

マルチモーダル情報を考慮し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

Learning Engagement Estimation Considering Multimodal Information for On-Demand Lecture Attendees

沖本 航大
Kodai Okimoto

広島市立大学大学院 言語音声メディア工学研究室
Language and Speech Research Laboratory, Graduate School of Information Sciences, Hiroshima City University

概要 本研究では、オンデマンド講義受講者の学習エンゲージメントを既存手法である表情、姿勢および視線に加え、生体情報から推定する手法を提案する。実験の結果、腕の加速度データによる推定が正解率 0.92 となり有効性が確認できた。また生体情報でも皮膚電気活動を用いた手法では既存手法とほぼ同等の結果が得られた。

1. はじめに

近年、MOOCs(Massive Open Online Courses)など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。また、そこから得た教育ログデータの分析により、教育・学習を支援する Learning Analytics という研究が盛んに行われている。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、オンライン講義は受講者の表情変化や言葉の抑揚などが伝わりにくく、対面型講義と比較してコミュニケーションが取りにくいことから、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、なおさら学習者の心理状態を推定するのは困難である。

そこで本研究では、オンデマンド講義受講者の表情、動作および生体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する。

2. 先行研究

本研究の目的は、学習者の学習エンゲージメントを推定することである。本節では、学習エンゲージメント推定についての先行研究を2つ紹介する。

Xiaoyang Ma ら [1]では OpenFace[2]を用いて算出した視線、頭部姿勢、Action Unitそして C3D モデルを用いて算出した姿勢情報を入力とし、学習器に NTM(Neural Turing Machine)を用いた 4 段階の学習エンゲージメントの推定手法を提案している。しかしこの研究では、RNN(Recurrent Neural Network)や LSTM(Long Short-Term Memory)など時系列の流れを考慮した推定処理が行われていない。

また、竹花ら[3]は双方向型講義において学習者の生体情報と心的状態の相関関係を抽出し形式化を試みている。この研究の中で、学習者が学習を楽しんでいる状態と相関のある生体情報として呼吸数や皮膚コンダクタンスが挙げられている。

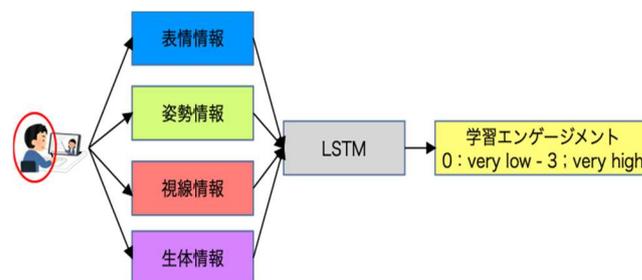


図 1. 提案手法の構成

表 1. 使用デバイス と収集する特徴量

デバイス	特徴量
OpenFace	顔の座標点や視線推定, 頭部姿勢推定, Action Unit(AU)の強度
OpenPose	上半身のキーポイント (鼻, 右目, 左目, 右耳, 左耳, 首, 右肩, 左肩)
JINS MEME	視線, 頭部姿勢, 加速度
E4 Wristband	皮膚電気活動, 心拍数, 加速度, 血液容積脈波, 体温
Garmin	呼吸数

3. 提案手法

本研究では、表情、視線、姿勢、生体情報といったマルチモーダルな情報を考慮し、さらにデータの時系列を考慮した手法を提案する。提案手法の構成を図 1 に示す。まず、学習者の様子をカメラで撮影した動画および複数のウェアラブルデバイスから、表情情報、姿勢情報、視線情報、生体情報を取得し、それらの情報を総合的に機械学習することで、学習者のエンゲージメントレベルを (0:very low, 1:low, 2:high, 3:very high) の 4 段階で推定する。

3.1 特徴量抽出

本研究では、生体情報や姿勢情報、視線情報を取得するため、web カメラからの映像と実験参加者のストレスになりにくく、かつ負担が少ないと考えられる 3 種類のウェアラブルデバイスを用いデータ収集並びに特徴量抽出を行った。表 1 に使用デバイスと収集する特徴量を示す。

表 1 のデバイスを選定した理由として、JINS MEME は Bixer ら [4]が視線情報から心ここにあらずの状態が検出可能なことを示唆しているためである。また E4 Wristband と Garmin は竹花ら [3]で学習者の心的状態と相関のある生体情報として呼吸数や皮膚コンダクタンスが報告されていたためである。

(1) OpenFace

OpenFace は顔の座標点や視線推定, 頭部姿勢推定, Action Unit(AU)の強度を検出可能なオープンソースツールである. Action Unit(AU)とは, 視認可能な顔の動きを包括的に測定するため Facial Action Coding System(FACS)[5]という表情理論において顔の表情を解剖学の知見を基に定義した最小の動作単位のことである. 本研究では, 計 709 次元の特徴量を使用した.

(2) OpenPose

OpenPose[6]は人間の骨格のキーポイント計 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. 本研究では, 上半身のキーポイント (鼻, 右目, 左目, 右耳, 左耳, 首, 右肩, 左肩) の計 8 点の 2 次元座標の 16 次元の特徴量を用いる.

(3) JINS MEME

JINS MEME はメガネ型のウェアラブルデバイスで, 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, また 3 点式眼電位センサにより瞬きの速度や, その間隔, 視線移動が認識可能である. 本研究では, 約 20Hz で 3 軸加速度や頭部姿勢を取得できる 14 次元データ, 15 秒の間隔で瞬き速度, 瞬き間隔, 視線移動を計測する 33 次元データを用いる.

