

皮膚電気活動計測による組み立て作業時の迷いの認識 Recognition of Confusion during Assembly Work by Measuring Electrodermal Activity

渡邊 昭信[†]
Akinobu Watanabe

村儀 天星[†]
Tensei Muragi

辻 愛里[‡]
Airi Tsuji

藤波 香織[‡]
Kaori Fujinami

1. はじめに

近年、工業製品の生産形態としては、従来主流であった同一仕様で大量生産する生産形態よりも、少量多品種の生産形態の割合が高まり、顧客ごとやロットごとに製品仕様は細かくカスタマイズされる傾向がある。このようなニーズに対応するためには、自動化による生産性向上だけでは効率的な対応ができず、現場作業員の製造組立スキルこそが重要であり生産効率と製品品質に大きく影響する。そのため、非熟練労働者でも簡単に組立作業に参加でき、熟練労働者に劣らない生産性を実現できるように、組立手順や注意事項などの情報を効果的に伝達する作業支援システムの研究が活発に行われている。

例えば、プロジェクターを用いた作業台へ作業指示内容を重畳して投影する手法 [1]や、ヘッドマウントディスプレイ (HMD) を用いた手法 [2]が提案されている。また、Funkらは、熟練作業者と未熟練作業者という2つの作業グループに対し、プロジェクターによる投影型支援システムを適用した場合の作業者の行動を長期間にわたって調査した[3]。それによると、常に指示内容が投影されていることや、長時間 HMD を装着し続けることを煩わしいと感じる作業者も存在することがわかり、作業者のスキルやコンディションに応じて、必要な支援をタイムリーに実施することができれば、作業支援自体が作業者の負担になる懸念を軽減しつつ効果的な支援を両立できると期待する。

本研究では、このようなタイムリーでオンデマンドな支援システムを実現するために、皮膚電気活動 (EDA) 信号を用いて、組立作業中における作業員の迷いの程度を認識する手法を提案する。

表 1 迷いレベルごとの評価基準

迷いレベル	評価基準
0	すぐわかりやすく、すらすらできた
1	やや分かりにくく、確認は少なかった
2	分かりにくく、確認が多かった
3	とても分かりにくく、ミスも多かった
4	とても分かりにくく、やりきることができない

[†] 東京農工大学大学院生物システム応用科学府生物機能システム科学専攻 Department of Bio-Functions and Systems Science, Graduate School of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology

[‡] 東京農工大学大学院工学研究院先端情報科学部門 Division of Advanced Information Technology and Computer Science, Institute of Engineering, Tokyo University of Agriculture and Technology

EDA は、認知努力や感情的覚醒などの認知プロセスおよび感情プロセスの変化を反映する特徴があり、手のひらなどの体表面の導電率によって測定できる[4]。迷っている状態とは、認知作業負荷が高い状態と考えられるため、EDA を計測することによって、迷いを認識可能であると考えられる。

本研究では、組立作業をシミュレーションした組立タスクを実施して EDA データを収集し、教師付き機械学習モデルを生成し、表 1 に示すような、迷いなしの状態を含む 5 つのレベルの迷いを分類する。本研究は、東京農工大学倫理審査委員会の承認を受け実施した (承認番号: 211201-0366)。

2. EDA 計測による迷い認識

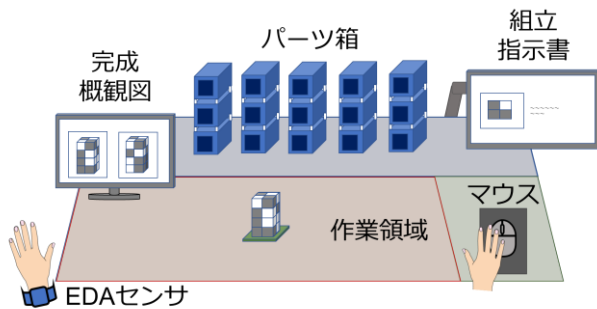
EDA 計測による特徴量と認識手法の研究として、EDA や皮膚コンダクタンス反応 (SCR) などの生体信号から得られる様々な特徴量と、それらを用いた統計手法、分析手法を比較評価した研究がおこなわれており、統計的な手法を用いたメル周波数ケプストラム (MFCC) 特徴量と Area Under the Curve (AUC) および Signal Magnitude Area (SMA) 特徴量が、一般的に使用される EDA 信号の SCR 特徴量を上回ることが報告されている[4]。このような特徴量や分析手法を用いることで、リアルタイムに動作する感情検出アプリケーションが実現可能であることを示唆している。また、EDA データには個人差も存在し、同じ刺激に対して異なる生理学的反応を有する傾向があることも報告されている。

