

脳波を教師とした体動からの認知作業負荷推定のための深層学習モデルの実装と評価 Implementation and evaluation of a deep learning model for estimating cognitive workload from body movements using EEG as a teacher

沼田遥[†] 島田径汰[†] 秋元弘太[†] 千葉慎二[‡] 矢入郁子[†]
Numata Haluka Simada Keita Akimoto Kouta Shinji Chiba Yairi Ikuko

1. 緒論

システム利用時のユーザエクスペリエンスやインタラクションから、対面講義での受講者の意欲や集中度合などを含めた、ユーザのタスクへのエンゲージメントの評価・分析が、近年、Human Computer Interaction(HCI)分野における注目テーマとなっている[1]。人が何らかのタスクに利用している脳の情報リソース量を表す認知作業負荷はエンゲージメントの一つであり、その定量化は医療や教育だけでなく、人にサービスを提供するあらゆる工学的システムやソフトウェアの設計・利用において重要な課題である。システムが人に的確なタイミングで介入して最適なサービスを提供するためには、認知作業負荷の連続値を実時間で定量化できる手法の実現が望ましい。現在、ElectroEncephaloGraphy(EEG)の計測済データから算出される PLI を用いて、人の認知作業負荷を事後に連続値として定量化する手法が実現されている[2]。実時間で定量化可能な手法が実現されたとしても、EEG 測定はユーザに電極を装着する必要があることと、音や振動などのノイズの影響を受けやすいことから、計測可能な環境や対象が限定される問題がある。そこで本研究は、脳波計測に比べ環境ノイズの影響を受けにくく、非装着で測定可能なイメージセンサを用いて得られる体動情報から、実時間で連続的に認知作業負荷を定量化する手法を提案する[3]。本稿は、聴覚定常応答(ASSR)による脳波からの認知作業負荷計測値を教師データとして、Azure Kinect DK によるタスク実行時の体動データから認知作業負荷を推定する深層学習モデルの実装・評価について報告する。

2. 関連研究

HCI 分野でエンゲージメントの測定に最も利用されている測定方法は、主観的なアプローチに分類される、質問用紙を用いた調査方法、アンケートである[1]。アンケートは簡単に適用することができ、詳細な分析が可能であるという利点がある一方で、回答者が過去を思い出しながら回答するため、リアルタイムなエンゲージメントの変動を測定することができないという欠点がある。この欠点を補う手法の 1 つが経験サンプリング法(ESM)であり、これは認知作業負荷の変化を計測することができる[1]。ESM は、瞬間毎の経験を繰り返し報告することでエンゲージメントの変化を捉えることができるが、ユーザは報告を行うたびに意識をタスクから離す必要があるため、タスクに集中し続けた状態でのエンゲージメントの変化を推定することができない。

エンゲージメントの客観的な測定方法として Behavioral

Trace Measurement が代表的である。この手法は Web ページでのマウスクリック数などを用いて、オンラインネットワークユーザ[4]やブログ訪問[5]でのエンゲージメントを測定する。しかし、この手法は人が何らかのシステム操作を必要としないタスクを行っている時のエンゲージメントの変化を推定することができない。

このようなタスクも含めて、人のエンゲージメントの変化を客観的に推定する方法として、心拍数、血流、EEG などの心理生理学的信号を利用して認知作業負荷を計測する方法が用いられている。EEG を利用した認知作業負荷推定方法には PI 成分の認知作業負荷の増加に伴い振幅が減少するという特性[6]を利用した方法があり、コンピュータゲームをプレイ中のユーザの認知作業負荷計測などの応用例がある[7]。深層学習を用いて認知作業負荷を推定した例の一つに、特殊な装着器具から得られる呼吸、心電図、フォトプレチスモグラム、皮膚温度信号を入力とした手法もある[8]。本研究は、心理生理学信号である EEG を教師データとして、画像センサからの認知作業負荷推定を試みる。

3. データセットの作成

3.1 実験

＜参加者＞ データ収集実験には 22 歳から 25 歳までの健康な男性 11 名が参加した。すべての参加者は右利きで聴覚に異常はなかった。本研究は上智大学「人を対象とする研究」に関する倫理委員会の承認を受けている。

