

# 接触型デバイスから取得した加速度情報を用いた オンデマンド講義受講者の学習エンゲージメント推定手法

沖本 航大<sup>†</sup> 目良 和也<sup>†</sup> 黒澤 義明<sup>†</sup> 竹澤 寿幸<sup>†</sup>

<sup>†</sup>広島市立大学大学院情報科学研究科

## 1. はじめに

近年, MOOCs (Massive Open Online Courses) をはじめとしたオンライン講座が発展し教育の情報化が進んでいる. オンライン講義における学習の質を向上させるには, 受講者の関心度を高め, 維持することが重要である. しかし, オンライン講義では受講者の表情や口調などが伝わりにくいため, 対面型講義と比較して受講者の心理状態を推定するのは容易ではない. 特に受講者が一方的に視聴する側となるオンデマンド講義では, なおさら心理状態の推定は困難である.

学習者の学習エンゲージメント (学習に関与する姿勢・態度) を推定する研究として, 視線, 頭部姿勢, 表情を用いる手法[1]や姿勢情報を用いる手法[2]などが提案されている. また, 映像以外から取得できる特徴量として, 複数のセンサから取得した生体情報や腕の 3 軸加速度情報を用いる手法[3]も提案されている.

そこで本研究では, オンデマンド講義受講者の表情や動作といった映像ベースの特徴量に加え, 接触型デバイスから取得した頭部および腕の加速度情報から, 時系列情報を考慮した機械学習手法を用いて学習エンゲージメントを推定する手法を提案する. さらに, 接触型デバイスを用いて計測した頭部加速度情報と映像解析によって推定した頭部加速度情報について学習エンゲージメント推定性能を比較する実験を行う.

## 2. 提案手法

### 2.1 システム構成

受講者の学習エンゲージメントを正確に推定するには, 表情, 姿勢, 視線, 頭部や腕の動きといったマルチモーダルな特徴量を扱うことと, 特徴量の時系列変化を考慮できることが重要である. そこで本研究では, 接触型デバイスによって計測された時系列データと映像解析によって得られた時系列データとの両方を機械学習器で処理する手法を提案する.

提案手法の構成を図 1 に示す. まず, 受講者の様子をカメラで撮影した映像の解析結果および複数の接触型デバイスの出力から表情情報, 姿勢情報, 視線情報, 加速度情報を取得し, それらの情報を単体もしくは組み合わせて機械学習することで, 受講者の学習エンゲージメントレベルを very low, low, high, very high の 4 段階で推定する. なお, 特徴量の時系列変化を考慮するため, 本研究では機械学習モデルとして, RNN (Recurrent Neural Network) の 1 種である LSTM (Long Short-Term Memory) を用いる.

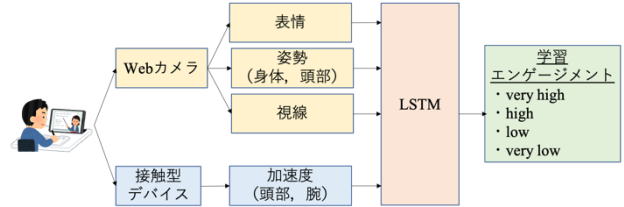


図 1. 提案手法の構成

表 1. 特徴量抽出ツールと特徴量

ツール・デバイス	モダリティ	主な特徴量	次元数	サンプリングレート
OpenFace	視線	視線ベクトル 視線角度	288	30fps
	頭部姿勢	頭部位置 頭部角度	6	
	表情	AU 強度	415	
	頭部加速度 (推定値)	3 軸加速度	204	
OpenPose	姿勢	鼻, 両目, 両耳, 首, 両肩の 2 次元座標	16	30fps
JINS MEME	頭部加速度 (計測値)	3 軸加速度	3	20Hz
E4 Wristband	腕の加速度 (計測値)	3 軸加速度	3	32Hz

### 2.2 特徴量の抽出

本研究では, 表情, 姿勢, 視線, 加速度情報を取得するため映像解析ツールおよび 2 種類の接触型デバイスを用い, データ収集並びに特徴量抽出を行う. 表 1 に本研究で用いる特徴量の詳細を示す.

