

ディープラーニングを利用した顔画像による着席状態の集中度解析

足利大学 大学院 工学研究科 情報・生産工学専攻 平石研究室 大学院 2 年

Shen Jinjian (シン キンケン)

【概要】本研究では、利用者の顔画像を元にディープラーニングを利用した着席者の集中度の解析を試みる。その際に、利用者の集中度の基準となるデータとして、簡易脳波センサと座圧センサを利用した。簡易脳波センサによって利用者の集中や散漫の状態を検出することが可能であり、座圧センサによって利用者の疲労の状態を検出することができる。本研究では、着席者の顔画像を収集し、各センサによって得られた集中度との関連性の学習を行なった。

1. はじめに

近年、コロナの影響で対面授業からオンライン授業の割合が多くなっている。オンライン授業中の学生の状態の把握は難しいのが現状である。集中や散漫等の状態を確認できれば、授業内容で補足説明が必要な場合や、学生の興味の状態などを把握することが可能となり、授業を効率的に実施することができるものと考えられる。これまで、座圧センサ[1]や脳波センサ[2,3]を利用して、着席者の疲労状態や集中状態の解析が行われている。本研究では、座圧センサや脳波センサなどの特別な装置を利用せず、顔画像のみから、着席者の状態を把握する方法を検討する。本研究では、脳波センサと座圧センサから得られる集中度や疲労度を基準として、着席者の顔画像との関連性の学習を行なった。

2. 簡易型脳波センサと座圧センサ

本研究では簡易型脳波センサとして NeuroSky 社によって開発された MindWave Mobile を利用した。このセンサでは、8 種類の脳波データを収集することができる。また、独自のモジュールにより、集中度とリラックス度を 0~100 の数値で取得することが可能である。それぞれのデータは1秒毎に収集することが可能である。座圧センサとして SR ソフトビジョンを利用した。この座圧センサでは、専用のソフトウェアを利用して圧力分布に加え、面圧中心位置を収集することができる。データを収集する頻度は自由に設定することが可能である。

3. 顔画像と視線

本研究では、顔認識と着席者の視線計測のために、Google によって開発された FaceNet アルゴリズムを基にした OpenFace を利用した。OpenFace では顔の立体的情報を取得し、眼、鼻、あごの輪郭、および、視線データを毎秒 30 枚顔画像収集することができる。図 1 は OpenFace の実行の様子である。

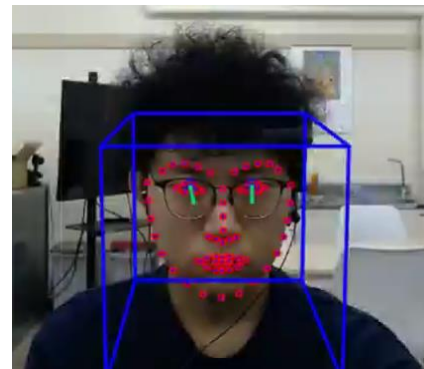


図 1: OpenFace の実行の様子

4. データ計測実験 A

研究 A では、脳波データと座圧データの関連性についての検証実験について説明する。本研究では1人の被験者(25 歳, 男子留学生)に対して 1 回の実験を行なった。被験者は脳波センサを装着し、椅子の上に座圧センサを敷いた状態で、30 分のゲーム(任天堂 switch の monster hunter rise)を実行してもらった。その時の座圧センサのデータと脳波データを収集した。座圧データに関しては 5 秒に 1 回のデータを収集するように設定し、脳波データも 5 秒毎にデータを保存するように設定した。

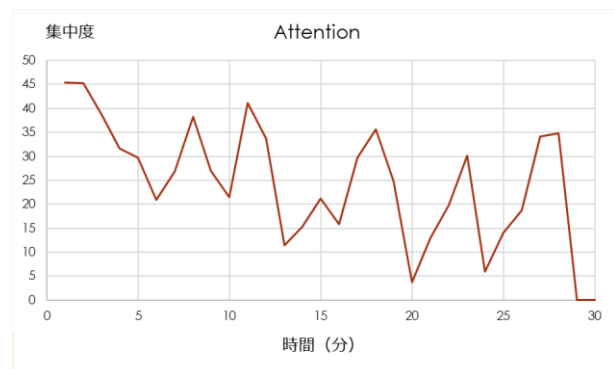


図 2: 集中度の変化の様子



図 3: 座圧データによる重心座標の標準偏差の変化の様子

図 2 は脳波センサによる集中度の変化の様子であり、図 3 は座圧センサによる重心座標の一分毎の標準偏差の変化の様子を示している。これらの結果を比較すると、時間が経過するにしたがって、脳波による集中度は下がる傾向があり、それに伴って、座圧データの値も大きなものが見られ、後半になるにしたがって疲れが見られる。

