

機械学習を駆使した顔型判定システムの開発

システム科学技術学部 情報工学科

2年 小林 陽

指導教員 システム科学技術学部 情報工学科

教授 猿田 和樹

助教 寺田 裕樹

1. はじめに

購入する眼鏡を決定する際に「その眼鏡が似合っているかどうか」を検討することは一般的であり、最も重視されているといっても過言ではない。しかし、この方法で評価することは主観的にも客観的にも難しく、この感覚的尺度を習得するためには多くの経験が必要となる。この問題を解決するため、大手眼鏡チェーンの株式会社ジンス（以下「JINS」と称する）では JINS BRAIN [1] という、眼鏡を着用した顔画像からその眼鏡が似合っているかどうかを判定できるデジタルサービスを提供している。このサービスには、JINS スタッフ 3000 人によって「似合う」または「似合わない」のラベルが付けられた約 60000 枚の眼鏡着用画像を学習させた深層学習モデルが使われている。判定結果として、いわゆる「似合い度」が百分率で得られ、眼鏡が似合っているかどうかを数値で知ることができる。確かに似合う眼鏡を見つけられない人にとっては眼鏡購入の手助けとなるサービスである。しかし、利用者はなぜ自分にこの眼鏡が似合うのかという根拠を知ることができず、先述した感覚的尺度を習得する手助けとはなり得ない。

こうした先進的な試みが行われる一方で、従来から知られている眼鏡の選び方がある。それは、顔の輪郭を 4 タイプに大別して得られる「顔型（面長顔、四角顔、丸顔、三角顔）」を基にして眼鏡を選ぶという方法である。それぞれの顔型に似合う眼鏡の形があらかじめ知られていて、自分の顔型に似合ういくつかの形の中から眼鏡を吟味することができる。また、自分の顔型が分かっているため、先述した感覚的尺度を習得するのも比較的容易になる。しかし、それぞれの顔型の定義は非常に曖昧で、結局自身の未熟な感覚に頼って顔型を推測する必要がある。加えて、4 タイプの顔型の特徴にぴったり当てはまる顔であることはまれであり、複数の顔型の特徴を併せ持つ顔であることが大半である。そこで、本研究課題では、似合う眼鏡を見つけることを容易にするため、機械学習を駆使し、入力された顔に対して顔型の混合割合を出力するシステムの開発を行う。

2. システム開発

2-1. 学習用データセットの選定と顔型のラベリング

本システムの中核を担うのは、上記の 4 タイプに顔画像を分類できるように学習した深層学習モデルである。このモデルの学習には、顔型がラベリングされた顔画像のデータセットが必要不可欠である。そこで、まずはこのデータセットを用意することから始めた。

顔型を正確に学習させるには、ラベリングする顔画像は顔の向きが正面あるいはそれに近い向きであることが有効であると考えた。この条件を満たす顔画像が多く含まれるデータセットが、

顔認識のベンチマークデータセットとして用いられる Labeled Faces in the Wild (LFW) であり、本研究では LFW に含まれる顔画像中の顔が直立するように処理されたデータセット [2] を使用した。このデータセットから機械学習ライブラリの一つである Dlib の顔検出ツールによって顔が一つだけ検出された画像のみを抽出し、また、学習のノイズとなり得る画像の除去も行った。

冒頭で述べた通り、顔型の定義は非常に曖昧である。そのため、定義に基づいて顔型をラベリングすることは不可能である。そこで、非階層的クラスタリングのアルゴリズムである k 平均法により顔画像を 4 つのクラスタに分類することで、ラベリングを試みた。ここで、 k 平均法を簡潔に説明する。 N 個のベクトル x_1, x_2, \dots, x_N と k 個のクラスタ中心ベクトル z_1, z_2, \dots, z_k があるとき、「 k 個の各クラスタへの N 個の各ベクトルの割り当て」と「クラスタ中心ベクトルの更新」を収束するまで繰り返すアルゴリズムである。 k 平均法においてクラスタ中心ベクトルの初期値はランダムに選ばれるが、ここでは 4 つの顔型のイラスト (図 1) を初期値として設定した。

