

読書行動識別のための Self-supervised Learning の実験的検討

坂本 周司[†] 山田 良博[†] 岩村 雅一[†] 岩田 基[†] 黄瀬 浩一[†]

[†] 大阪府立大学大学院工学研究科
〒599-8531 大阪府堺市中区学園町 1-1

E-mail: {sakamoto,yamada}@m.cs.osakafu-u.ac.jp, {masa,iwata,kise}@cs.osakafu-u.ac.jp

あらまし 近年、眼や頭の動きを用いて読書行動を識別する研究が行われている。特徴量として、眼電位 (EOG) や加速度、角速度の平均値や分散値が用いられている。しかし、統計量を用いる方法では読書行動識別に有効な特徴を十分に表現できず、識別精度を改善できないという問題がある。一方で、Deep Learning を用いて有効な特徴を得るには大量のラベル付きデータが必要となるが、そのようなデータを得ることは一般に困難であるという問題もある。本研究では、これらの問題を解決するため、Self-supervised Learning を用いる手法を提案する。これにより、少量のラベル付きデータで高精度な識別が期待できる。本稿では、Self-supervised Learning として、EOG と加速度、角速度に施した変形の種類を識別することで有効な特徴表現を学習する。学習した特徴表現を用いて、読書行動識別を行う。実験の結果、実験参加者 10 人の読書行動の 4 クラス平均識別精度は 57.3% であった。

キーワード Self-supervised Learning, CNN, 読書行動, JINS MEME

1. まえがき

近年、日々の読書行動を解析し、読書量を自動的に記録するという研究が行われている。記録した読書量を可視化し、学習者にフィードバックすることによって、自らの読書習慣を見直すことが可能となる。これにより、読書に対するモチベーションの向上が見込まれ、読書量の向上が期待される。

読書行動を解析する方法として、眼や頭の動きを用いる方法がある。眼や頭の動きを測定するために、JINS MEME と呼ばれる眼鏡型ウェアラブルデバイスが用いられている。JINS MEME には眼電位 (EOG) センサ、加速度センサ、角速度センサが搭載されており、特徴量として各センサから得られるデータの統計量を算出することで、読書行動識別が行われている [1]。しかし、統計量の特徴量として用いることで読書パターンに関する情報が十分に得られず、識別精度が不十分であるという問題がある。

この問題を解決する特徴抽出法として Deep Learning を用いる方法が考えられる。しかし、Deep Learning を用いるには、読書行動に関するラベルを付与した大量の生体データが必要となる。さらに、学習者が自ら手動で生体データに対してラベルを付与しなければならず、非常に手間がかかり容易ではない。

本研究では、これらの 2 つの問題点を解決するため、Self-supervised Learning を用いた手法を提案する。Self-supervised Learning では、自動で付与できるラベルを用いて生体データの特徴表現を行う。これにより、ラベル付けされた生体データが少量であっても高精度な読書行動識別が期待できる。本研究では、EOG、加速度、角速度に拡大や伸縮といった変形を施し、どの変形が施されているかを推定するというタスクを解くこと

により、読書パターンに関する特徴表現を学習できるのではないかと考える。学習した特徴表現を読書行動識別に用いることで、精度の向上を目指す。

本研究では、EOG に 7 種類の変形、加速度、角速度に 8 種類の変形を施し、どの変形が施されているか、また変形が施されていないかを識別することによって有効な特徴表現を学習する。その後、学習した特徴表現を用いて読書行動の認識モデルを作成する。実験の結果、実験参加者 10 人に対する EOG に施された変形の 8 クラス平均分類精度、加速度、角速度に施された変形の 9 クラス平均分類精度はそれぞれ 82.3%、95.9%、95.9% であり、実験参加者 10 名の 4 種類の読書行動の平均識別精度は 57.3% であった。結果から、従来手法と比較して提案手法は高精度な読書行動識別を達成していることを確認した。なお、本研究は、大阪府立大学工学研究科倫理委員会の承認を得ている事を付記しておく。