#### 2.2.1 計測加速度

JINS MEME[4]はメガネ型の接触型デバイスで, 3 軸加速度センサと 3 軸ジャイロセンサにより回転方向や回転角度が認識できる. 本研究では受講者の頭部の動きを捉えるため x,y,z 軸の 3 軸加速度を用いる.

E4 wristband[5]は腕時計型の接触型デバイスで, 人間の生理信号と腕の 3 軸加速度を取得可能である. 本研究では, 受講者の腕における動きを捉えるため x,y,z の 3 軸加速度を用いる.

本稿では接触型デバイスによって計測される特徴量をまとめて“計測加速度”と定義する.

#### 2.2.2 映像特徴量

OpenFace[6]は視線, 頭部姿勢, 表情情報を検出可能なオープンソースツールである. 視線情報としては, 目のランドマーク 56 点の 2 次元および 3 次元座標, 両目の 3 次元視線ベクトル, そして両目で平均化された視線角度の計 288 次元を推定可能である. 頭部姿勢情報としては, 3 次元空間での顔の位置や Pitch, Yaw, Roll の 3 成分に変換した顔の角度の計 6 次元を推定可能である. 表情情報としては, 目, 鼻, 口, 眉, 輪郭のランドマーク計 68 点の 2 次元および 3 次元座標や 18 種類の Action Unit[7]の強度が推定可能である.

<sup>†</sup> A Method to Estimate Learning Engagement of On-Demand Lecture Attendees by Using Acceleration Information Obtained from Contact Devices: Kodai OKIMOTO, Kazuya MERA, Yoshiaki KUROSAWA, Toshiyuki TAKEZAWA, Graduate School for Information Sciences, Hiroshima City University

OpenPoseは人間の骨格キーポイント計18点の2次元座標を推定することが出来るオープンソースツールである。本研究における受講者の映像データはバストショットであるため、上半身のキーポイント（鼻、両目、両耳、首、両肩）8点の2次元座標の計16次元の特徴量を用いる。

本稿では映像解析ツールによって算出される特徴量をまとめて“映像特徴量”と定義する。

### 3. 機械学習実験

本節では、接触型デバイスによって得られる計測加速度を用いた手法と、映像から取得できる映像特徴量（視線、頭部姿勢、身体姿勢、表情）のみを用いた手法、計測加速度と映像特徴量の両方を用いた手法について学習エンゲージメント推定実験を行う。

さらに、計測加速度の代わりに映像解析ツールを用いて推定される加速度（推定加速度）を用いた実験も行う。推定加速度は、OpenFaceから算出した68点の3次元顔ランドマーク座標それぞれの変化量より算出する。

#### 3.1 実験用講義視聴データの収集

実験参加者7名（男性5名、女性2名）にJMOOC[8]で公開されている18本のオンデマンド講義について、概要を読んでもらった上で自身の関心度を5段階で評価してもらった。そして、評価が最大・最小となった講義を人毎に1本ずつ計2本選定し、講義を約10分間視聴してもらった。その際webカメラを用いて視聴している際の表情、姿勢情報及び視線を30fpsで記録した。頭部と腕の加速度情報の計測には2.1節で述べた接触型デバイスを用いた。

オンライン講義視聴後、実験参加者は話題が移り変わるタイミングで区切った講義のセクション毎に学習エンゲージメントをvery low, low, high, very highの4段階で自己評価した。実験参加者一人あたりのセクション数は平均4件（最小2件、最大6件）、平均セクション時間は4分36秒（最小1分15秒、最大10分）、各ラベル毎のセクション数はそれぞれ、very low=6件、low=8件、high=9件、very high=6件となった。

#### 3.2 機械学習器のパラメータ

本研究では、100ユニットからなる1つの隠れ層と1つの全結合層からなるLSTMを用いて4クラスを分類する。予備実験の結果に基づき、ハイパーパラメータは映像特徴量のみによる手法のバッチサイズを128、その他はバッチサイズを64に設定した。エポック数は10,000に設定した。また、中間層は1層、最適化関数は確率的勾配降下法、学習率は0.001、データ分割はホールドアウト法を用い、train:validation:test=6:2:2として実験を行なった。