次にクラスタリングに用いる顔画像の特徴量について説明する。一般に顔型の違いを説明する際に、顔の縦横比がよく用いられる。例えば、丸顔は縦横比が 1:1 程度、などである。顔型の定義としては不十分だが、理屈としては理解できる。このことから、この「顔の縦横比」を基に、顔の輪郭の特徴をさらに捉えられるよう改良すれば、クラスタリングによって顔画像をそれぞれに共通する特徴で 4 つの顔型に分類できると考えた。そこで特徴量の抽出に Dlib で利用できる shape predictor 68 face landmarks [4] という学習済みモデルを使用した。このモデルを利用することで、顔画像から 68 点のランドマーク (図 2) とそれらの座標が得られる。これらのランドマークのうち、顔の輪郭に相当する 1~17 番の点と、2 点間の距離が顔の縦の長さに相当する 9 番と 28 番の点に着目する。9 番と 28 番を通る直線を対象軸として考えたときに、対応する輪郭上のそれぞれの 2 点間の距離 (1 番と 17 番, ..., 8 番と 10 番) と顔の縦の長さとの比を取ることで、合計 8 個の特徴量を得た。この特徴量を抽出する処理を 9676 枚の顔画像に対して行った後、図 1 の 4 つの顔型の特徴ベクトルをクラスタ中心ベクトルの初期値としてクラスタ

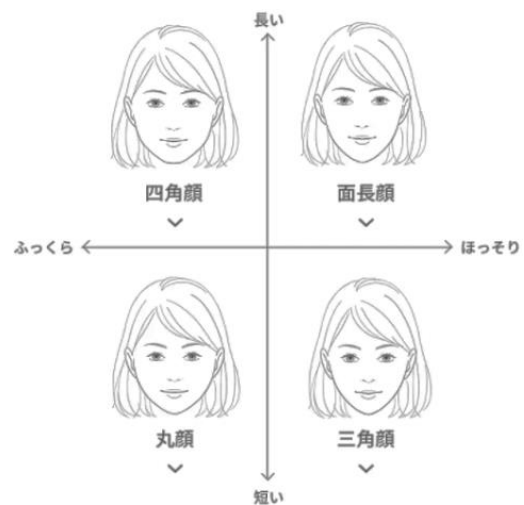


図 1 4 つの顔型のイラスト [3]

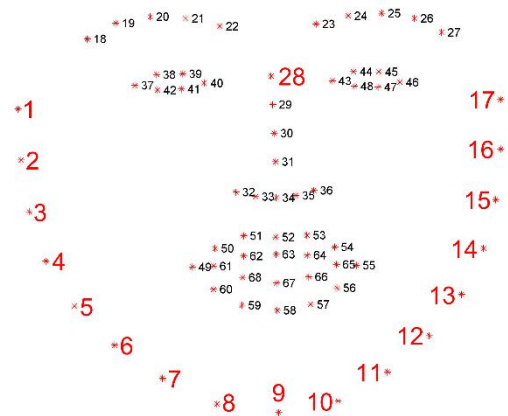


図 2 得られるランドマークの番号 [5]

リングを行った。クラスタリングの結果として、四角顔クラスタに 3461 枚、三角顔クラスタに 3217 枚、面長顔クラスタに 1734 枚、丸顔クラスタに 1264 枚の顔画像が割り当てられた。後述する深層学習モデルでこの 9676 枚の顔画像を学習したところ、面長顔と丸顔の識別精度が著しく低くなった。面長顔と丸顔クラスの顔画像枚数が他の 2 クラスよりも少ないことで過学習が引き起こされたと考えられる。文献 [6] にも「訓練サンプルの量が足りないことは過適合を引き起こす最

