

# 表情認識による集中状態判断

高澤奈生<sup>†1</sup> 坂本彩<sup>†1</sup> 大塚亜未<sup>†1</sup> 青柳達也<sup>†1</sup>

**概要:** 本論文では、転移学習のアプローチを用いて、顔画像から集中している表情を認識する手法について述べる。転移学習とは、ある分野で学習済みのモデルを、別の分野で再利用する技術であるが、今回は Microsoft Azure の Face API を学習済みのモデルとして利用している。Face API とは、ある一枚の画像をデータとして渡すと、そこに映っている人物らの年齢や性別、感情の度合いなど、顔に関する様々な属性を推測してくれる API である。この Face API のモデルを再利用することによって、「顔画像から集中している表情を識別する」問題を効率的に解決することを目指した。識別器の作成にあたっては、20代の女子大学生7名の「集中している状態」と「集中していない状態」の顔画像を収集した。また、識別器の精度を高めるため、モデル選択、ラベル付けの方法、特徴量選択、ハイパーパラメータ選択の4つの観点において実験を行った。その結果、 $C=1000$ 、 $\gamma=0.1$  の RBF カーネルを利用した SVM 識別器が 95.80% と最も精度が高くなることがわかった。そして、この識別器を利用し、ユーザの集中状態を判定するアプリを作成した。

## 1. はじめに

本論文では表情認識の中でも特に「集中」の表情認識について研究を進めた。「集中」の表情に論点を絞った理由は、「集中」の状態を自分自身で判定することが困難なためである。「喜び」や「悲しみ」の状態であれば、自分自身で感じることができる。しかし、集中している時に、「今現在、自分は集中しているな」と感じることは少ない。

では、人間の「集中」している状態を機械的に判定することは、我々の生活に如何なる利益をもたらすか。この点、「集中」状態の判定は、学習活動全般において役に立つ。「集中していない」と判定された場合には機械側から注意が寄せられるなどの形で、学習効果の向上が期待できる。

このような点を踏まえ、本研究ではユーザの集中状態を判定することを目指した。

## 2. 関連技術

関連技術として、パナソニック株式会社が開発した「感情・体調センシング」[1]、オムロン株式会社が開発した「ドライバー見守り車載センサー」[2]が挙げられる。

「感情・体調センシング」は、ユーザの表情と生体情報から、感情、眠気、体調を推定することができる技術である。「カメラ」と「サーモカメラ」の2種類のカメラが使われており、「カメラ」によって表情、まばたき、脈拍の情報を取得する一方、「サーモカメラ」によって皮膚温度、放熱量の情報を取得する。これらの情報を組み合わせ、ユーザの感情、眠気、体調を推定するものである。この技術は、表情分析によってユーザの状態を把握できるという点において、本論文で提案したアプリと類似している。しかし、ユーザの集中状態は判別しない。

「ドライバー見守り車載センサー」は、ドライバーの視線や顔の向きから、ドライバーの運転への集中度を判別するものである。マスクやサングラスを着用していても、目の開閉や視線の方向を高精度に識別できるという。この技術は、ユーザの集中状態を判別するという点においては、

本論文で提案したアプリと類似しているが、集中状態を判別する方法が異なっている。「ドライバー見守り車載センサー」は、ドライバーの「視線」が重要な判定要素になっているのに対して、本研究では、視線ではなく、ユーザの表情をもって集中状態を判別することを目標としている。机に向かって下を向いているユーザや、正面を向いて話者の話を聞いているユーザの集中状態も判別可能なアプリを作成した。

## 3. 研究の目的

「2. 関連研究」でも記述した通り、ユーザの視線に関わらず、ユーザの集中状態を判定することを本研究の目的とする。ユーザの表情をもって集中状態を判断したい。

## 4. データセットの作成

20代の女子大学生7名を被験者とし、集中している顔画像を1618枚、集中していない顔画像を4455枚、計6073枚のデータセットを作成した。

集中しているかどうかを、その対象への視線に関わらず、ユーザの表情によって判別することを目標としたため、以下のような複数の状況下でデータの収集を行った。

### 集中している状況

- ・スマートフォンのゲームをしている
- ・パソコンのゲームをしている
- ・パソコンで被験者が興味のある動画を見ている
- ・他者の話（被験者が興味のある内容）を聞いている

### 退屈している状況

- ・パソコンで被験者が興味のない動画を見ている
- ・他者の話（被験者が興味のない内容）を聞いている

## 5. 識別器の作成

本研究では識別器の作成にあたり、転移学習的なアプローチをしていく。

識別器に渡す特徴量としては、Microsoft Azure の「Face API」から得られる、顔に関するデータを利用した。一般的には、SIFT 特徴量や SURF 特徴量などの画像特徴量を抽出するが、今回は「Face API」から得られるデータを利用することとした。「Face API」とは、ある一枚の画像をデータとして渡すと、そこに映っている人物らの年齢や性別、感情の度合いなど、顔に関する様々なデータを推測してくれる API である。転移学習的なアプローチとして「Face API」のモデルを再利用し、より効率的に集中状態を判定する。このデータを特徴量として識別器に渡し(表 1)、顔画像を「集中している」「集中していない」の 2 つに分類できるように学習させた。