4.1 まとめ

実験結果より、脳波データによる集中度が下がると同時に、座圧センサによる疲労度は上がる傾向を確認することができた。

5. データ計測実験 B

研究 B では 1 人の被験者(26 歳, 男子留学生)に対して 2 回の実験を行なった。被験者は脳波センサを装着し、椅子の上に座圧センサを敷いた状態で、30 分のドラマ(アメリカン・ホラー・ストーリー)を視聴してもらい、その時のセンサデータと顔画像を収集した。座圧データに関しては 5 秒に 1 回のデータを収集するように設定し、顔画像も 5 秒毎にデータを保存するように設定した。図 4 は 1 回目の実験の結果であり、図 5 は 2 回目の実験の結果を示している。それぞれ上のグラフは脳波データによる集中度であり、下のグラフは座圧データの 1 分毎の標準偏差の値を示している。座圧データに関しては、値が大きいほど疲れていることを意味する[1]。

図 4 の 1 回目の実験結果を見ると、時間が経過するにしたがって、脳波による集中度は下がる傾向があり、それに伴って、座圧データの値も大きなものが見られ、疲れる傾向が見られる。一方、図 5 の 2 回目の実験結果では、1 回目とは逆に、徐々に集中度が高くなる傾向が見られ、また、座圧データの値は、徐々に小さくなる傾向がみられる。つまり、1 回目の実験では、最初に集中し、徐々に疲れて集中が減少するパターンであり、2 回目の実験では、最初はあまり集中できず、徐々に集中していくパターンである。

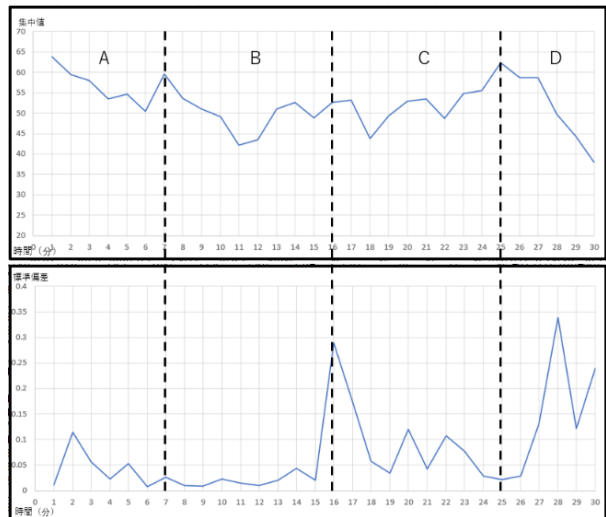


図 4: 1 回目の脳波データ(上)と座圧データ(下)

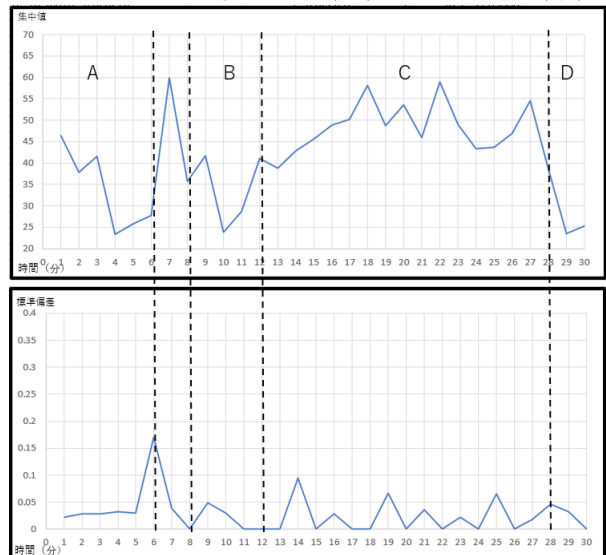


図 5: 2 回目の脳波データ(上)と座圧データ(下)

次に、脳波データと座圧データを元に、収集した顔画像を、ディープラーニングによる分析を行うために、集中状態と散漫状態に分類する。ここで、2 回目の実験において、A 区間(0-7 分)は集中していると同時に疲れてない状態であり、B 区間(7-16 分)はやや疲れが見られる状態である。C 区間(16-25 分)は集中しているが疲れた状態であり、D 区間(25-30 分)は、集中も途切れ、疲れも見られる状態を表している。したがって、A 区間は疲れもなく完全に集中した状態であり、逆に、D 区間は完全に疲れていて集中が途切れ状態であるため、A 区間の顔画像データ(84 枚)を集中状態とし、D 区間の顔画像データ(60 枚)を散漫状態として分類を行なった。

2 回目の実験では、C 区間(12-28 分)は集中していると同時に疲れてない状態であり、A(0-6 分)、B(8-12 分)、D(28-30 分)区間は、集中も途切れ、疲れも見られる状態を表しているため、C 区間の顔画

像データ(192 枚)を集中状態とし, A, B, D 区間の顔画像データ(108 枚)を散漫状態として分類した.