＜実験方法＞ 本実験は PC を用いた Introduction, Task1, Task2 の 3 つのタスクから構成される。Introduction は 6 分間何も考えないように指示され PC の白いスクリーンを見続けるタスク、Task1 は Web テストをブラウザ上で行うタスク、Task2 は Kindle for Mac 上で本を読むタスクである。実験中、2 台の Azure Kinect DK が参加者の正面(Master Kinect)と左手側(Sub Kinect)に設置され、体動の計測が行われた。また参加者の頭部には、拡張 10-20 法に基づいて Fpz, FC3, FC4, FCz, O1, O2 の 6 箇所脳波計の電極が設置され、脳波の計測が行われた。

＜実験手順＞ 参加者は指定の椅子に座り、脳波キャップを用いて前述の 6 箇所にアクティブ電極、左乳様突起に基準電極、右乳様突起にグラウンド電極を装着した。電極のインピーダンスを下げるために、皮膚前処理剤で処理したのち、基準電極とグラウンド電極はエールローデを、アクティブ電極はペースト剤を用い装着された。各電極のインピーダンスが確認された後、参加者はイヤフォンを装着し、実験中に入力される 40Hz の音声刺激が正常に聞こえるようボリュームの調整を行った。参加者は Introduction を 6 分間行った後、Task1 の操作方法を確認し、Task1 を 21 分間行った。30 分以上 Task1 を行った参加者もいた。5 分の休憩後、参加者は Task2 の操作方法を確認したのち、Task2 を 21 分間行った。Task1 同様、30 分以上 Task2 を行った参加者もいた。

[†] 上智大学 Sophia University

[‡] 日本マイクロソフト Microsoft Japan Co., Ltd.

<実験結果> 本実験により, 11 名分の長さ 6 分間の Introduction データ, 10 名分の長さ 21 分間の Task1 データ, 9 名分の長さ 21 分間の Task2 データを得た. ここで, Task1 で 2 名, Task2 で 1 名実験中に脳波計と PC 間の Bluetooth 接続が切れてしまったため, 2 つのファイルに分け合計 21 分間のデータを得た.

3.2 データの処理

<認知作業負荷>初めに脳波データから関東の電源ノイズ 50Hz の周波数成分を除去した. その後, 時系列の脳波データから 1 秒ごとに長さ 60 秒の部分時系列データを生成し, (1)式を用いてチャンネルごとの PLI を算出した. ここで, F は周波数特性, f は周波数, ch はチャンネル, k は試行, n は平均する試行数である. 先行研究に基づき周波数は 40Hz, チャンネルは FC3, FC4, FCz の電極のみを利用し 1 回の試行は 1 秒間隔で 1000ms の脳波を用いた. 1 回の試行に対してフーリエ解析を行い, その位相に関して 60 回分の試行を平均することで PLI を算出した. 認知作業負荷は FC3, FC4, FCz から得られた PLI を z-score で標準化し, その平均を 1 から引いたものと定義した.

$$PLI(f, ch) = \frac{1}{n} \left| \sum_{k=1}^n \frac{F(f, ch, k)}{|F(f, ch, k)|} \right| \quad (1)$$

<体動データ> Azure Kinect DK では深度センサを用いて骨格表現の 32 箇所の関節 (PLEVIS, SPINE_NAVAL, SPINE_CHEST, NECK, CLAVICLE_LEFT, SHOULDER_LEFT, ELBOW_LEFT, WRIST_LEFT, HAND_LEFT, HANDTIP_LEFT, THUMB_LEFT, CLAVICLE_RIGHT, SHOULDER_RIGHT, ELBOW_RIGHT, WRIST_RIGHT, HAND_RIGHT, HANDTIP_RIGHT, THUMB_RIGHT, HIP_LEFT, KNEE_LEFT, ANKLE_LEFT, FOOT_LEFT, HIP_RIGHT, KNEE_RIGHT, ANKLE_RIGHT, FOOT_RIGHT, HEAD, NOSE, EYE_LEFT, EAR_LEFT, EYE_RIGHT, EAR_RIGHT) の三次元座標, 四元数, そして信頼度を 15fps で測定した. 信頼度は測定値の信憑性を 3 段階で評価したもので, 測定した関節が物体に隠れることなく直接測定できている場合は 2, 体や物体により隠れており近くの関節から位置を推定している場合は 1, 推定が行えなかった場合は 0 と, 記録される. 本研究は深層学習の学習用データとして, Master Kinect による実験参加者の正面から測定したデータを採用した.