また、LSTMでは複数のモデルを学習するにあたりサンプリングレートを統一しシーケンス長を同一にする必要があるため、本稿ではサンプリングレートが64Hzとなるように全特徴量の時系列データを線形補間した。

#### 3.3 実験結果

機械学習実験の結果を表2に示す。表より、提案手法である計測加速度を用いた機械学習器が正解率0.98となり最も高い結果であった。また、映像特徴量に計測加速度を加えた手法でも正解率0.93と、従来手法より大きく性能が向上することが確認できた。このことから接触型デバイスから取得した計測加速度情報は学習エンゲージメントの推定に有効であることが確認できた。

表2. 各特徴量セットによる機械学習正解率

映像特徴量のみ (従来手法)	計測加速度 のみ	計測加速度 +映像特徴量	推定加速度	推定加速度 +映像特徴量
0.58	<b>0.98</b>	<b>0.93</b>	0.65	0.73

また、計測加速度の代わりに推定加速度を用いた実験の正解率は0.65となり、計測加速度より劣るものの従来手法より良い結果が得られた。このことから、接触型デバイスの代用として映像から推定した加速度情報を用いる手法も従来手法より有効であるといえる。

実際に収録した映像を確認すると、学習エンゲージメントが高いと評価されたデータに関しては腕や頭部の動きが少なくなる傾向があり、逆に学習エンゲージメントが低いと評価されたものに関しては、腕や頭部の動きが多くなる傾向が確認された。逆にy軸頭部加速度については高群で平均加速度が大きくなっていったが、これは講義を聞きながらうなずく動作が影響しているものと考えられる。

#### 4. おわりに

本研究では、オンデマンド講義受講者の表情、姿勢、視線及び接触型デバイスから取得した加速度情報から受講者の講義に対する学習エンゲージメントを推定するための手法について提案した。結果として、接触型加速度で最も高い正解率0.98が得られ、従来手法と接触型加速度を組み合わせた提案手法では、正解率0.93となり、腕、頭部の加速度情報を加えることで性能向上が確認できた。

今後は、機械学習器のチューニングによる性能向上を目指すとともに、接触型デバイスで計測される加速度と同等の加速度情報を映像から推定する手法についても検討する予定である。

#### 謝辞

本研究は国立研究開発法人科学技術振興機構（JST）の「COIプログラム令和4年度加速支援（ Grant 番号 JPMJCA2208）」の支援によって行われた。

#### 参考文献

- [1] X. Ma, et al., “Automatic Student Engagement in Online Learning Environment Based on Neural Turing Machine”, International Journal of Information and Education Technology, Vol. 11, No. 3, pp.107-111, 2021.
- [2] B. Zhu, et al., “Multi-rate Attention Based GRU Model for Engagement Prediction”, Grand Challenge Paper: Emotion Recognition in the Wild Challenge ICMI 20, 2020.
- [3] N. Gao, et al., “N-Gage: Predicting in-Class Emotional, Behavioural and Cognitive Engagement in the Wild”, in Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., Vol. 4, No. 3, Article 79, pp.1-26, 2020.
- [4] JINS MEME : <https://jinsmeme.com/> (2023/1/6 アクセス)
- [5] Empatica E4 Wristband : <https://www.empatica.com/research/e4/> (2023/1/6 アクセス)
- [6] T. Baltrusaitis, et al., “OpenFace 2.0: Facial Behavior Analysis Toolkit”, in Proc. International Conference on Automatic Face and Gesture Recognition (FG), IEEE Computer Society, pp.59-66, 2018.
- [7] P. Ekman, et al., “Facial Action Coding System: a Technique for the Measurement of Facial Movement”, Palo Alto: Consulting Psychologists Press, 2002.
- [8] JMOOC : <https://www.jmooc.jp/> (2023/1/13 アクセス)