

主体的学習における表情からのエンゲージメント推定 のためのデータセット構築に関する検討

Data Set Construction for Engagement Estimation from Facial Expressions in Self-directed Learning

長谷川 忍^{*1}, ZHENG Xianwen^{*2}, 太田 光一^{*2}, 卯木 輝彦^{*3,*4}
Shinobu HASEGAWA^{*1}, Xianwen ZHENG^{*2}, Koichi OTA^{*1}, Teruhiko UNOKI^{*3,*4}

^{*1} 北陸先端科学技術大学院大学 情報社会基盤研究センター

^{*1} Research Center for Advanced Computer Infrastructure, JAIST

^{*2} 北陸先端科学技術大学院大学 先端科学技術研究科

^{*2} Graduate School of Advanced Science and Technology, JAIST

^{*3} IMAGICA Group, ^{*4} フォトロン

^{*3} IMAGICA Group, ^{*4} Photron,

Email: {hasegawa, s1810428, ota}@jaist.ac.jp, unoki@photron.co.jp

あらまし：本稿では、学習者がそれぞれのペースで学習を行う主体的学習において、ノート/タブレット PC に内蔵されているカメラの映像を活用して、学習プロセスにおける学習者の状態を推定する際に、表情要素などの学習者の外的状態や、エンゲージメントなどの心的状態、学習者が行うタスクなど、より効果的なデータセットを構築する上で検討すべきポイントについて議論する。

キーワード：主体的学習, 学習状態推定, エンゲージメント, データセット, PC 内蔵カメラ

1. はじめに

e-learning やアクティブラーニングといった、学習者それぞれの主体的な活動が重視される学習環境では、学習プロセスが個別化されており、個々の学習者の学習活動に対する意欲や取り組み方、理解状態などを包括した「エンゲージメント」を推定して適切な教育的支援を行うことは容易でない。一方で、こうした学習環境においてノート/タブレット PC が利用されるようになってきているが、これらの多くに標準で内蔵されているカメラは、遠隔会議等の特定用途のみでしか活用されていない。

本研究の最終的な目的は、こうしたノート/タブレット PC を活用する主体的学習活動において、内蔵カメラで収録した顔映像を利用して学習者のエンゲージメントを推定するためのシステムを開発することである⁽¹⁾。こうした推定を実現する上では、対象となる学習活動を的確に反映したデータセットを準備して、適切な特徴量の抽出や提案したアルゴリズムの妥当性についての評価を行う必要がある。そこで本稿では、対象となる主体的学習の構成要素について議論するとともに、先行研究で収集・利用されたデータセットについて整理し、本研究で目指すべきデータセットの在り方についてまとめる。

2. 主体的学習の構成要素

本研究における主体的学習とは、ある学習タスク（目的や空間・時間を内包）を達成するために、ノート/タブレット PC 上でツール（Web ブラウザやアプリケーション、支援システム等）を介して、学習空間を選択的に学習し、そのプロセスを内省しながら知識として構造化することを繰り返す学習過程

として定義する。これを通じて、学習者はそれぞれの目的やスタイル、エンゲージメントや理解度等の心的状態に応じて多種多様な情報を、それぞれのペースで概念や概念間の関係として学習する。

こうした学習過程においては、従来のゴールが明確な教育的なタスクと比較して解が一意ではなく、テストなどでその成果を測定することは困難である。そのため、表情に代表される測定可能な学習者の外的状態の時系列変化は、主体的学習における心的状態を推定する上での重要な情報源となり得る。ここでの主体的学習タスクの例としては、e-Learning や Web 調べ学習、マインドマッピングなどが挙げられよう。

3. 先行研究におけるデータセットの分析

ここでは、顔映像を利用した学習者のエンゲージメント推定に関する近年の先行研究⁽²⁻⁶⁾で利用されたデータセットについて整理を行い、主体的学習におけるエンゲージメントの推定に適用可能かどうかについて検討する。

前節で述べた主体的学習の構成要素に基づいて、本研究では、(a) 学習タスク、(b) 学習(データ)時間、(c) 学習状態の尺度、(d) データの収集方法、(e) データ数、(f) その他の観点から表 1 のように整理した。

先行研究で調査対象にした「エンゲージメント」は日本語に一对一で対応する概念がなく、本研究で学習者の内的状態として活用することができるかどうかについては検討の余地があるが、これらの研究では 2 段階（エンゲージメントあり/なし）から 4 段階（エンゲージメント：なし、弱い、あり、強い）の評価が行われている。これらのうち、DAISEE⁽⁷⁾

