

ディープニューラルネットワークを用いた 顔画像による野球選手・サッカー選手の分類

谷聖一 研究室 田名網 勇希
Yuuki Tanaami

概要

本演習では、機械学習を用いて、顔画像による野球選手・サッカー選手の分類を試みた。その際、顔画像による性的指向の分類において、人間の分類精度を超えたとする論文 [Wang, Y., & Kosinski, M. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of Personality and Social Psychology*. (in press)] の分類手法を参考にした。

1 はじめに

子供の頃、「○○は野球顔だね」、「△△くんはサッカー顔っぽい」などという話をした経験がある人もいるだろう。野球顔は丸顔で丸い目、サッカー顔は縦長な顔で細長い目という漠然としたイメージが報告者にはある。しかし、野球選手・サッカー選手の顔を見ても野球顔なのかサッカー顔なのか判断できない。はたして顔画像から野球選手とサッカー選手を分類できるのだろうか。

2017 年 9 月にスタンフォード大学の Yilun Wang, Michal Kosinski らが、顔画像による性的指向の分類において、人間の精度を超えたとする論文を発表した ([1])。[1] では、18 歳から 40 歳の白人を対象に、アメリカの出会い系サイトの利用者 14,776 人の顔画像 35,326 枚を用いて、同性愛者と異性愛者への分類を行なった。[1] では、顔画像が 1 人あたり 1 枚の場合の精度は、男性 81%、女性 71%、顔画像が 1 人あたり 5 枚の場合では、男性 91%、女性 81%という実験結果が得られている。同顔画像を人間が分類した場合の精度は、男性 61%、女性 54%であった ([1])。また [1] では、分類精度が顔画像の背景に依存していないか、出会い系サイト以外の顔画像でも分類可能か検証を行なっている。

[1] では、ディープニューラルネットワーク (Deep Neural Network: DNN) を用いた機械学習 (Machine Learning: ML) を行なっている。機械学習とは、「明示的にプログラムしなくても学習する能力をコンピュータに与える研究分野」 ([2]) である。ニューラルネットワーク (Neural Network: NN) とは、生物の脳神経系を模した情報処理システムである ([3])。ニューラルネットワークを多層構造化したネットワークを、ディープニューラルネットワークという。

本演習では、[1] の分類手法を用いて、顔画像による野球選手・サッカー選手の分類を試みる。

2 準備

この節では、使用したニューラルネットワークのモデル・評価指標・評価方法を説明する。

2.1 VGG Face

VGG は、オックスフォード大学の Visual Geometry Group [4] が開発したディープニューラルネットワークである。全 16 層で構成され、ILSVRC2014 の分類問題では 2 位を獲得した。VGG を顔画像に特化させて学習させたものが、VGG Face である。

2.2 AUC

AUC (Area Under the Curve) とは、二値分類における評価指標の一つである。AUC は 0.0~1.0 を取り、AUC が 1.0 のときは完全な分類・0.5 のときはランダムな分類・0.5 未満のときはランダムな分類より劣っていることを示す。

2.3 K-分割交差検証法

K-分割交差検証法 (K-Fold Cross Validation) とは、機械学習の評価方法の 1 つである。データセットを K 個に分割し、K-1 個のデータを学習データ、残り 1 個のデータをテストデータとして分類を行う。これを K 個のデータそれぞれで行い、K 個の評価の平均を分類の評価として用いる。

3 演習方法

以下に、本演習で利用したサービス・言語・ライブラリを示す。

- サービス

- Bing Image Search API [5]

野球選手・サッカー選手の画像収集に用いる。

- Face++[6]
画像から顔の情報を取得するために用いる.
- 言語
 - Python
- ライブラリ
 - Keras[7]
 - keras-vggface[8]
VGG Face による特徴抽出を行う際に用いる
 - scikit-learn[9]
分類を行う際に用いる

演習の手順を以下に示す.

1. 画像収集: Bing Image Search API[5] を使用して画像を収集する
2. 画像の前処理: Face++[6] を使用して, 収集した各画像から取得した顔の情報を元に画像の前処理を行う
3. VGG Face による特徴抽出
4. データの次元削減
5. 分類と評価

以下に, 各手順の詳細を示す.

3.1 画像収集

Bing Image Search API[5] を用いて, 以下を満たす選手の画像を収集する.

- 野球選手
NPB[10] に所属する日本人選手 (2017/12/22)
- サッカー選手
J1・J2・J3[11] に所属する日本人選手 (2017/12/22)

本演習での日本人選手とは, 出身地が日本国内かつ, 登録名にカタカナが含まれない選手のことを言う. 野球選手は 853 人, サッカー選手は 1,886 人, 合計 2,739 人の画像を 128,715 枚収集した.

3.2 画像の前処理

収集した画像には, 分類を行うにあたってノイズとなる画像が多く含まれる. そこで, 以下の手順で画像の前処理を行う.