2. 関連研究

本章では、読書行動識別に関する研究と Self-supervised Learning に関する研究について述べる。

2.1 読書行動識別に関する研究

読書行動識別には、眼の動きを計測するデバイスとしてアイトラッカや JINS MEME が用いられている。アイトラッカを用いて読書行動識別を行った研究に、Kelton らの研究 [2]、Landsmann らの研究 [3] がある。Kelton らは、ニュース記事の読み方に対して、通常の読み方であるか、ざっと読んでいるかを分類している。この研究では、ディスプレイに装着したアイトラッカを用いて学習者の視点を計測している。視点を fixation と saccade に分類し、特徴量として fixation の平均持



図 1 JINS MEME の概観

表 1 読書行動のクラスの分類とデータセットのサンプル数

クラス	読書行動	サンプル数 (ノイズ除去前)	サンプル数 (ノイズ除去後)
0	何も読んでいない	35,228	32,708
1	英語を読んでいる	5,724	5,340
2	横書きの日本語を読んでいる	6,166	5,792
3	縦書きの日本語を読んでいる	6,161	5,798

続時間や saccade の平均長などを用いている。fixation とは注視点のことであり、saccade とは注視点間の素早い眼の動きのことである。Landsmann らは、ディスプレイ装着型のアイトラッカを用いて、文書を読んでいるか、読んでいないかの識別を行っている。特徴量として、fixation の垂直方向の距離の平均や標準偏差が用いられている。

Kelton らの研究、Landsmann らの研究では、ディスプレイ上の文書を読むという実験的な環境で眼の動きを測定しているため、読書行動が制限される。そこで、Ishimaru らは、JINS MEME を用いることでディスプレイ上だけでなく日常生活における読書行動識別を行っている [1]。JINS MEME は、図 1 のような眼鏡型のウェアラブルデバイスであり、EOG センサ、加速度センサ、角速度センサが搭載されている。これらのセンサを用いて日常生活における読書行動識別を行っている。この研究では、得られたデータを 30 秒ごとのウィンドウに区切り、各ウィンドウに対して特徴量を抽出している。特徴量については、EOG の平均値と分散値、加速度、角速度それぞれの分散値を算出している。その後、Support Vector Machine (SVM) を用いて 30 秒ごとに 4 種類の読書行動を分類している。4 種類の読書行動とは、表 1 の通りである。結果として、実験参加者 10 人の読書行動の 4 クラス平均分類精度は 32% であったと報告されている。

Ishimaru らの研究は、Kelton らの研究や Landsmann らの研究と比較して読書行動に制限を加えていないため、読書行動識別はより困難であるといえる。また、Ishimaru らの研究において、特徴量として統計量を用いる方法では、日常生活における読書行動識別は難しいことが考えられる。そこで、Deep Learning を用いて読書行動識別を行った研究として、Mozaffari らの研究 [4]、Segawa らの研究 [5] がある。Mozaffari らは、Long Short-Term Memory を用いて、文書の読み方について通常の読み方であるか、ざっと読んでいるか、ざっと目を通しての 3 クラス分類を行っている。入力データとして、連続した N 個の fixation の持続時間、saccade の長さ、角度、速度を用いている。読書行動の 3 クラス分類精度は $N = 10$ のとき 92.5% であったと報告されている。Segawa らは、一人称視

点画像を用いて本を識別することで、読書行動の検出を行っている。画像中の本の識別には CNN が用いられており、読書検出の適合率は 79.0%、再現率は 94.2% であったと報告されている。

以上より、Deep Learning を用いることで、高精度な読書行動識別が期待できる。JINS MEME から得られる EOG や加速度、角速度に対して Deep Learning を適用し、読書行動識別を行った研究はこれまで確認されていないが、Deep Learning を用いる手法が有効である可能性がある。