(4) E4 wristband

E4 wristband は腕時計型のウェアラブルデバイスで, 人間の生理信号を取得可能である. 本研究では, 32Hz で取得可能な 3 軸加速度や 1Hz で取得可能な心拍数, 32Hz で取得可能な血液容積脈波, 4Hz で取得可能な皮膚電気活動, 体温といった特徴量を用いる.

(5) Garmin Venu Sq Music

Garmin は光パルスレート技術を使用して 1 分間当たりの呼吸数を取得することが出来る. 本研究では 1 分間の呼吸数を用いる.

3.2 講義視聴データの収集

事前に選定した 18 本のオンデマンド講義 (情報分野, 語学系など) の中から実験参加者 7 名 (男性 5 名, 女性 2 名) に自身に関心を持ってそうな講義, 持てなさない講義を概要を読んでもらった上で選定してもらい, 選定した講義を約 10 分間視聴してもらった. その際に web カメラ(Anker PowerConf C300)を用い視聴している際の表情および姿勢情報の記録を 30fps で行った. 視線情報並びに生体情報は, 3.1 節で述べたウェアラブルデバイスを用い記録を行った.

オンライン講義視聴後, 事後アンケートとして実験参加者は視聴した各講義に対する学習エンゲージメントを話題毎で区切った講義のセクション毎に 0~3(0:very low, 1:low, 2:high, 3:very high)で回答した. 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.

表 2. 各特徴量による機械学習結果

特徴量 (計測機器)	正解率	適合率	再現率	F 値	
顔座標点, 視線, 頭部姿勢, AU (OpenFace)	0.71	0.59	0.66	0.61	
上半身の部位座標 (OpenPose)	0.62	0.35	0.45	0.37	
JINS	3 軸加速度, 頭部姿勢	0.88	0.88	0.87	0.87
	瞬き速度&間隔, 視線移動	0.62	0.65	0.54	0.56
E4	加速度	0.92	0.91	0.92	0.91
	血液容積脈波	0.37	0.27	0.28	0.21
	皮膚電気活動	0.55	0.58	0.53	0.52
	心拍数	0.34	0.26	0.30	0.27
	体温	0.42	0.20	0.31	0.24
呼吸数 (Garmin)	0.33	0.18	0.27	0.20	

4. 学習実験

本研究の最終的な目的は, 表情, 視線, 姿勢, 生体情報といったマルチモーダル情報を考慮して学習エンゲージメントを推定することであるが, 本稿ではその前段階であるシングルモーダルで機械学習実験を行い比較, 考察を行う.

4.1 ネットワーク

本研究では, 学習者の状態の時系列変化を考慮するため, RNN(Recurrent Neural Network)モデルの 1 種である LSTM(Long Short-Term Memory)モデルを採用した. 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 4 クラスを分類する.

4.2 実験条件

予備実験の結果に基づき, ハイパーパラメータはバッチサイズが OpenFace のみ 256, その他は 64 で設定した. エポック数は loss が下がりきっていなかった皮膚電気活動, 心拍数に関してはエポック数は 50,000 に設定し, その他は 1,000 である. 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 0.0001, データ分割 train:validation:test = 6:2:2 として実験を行なった.

4.3 実験結果

各特徴量抽出器から取得した特徴量で学習させた結果を表 2 に示す. 表 2 より, 最も性能が良かったのは正解率が 0.92 の加速度データであった. このことから, 受講者の腕の動きを加速度センサにより取得することは学習エンゲージメントの推定に効果的だと考えられる. また, 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定されたデータに関しては腕の遊びが少なくなる傾向が確認できた. 逆に学習エンゲージメントが低いと評定されたものに関しては, 腕の遊びが多くなる傾向が確認できた.

また既存手法では扱われていなかった生体情報については[5]で報告されている通り, 皮膚電気活動が学習実験で使用した生体情報の中でも最も高く, 正解率が 0.55 となった. これは, 既存手法と同等の結果である. 一方, 呼吸数は正

解率 0.33 と低い結果となった。これは、学習エンゲージメントが低い状態と高い状態との間であまりデータに違いがないことが原因だと考えられる。実際、データを比較してみても両群の間に顕著な違いはみられなかった。

5. おわりに

本研究では、オンデマンド講義受講者の表情、動作および生体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、腕の加速度データで最も高い正解率 0.92 が得られ、生体情報に関しては皮膚電気活動が正解率 0.55 と既存手法と同等程度の識別性能を見せた。

今後は、ハイパーパラメータのチューニングや現在シングルモーダルで行なっている実験を特徴量を組み合わせ行うマルチモーダル実験を予定している。

参考文献

- [1] Xiaoyang Ma , Min Xu, Yao Dong, and Zhong Sun .Automatic Student Engagement in Online Learning Environment Based on Neural Turing Machine, International Journal of Information and Education Technology, Vol. 11, No. 3, March 2021
- [2] Tadas Baltrušaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. Openface 2.0: Facial behavior analysis toolkit. In Proc. International Conference on Automatic Face and Gesture Recognition (FG), pp. 59–66. IEEE Computer Society, may 2018.
- [3] 竹花和真, 田和辻可昌, 村松慶一, 松居辰則, 学習時における学習者の生体情報と心的状態の関係の形式化の試み, 人工知能学会 第74回 先進的学習科学と工学研究会 2015
- [4] Robert Bixler¹ , Sidney D’Mello, Toward Fully Automated Person-Independent Detection of Mind Wandering , International Conference on User Modeling, Adaptation, and Personalization UMAP 2014: User Modeling, Adaptation, and Personalization pp 37-48
- [5] Ekman Paul, Friesen Wallace, and Hager Joseph, Facial action coding system: a technique for the measurement of facial movement, Palo Alto: Consulting Psychologists Press, 2002.
- [6] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Realtime multi-person 2d pose estimation using part affinity fields. In Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 1302-1310.