[†]

[†]関西学院大学大学院理工学研究科 〒669-1330 兵庫県三田市学園上ヶ原 1

^{††}長崎県立大学情報システム学部 〒851-2130 長崎県西彼杵郡長与町まなび野 1-1

E-mail: [†]{hqr57902, toga.m, nagata}@kwansei.ac.jp, ^{††}tobitani@sun.ac.jp

あらまし 本研究は衣服のテクスチャ柄に対する感性的質感に着目し、感性の階層構造に基づいて印象と感情を推定するモデルの構築を目的とする。柄画像を対象とした主観評価実験で印象と感情を定量化し、マルチタスク学習を用いて感性推定モデルを構築する。マルチタスク学習を用いて多様な印象を同時に学習し感情層に繋げることで、感性の階層構造における印象と感情の関係性の表現を可能とする。シングルタスクモデルとの精度比較により、マルチタスク学習の有用性が確認された。

キーワード ファッション, 推薦システム, 画像認識, マルチタスク学習

Hierarchical textile pattern aesthetic prediction using multitask CNN

Nonomi YAMASHITA[†] Kensuke TOBITANI^{††} Miyuki TOGA[†] and Noriko NAGATA[†]

[†]Graduate School of Science and Technology, Kwansei Gakuin University 1 Gakuenuegahara, Sanda-shi, Hyogo, 669-1330 Japan

^{††}Faculty of Information Systems, University of Nagasaki 1-1 Manabino, Nagayo-cho, Nishisonogi-gun, Nagasaki, 851-2130 Japan

E-mail: [†]{hqr57902, toga.m, nagata}@kwansei.ac.jp, ^{††}tobitani@sun.ac.jp

Abstract The aim of this study is to construct a model for estimating impressions and emotions based on the hierarchical structure of emotions, focusing on the emotional quality of textured patterns on clothing. Impressions and emotions are quantified in subjective evaluation experiments on pattern images, and a sensitivity estimation model is constructed using multitask learning. By learning various impressions simultaneously using multitask learning and connecting them to the emotion layer, it is possible to express the relationship between impressions and emotions in the hierarchical structure of sensitivity. Comparison of the accuracy with single-task models confirms the usefulness of multitask learning.

Keywords Fashion, recommendation systems, image recognition, multi-task learning

1. はじめに

感性的質感認知は価値判断や意思決定において重要な機能となる [1]. 機械学習の分野では、審美性を機械学習モデルで表現する研究が盛んに行われているが、感性的質感認知における印象から感情、価値へのプロセスを扱った研究は少数である。プロダクトデザイン分野では、ユーザーニーズの多様化に伴いプロダクトのカスタマイズが注目されているが、専門的な知識や経験を要し労力や時間の面でも課題となっている。そのため、感性的質感を分析しユーザーが簡単にイメージに合ったプロダクトを入手できれば、ユーザー満足や Well-being の向上だけでなく、大量生産の無駄削減や持続可能な社会の実現にも寄与すると考える。また、多くの EC サイトではユーザーの意思決定を補助するため推薦システムが使用されている。しかし、これらの手法には購入履歴が必要といった課題や似たような製

品ばかりが提案されるといった課題がある。そのため、少量のデータからユーザーの好みやニーズを推定する技術の開発が求められている。

そこで本研究では、本研究では、衣服の柄を対象に感性的質感と物理特徴との関係性をモデル化し、感性的質感の構造に基づいた柄の印象推定モデルの構築を目的とする。

2. 先行研究

複数の評価を統合して判断するための学習モデルの代表的なものとして、マルチタスク学習があげられる。マルチタスク学習は、単一のモデルで複数のタスクを解く手法であり、複数のタスクを同時に学習することで、タスク間で共通の要因を獲得できる [2][3].