2.2 Self-supervised Learning に関する研究

Deep Learning を用いて高精度な読書行動識別を行うには、大量のラベル付きデータが必要となる。この問題に対して、少量のラベル付きデータでも高精度な認識を可能とする手法として、Self-supervised Learning が用いられている。Self-supervised Learning とは、自動で生成できるラベルを用いるタスク (Pre-text Task) を解くことで、本来解きたいタスク (Target Task) のための特徴表現の学習を行う手法である。Self-supervised Learning は、画像認識分野で多く利用されている。主要な Pre-text Task として、画像の幾何変換や加工に関するタスクや画像と音声を用いるタスク、複数枚の画像間に関するタスクがある。画像の幾何変換に関するタスクには、画像の回転角を推定するタスク [6] や、回転に加えて左右の反転も推定するタスク [7] がある。画像の加工に関するタスクには、分割した画像を並び替え、正しい並びを推定する「ジグゾーパズル」を解くタスク [8] や、画像の一部を欠損させ、欠損部を推定するタスク [9]、グレースケール化した画像に対し着色を行うタスク [10] がある。画像と音声を用いるタスクには、画像と音声の組み合わせが正しいかを推定するタスク [11] がある。複数枚の画像間に関するタスクには、2 枚の画像からカメラの動きを推定するタスク [12] や、動画中の連続した画像の時系列が正しいかを推定するタスク [13] がある。このようなタスクを解くことにより、画像認識のための有効な特徴表現を獲得している。

一方で、センサデータを対象とした研究に対しても、Self-supervised Learning を用いた手法が提案されている。Human Activity Recognition (HAR) 分野において、Saeed らは加速度と角速度に拡大や回転といった 8 種類の変形を施し、どの変形が施されているかを識別するという手法を提案している [14]。この手法により学習した特徴表現を行動識別に用いることで、高精度な認識を行うことができると述べられている。HAR 分野において、識別対象となる行動は歩く、走るといった身体的行動であり、読書行動は含まれていない。

そこで本研究では、Saeed らの手法を読書行動に適用する。加速度、角速度だけでなく EOG に対しても変形を施し、施した変形の種類を識別することで、読書行動に有効な特徴表現を学習する。これにより、従来手法である Ishimaru らの手法よりも高精度な読書行動識別を目指す。

3. 提案手法

本稿では、Human Activity Recognition で提案されている

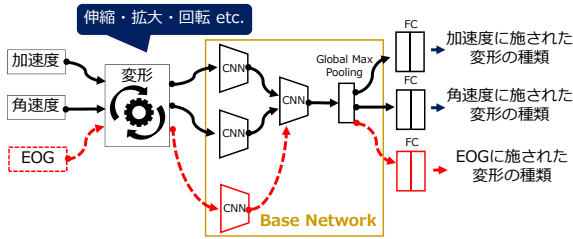


図2 本研究の特徴抽出ネットワーク

Saeed らの手法 [14] を拡張して、読書行動識別に有効な特徴を抽出する手法を提案する。提案手法は以下の2段階から構成される。まず、EOG、加速度、角速度に対して拡大や伸縮といった変形を施し、施されている変形の種類の識別を行う。その後、学習した特徴を用いて読書行動識別モデルの作成を行う。以下に詳細を述べる。

3.1 特徴抽出器の学習

用いるデータは、JINS MEME から得られる EOG、加速度、角速度である。本研究では、Saeed らの手法を拡張することにより、読書行動に有効な特徴表現の学習を行う。以下、Saeed らの特徴抽出法について詳細を述べる。

3.1.1 Saeed らの特徴抽出法

Saeed らの手法は、加速度や角速度のセンサデータに拡大や伸縮といった変形を施し、どの変形が施されているかを識別することで有効な特徴表現を学習する Self-supervised Learning の手法である。Saeed らの手法で提案されている特徴抽出のネットワークは図2における Base Network のうち、EOG に関する部分を除いたものである。Base Network の構成は、加速度と角速度から特徴を抽出する2つの CNN ブロックと、共有した特徴を抽出する CNN ブロック、Global Max Pooling 層で成り立つ。入力として、加速度、角速度に表2に示す8種類の変形を施したデータを用いている。図2のように Base Network に全結合層を付加することで、入力データに施されている変形の種類の識別を行っている。