5.1 ディープラーニングによる分析

本研究では機械学習ソフトウェアの Weka を利用して, 1回目の実験の顔画像を利用して分類モデルの生成を行ない, 2回目のデータを利用して, その分類モデルの評価を実施した. Weka では, 様々なディープラーニングのモデルを利用することができるが, 本研究では, 実行可能であった3つのモデルを利用して評価を実施した.

表1は, それぞれのモデルの分類精度を示している. LeNet は非常に精度が高く, FaceNet と SqueezeNet の精度は低い結果となった. FaceNet と SqueezeNet の精度の低い原因は, 識別の特徴量が多いため, データ量が少なくいために, 学習の実行が十分でなかったことが考えられる.

表1:実験結果

モデル	分類精度
LeNet	99.12
FaceNet	57.68
SqueezeNet	52.62

5.2 まとめ

実験中の着席者の集中度の状態を計測することを目指して, 脳波センサと座圧センサを利用し, 顔画像の分析を行なった. 今回の実験において, 精度が低い原因は以下の3つの要因が考えられる.

本研究では, 実験の容易さから 30 分のドラマを視聴する実験を行なった. しかし, オンライン授業等を考えた場合には 30 分は短い. 今後は実際のオンライン授業を想定し, 90 分程度のデータを分析を実施する. また, 実験における実験者の数を増やし, 集中度の識別を目標として実験をする.

6. データ計測実験 C

研究 C では 2 人の被験者(A: 26 歳男子留学生, B: 24 歳男子留学生)に対して1回の実験を行なった. 被験者は脳波センサを装着し, 椅子の上に座圧センサを敷いた状態で, 40 分の授業(中国の近代歴史)を視聴してもらい, その時のセンサデータと顔画像を収集した. 座圧データ(面圧中心位置)に関しては1秒に1回のデータを収集するように設定し, 顔画像も1秒毎にデータを保存するように設定した.

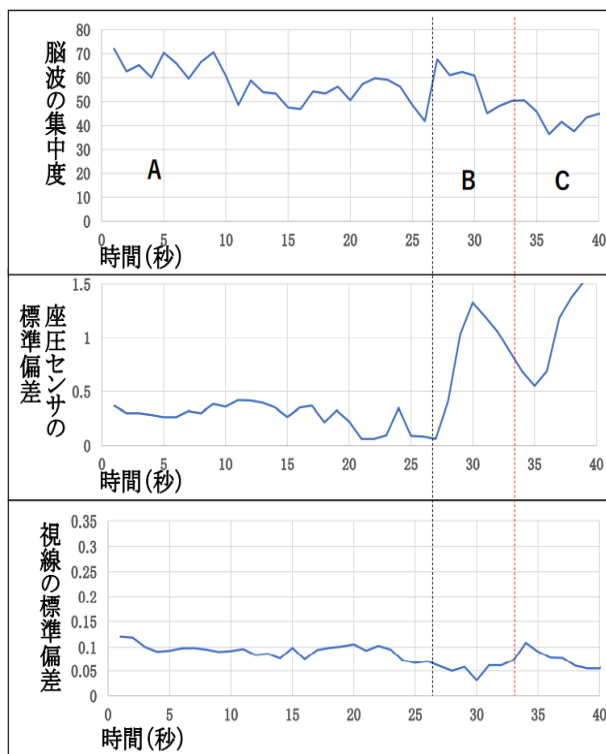


図 4: 被験者 A の脳波データ(上)座圧データ(中)と視線データ(下)

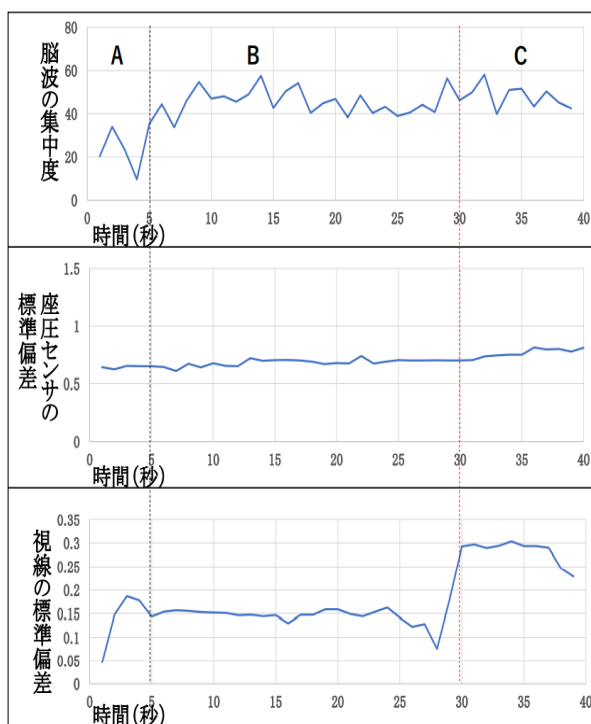


図 5: 被験者 B の脳波データ(上)座圧データ(中)と視線データ(下)