体動データの正規化前処理として, 関節の位置を表す三次元座標系の原点を, カメラから脊髄と胸の境界の位置にある SPINE_CHEST に移動した. これにより SPINE_CHEST の座標が (0,0,0) になり固定されるため, 学習に使うデータセットから SPINE_CHEST 情報を削除した. 次に, SPINE_CHEST と, 脊髄と膺の境界にあたる SPINE_NAVAL の関節間の距離に基づいてスケールリングを行った. この 2 段階の前処理は [9] を参考にした.

上記の処理を行った体動データから 1 秒ごとに長さ 60 秒の部分時系列データを生成した. <認知作業負荷> と <体動データ> より, 60 秒間の体動データとその時間に計測された認知作業負荷とが対となっているデータセットが作られた.

3.3 学習用データセット

本研究は実験によって得られた認知作業負荷と体動データから, 8 つのデータセットを作成した. 但し, 体動データに信頼度が 0 の測定値が含まれているデータポイントは削除した.

認知作業負荷をもとに回帰問題とクラス分類問題用の教師ラベルを作成した. <認知作業負荷> の処理を行った認知作業負荷のデータを回帰問題用の教師データとして利用した. クラス分類問題用の教師データは個人毎の認知作業負荷データを数値が大きい順にクラス 2, クラス 1, クラス 0 と 3 つのラベルを付すことで作成した. クラス分類問題用データセットはクラス 2 と 1 との境界, クラス 1 と 0 との境界でのデータの混在が大きいことから, クラス 1 の分類精度が低くなることが予想される. 以降, 回帰問題用データセットを R, クラス分類問題用データセットを C と呼ぶ.

一つ目のデータセットは, R と C に対して <体動データ> の前処理を行ったデータで, RF, CF とした. 二つ目のデータセットとして, <体動データ> の前処理を行った体動データに含まれている関節のうち, 信頼度が低い関節を削除したデータセットを作成した. 具体的には机によって隠れている下半身の関節 (HIP_LEFT, KNEE_LEFT, ANKLE_LEFT, FOOT_LEFT, HIP_RIGHT, KNEE_RIGHT, ANKLE_RIGHT, FOOT_RIGHT), 腕を伸ばしたときに Azure Kinect DK の視野外に出てしまう手の関節 (WRIST_LEFT, HAND_LEFT, HANDTIP_LEFT, THUMB_LEFT, WRIST_RIGHT, HAND_RIGHT, HANDTIP_RIGHT, THUMB_RIGHT), 頭に含まれている関節のうち, その中で信頼度が最も高い HEAD 以外の関節 (NOSE, EYE_LEFT, EAR_LEFT, EYE_RIGHT, EAR_RIGHT) を取り除いた. この処理では [3] で認知作業負荷に相関が高いとされる顔, 首, 右肩の動きが観測できるよう, その動きを観測するのに必要な関節が残された. この処理により生成されたデータを RI, CI とした. 三つ目のデータセットとして, <体動データ> の前処理を行った 15fps の体動データを平均して 1 秒平均処理を行い, 1fps のデータを作成した. この処理により生成したデータを RA, CA とした. さらに, 四つ目のデータセットは関節の削除と 1 秒平均処理の両方を行って生成し, このデータセットを RIA, CIA とした.

4. 深層学習モデルの実装と評価

本章では, LSTM モデルを用いて Azure Kinect DK で得た関節の位置データから認知作業負荷の推定を行った. また, 学習では 10-fold cross validation を用いて学習データとテストデータに分割した. さらに, 学習データの 10% を検証データとして利用した.