表 1 識別器に使用するデータ

情報名	説明
age	年齢.
smile	笑顔の度合い.
roll ( headPose )	顔を正面から見たときの傾き.
yaw ( headPose )	顔の左右方向への向き.
anger ( emotion )	怒りの度合い.
contempt ( emotion )	軽蔑の度合い.
disgust ( emotion )	嫌悪感の度合い.
fear ( emotion )	恐怖の度合い.
happiness ( emotion )	喜びの度合い.
neutral ( emotion )	中立の度合い.
sadness ( emotion )	悲しみの度合い.
surprise ( emotion )	驚きの度合い.
blur	画像のぼやけ度合い.
exposure	顔の露出度合い.
noise	画像のノイズ度合い.

このとき、収集した顔画像の枚数については、「集中している」顔画像が全体の 26.6 %、「集中していない」顔画像が全体の 73.4 %を占めており、データに偏りがある。そのため、クラスに重み付けを行い、少数である「集中している」顔画像が与える影響を大きくする。具体的には、scikit-learn の SVC モデルの「class\_weight」を「balanced」に設定した。

また、全画像に対して「集中している」「集中していない」のラベル付けを人間が行う。被験者が集中しているように見える時の顔画像を「集中している」、被験者が退屈しているように見える時や笑っている時の顔画像を「集中していない」と定義し、正解ラベルとした。

### 5.1 人間によるラベル付け

人間が目視でラベル付けをすることに関しては、

Whitehill らが論じている[3]。Whitehill らは、Engagement について論じているが、ここでは「Engagement」「没頭していること」を「集中していること」と同義であるとして考える。Whitehill らはまず、同じ画像に対して、複数の人間によるラベル付けが一致するかどうか調べた。その画像が「集中している」か「集中していない」かの二値分類を複数の人間が行った結果、コーエンのカップ係数は 0.96 と高い一致度になった。さらに詳しく 4 つの区分に被験者の集中状態を分類した場合でも、カップ係数は 0.56 と高い一致度であったため、この評価方法は信頼性があると判断される。

次に Whitehill らは「人間によるラベル付け」と「テストのスコア」を比較した。「人間によるラベル付け」としては「1」「2」「3」「4」の 4 つの値が利用された。1 に近ければ近いほど集中しておらず、4 に近ければ近いほど集中しているように見える、としてラベル付けを行った。「テストのスコア」としては「Set」というカードゲームのスコアを利用している。被験者が「Set」を 3 分間行い、そのゲームのスコアを「テストのスコア」としている。

この実験における「人間によるラベル付け」と「テストのスコア」の相関係数は、ラベルが「1」の場合は-0.39、「2」の場合は-0.32、「3」の場合は-0.34、「4」の場合は 0.57 となった。つまり、被験者が強い集中状態にある場合、人間によるラベル付けが適切だということが、この実験結果により証明された (Whitehill et al., 2014)。

### 5.2 モデル選択

機械学習には様々なモデルが存在する。顔に関するデータを基に「集中している」「集中していない」を判定するためにはどのモデルが適当なのか、複数のモデルについて検証した。今回使用したモデルは、「ロジスティック回帰」、「サポートベクターマシン(SVM)」、「ランダムフォレスト」、そして前述した 3 つのモデルから構成される「アンサンブル学習」を利用したモデルの計 4 つである。どのモデルにおいても、全データを 5 分割したクロスバリデーションを行い、識別率を求めた。

表 2 ハイパーパラメータによる識別率の違い

カーネル名	スコアの平均値
ロジスティック回帰	0.8264
SVM	0.9682
ランダムフォレスト	0.9678
アンサンブル学習	0.9662

以上のことから、識別率の高い「SVM」を本研究においては使うこととする。

### 5.3 ハイパーパラメータ選択

より精度が高くなるように、適切なハイパーパラメータを選択する。今回は SVM を使うこととしたため、どのカー

ネルを使うか、そのカーネルに付随するハイパーパラメータはいくつが適切か、グリッドサーチ（全探索）を用いて決める。使用するカーネルは「線形カーネル」「RBFカーネル」「多項式カーネル」の3つである。

今回、「線形カーネル」のCの値については、10, 100, 1000, 「RBFカーネル」のCの値については、10, 100, 1000,  $\gamma$ の値については0.01, 0.1, 1, 「多項式カーネル」のCの値については、10, 100, 次数は2, 5,  $\gamma$ の値については0.01, 0.1, 1の中から最も適切なカーネルとハイパーパラメータを選択する。