一方、感性的質感認知に関する研究では、古くからテクスチャ認知・生成、オノマトペや印象評価語と画

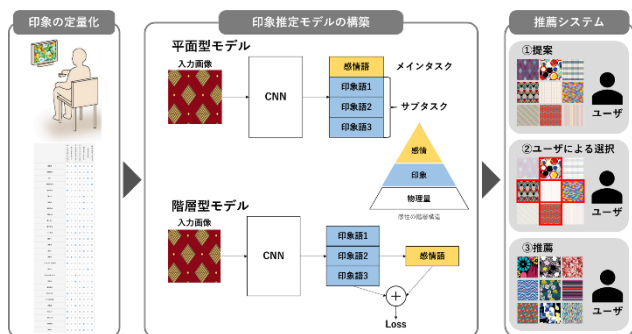


図 1 提案手法の概念図

像特徴量の関係、視覚特性や神経基盤の解明など多方面から行われてきた。最近では機械学習の分野で、好み、美的良さや感情など審美性 (aesthetic scores, aesthetic values) を CNN でモデル化し画像理解・推薦に応用する研究が盛んに行われている[4-7]。しかし、心理学や感性工学の分野では、対象に対する評価者の反応(美しい、好き等の感情)は、対象そのものが備える性質(派手な、スリムな等の印象、美的概念, aesthetic concepts)とは明確に区別され、感情は印象(に関する情報処理)を経て形成される評価であると説明されている [8]。対象に関する多様な印象を CNN モデルで表現した研究はまだ少数である。

寸田らは、衣服の柄の検索や生成のために画風変換アルゴリズム Neural Style Transfer [9] における Style と呼ばれるヒトの直感によく合致する質感特徴量を用いて、柄を表現する 41 の印象のモデル化を行った [10]。この中で「かわいい」のような高次の感性的質感には個人差が見られたことから、個人差を考慮したモデルへの拡張が課題として挙げられた。

3. 提案手法

本研究では、衣服の柄を対象に感性的質感と物理特徴との関係性をモデル化し、感性的質感の構造に基づいた柄の印象推定モデルの構築を目的とする。1 つは多様な感性的質感を扱う枠組みとして Multi-task CNN モデルを用い、感情語 (1 語) をメインタスク、印象語 (28 語) をサブタスクとして構成し、同時に解くことにより個人差の表現を可能とする。もう 1 つは感性的質感の構造をモデルで表現するため階層的なモデルの構築を行う。提案手法の概念図を図 1 に示す。(1) はじめに、主観評価実験を行い柄に対する感性的質感を定量化する。(2)次に、感性的質感を推定するモデルを構築する。(4)全ての柄に対してこのモデルを用いて感性的質感を推定し、推定値が高い柄をユーザーに推薦する。



図 2 柄画像の例

表 1 印象語 28 語

かわいい	きれいな	やわらかい	ガチャガチャした	コントラストの高い	シックな	センスのある
不揃いな	冷たい	凝った	単色の	古い	古風な	均等な
対称的な	平行な	幾何学的な	怪しげ	洋風な	派手な	特徴的な
穏やかな	細かい	艶やかな	華々しい	規則的な	集合体の	高価な

3.1. 印象の定量化

柄から喚起される視覚的な印象を定量化するため、印象評価実験を行った。実験には、エキスパートと非エキスパートを問わない 4440 名が参加した。実験に用いた柄画像は 2878 枚 (図 2) で、画像 1 枚あたり 20 名分の評価が得られるように実験を行った。評価項目には、表 1 の印象語 28 語と「好き-嫌い」「良い-悪い」「覚醒-沈静」「快-不快」の感情語 4 語を用いた。

実験方法は、被験者に画面に表示された柄の画像を観察してもらい、各評価項目についてどの程度あてはまるか評価してもらった。印象語 28 語と「好き-嫌い」「良い-悪い」の感情語 2 語については、「全くあてはまらない」から「非常にあてはまる」の評価尺度からなる 7 件法で評価を行い「全くあてはまらない」を -3 点、「非常にあてはまる」を 3 点として各評価尺度に対して 1 点刻みの点数化を行った。「覚醒-沈静」「快-不快」の感情語 2 語については、アフェクトグリッド法で評価を行った。横軸を「覚醒-沈静」、縦軸を「快-不快」とする 2 次元のアフェクトグリッドを用い、各次元について 1 から 9 までの 9 段階で回答を求めた。

得られた評価データに対し、評価内容や回答時間からクリーニングを行った。実験刺激の中にダミー画像を含め、この画像に対する評価が指示と異なった被験者を不誠実回答者とした。この不誠実回答者の中で、回答時間が最頻値よりも短い被験者を短時間回答者とした。また、28 個の評価項目のうち、いずれかの項目で全ての画像に対する評価値が同一であった被験者は、次節で相関係数が算出できないため除外した。クリーニングを行った結果、採用した有効な回答者は 3347 名であった。