3.1.2 本研究の特徴抽出法

Saeed らの特徴抽出法は、読書行動識別のためのものではない。そこで、読書行動に関する特徴を抽出するために、加速度、角速度だけでなく、EOG にも変形を加え、特徴を抽出する CNN ブロックを追加した図2のようなネットワークを用いる。図2に入力するデータには、EOG、加速度、角速度に対して、表2に示す8種類の変形を施したものをを用いる。表2におけるパラメータ a , b , c , d についてはそれぞれ、 $-100 \leq a \leq 100$, $5 \leq b \leq 10$, $c = 2$, $d = 90$ と設定する。ここで、EOG の変形について、EOG には加速度や角速度のような回転軸が存在しないため、回転を除く7種類の変形を施す。変形を施した EOG、加速度、角速度を図2のネットワークに入力し、それぞれのデータに対して、表2に示す8種類の変形のうちのどの変形が施されているか、または施されていないかの識別を行う。加速度、角速度に対しては9クラス分類を行い、EOG に対しては8クラス分類を行う。

表2 センサデータに施す変形

ランダムノイズ a の付加
b 倍に拡大
垂直方向に反転
水平方向に反転
時系列を c 倍伸長 / c 倍収縮
z 軸を基準に d 回転
時系列の並び替え
次元の並び替え

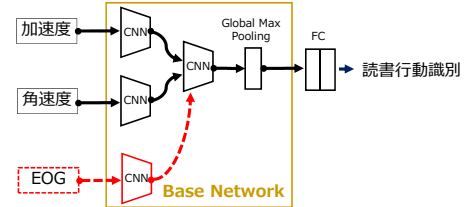


図3 読書行動識別ネットワーク

以下、本研究で提案する図2の特徴抽出ネットワークの詳細について述べる。EOG、加速度、角速度から特徴を抽出する CNN ブロックは、それぞれ3層の1次元 CNN 層で成り立つ。3層の1次元 CNN のユニット数はそれぞれ32, 64, 96であり、カーネルサイズはそれぞれ24, 16, 8である。各 CNN 層の後には Batch Normalization が施され、最後に Max Pooling 層と訓練データに対する Dropout が適用されている。Dropout の確率は0.1と設定している。特徴を結合する CNN ブロックは、EOG、加速度、角速度の特徴を結合する結合層と1層の CNN 層、Batch Normalization で成り立つ。CNN 層のユニット数は128、カーネルサイズは4である。最後に、変形の種類の識別するための全結合層を EOG、加速度、角速度それぞれに対して2層ずつ使用し、2層の全結合層のユニット数はそれぞれ256, 512である。各 CNN 層、全結合層には活性化関数として ReLU を用いている。出力層は Softmax 関数である。

3.2 読書行動識別モデルの学習

学習した図2のネットワークのうち、Base Network を用いる。Saeed らの手法では Base Network の重みを固定する方法を用いているが、予備実験の結果、読書行動識別に対しては十分な精度が得られなかった。そこで、Base Network の重みを再学習させることで読書行動識別を行う。ここで、識別する読書行動は表1に示す4種類である。図3のように Base Network に全結合層を付加することで、読書行動の4クラス分類を行う。

以下、本研究で用いる図3の読書行動識別のネットワークの詳細について述べる。図3の Base Network は、図2と同じものを用いる。Base Network から得られた特徴に対し、読書行動識別のための全結合層を使用する。全結合層のユニット数は1024であり、活性化関数として PReLU を用いる。出力層は Softmax 関数である。