以下に、グリッドサーチを行った結果を示す。

```
0.659 for {'svm_C': 10, 'svm_kernel': 'linear'}
0.658 for {'svm_C': 100, 'svm_kernel': 'linear'}
0.658 for {'svm_C': 1000, 'svm_kernel': 'linear'}
0.694 for {'svm_C': 10, 'svm_kernel': 'rbf', 'svm_gamma': 0.01}
0.744 for {'svm_C': 10, 'svm_kernel': 'rbf', 'svm_gamma': 0.1}
0.761 for {'svm_C': 10, 'svm_kernel': 'rbf', 'svm_gamma': 1}
0.727 for {'svm_C': 100, 'svm_kernel': 'rbf', 'svm_gamma': 0.01}
0.757 for {'svm_C': 100, 'svm_kernel': 'rbf', 'svm_gamma': 0.1}
0.760 for {'svm_C': 100, 'svm_kernel': 'rbf', 'svm_gamma': 1}
0.752 for {'svm_C': 1000, 'svm_kernel': 'rbf', 'svm_gamma': 0.01}
0.769 for {'svm_C': 1000, 'svm_kernel': 'rbf', 'svm_gamma': 0.1}
0.760 for {'svm_C': 1000, 'svm_kernel': 'rbf', 'svm_gamma': 1}
0.700 for {'svm_degree': 2, 'svm_C': 10, 'svm_kernel': 'poly', 'svm_gamma': 0.01}
0.705 for {'svm_degree': 2, 'svm_C': 10, 'svm_kernel': 'poly', 'svm_gamma': 0.1}
0.723 for {'svm_degree': 2, 'svm_C': 10, 'svm_kernel': 'poly', 'svm_gamma': 1}
0.733 for {'svm_degree': 5, 'svm_C': 10, 'svm_kernel': 'poly', 'svm_gamma': 0.01}
0.732 for {'svm_degree': 5, 'svm_C': 10, 'svm_kernel': 'poly', 'svm_gamma': 0.1}
0.726 for {'svm_degree': 5, 'svm_C': 10, 'svm_kernel': 'poly', 'svm_gamma': 1}
0.702 for {'svm_degree': 2, 'svm_C': 100, 'svm_kernel': 'poly', 'svm_gamma': 0.01}
0.716 for {'svm_degree': 2, 'svm_C': 100, 'svm_kernel': 'poly', 'svm_gamma': 0.1}
0.724 for {'svm_degree': 2, 'svm_C': 100, 'svm_kernel': 'poly', 'svm_gamma': 1}
0.732 for {'svm_degree': 5, 'svm_C': 100, 'svm_kernel': 'poly', 'svm_gamma': 0.01}
0.709 for {'svm_degree': 5, 'svm_C': 100, 'svm_kernel': 'poly', 'svm_gamma': 0.1}
0.726 for {'svm_degree': 5, 'svm_C': 100, 'svm_kernel': 'poly', 'svm_gamma': 1}
```

図1 ハイパーパラメータによる識別率の違い

グリッドサーチの結果、C=1000,  $\gamma=0.1$ の「RBFカーネル」が最も精度が高くなることがわかった。

## 6. 結果

以上のことから識別器を調整した結果、 $\gamma=0.1$ , C=1000のRBFカーネルを利用したSVM識別器の識別率は91.23%となった。

## 7. アプリの概要

手軽に識別器を使用することができるよう、スマートフォンのアプリを開発した。スマートフォンのカメラで顔を撮影すると、識別器によって集中状態か判断され、顔の上にその結果が表示される(図1)。



図2.実際のスマートフォンの画面

## 8. おわりに

本論文では、表情認識を用いた集中状態を判断するため、集中している状態と集中していない状態の顔画像を収集し、識別器を作成した。

結果として、 $\gamma=0.1$ , C=1000, のRBFカーネルを利用した識別率は95.80%となり、視線に関わらずユーザの表情をもって集中状態を判別したいという当初の目標は達成できたと考えられる。しかし、被験者が集中している状態なのに「集中していない」と判定されることが多々あった。それらの顔画像を見たところ、半目になっていて、退屈そうに見える画像も存在した。これらの画像は、瞬きの最中であるために半目になっているだけなのだが、現状、被験者が瞬きをしているかどうか判別することはできない。そのため、瞬きの前後の画像は使用しないようにすることを今後の課題としたい。

また、使用しているデータセットにも偏りがあり、被験者によって、収集した画像の枚数に大きな差がある。そのため、今後はより多くの被験者に協力してもらい、全員が同程度の枚数になるよう調整したい。

## 参考文献

- [1] “感情・体調センシング”, パナソニック,  
[https://industrial.panasonic.com/jp/products-ex/ceatec2017co/sensing\(参照2018-1-20\)](https://industrial.panasonic.com/jp/products-ex/ceatec2017co/sensing(参照2018-1-20))
- [2] 世界初「ドライバ―見守り車載センサー」を開発, OMRON,  
[https://www.omron.co.jp/press/2017/09/c0927.html\(参照2019/01/10\)](https://www.omron.co.jp/press/2017/09/c0927.html(参照2019/01/10))
- [3] Jacob Whitehill, Zewelanj Serpell, Yi-Ching Lin, Aysha Foster, and Javier R. Movellan. (2014) “The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions”.
- [4] Aurelien Geron, 「Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems」Oreilly & Associates Inc, 2017.
- [5] Face, Microsoft Azure,  
[https://azure.microsoft.com/ja-jp/services/cognitive-services/face/\(参照2019/01/10\)](https://azure.microsoft.com/ja-jp/services/cognitive-services/face/(参照2019/01/10))