3.2. 印象推定モデルの構築

本研究では、マルチタスク学習を用いて、感情語「好

表 1 学習モデル比較結果

評価語	平面型		階層型	
	平均		分布	
	シングル	マルチ	シングル	マルチ
かわいい	0.56	0.72	0.72	0.69
きれいな	0.54	0.65	0.64	0.62
やわらかい	0.75	0.80	0.82	0.81
ガチャガチャした	0.75	0.86	0.87	0.87
コントラストの高い	0.73	0.80	0.85	0.84
シックな	0.50	0.61	0.69	0.67
センスのある	0.44	0.52	0.46	0.44
不揃いな	0.68	0.78	0.79	0.78
冷たい	0.73	0.71	0.81	0.80
凝った	0.78	0.82	0.85	0.83
単色の	0.81	0.85	0.91	0.91
古い	0.56	0.63	0.66	0.65
古風な	0.58	0.65	0.68	0.66
均等な	0.72	0.78	0.81	0.81
対称的な	0.67	0.74	0.77	0.75
平行な	0.73	0.81	0.85	0.85
幾何学的な	0.72	0.77	0.81	0.79
怪しげ	0.77	0.81	0.83	0.82
洋風な	0.49	0.58	0.64	0.60
派手な	0.83	0.85	0.88	0.87
特徴的な	0.74	0.85	0.91	0.90
穏やかな	0.77	0.82	0.82	0.81
細かい	0.75	0.78	0.81	0.79
艶やかな	0.64	0.71	0.72	0.69
華々しい	0.66	0.75	0.78	0.77
規則的な	0.67	0.76	0.82	0.81
集合体の	0.68	0.74	0.75	0.74
高価な	0.57	0.67	0.67	0.66
印象語28語平均	0.67	0.74	0.77	0.76
好き - 嫌い	0.63	0.65	0.62	0.62

き - 嫌い」の回帰タスクと印象語 28 語の回帰タスクを同時に解かせる印象推定モデルを構築する。マルチタスク学習によって関連する複数の課題を同時に学習することで、課題間で共通の要因を獲得でき、予測精度の向上が期待できる。特徴量抽出器として学習済みの ResNet-50 を用いた。多出力なモデルであるため、全結合層を出力サイズが 1 の全結合層 29 個に変更した。損失関数には二乗平均誤差(MSE)を用い、各タスクの損失の合計を最小化するようにした(式 1)。メインタスクを感情語「好き - 嫌い」の回帰タスク、サブタスクを印象語 28 語の回帰タスクとし、サブタスクの重みを 0.5 から 1.0 の範囲でパラメータチューニングを行った。

$$L_{all} = L_{main} + \lambda_{sub1}L_{sub1} + \lambda_{sub2}L_{sub2} + \dots + \lambda_{subn}L_{subn} \quad (1)$$

4. 結果・考察

学習データやモデル構造などを変更したモデルを複数構築し精度比較を行う。まず、マルチタスクモデ

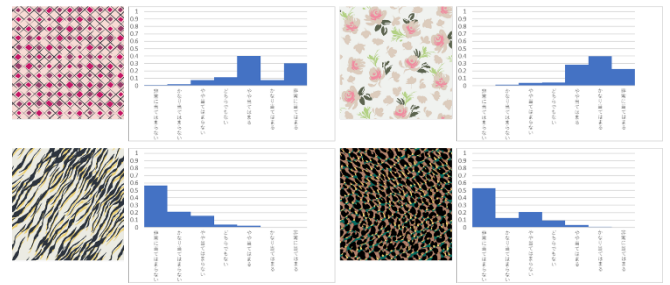


図 3 モデルの出力結果例
「かわいい」

ルの有用性を調べるため、シングルタスクモデルとの比較を行った。次に、感情と印象の関係性に基づいて分割したデータが、好みの違いを考慮したモデル構築につながるか調べるため、複数の学習データ間で精度比較を行った。続いて、印象分布を学習データとして用いたモデルとの比較を行った。最後に、感性の階層構造をモデルで表現するため階層的なモデルを構築し比較を行った。

推定精度の評価指標には、実測値と推定値との相関係数を用いた。印象分布を用いた分類モデルでは、印象分布の期待値と推定結果(推定印象分布)の期待値との相関係数を算出した。

マルチタスクモデルとシングルタスクモデルとの比較結果を表 1 (シングル・マルチ) に示す。ほぼ全ての評価語において、シングルタスクモデルよりもマルチタスクモデルの精度が高かった。この結果と 1 つのモデルで複数のタスクを同時に解くことができる効率性から、マルチタスクモデルの有用性を確認した。