表1 エンゲージメント推定のためのデータセットの整理

研究	(a) タスク	(b) 時間	(c) 学習状態	(d) 収集方法	(e) データ数
(2)	ビデオ視聴	約 10 秒	2-levels/3-levels	ラベル書き換え	112 名, 9068 ビデオ
	DAiSEE ⁽⁷⁾ のデータセットから 2 段階と 3 段階のエンゲージメントをタグ変更により付与				
(3)	ビデオ視聴	約 5 分	4-levels	5 名の専門家	197 ビデオ
	場所等の制約なしでデータを収集し, 5 名の専門家がビデオのみを見てエンゲージメントレベルを評価				
(4)	ビデオ視聴	約 5 分	4-levels	5 名の専門家	262 ビデオ
	EmotiW 2018 ⁽⁸⁾ の Engagement in the Wild データセットを利用				
(5)	テキスト/図の学習	約 35 分	2-levels	8 名の学生	47 名
	カメラ映像に加えて, マウスとキーボードの情報も取得して特徴量抽出				
(6)	エッセイ執筆	約 60 分	2-levels	自己申告	23 名
	2 分おきのエンゲージメントの申告と, 1 週間後にビデオを見て再度のエンゲージメント評価				
(1)	問題回答(CAB)	約 30 秒	5-levels	自己申告	19 名, 453 ビデオ
	問題毎の難しさ, 面白さ, 疲労度, 集中度を, 問題回答後に問題を見ながら自己申告				

や EmotiW2018⁽⁸⁾などのデータセットは公開されているが, 学習対象がビデオ視聴となっており, 主体的学習としてはやや限られている. 他の研究では, テキストや図をベースにしたコンテンツの学習やエッセイの執筆など主体的な活動を含むタスクが行われているが, データ数はやや限られている.

心的状態の収集方法については, 専門家によるビデオアノテーションが行われる場合と自己申告に基づく場合(定期的に質問/タスク終了後に質問)がある. エンゲージメントのような抽象的な概念では専門家による評価は適切であると考えられるが, 難易度や面白さ, 疲労度などより詳細な項目を分析するためには自己申告も有効である. 一方で, 自己申告では学習者への負担や主観的な妥当性の弱さなど注意すべき点もある.

4. おわりに

本稿では, 学習空間を選択的に学習し, そのプロセスを内省しながら知識として構造化することを繰り返す学習過程としての主体的学習の構成要素についてモデル化し, 主体的学習における心的状態の推定を, 顔映像を利用して行う先行研究の整理を通じて, データセットの利用可能性と収集手法について検討した. 今後は, 公開されているデータセットのより具体的な調査と, 本研究を評価する上で必要なデータセットの構築に向けた要件の検討が挙げられる.

参考文献

- (1) 長谷川忍, 平子温, 卯木輝彦: “PC カメラ映像を利用した学習状態推定手法に関する研究”, 教育システム情報学会研究報告, Vol.34, No.1, pp.41-48 (2019).
- (2) M. Ali Akber Dewan, Fuhua Lin, Dunwei Wen, Mahbub

- Murshed, and Zia Uddin: “A Deep Learning Approach to Detecting Engagement of Online Learners,” 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation, DOI: 10.1109/SmartWorld.2018.00318, (2018).
- (3) Cheng Chang, Cheng Zhang, Lei Chen, and Yang Liu: “An Ensemble Model Using Face and Body Tracking for Engagement Detection,” Proceedings of the 20th ACM International Conference on Multimodal Interaction, pp.616-622, (2018).
- (4) Xuesong Niu, Hu Han, Jiabei Zeng, Xuran Sun, Shiguang Shan, Yan Huang, Songfan, Xilin Chen: “Automatic Engagement Prediction with GAP Feature,” Proceedings of the 20th ACM International Conference on Multimodal Interaction, pp.599-603, (2018).
- (5) Zhaoli Zhang, Zhenhua Li, Hai Liu, Taihe Cao, and Sannyuya Liu: “Data-driven Online Learning Engagement Detection via Facial Expression and Mouse Behavior Recognition Technology,” Journal of Educational Computing Research, <https://doi.org/10.1177/0735633119825575>, (2019).
- (6) Hamed Monkarezi, Nigel Bosch, Rafael A. Calvo, and Sidney K. D’Mello: “Automated Detection of Engagement Using Video-Based Estimation of Facial Expressions and Heart Rate,” IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, VOL. 8, NO. 1, pp.15-28, (2017).
- (7) A. D’Cunha, A. Gupta, K. Awasthi and V. Balasubramanian, “DAiSEE: Towards user engagement recognition in the wild,” in arXiv preprint arXiv:1609.01885, (2016).
- (8) Abhinav Dhall, Amanjot Kaur, Roland Goecke, and Tom Gedeon: “EmotiW 2018: Audio-Video, Student Engagement and Group-Level Affect Prediction,” In ACM International Conference on Multimodal Interaction 2018, DOI: 10.1145/3242969.3264993, (2018).