

顔画像からの集中状態判定

高澤奈生ⁱ 大塚亜未ⁱⁱ 青柳龍也ⁱ

概要： 本論文は、ある人物の顔画像から得られる情報を利用して、機械学習でその人物の集中状態を判定することを目的としたものである。使用したデータセットは 20 代の女子大学生 6 名の「集中している状態」と「集中していない状態」の時の顔画像である。2 つの手法でこれらの顔画像の集中状態を判定し、それら手法について比較した。1 つめの手法は、dlib から得られる顔特徴点の座標を特徴量とする手法であり、2 つめの手法は、Microsoft Azure の Face API から得られる表情の表出度合いを特徴量とする手法である。

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。

集中状態を判定するための情報としては、dlib から得られる顔特徴点データと Microsoft Azure の「Face API」から得られる、感情などの顔に関するデータを使用する。これらのデータを使って識別器を訓練し、どちらのデータを使う手法の方が精度が高くなるのか、比較検討した。

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したい。

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。

4. データセットの作成

20 代の女子大学生 6 名を被験者とし、集中している顔画像を 1186 枚、集中していない顔画像を 1186 枚、計 2372 枚のデータセットを作成した。

表 1 ユーザごとのデータ数

被験者id	集中	集中以外
0	249	249
1	413	413
2	246	246
3	97	97
4	115	115
5	66	66
合計	1186	1186

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

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った。被験者が集中しているよう

ⁱ 津田塾大学 大学院 理学研究科
ⁱⁱ 津田塾大学

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。

4.1 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. 識別器の作成

本研究では識別器の作成にあたり、dlib から得られる顔特徴点データと Microsoft Azure の Face API から得られる、感情などの顔に関するデータを特徴量として使用する。そのため、これら 2 つのデータのための識別器をそれぞれ作成した。

5.1 dlib による顔特徴点データ

dlib から得られる顔特徴点は、以下の画像 (図 1) のとおり全 68 点である。この 68 点の顔特徴点の xy 座標を 0 から 1 の間に収まるような値に正規化した後、これを使用して識別器を作成した。

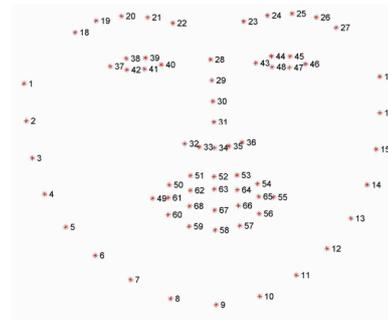


図 1 dlib から得られる顔特徴点[4]

5.2 Face

API による感情などの顔に関するデータ

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。「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(表 1)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。

表 2 Face API から得られるデータ (感情)

情報名	説明
anger	怒りの度合い。
contempt	軽蔑の度合い。
disgust	嫌悪感の度合い。
fear	恐怖の度合い。
happiness	喜びの度合い。
neutral	中立の度合い。
sadness	悲しみの度合い。
surprise	驚きの度合い。
smile	笑顔の度合い。

Face API から得られるデータ全 26 個のうち、どのデータを使用するか、3 つのケースに分けて検証した。よって、以下の 5 ケースにおいてそれぞれ識別器を作成した。

表 3 作成した特徴量のケース

	使用するデータ	説明変数の大きさ	データ取得元
ケース 1	顔に関する全データ	23	Face API
ケース 2	感情+頭部の姿勢+年齢	12	Face API
ケース 3	感情+頭部の姿勢	11	Face API
ケース 4	感情	9	Face API
ケース 5	顔特徴点	136	dlib

このデータを特徴量として識別器に渡し(表 2), 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。

6. 結果と考察

ランダムフォレストを使用し、正答率, 再現率, 適合率,

F 値を求めた。ランダムフォレストと SVM で識別器を作成したが、どちらも同じような傾向だったため、ここではランダムフォレストの結果を示す。識別器を訓練する方法としては 2 通り試した。1 つめは全被験者のデータを使用し、5 分割のクロスバリデーションを行う方法であり、2 つめはある一人の被験者のデータをテストデータとし、残りの被験者のデータを訓練データとして訓練する方法である。2 つめの方法で得られた正答率、再現率、適合率、F 値は全被験者の平均を取った。

表 4 使用した特徴量ごとの正答率

	方法 1	方法 2
ケース 1	0.775	0.591
ケース 2	0.764	0.556
ケース 3	0.753	0.495
ケース 4	0.712	0.537
ケース 5	0.900	0.470

表 5 使用した特徴量ごとの再現率

	方法 1	方法 2
ケース 1	0.796	0.645
ケース 2	0.783	0.552
ケース 3	0.781	0.505
ケース 4	0.795	0.592
ケース 5	0.910	0.291

表 6 使用した特徴量ごとの適合率

	方法 1	方法 2
ケース 1	0.765	0.582
ケース 2	0.754	0.584
ケース 3	0.740	0.554
ケース 4	0.682	0.547
ケース 5	0.891	0.443