平均値で学習したモデルと印象分布で学習したモデルの比較を行った結果を表 1 (平均・分布) に示す。図 3 にモデルが推定したかわいいの印象分布を示す。印象分布は印象評価における評価のばらつきで、各評価点の人数を合計人数で割った割合とする。印象語 28 語の精度の平均は全参加者の印象分布のデータで学習させたモデルが最も高かった。この理由として、平均値と比べて画像 1 枚あたりに付与された情報量が多いためと考えられる。平均値では、他の参加者の評価は平均を取ることで潰れてしまうが、印象分布を用いることで他の参加者の評価の傾向も学習時に考慮することができるため、精度が高くなったと考える。

感性の階層構造を印象推定モデルの構造で表現するため、印象語を推定する層の後に感情語を推定する層を置いたモデルを構築した。モデルの構造を図 2 に示す。印象分布を学習データとした階層型モデルと平面型のマルチタスクモデルとの比較を表 1 (階層型・平面型) に示す。平面型モデルと比較した結果、印象語 28 語は精度に大きな差はないがやや平面型モデルの精度が高かった。感情語の精度は大きな差は出な

ったが、階層型モデルの方がやや精度が高くなった。この理由として、印象語は平面型モデルと構造自体は変わっていないためであると考えられる。しかし、感情語はやや精度が高くなっていることから各印象語の重みを工夫することでより感情語の精度を上げられる可能性があると考えられる。この結果から、28印象語のマルチタスク出力から感情語の出力を得る構造の工夫をする必要があると考えられる。

5. まとめ

本研究では、衣服の柄を対象に感性的質感と物理特徴との関係性をモデル化し、感性的質感の構造に基づいた柄の印象推定モデルの構築を行った。1つは多様な感性的質感を扱う枠組みとして Multi-task CNN モデルを用い、感情語（1語）をメインタスク、印象語（28語）をサブタスクとして構成し、同時に解くことにより個人差の表現を可能とした。もう1つは感性的質感の構造をモデルで表現するため階層的なモデルの構築を行った。はじめに、主観評価実験を行い柄に対する感性的質感を定量化する。次に、柄に対する感性的質感を推定するモデルを構築し各モデルを比較した。シングルタスクモデルとの比較により、マルチタスクモデルの有用性が確認された。また、感性評価構造を意識したモデルの構造が重要であることがわかった。

今後は、単一のモデルでより多様な感性的質感の個人差を表現できる手法を検討する。

文 献

- [1] 小松英彦, “ 質感の科学への展望, ” 映像情報メディア学会誌, vol. 66, no. 5, pp.331– 337, 2012.
- [2] Caruana, Rich. “Multitask learning.” *Machine learning*, 28 41-75, 1997.
- [3] Tang, H., Liu, J., Zhao, M., and Gong, X., “Progressive layered extraction (ple): A novel multi-task learning (mtl) model for personalized recommendations”, *Proceedings of the 14th ACM Conference on Recommender Systems*, pp. 269–278, 2020.
- [4] Soydaner, Derya, and Johan Wagemans, "Multi-task convolutional neural network for image aesthetic assessment", *arXiv preprint arXiv:2305.09373*, 2023.
- [5] Li, L., Zhu, H., Zhao, S., Ding, G., and Lin, W., “Personality-assisted multi-task learning for generic and personalized image aesthetics assessment”, *IEEE Transactions on Image Processing*, 29, pp.3898-3910, 2020.
- [6] Chen, Y., Pu, Y., Zhao, Z., Xu, D., Man, and Qian, W., “Image Aesthetic Assessment Based on Emotion-Assisted Multi-Task Learning Network”, In *Proceedings of the 2021 6th International Conference on Multimedia Systems and Signal Processing*, pp. 15-21, 2021.
- [7] Li, L., Zhu, H., Zhao, S., Ding, G., & Lin, W., “Personality-assisted multi-task learning for generic and personalized image aesthetics assessment.” *IEEE Transactions on Image Processing*, 29 3898-3910, 2020.
- [8] S. Hashimoto, A. Yamada, and N. Nagata, "A quantification method of composite impression of products by externalized evaluation words of the appraisal dictionary with review text data," *International Journal of Affective Engineering*, vol.18, no.2, pp.59–65, 2019.
- [9] L. A. Gatys, A. S. Ecker, and M. Bethge, “Image style transfer using convolutional neural networks” *The IEEE Conference on Computer Vision and Pattern Recognition 2016*, pp.2414–2423, 2016.
- [10] N. Sunda, K. Tobitani, I. Tani, Y. Tani, N. Nagata, and N. Morita, “Impression estimation model for clothing patterns using neural style features” *HCI International*, pp.689–697, 2020.