大の原因」と記載がある。このモデルの学習時に、非常に少ないエポック数の段階から検証損失と訓練損失の乖離が確認されたことから過学習が生じたと考えられる。そこで、面長顔と丸顔クラスに対してデータ拡張を行った。データ拡張の詳細としては、面長顔と丸顔クラスの顔画像に対して鏡映反転を行い、また、丸顔クラスの顔画像に対してガンマ補正 ($\gamma = 1.5$) を行った。データ拡張の結果として、四角顔クラスが 3461 枚、三角顔クラスが 3217 枚、面長顔クラスが 3459 枚、丸顔クラスが 3774 枚、計 13911 枚の学習データが得られた。さらに、この 13911 枚の画像の顔のサイズはそれぞれ異なるため、顔の位置と大きさを統一する正規化処理を行い、 250×250 ピクセルの 13911 枚の正規化顔画像を得た。

2-2. モデルの構築と学習

続いて、本研究で構築したモデルについて説明する。図 3 に構築したモデルの構造を示す。プーリング層では最大プーリングを計算している。また、畳み込みの前に入力画像に対してゼロパディングを行っている。ここで、入力画像は 250×250 ピクセルのグレースケール画像である。

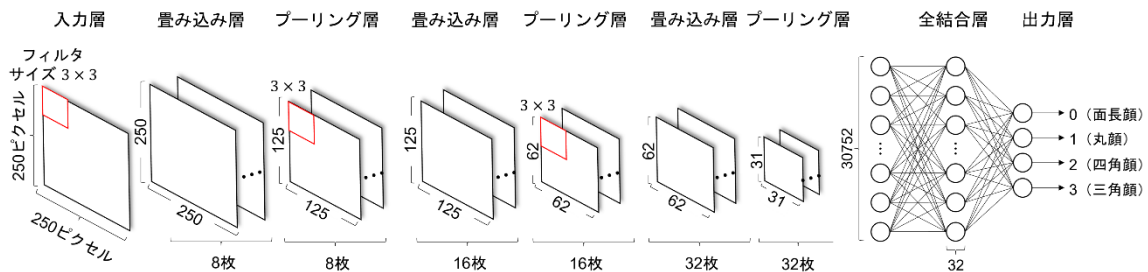


図 3 ネットワーク構成

畳み込み層および全結合層での活性化関数として正規化線形関数を使用し、プーリング層と全結合層の出力でドロップアウト ($p = 0.23$) を行っている。また、出力層の活性化関数としてソフトマックス関数を使用した。なお、このモデルは `tensorflow.keras` モジュールによって構築した。モデルの学習であるが、13911 枚の学習データのうち、80%にあたる 11128 枚を訓練データ、10%にあたる 1391 枚をテストデータ、残りの 1392 枚を検証データとした。なお、最適化アルゴリズムとして Adam、損失関数として交差エントロピー誤差、評価指標として正答率を使用し、バッチサイズは 32 とした。図 4 にこの学習済みモデルの混同行列を示す。学習には 16 エポック（早期打ち切り）を要し、テストデータでの損失は約 0.873、正答率は約 0.611 となった。図 4 より、再現率の平均は約 0.605、適合率の平均は約 0.604 となる。再現率と適合率の調和平均 F_1 値は約 0.605 となった。学習済みモデルに画像を入力した時の出力結果を図 5 に示す。図 1 を参考にし、図 5 の入力と出力を見比べると、妥当な出力結果であると考えられる。図 5 に示した画像以外