4.1 深層学習モデル

本研究では LSTM 層と, 2 層の全結合層からなる 3 層の深層学習モデルを使用した. 深層学習モデルには精度の向上と時間短縮のために EarlyStopping と ReduceLROnPlateau を導入した.

隠れ層のノード数, オプティマイザ, 学習率, Dropout 率, 及び隠れ層の決定係数のパラメータチューニングを行った. ハイパーパラメータチューニングには少ない探索回数で最適化できるベイズ最適化を用いた [10]. ここで, ベイズ最適化によるハイパーパラメータチューニングを行うためのデータセットには, 学習の時間短縮のために正規化を行う前の CF からランダムに 1/10 のデータを取り出して使用した. 隠れ層のノード数の最適化は 10~4500 の範囲を

分割しながら行った。結果、ノード数が 3524 の時、Accuracy が最高値の 58% になったので、隠れ層のノード数は 3500 にした。オプティマイザの最適化には Rmsprop, Adagrad, Adadelata, Adam, Admax, Nadam, 及び SGD を使った。それぞれのオプティマイザに対して学習率の最適化を、Adadelata は 0~1, Rmsprop と Adagrad は 0~0.1, Adam と Nadam と SGD と Admax は 0~0.01, の範囲で分割しながら行った。その結果、オプティマイザに Admax, 学習率に 0.00005 付近の数値を使った時、約 70% と最も高い精度が出た。よって、モデルのオプティマイザは Admax, 学習率は 0.00005 とした。入力層と隠れ層の Dropout 率は 0~0.5 の範囲で行った。その結果、入力層の Dropout 率が 0.05 付近、隠れ層の Dropout 率が 0.1 付近の数値を使った時、高い精度になったため、これらの値を用いた。また、隠れ層の活性化関数の探索には RF の全データを使用した。Relu, Elu, Selu, 及び Softplus で学習を行い比較した結果、それぞれの決定係数は、Relu が 0.9307, Elu が 0.9117, Selu が 0.9107, Softplus が 0.9056 であったため、Relu を採用した。

4.2 クラス分類問題

クラス分類問題のモデルには出力層の活性化関数に Softmax, 損失関数に category crossentropy を利用した。また、モデルの評価には Accuracy と F 値を利用した。表 1 は CF, CI, CA, CIA で学習を行った結果を示す。表 1 から、信頼度の低い関節の削除と 1 秒平均処理により精度が上がっていることが確認できる。また、CIA の学習により得られたヒートマップを図 1 に示す。図 1 から、クラス 2 とクラス 0 を正しく分類できていることがわかる。また、CF と CA の学習時の学習データと検証データの損失の変化を図 2, 図 3 に示す。図 2, 図 3 では学習データでの損失を青色、検証データでの損失をオレンジ色でプロットした。図 2 と図 3 の学習データと検証データの損失の差を比較すると図 3 の差が図 2 の差に比べても大きくなっているのがわかる。これより 1 秒平均処理を行うことで過学習が大きくなったことがわかる。

4.3 回帰問題

回帰問題のモデルには出力層の活性化関数に linear, 損失関数に平均絶対誤差(MAE)を利用した。また、モデルの評価は二乗平均平方誤差(RMSE)と決定係数を利用した。表 2 に RF, RI, RA, RIA で学習を行った結果を示す。表 2 より、全てのデータセットでの学習で決定係数が 0.9 を超えていることがわかる。また、表 2 の RMSE より信頼度の低い関節の削除と 1 秒平均処理により誤差が減ったことが確認できる。

4.4 キャリブレーションを行う場合の予測精度の遷移

本章では CIA と RIA を用いて未知のユーザの予測をするときの、キャリブレーション用に必要なデータの割合を求めた。まず、leave-one-subject-out(LOSO)により CIA をテストデータと学習データに分けた。その後、テストデータの一部を学習データに移動することで学習データに含まれていなかったユーザのデータを学習データに加えた。テストデータから学習データへ移動するデータの量を変化させながら、モデルの精度の変化を確認した。移動するデータの割合は 0% から 10% までは 1% 刻み、10% から 90% まで