1. 機械的な画像の除去

2. 顔領域の切り取り
 3. 手作業による画像の除去
 4. 野球選手とサッカー選手の人数を合わせる
- 1 では, Face++[6] を使用して取得した各画像の顔の情報から, 以下の条件を満たす画像を除去する.
- 検出した顔の数が 0 または複数
 - 瞳孔間の距離が 40 ピクセル未満
 - 顔が上下を 10° 以上 10° 以下向いている, または左右を 15° 以上 15° 以下向いている

次に除去後の画像のみから, 顔領域の切り取りを行う. 3 では手作業で画像を確認し, 以下の条件を満たす画像を除去する.

- 収集した人物と一致していない
- 顔の輪郭・眉毛が隠れている
- 顔の領域内に顔以外のもの (眼鏡・手・飲食物等) が写っている
- 野球選手・サッカー選手であることがわかる (ボール・グラブ・ヘルメット・帽子・帽子の影等の写り込み)

手作業による画像の除去を行った後, 野球選手とサッカー選手の人数を 1 歳単位で合わせる. その際に分類に使用する選手は, Python の random ライブラリを用いて選定する. 表 1 は各年齢ごとの選手の人数, 表 2 は各年齢ごとの画像の枚数を表す. 分類には, 野球選手・サッカー選手合わせて 1,252 人, 合計 3,242 枚の画像を使用する.

表 1: 前処理後の選手の人数

年齢	野球選手	サッカー選手
16～19	68	68
20～29	437	437
30～39	116	116
40～49	5	5
合計	626	626

表 2: 前処理後の画像の枚数

年齢	野球選手	サッカー選手
16～19	156	144
20～29	1,101	1,128
30～39	297	384
40～49	16	16
合計	1,570	1,672

3.3 VGG Face による特徴抽出

VGG Face に, 224×224 にリサイズした画像を入力として入れ, 中間層 (図 1: fc6) から 4096 次元の特徴を抽出する.



図 1: VGG Face

3.4 データの次元削減

VGG Face の中間層から抽出した 4096 次元の特徴の次元削減を行う. 4096 次元は次元数が大きいため, 分類の際に計算時間が大きくなってしまふ. 本演習では, 一般的な次元削減の手法である特異値分解 (Singular Value Decomposition: SVD) を使用し, 4096 次元から 500 次元への次元削減を行う.

3.5 分類と評価

分類にはロジスティック回帰 (Logistic Regression: LR) と SVC (Support Vector Classification) を使用し, 野球選手・サッカー選手の二値分類を行う.

- ロジスティック回帰
 - C: 0.0001
 - solvers: liblinear

- SVC

scikit-learn のデフォルト値を使用

評価指標は, AUC と accuracy を使用し, 20-Fold Cross Validation で評価を行う.

4 演習結果

分類結果を表 3 に示す. SVC よりもロジスティック回帰の方が精度が高く, ランダムな分類による精度 (AUC=0.5) を超えることができた.

表 3: 分類結果

	AUC	accuracy
LR	0.82	0.84
SVC	0.73	0.75

5 今後の課題

今後の課題は 3 点ある. 1 点目は, 手作業による画像の除去を複数人で行うことである. 本演習では報告者 1 人のみで行ったが, 複数人の第三者による除去が必要である. 2 点目は, 画像の背景・顔の表情・写真撮影時の年齢と現在の年齢との違い等が, 分類結果に影響を与えるか検証を行うことである. 3 点目は, 今回使用した顔画像を人間が分類した場合の精度と比較, また他の手法で分類した場合の精度と比較することである.

参考文献

- [1] Wang, Y., & Kosinski, M. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of Personality and Social Psychology*. (in press)
<https://osf.io/zn79k/>(参照: 2018/2/12)
- [2] Arthur Samuel(1959)
https://www.ibm.com/developerworks/community/blogs/jfp/entry/What_Is_Machine_Learning?lang=en(参照: 2018/2/12)
- [3] 村上研究室 ニューラルネットワーク - 画像処理・理解研究室 - 愛媛大学
<http://ipr20.cs.ehime-u.ac.jp/column/neural/index.html>(参照: 2018/2/12)
- [4] Viual Geometry Group
<http://www.robots.ox.ac.uk/~vgg/>(参照: 2018/2/12)
- [5] Bing Image Search API
<https://www.faceplusplus.com/>(参照: 2018/2/12)
- [6] Face++
<https://www.faceplusplus.com/>(参照: 2018/2/12)
- [7] Keras
<https://keras.io/ja/>(参照: 2018/2/12)

- [8] keras-vggface
<https://github.com/rcmalli/keras-vggface>(参照: 2018/2/12)
- [9] scikit-learn
<http://scikit-learn.org/stable/>(参照: 2018/2/12)
- [10] NPB
<http://npb.jp/>(参照: 2018/2/12)
- [11] J リーグ
<https://www.jleague.jp/>(参照: 2018/2/12)