図 4 は被験者 A の実験の結果であり, 図 5 は被験者 B の実験の結果を示している. それぞれ上段のグラフは脳波データによる集中度, 中段のグラフは座

座圧データの1分毎の標準偏差の値を、下段のグラフは顔画像の視線データ(視線の座標)の1分毎の標準偏差の値である。座圧データに関しては、値が大きいほど疲れていることを意味する[1]。図4の結果を見ると、時間が経過するにしたがって、脳波による集中度と視線の標準偏差は下がる傾向があり、それに伴って座圧データは大きな値が見られ、疲れる傾向が見られる。一方、図5の結果では、被験者Aとは逆に、徐々に集中度と視線の値が高くなる傾向が見られ、座圧データは、さほど変化は見られない。つまり、被験者Aは、最初に集中し、徐々に疲れて集中が減少するパターンであり、被験者Bは、最初はあまり集中できず、徐々に集中していくパターンである。

6.1 学習結果

次に、脳波データと座圧データを元に、収集した顔画像を、ディープラーニングによる分析を行うために、集中状態と散漫状態に分類した。ここで被験者Aにおいて、A区間(0-27分)は集中していると同時に疲れてない状態であり、B区間(27-34分)はやや疲れが見られる状態である。C区間(34-40分)は、集中も途切れ、疲れた状態を表している。したがって、A区間は疲れもなく完全に集中した状態であり、逆に、C区間は完全に疲れていて集中が途切れた状態であるため、A区間の顔画像データ(1208枚)を集中状態とし、C区間の顔画像データ(420枚)を散漫状態として分類を行なった。被験者Bでは、BとC区間(5-40分)は集中していると同時に疲れてない状態であり、A(0-5分)区間は、集中できず、疲れてない状態を表しているため、BとC区間の顔画像データ(2100枚)を集中状態とし、A区間の顔画像データ(300枚)を散漫状態として分類した。

まず、研究Bと同様に脳波データと座圧データによる分類をもとに、ディープラーニングによる顔画像の分類を行なった。ここで前章の結果からLeNetの精度がもっとも良かったため、LeNetのみ分析を行なった。分類精度は82.427%であった。

表1: Lenetの混合行列

	集中	散漫
集中	3299(TP)	2(FN)
散漫	704(FP)	23(TN)

しかし、表2に示した混合行列を見ると、ほとんどの画像が集中として判断されており、散漫を散漫と判断

できた割合が23/727で、3.16%とかなり低い結果となっていた。したがって、今回の実験においてLeNetによる学習は、うまく機能していない結果となった。

6.2 視線データによる分析

ここでは、前節のように顔画像を分類するのではなく、視線データの分類を行なった。画像データは1秒ごとのデータであったが、視線データは1秒間で30データの保存されている。分類するデータが画像データではないため、ここでは機械学習手法としてランダムフォレストを利用した。そして、10-fold cross validationによる評価を実施した。その結果、分類精度は94.87%であった。表3は混合行列を示している。前節の結果では、散漫の状態の分類ができなかったが、今回の結果では8441/11607で72.7%であり、ランダムフォレストによる学習は、うまく機能したものと考えられる。

表3: ランダムフォレストの混合行列

	集中	散漫
集中	90585(TP)	2182(FN)
散漫	3166(FP)	8441(TN)

7. まとめ

本研究では、脳波センサと座圧センサの値を基準として、顔画像と視線の分析を行なった。顔画像のLeNetの機械学習の精度82.427%であるが、散漫の画像の分類は3.16%とかなり低い結果となり、LeNetによる学習はうまく機能しなかった。しかしながら、視線データを利用したランダムフォレストによる分類精度94.87%と高く、散漫の状態の分類も72.7%と比較的高い精度での分類が可能であった。

8. 参考文献

- [1] Li Yuchuan, 座圧センサを利用した着席者の疲労状態の解析, 第17回「学生&企業研究発表会」代替事業, 2020年度大学コンソーシアムとちぎ 研究発表要旨集, pp.74-75, 2021.1.
- [2] Shen jinjian 平石広典, ディープラーニングを利用した顔画像による着席状態の集中度解析, 情報処理学会第84回全国大会 IPSJ-Z84-1S-06, pp.389-390, 2022.3.
- [3] 渡辺 裕介 吉野 碧 菱山 玲子, 集中度推定システムによる社会的な手抜き抑制効果に関する分析 第80回全国大会講演論文集 IPSJ-Z80-3V-07, pp.389-390, 2018.3.