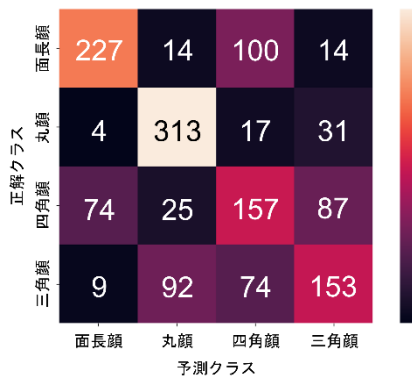


図 4 混同行列

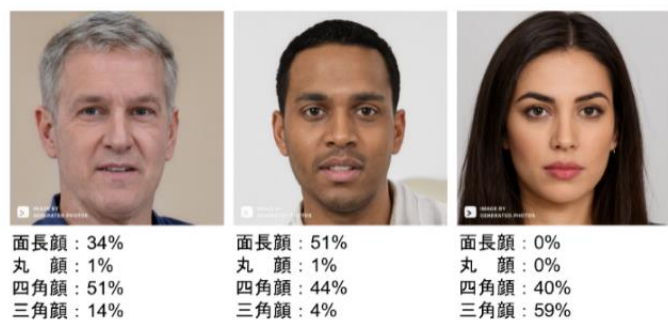


図 5 システムへの入力画像と出力の顔型混合割合

(画像: [Generated Photos](#))

でも同様に、妥当と判断できる出力が得られた。一方で、明らかに顔型の混合割合が推測できるような画像を入力した際に、予想に反する出力が得られたケースもあった。

3. 考察

2-2で、学習済みモデルの評価指標として正答率・適合率・再現率・ F_1 値を算出した。どの指標に着目しても、一定の性能は得られているものの、決して高精度な分類は実現できていないといえる。その原因として、学習データの顔の向きが統一されていなかったことが考えられる。明らかに正面顔ではないと判断した画像は除外したが、データ数確保のために多少横を向いていても正面顔とみなしていた。その結果、正面顔とみなした画像がクラスタリングで本来とは異なるクラスに割り当てられた可能性がある。本システムでは顔の向きに自由度を持たせることは考えていないため、純粋な正面顔のみを集めたデータセットを作成して再びモデルを学習させ、性能を評価したい。一方で、既存の評価指数で本研究のモデルを正しく評価することはできるのかと疑問に感じた。システムの目的は、ある顔の顔型混合割合を出力することであり、入力のある特定のクラスだと明確に判別することではない。この点に関しては、出力の混合割合が感覚と一致するか等の実験が有効であると考えられ、今後取り組みたい。また、クラスタリングで使用する特徴量についても検討し直す必要がある。例えば、顔型の印象に影響を与える額の形状も特徴量として扱う、などである。他にも、画像に付与するラベルの種類を増やすなど、深層学習モデルの改善策が多数考えられるので、これらは今後の課題としたい。

4. おわりに

本研究課題では顔型判定システムのプロトタイプを開発し、構想段階で目的とした機能を実現できた。今後の構想としては、深層学習モデルを改善し、顔型判定システムを GUI アプリケーションとして実装することが挙げられる。

5. 参考文献

- [1] ABOUT | JINS BRAIN | JINS-眼鏡 (メガネ・めがね),
<https://brain.jins.com/about/> (参照 2022-03-8)
- [2] Gary B. Huang, Marwan A. Mattar, Honglak Lee, Erik Learned-Miller, "Learning to Align from Scratch", *Advances in neural information processing systems 25*, pp.764-772, 2012
- [3] メガネの選び方 (似合うメガネを選ぶ) | JINS - 眼鏡 (メガネ・めがね),
https://www.jins.com/jp/guide/eyewear/select_glasses.html (参照 2022-03-09)
- [4] GitHub - davisking/dlib-models: Trained model files for dlib example programs.,
https://github.com/davisking/dlib-models#shape_predictor_68_face_landmarksdatbz2
(参照 2022-03-09)
- [5] i·bug - resources - Facial point annotations,
<https://ibug.doc.ic.ac.uk/resources/facial-point-annotations/> (参照 2022-03-09)
- [6] 岡谷貴之, 深層学習, 講談社, p.35, 2015