4. 実験

本章では、提案手法を用いた実験について述べる。実験には、Ishimaru らの研究で用いられたデータセットを用いる。この

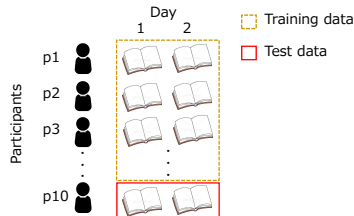


図4 評価方法

データセットを以後、OPU_RD と呼ぶ。

4.1 OPU_RD

OPU_RD は、日常生活において、JINS MEME から EOG、加速度、角速度を記録したもので構成される。JINS MEME の EOG センサは図 1 のように Left, Right, Bridge の 3 つの電極部分で成り立ち、水平方向と垂直方向の目の動きによる電位変化が記録される。Left, Right, Bridge のそれぞれの電位を L, R, B とすると、水平方向の EOG は $L - R$ 、垂直方向の EOG は $B - \frac{L+R}{2}$ で算出される。また、加速度、角速度についてはそれぞれ x, y, z 軸方向のデータが記録されている。EOG センサ、加速度センサ、角速度センサのサンプリングレートはそれぞれ 100Hz である。また、実験参加者は 10 名であり、1 人あたり 1 日約 12 時間、JINS MEME を装着して 2 日間生活した。1 日の間に各実験参加者は、英語の文書、日本語の縦書きの文書、日本語の横書きの文書をそれぞれ約 1 時間読んでいる。実験参加者は Narrative Clip^(注1) と呼ばれる小型のカメラを衣服に取付け、30 秒ごとに撮影を行い、実験参加者の行動を記録した。1 日の実験が終わると、実験参加者は撮影した 30 秒ごとの写真に対して、表 1 のうちのどの読書行動を行っていたかのラベル付けを行った。ラベルを付与した写真が撮影される 30 秒前の全ての EOG、加速度、角速度のデータに対して、写真に付与したラベルと同一のラベルを付与した。

4.2 データの前処理

OPU_RD に対し、ネットワークに入力するために、ウィンドウに分割した。ウィンドウサイズは 30 秒と設定し、オーバーラップを 15 秒とした。各ウィンドウにおいて読書行動のラベルは多数派のものとし、30 秒ごとのウィンドウに対して読書行動識別を行った。また、OPU_RD に含まれる EOG、加速度、角速度には、JINS MEME の装着具合の違いにより、基準値のずれが生じている。そこで、各ウィンドウのデータに対してウィンドウの平均値を差し引くことによって、基準値の修正を行った。さらに、OPU_RD のうち EOG には、JINS MEME に搭載されている EOG センサの接触不良等の原因で、数秒程度のバースト性のノイズが現れる。そこで、EOG データに対してノイズの含まれるウィンドウを OPU_RD から除去した。ノイズの判定方法については、 $T[\mu V]$ を上回る値または $-T[\mu V]$ を下回る値が t 秒以上連続して確認された場合、ノイズと判定した。本稿では、ノイズ判定において、 $T = 1000$ 、 $t = 2$ と設定した。図 2 と図 3 のネットワークの Optimizer については

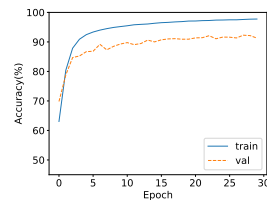


図5 EOG に施した変形種類の識別精度の推移

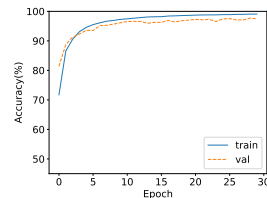


図6 加速度に施した変形種類の識別精度の推移

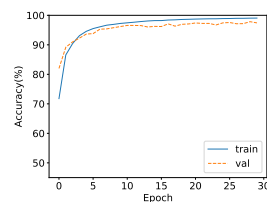


図7 角速度に施した変形種類の識別精度の推移

Adam を用い、学習率は 0.001 とした。学習回数は 30 回とし、学習の際、バッチサイズを 24 とするバッチ処理を行った。

4.3 評価方法

実験の評価方法について、特徴抽出の実験、読書行動識別の実験ともに、図 4 に示す方法を用いた。具体的には、実験参加者 10 名のうち、1 名のデータをテストデータ、残り 9 名のデータを訓練データとすることで学習を行い、Leave-One-Participant-Out Cross-Validation を用いて識別を行った。さらに、訓練データのうち、ランダムに選択した 20% のデータを識別モデルの評価を行うためのバリデーションデータとして用いた。