表 7 使用した特徴量ごとの F 値

	方法 1	方法 2
ケース 1	0.780	0.543
ケース 2	0.768	0.502
ケース 3	0.760	0.466
ケース 4	0.734	0.536
ケース 5	0.901	0.308

以上のことから、方法 1 においては、dlib から得られる顔特徴点を使用したケース 5 が、どの数値も高くなることがわかった。ここで、ケース 5 の特徴量を使用した場合の特徴量の重要度を調べると、輪郭の顔特徴点が最も識別に影響を与えていることがわかった (図 3)。

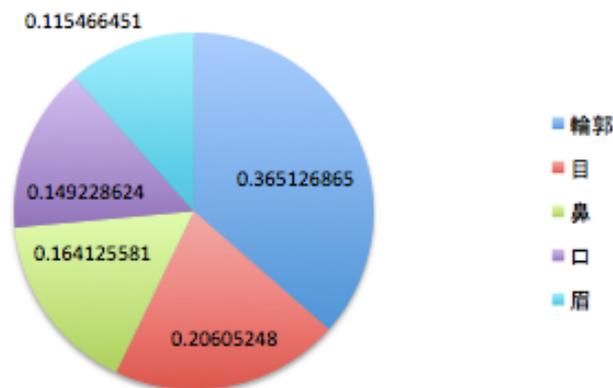


図 3 各顔特徴点の重要度

確かに、俯いている被験者は集中していないように見えることが多いが、輪郭のデータが全体の 37% の重要度を占めるとは考えづらい。また、正答率、再現率、適合率、F 値のどの値においても、方法 1 と方法 2 を比較すると、ケース 5 の値が最も精度が下がっている。このため、ケース 5 の識別器は過学習してしまった可能性が高い。

Face API から得られるデータを使用する、ケース 1・2・3・4 では、ケース 3・4 と比べると、正答率・適合率においては、ケース 1・2 の方が値が高かった。ここで、ケース 1 の特徴量を使用した場合の特徴量の重要度を調べると (図 4)、中立のデータが 0.2624 と最も高く、2 番目はロール (顔を正面から見たときの傾き) のデータが 0.139、ついで年齢のデータが 0.1165 となった。

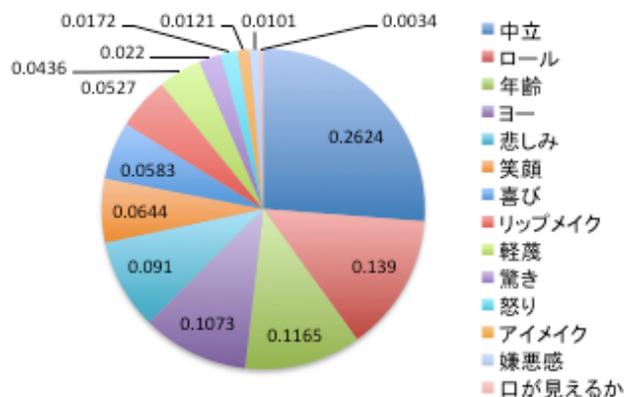


図 4 顔に関するデータの重要度

このことから、ケース 1・2 のデータの方が正答率・適合率において値が高いのは、被験者を特定できるデータが、特徴量の中に含まれているためだと考えられる。「推定年齢」などの要素が識別に影響を与えている可能性がある。

7. おわりに

dlib から得られる顔特徴点を使用した識別器が最も精度が高かったが、方法 1 と方法 2 を比較してみると、ケース 5 が最も精度の差が顕著だった。ついで、ケース 1・2 などの被験者を特定できるようなデータが含まれているデータを使用した場合もスコアが高かった。これは、過学習して

しまっている可能性が高いと考える。そのため、データが汎化してある、ケース3・4のデータを使用した識別が正しい識別であるといえる。このことから、dlib から得られる顔特徴点データを使用するよりも、データが汎化されるFace APIの顔情報データを使用すべきだと考える。

今後の課題としては、データをさらに収集することが挙げられる。現時点では被験者ごとのデータの偏りや、データ数の少なさが問題となっているので、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。また、画像を直接特徴量としてCNNなどのディープラーニングを使用して集中状態が判定できないか試していきたい。

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 2019/12/16)
- [2] 世界初「ドライバー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 2019/12/16)
- [3] Jacob Whitehill, Zewelanjhi Serpell, Yi-Ching Lin, Aysha Foster, and Javier R. Movellan. (2014) “The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions”.
- [4] Christos Sagonas, Georgios Tzimiropoulos, Stefanos Zafeiriou and Maja Pantic. (2013) “300 Faces in-the-Wild Challenge: The first facial landmark localization Challenge”.
- [5] Aurelien Geron, 「Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems」Oreilly & Associates Inc, 2017.
- [6] Face, Microsoft Azure,
<https://azure.microsoft.com/ja-jp/services/cognitive-services/face/>(参照 2019/12/16)