表 1 クラス分類問題の評価結果

Data set	Accuracy (%)	F-score
CF	86.8	0.868
CI	87.5	0.875
CA	87.8	0.878
CIA	88.8	0.888

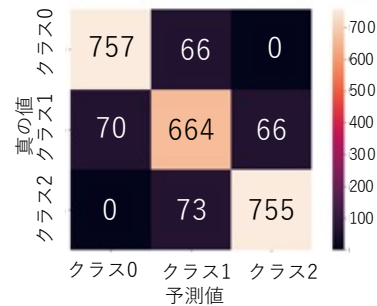


図 1 CIA のヒートマップ

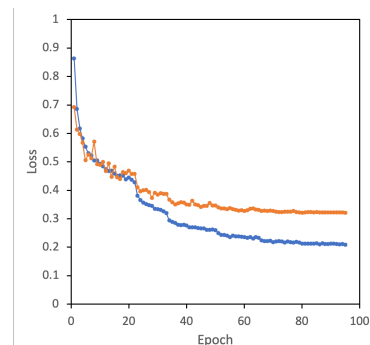


図 2 CF の Loss

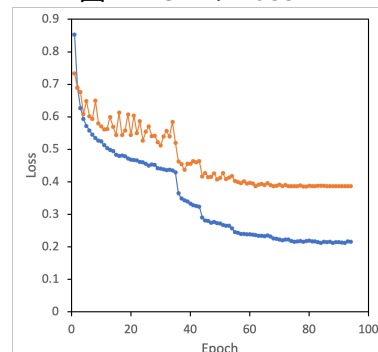


図 3 CA の Loss

表 2 回帰問題の評価結果

Data set	RMSE	R squared
RF	0.252	0.931
RI	0.251	0.932
RA	0.233	0.943
RIA	0.210	0.955

は 10% 刻みで行った。これを、テストデータとする実験参加者を変えて計 11 回繰り返す、平均値を算出した。図 4 はテストデータから学習データに移動するデータの割合に対する、精度の平均値を示す。同様の実験を RIA のデータ

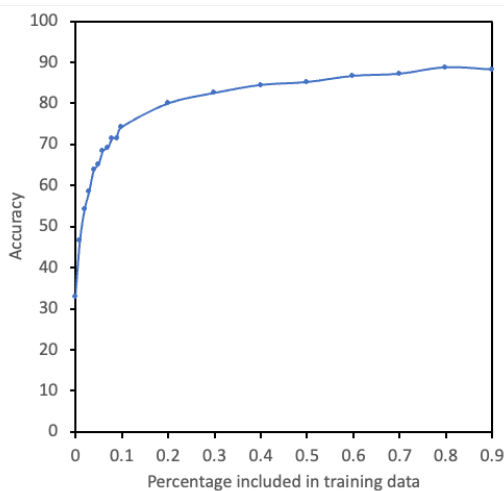


図 4 LOSO からのデータ移行による精度遷移(C)

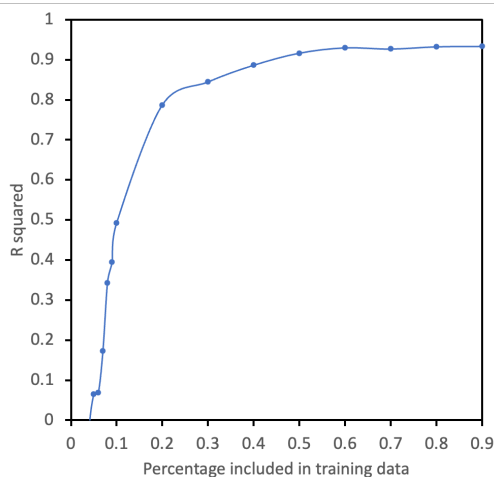


図 5 LOSO からのデータ移行による精度遷移(R)

セットで行った結果が図 5 である。図 4 より、RIA ではテストデータから学習データへ約 20%のデータを移動した時点で決定係数が 0.6 を超えることが確認できた。また、図 5 より CIA ではテストデータから学習データへ約 20%のデータを移動した時点で Accuracy が 80%を超えることが確認できた。