4.4 特徴抽出の実験

4.4.1 実験条件

OPU_RD から、EOG、加速度、角速度に表 2 に示す変形を施し、どの変形が施されているか、または変形が施されていないかの識別を行った。加速度、角速度に対しては 9 クラス分類を行い、EOG に対しては 8 クラス分類を行った。各変形を施したデータをそれぞれ同じ割合で用意し、EOG の 8 クラス分類のチャンスレートは 12.5%、加速度と角速度の 9 クラス分類のチャンスレートは 11.1% であった。

4.4.2 実験結果

EOG、加速度、角速度に施された変形の種類を識別した結果、テストデータに対する実験参加者 10 人の平均識別精度は EOG が 82.3%、加速度が 95.9%、角速度が 95.9% であった。図 5、図 6、図 7 に、それぞれ EOG、加速度、角速度の変形の種類を識別したときの実験参加者 10 人の平均学習精度と平均バリデーション精度の推移を示す。図 5、図 6、図 7 より、学習を進めていくと訓練データ、バリデーションデータともに精

(注1) : <http://getnarrative.com/>

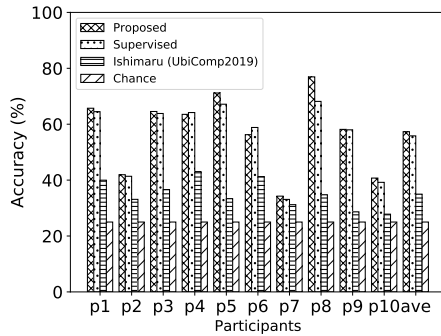


図 8 各実験参加者の読書行動識別精度

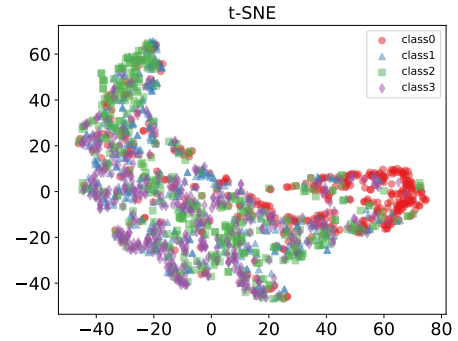


図 9 Ishimaru らの手法の p8 の特徴可視化

度が向上していることがわかる。

4.5 読書行動識別の実験

4.5.1 実験条件

4.4 節で学習したネットワークのうち、図 3 のように Base Network を利用して再学習を行うことにより、読書行動識別を行った。OPU_RD に含まれている各クラスのサンプル数を表 1 に示す。表 1 より、OPU_RD にはラベルの偏りが存在する。そこで、ラベルの偏りを除去するために、訓練データ、テストデータ両方に対してアンダーサンプリングを行った。具体的には、各クラスのサンプルの割合が均等になるように、多数派クラスのサンプルに対してランダムに選択し、OPU_RD から除去した。これにより、読書行動識別の 4 クラスの分類のチャンスレートは 25% となる。

また、比較手法として、Ishimaru らの手法と Supervised Learning の手法を用いた。Ishimaru らの手法とは、2 章でも述べたように特徴量として統計量を用いる従来手法である。Supervised Learning の手法とは、Self-supervised Learning を行わずに、図 3 のネットワークにより、ラベルありデータのみを用いて学習を行う教師あり学習の手法である。Self-supervised Learning を用いた提案手法は、Supervised Learning と異なり、ラベルありデータだけでなく、ラベルなしデータを特徴抽出に用いることができるという利点がある。提案手法を Ishimaru らの手法、Supervised Learning の手法と比較することによって、提案手法の有効性を検証した。

4.5.2 結果・考察

各実験参加者について、読書行動識別を行った結果を図 8 に示す。図 8 で、Proposed は提案手法の結果である。Supervised は Self-supervised Learning を行わずに教師あり学習を行った結果、Ishimaru は Ishimaru らの手法の結果を示している。なお、Chance は、チャンスレートを示している。実験参加者 10 名の平均識別精度は、提案手法が 57.3%、Supervised が 55.8%、Ishimaru が 35.0%、Chance が 25.0% であった。結果から、Ishimaru らの手法やチャンスレートよりも高い識別精度を示した。一方で、提案手法は Supervised と同程度の識別精度を示した。

