

Multi-task CNN による感性的質感の個人差を考慮した 柄推薦システムの構築

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 本研究では、感性的質感の個人差に着目し、少量のデータからユーザの求めるイメージに合った衣服の柄を推薦するシステムの構築を目的とする。はじめに、主観評価実験で柄に対する感性的質感を定量化し、被験者の好みを明らかにするためクラスタリングを行った。次に、シングルタスクモデル、被験者全員分のデータを用いたモデル、クラスタ毎にモデル構築を行い、精度を比較した。学習に用いるデータについて、全てのデータを用いた場合と画像ごとの平均値を用いた場合の精度の比較も行った。その結果、マルチタスクモデルを用いることで、シングルタスクモデルよりも精度が向上した。

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

1. はじめに

プロダクトデザインの分野では、感性的質感への関心が高まっている。感性的質感とは、素材の表面性状（テクスチャ）から喚起される感情や印象のことで、人がものの良し悪しや好ましさを評価・判断するうえで重要な要素となる。しかし、感性的質感には個人差があるため、この個人差をプロダクトに活かしていくことが求められている。特に EC の普及によるユーザニーズの多様化に伴い、ファッション業界では衣服のオーダーメイドサービスといった、プロダクトのカスタマイズ化やパーソナライズ化が注目されている。多くの選択肢からユーザの意思決定を補助するため、EC サイトではユーザが興味を持つと思われる製品を推薦するシステムが用いられている。しかし、購入履歴がなければ推薦できない問題や似たような製品ばかりが推薦されてしまう問題がある。そのため、ユーザー一人一人の好みやニーズを、少量のデータから推定する技術が求められている。感性的質感の個人差を分析し、誰でも簡単に自分の好みに合ったプロダクトを手にすることができれば、ユーザの満足や Well-being の向上のみならず、大量生産による無駄の削減と持続可能な社会の実現にもつながると考える。そこで本研究では、感性的質感の個人差に着目し、少量のデータからユーザの求めるイメージに合った衣服の柄を推薦するシステムの構築を目的とする。

2. 先行研究

推薦システムのような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 [1]。

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 [2-4]。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。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 [5] 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 [6]。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、衣服の柄を推薦するシステムの構築において、個人の柄に対する好みの違いを考慮した推薦を実現することを目的とする。多様な感性的質感を扱う枠組みとして Multi-task CNN モデル用い、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。また、個人差を扱う最初の試みとして個人のタイプ分析を行い、タイプ毎にモデル構築を行う。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、(1)の結果を基に被験者が柄の好ましさを評価するうえで、どの印象に注目しているのかを明らかにするため、クラスタリングを行う。(3)(2)で分類したクラスタごとに、柄に対する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に推薦する。

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

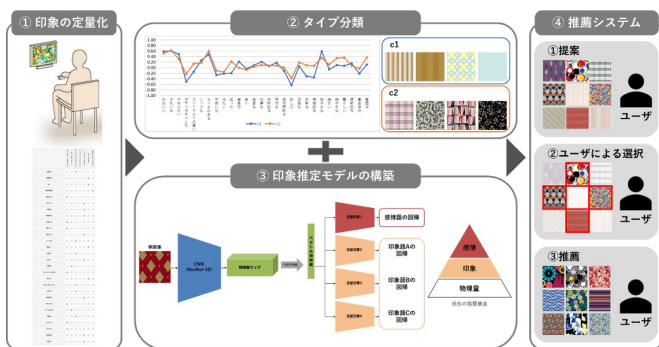


図1 提案手法の概念図

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

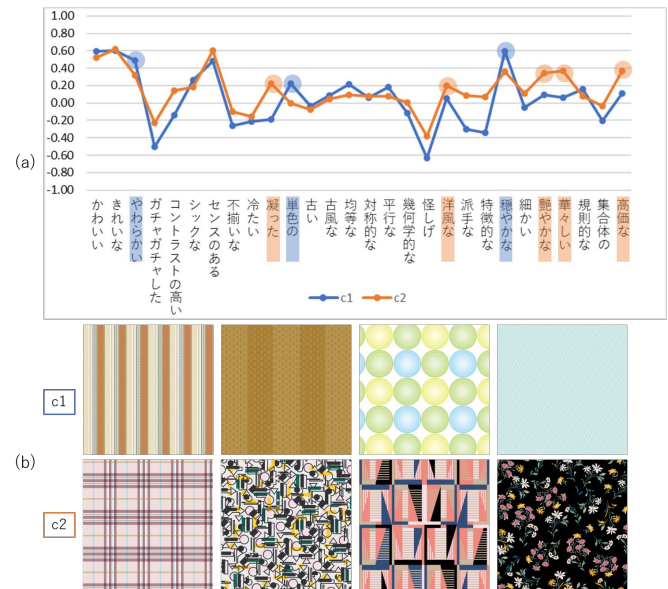


図2 (a)クラスタ毎の好き-嫌いとの相関 (b)クラスタ毎の好き-嫌いの得点が高い柄

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. クラスタリング

モデルを学習させるデータを、感情と印象の関係性に基づいて分割するため、柄に対する印象と好みの類似性に基づき、被験者のセグメンテーションを行った。実験で得られたデータを「好き-嫌い」と印象語28語との相関係数に基づき、ward法を用いてクラスタリングを行った。クラスタ毎の解釈のしやすさを考慮して、被験者を2つのクラスタに分割した(図2)。