5. 考察

信頼度の低い関節を削除して 1 秒平均処理を行うことで、クラス分類問題と回帰問題の両方で精度が向上した。また、クラス分類問題ではクラス 0 とクラス 2 の分類を正しく行っていたため、タスクを中断する判断など、認知作業負荷が高い状態であることを認識する手法を活用できる場面 [11]で、この手法は利用できると考えられる。また、回帰問題では RIA を用いた学習の決定係数が 0.9549 であり、モデルがデータによく当てはまっていることがわかる。しかし、決定係数が 0.9 を超えていることと図 3 の学習データと検証データの損失の差から過学習が起きていることがわかる。この原因には本研究で利用したデータセットでは計測時間が近いデータ内に同じ体動が含まれていることが挙げられる。つまり異なるデータに同一の体動が含まれていることが原因の 1 つであると考えられる。

また、キャリブレーションから、未知のユーザの全バリエーションのうち、約 20%のデータをキャリブレーション

用に収集することで実用可能なレベルの予測が可能になることを確認した。

6. 結論

本稿は脳波から測定した認知作業負荷を教師データとして、Azure Kinect DK の深度センサによる体動データから認知作業負荷を推定する、深層学習モデルの実装・評価を行った。複数のデータセットを作成して LSTM モデルで学習を行った結果、クラス分類問題では Accuracy:88.78%、F 値:0.8879、回帰問題では RMSE:0.215、決定係数:0.959 という精度が得られた。これにより、計測環境に制限のある EEG に変わり、イメージセンサから認知作業負荷を推定する手法の実現可能性が示された。

謝辞

実験に参加して下さった皆様に深く感謝を申し上げます。本研究は JSPS 科研費基盤研究(B)20H04476 の助成を受けたものです。

参考文献

- [1] Doherty, K., Doherty, G., "Engagement in HCI: Conception, Theory and measurement", ACM Computing Surveys, Vol.51, No.99(2019).
- [2] Yokota, Y., Tanaka, S., Miyamoto, A., Naruse, Y., "Estimation of Human Workload from the Auditory Steady-State Response Recorded via a Wearable Electroencephalography System During Walking", Frontiers in Human Neuroscience, Vol.111, pp.223,228 (2017).
- [3] Simada, K., Chiba, S., Yokota, Y., Naruse Y., Yairi, I., "Correlation analysis between the learning concentration estimated by EEG and the body motion measured by image sensors", National Institute of Information and Communications Technology (2021).
- [4] Hennig, P., Berger, P., Meinel, C., Graber, M., Hildebrandt J., Lehmann, S., Ramson, C., "Tracking visitor engagement in the blogosphere for leveraging rankings", UEO'13: 2013 International Conference on Social Computing. IEEE, (2013).
- [5] Lehmann, J., Lalmas, M., Baeza-Yates, R., Yom-Tov, E., "Networked User Engagement", Workshop on User Engagement Optimization, (2013).
- [6] Pratt, N., Willoughby, A., Swick, D., "Effects of working memory load on visual selective attention: behavioral and electrophysiological evidence", Front. Hum. Neurosci., 13, June (2011).
- [7] Allison Brendan Z., Polich, John, "Workload assessment of computer gaming using a single-stimulus event-related potential paradigm", Biological Psychology, Vol.77, Issue.3 (2008).
- [8] Momeni, N., Dell'Agnola, F., Arza, A., Atienza, D., "Real-Time Cognitive Workload Monitoring Based on Machine Learning Using Physiological Signals in Rescue Missions", Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 41st, pp.3779,3785 (2019).
- [9] Amir, S., Jun, L., Tian-Tsong, N., Gang, W., "NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis", Conference on Computer Vision and Pattern Recognition (CVPR), (2016).
- [10] Jiménez, J., Ginebra, J "pyGPGO: Bayesian Optimization for Python", Journal of Open Source Software, Vol.2, Issue.19, pp.431 (2017).
- [11] Iqbal, S., Bailey, B., "Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption", Conference on Human Factors in Computing Systems, (2005).