以下、提案手法と Ishimaru らの手法を比較して、識別精度が向上した p8 について考察する。まず、p8 について、t-SNE [15] と呼ばれる特徴を可視化する手法を用いた。図 9 と図 10 に

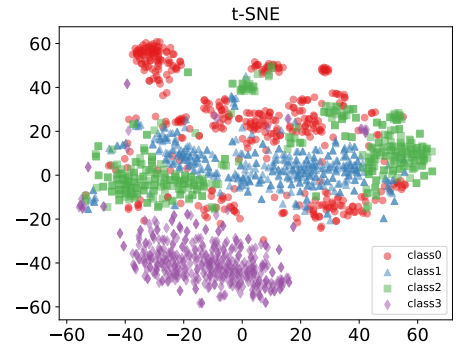
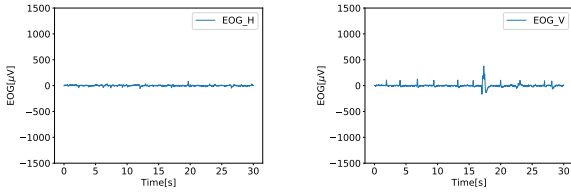


図 10 提案手法の p8 の特徴可視化

それぞれ Ishimaru らの手法と提案手法の特徴を可視化した結果を示す。図 10 は、図 3 の学習したネットワークのうち、Global Max Pooling 層の出力を可視化したものである。図 9 より Ishimaru らの手法で抽出された特徴について異なるラベルが混在している。一方で、図 10 より提案手法で抽出された p8 の特徴については、ラベルごとにまとまっている部分が多く見られる。このことから、提案手法で学習した特徴は、Ishimaru らの手法に比べて読書行動識別に有効であったことがわかる。

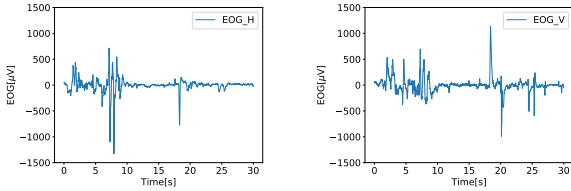
一方、提案手法の中でも識別精度が低い p7 について考察する。p7 のデータのうち、クラス 1、クラス 2、クラス 3 のサンプルがクラス 0 と多く誤認識されていた。図 11 と図 12 にそれぞれ、クラス 2 と正しく認識された EOG データ、クラス 0 と誤認識されたクラス 2 の EOG データを示す。図 12 のデータには、本稿で用いたノイズ判定では除去できないノイズが含まれており、図 11 とは異なるパターンであると判断された可能性がある。OPU_RD は日常生活におけるデータセットであるため、ノイズが大量に含まれている。ノイズを含むウィンドウを全て取り除くと、データ数が極めて少なくなり、識別精度が低下する可能性がある。そこで今後は、ノイズが含まれるデータセットにおいて、ノイズに頑健な読書行動識別の手法が必要である。

また、提案手法が Supervised と同程度の識別精度であった原因について考察する。Self-supervised Learning は、一般にラベルなしデータがラベルありデータに比べて大量にある場合に有効な手法である。ところが、本稿の Self-supervised Learning で用いたデータ量は Supervised Learning で用いたデータ量の 2 倍ほどと大量とはいえないレベルに留まった。このため