クラスタ1は「好き-嫌い」と「きれいな」「穏やかな」「かわいい」「やわらかい」「センスのある」との間に中程度の正の相関、「シックな」「単色の」「均等な」との間に弱い正の相関が見られた。その中で、クラスタ2と比べて「穏やかな」「単色の」「やわらかい」との相関が特に高かった。クラスタ2は、「好き-嫌い」と「きれいな」「センスのある」「かわいい」との間に中程度の正の相関、「高価な」「華々しい」「穏やかな」「艶やかな」「やわらかい」「凝った」「洋風な」との間に弱い正の相関が見られた。その中で、クラスタ1と比べて「凝った」「華々しい」「高価な」

「艶やかな」「洋風な」との相関が特に高かった。各クラスタの人数は、クラスタ 1 が 1431 名、クラスタ 2 が 1916 名となった。

3.3. 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。データセットを Train : Validation : Test = 8 : 1 : 1 に分割し学習と検証を行った。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した。

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

モデルの学習には、画像 1 枚あたりに付けられた全被験者の評価の平均値を用いた。まず、マルチタスクモデ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。被験者全員分のデータで学習させたモデル、クラスタリングで分割した 2 つのクラスタ毎のデータで学習させたモデル、ランダムに 1431 名と 1916 名に 2 分割したデータで学習させたモデルを比較した。

4. 結果・考察

テストデータに付けられた評価の平均値とモデルから得られた推定値との相関係数を推定精度の指標とする。

被験者全員分のデータで学習させ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を表 2 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方が精度が高かった。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。

次に、マルチタスク学習を用いて被験者全員分のデータで学習させたモデル、クラスタ毎のデータで学習させたモデル、ランダムに分割したデータで学

表 2 マルチタスクモデルとシングルタスクモデルの比較

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

習させたモデルの比較を表 3 に示す。全てのモデルにおいて評価値と推定値との間に正の相関が見られた。印象語 28 語の精度の平均は、全被験者のデータで学習させたモデルが最も高かった。また、印象語 28 語の精度の平均をクラスタ 1 と 2 で比較すると、クラスタ 2 の方が高かった。この理由として、クラスタ 2 はクラスタ 1 と比べて、「好き-嫌い」との正の相関が強い印象語は少ないため、「好き-嫌い」との類似性の特徴があまりデータに反映されなかったが、構成人数がクラスタ 1 よりも多いため、全体的な印象語の精度は上がったと考える。全被験者のデータで学習させたモデルの精度が最も高かったのも、ランダムに 2 分割したデータで学習したモデルよりも精度が高いことから、学習データの数が多かったためと考える。「好き-嫌い」では、全被験者とクラスタ 1 のデータで学習させたモデルの精度が高かった。「好き-嫌い」もクラスタ 1 と 2 で比較すると、クラスタ 1 の方が高かった。この理由として、クラスタ 1 はクラスタ 2 よりも「好き-嫌い」と印象語との正の相関が強いものが多く「好き-嫌い」との類似

表 3 学習データの比較

評価語	全被験者	クラスタ 1	クラスタ 2	ランダム 平均
かわいい	0.72	0.57	0.65	0.55
きれいな	0.65	0.59	0.59	0.53
やわらかい	0.80	0.67	0.74	0.66
ガチャガチャした	0.86	0.78	0.77	0.78
コントラストの高い	0.80	0.67	0.74	0.71
シックな	0.61	0.51	0.56	0.50
センスのある	0.52	0.37	0.52	0.37
不揃いな	0.78	0.67	0.77	0.71
冷たい	0.71	0.68	0.73	0.55
凝った	0.82	0.72	0.77	0.70
単色の	0.85	0.82	0.85	0.71
古い	0.63	0.49	0.52	0.49
古風な	0.65	0.51	0.60	0.53
均等な	0.78	0.64	0.77	0.70
対称的な	0.74	0.64	0.75	0.66
平行な	0.81	0.72	0.78	0.69
幾何学的な	0.77	0.65	0.70	0.65
怪しげ	0.81	0.76	0.76	0.75
洋風な	0.58	0.35	0.57	0.46
派手な	0.85	0.79	0.80	0.76
特徴的な	0.85	0.70	0.82	0.69
穏やかな	0.82	0.70	0.74	0.73
細かい	0.78	0.64	0.75	0.68
艶やかな	0.71	0.59	0.65	0.48
華々しい	0.75	0.71	0.71	0.67
規則的な	0.76	0.70	0.76	0.68
集合体の	0.74	0.55	0.62	0.61
高価な	0.67	0.58	0.62	0.49
印象語 28 語平均	0.74	0.63	0.70	0.63
好き・嫌い	0.65	0.65	0.50	0.53

性の特徴がデータに反映されていたため、クラスタ 1 はクラスタ 2 よりも構成人数は少ないが精度は上がったと考えられる。

5. まとめ

本研究では、感性的質感の個人差に着目し、少量のデータからユーザの求めるイメージに合った衣服の柄を推薦するシステムの構築を行った。(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、(1)の結果を基に被験者が柄の好ましさを評価するうえで、どの印象に注目しているのかを明らかにするため、クラスタリングを行う。(3)(2)で分類したクラスタ毎に、柄に対する感性的質感を推定するモデルを構築する。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示唆された。また、モデルの精度は学習データの数によって変化するが、データにクラスタの特徴が反映されていれば少量でも精度が上がる可能性があることがわかった。

今後は、クラスタリングによって個人の好みの違いを表現するのではなく、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] 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)
- [2] Soydaner, Derya, and Johan Wagemans, "Multi-task convolutional neural network for image aesthetic assessment", arXiv preprint arXiv:2305.09373 (2023)
- [3] 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)
- [4] 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)
- [5] 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)
- [6] 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 2020, pp.689–697 (2020)