(a) 水平方向の EOG (b) 垂直方向の EOG

図 11 p7 のクラス 2 と正しく認識された EOG データ



(a) 水平方向の EOG (b) 垂直方向の EOG

図 12 p7 のクラス 0 と誤認識されたクラス 2 の EOG データ

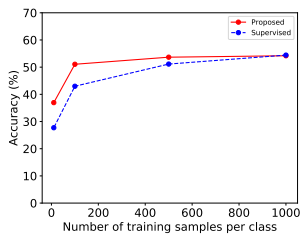


図 13 各クラスのサンプル数の変更に伴う平均識別精度

に、提案手法が Supervised Learning と同程度の識別精度になったと考えられる。そこで、Supervised Learning に比べて Self-supervised Learning が有効であるか検証するために、読書行動識別の学習に用いる各クラスの訓練データに用いるサンプル数を 1000, 500, 100, 10 と制限した。結果を図 13 に示す。この図より、各クラスの訓練データに用いるサンプル数が 10 のときの平均識別精度は、提案手法が 37.0%, Supervised が 27.7% であり、提案手法が Supervised を 9.3% 上回った。これにより、ラベルありデータが少ない場合でも、Self-supervised Learning による手法が有効である可能性が示された。

5. まとめ・今後の課題

本稿では、Saeed らの手法を拡張し、EOG、加速度、角速度に変形を施し、施した変形の種類を識別することで読書行動識別のための特徴表現を学習した。EOG に対して 8 クラス分類、加速度、角速度に対して 9 クラス分類を行った平均識別精度はそれぞれ 82.3%, 95.9%, 95.9% であった。また、学習した特徴表現を用いて読書行動識別を行った結果、実験参加者 10 名の平均識別精度は 57.3% であった。

今後の課題として、ノイズに頑健な読書行動識別の手法を提案すること、Self-supervised Learning に大規模ラベルなしデータセットを用いることで Self-supervised Learning の有効性の検証を行うことが挙げられる。

謝辞 本研究の一部は、JST CREST (Grant No. JP-MJCR16E1), ならびに大阪府立大学キーププロジェクトの補

助による。

文 献

- [1] Shoya Ishimaru, Takanori Maruichi, Manuel Landsmann, Koichi Kise, and Andreas Dengel. Electrooculography dataset for reading detection in the wild. In *Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, pp. 85–88. ACM, 2019.
- [2] Conor Kelton, Zijun Wei, Seoyoung Ahn, Aruna Balasubramanian, Samir R Das, Dimitris Samaras, and Gregory Zelinsky. Reading detection in real-time. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, p. 43. ACM, 2019.
- [3] Manuel Landsmann, Olivier Augereau, and Koichi Kise. Classification of reading and not reading behavior based on eye movement analysis. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, pp. 109–112. ACM, 2019.
- [4] Seyyed Saleh Mozaffari, Federico Raue, Saeid Dashti Hossanzadeh, Stefan Agne, Syed Saqib Bukhari, and Andreas Dengel. Reading type classification based on generative models and bidirectional long short-term memory. 2018.
- [5] Yuta Segawa, Kazuhiko Kawamoto, and Kazushi Okamoto. First-person reading activity recognition by deep learning with synthetically generated images. *EURASIP Journal on Image and Video Processing*, Vol. 2018, No. 1, p. 33, 2018.
- [6] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Un-supervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728*, 2018.
- [7] Ozsel Kilinc and Ismail Uysal. Learning latent representations in neural networks for clustering through pseudo supervision and graph-based activity regularization. *arXiv preprint arXiv:1802.03063*, 2018.
- [8] Mehdi Noroozi, Ananth Vinjimoor, Paolo Favaro, and Hamed Pirsiavash. Boosting self-supervised learning via knowledge transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9359–9367, 2018.
- [9] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2536–2544, 2016.
- [10] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *European conference on computer vision*, pp. 649–666. Springer, 2016.
- [11] Relja Arandjelovic and Andrew Zisserman. Look, listen and learn. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 609–617, 2017.
- [12] Pulkit Agrawal, Joao Carreira, and Jitendra Malik. Learning to see by moving. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 37–45, 2015.
- [13] Basura Fernando, Hakan Bilen, Efstratios Gavves, and Stephen Gould. Self-supervised video representation learning with odd-one-out networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3636–3645, 2017.
- [14] Aaqib Saeed, Tanir Ozcelebi, and Johan Lukkien. Multi-task self-supervised learning for human activity detection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 3, No. 2, p. 61, 2019.
- [15] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, Vol. 9, No. Nov, pp. 2579–2605, 2008.