本研究では、固定長の時間範囲を持たせたウィンドウで切り出すサイズのデータを 1 単位として、ある固定時間ずつ時間範囲をずらして、時系列の EDA データからサンプリングした。例えば、時刻 T から時刻 T+10 までの 10 秒間計測した EDA データから、ウィンドウサイズ 5 秒でサンプリングする場合、隣接ウィンドウ間で重複する時間範囲をウィンドウサイズの 50% とするとき得られるデータとしては、時刻 T から時刻 T+5、時刻 T+2、5 秒から時刻 T+7、5 秒、時刻 T+5 秒から時刻 T+10 秒、の 3 データを取得できる。このようにして得られる EDA データから、機械学習の分類器に入力する多次元の特徴量ベクトルを生成する。特徴量ベクトルは、迷いのレベルを特徴付けるものとするため、一般的に使用される統計指標に加え、経時変化を反映させる指標の 2 種類を組み合わせた。一般的な統計指標としては、平均、範囲、標準偏差、分散、歪度、および尖度とした[4]。さらに、時間的変位を反映するための指標として、EDA データの 1 次近似によって取得される直線の切片と傾き、第 1、第 2、および第 3 四分位数を導入した。このようにして得られる、合計 11 種類の時間領域 EDA 特徴量を、迷い認識器 (分類器) に与える特徴ベクトル (11 次元) として採用した。

3. データ収集

EDA データは、視線遷移や手の動きなどのマルチモーダルデータを使用して組立作業者の迷いを認識するプロジェクトの一環として収集した。本報告では、EDA データ収集に関連する部分に焦点を当てる。

まず、データ収集環境について説明する。図 1 に示すように、製造現場の組立ラインを模した環境として、部品棚、組立説明書、作業台、完成品画像からなる模擬組立環境を構築した。



(EDA データ収集に関連するコンポーネントのみを表示)

図 1 データ収集環境

組み立てる対象製品を構成する部品として、レゴブロックを使用した。作業台上の配置としては、レゴブロックを積み上げる領域を正面手前に配置し、奥に部品棚を配置した。部品棚の両脇に組立説明書と完成品画像を、液晶モニタを用いて表示した。作業台の利き手側にはマウスを配置した。このマウスは、作業ステップごとの説明書のページ送りや悩みレベルの評価記録のために使用する。EDA データを計測するデバイスとしては、リストバンド型の生体信号センサーである Empatica E4 センサーを使用した。EDA センサーを参加者の非利き手首に装着し、EDA センサーを装着した上腕を作業台上に静止した状態を、実験を通して維持した。組立作業とマウス操作は、EDA センサーを装着していない利き手のみで実施した。

次に、データ収集作業の概要を説明する。レゴブロックを特定の形状に組み立てる模擬組立作業を実行する参加者として、20 代の学部生・大学院生 16 名を募集した。彼らには 4 パターンの組立タスクを 3 回ずつ、計 12 回実施してもらった。ただし、12 回連続して実施すると、作業時間が長くなり、参加者の疲労が蓄積し疲労の度合いが取得データに影響するリスクを考慮し、2 パターンずつ 2 つの時間帯に分けて実施し、2 つの時間帯の間には 1 時間以上の時間を置いた。また、各タスク間にも数分程度の休憩を適宜とってもらった。

6 セット分のデータ収集実験のフローを図 2 に示す。実験開始の冒頭で、本実験への協力に対する感謝と、実験内容と取得データの取り扱いの説明および承諾の確認を、テスターが参加者に対して実施した。参加者の同意が得られた後、参加者に視線センサーおよび EDA センサーを装着してもらい、液晶モニタ上に表示される作業手順書に従って作業を実施してもらった。

1 回のタスクの作業フローとしては、指定された部品を部品棚から見つける工程と、作業台上で部品を取り付ける工程をセットで 1 ステップとし、合計 20 ステップで 1 つの製品を完成させる構成の作業フローとした。

作業手順書は右側のモニタ上に表示され、完成画像が左側のモニタ上に表示される。作業手順書の最初のページで取り出す部品を部品番号と数量で指定したのち、次のページで部品の取り付け位置と向きを指示した。作業手順書のページ送りは、利き手でマウスをクリックする操作で実行してもらった。これら 2 ページを 1 ステップ分の指示内容とし、20 ステップ分で 40 ページの作業指示書とした。なお、完成品画像は左側のモニタに常に表示した。

参加者は、各ステップの部品取り付け作業の後、次のステップに進む前に、表 1 に示す悩みレベルの定義に基づき、当該ステップの部品取り出しと取り付け作業で知覚した悩みレベルを 5 段階で評価し、マウスクリックで入力した。

1 セットの作業行程中であれば、マウスクリックで段階的に前後のステップに移動することができ、作業手順書のページを戻して組立作業をやり直すことや、悩みレベルを更新することも可能とした。

20 ステップの作業が終わった後、視線センサーのカメラで撮影した一人称視点の動画を見ながら、テスターが参加者に悩みの内容をヒアリングすることで、ステップごとに入力した悩みレベルに認識ずれがないかレビューした。

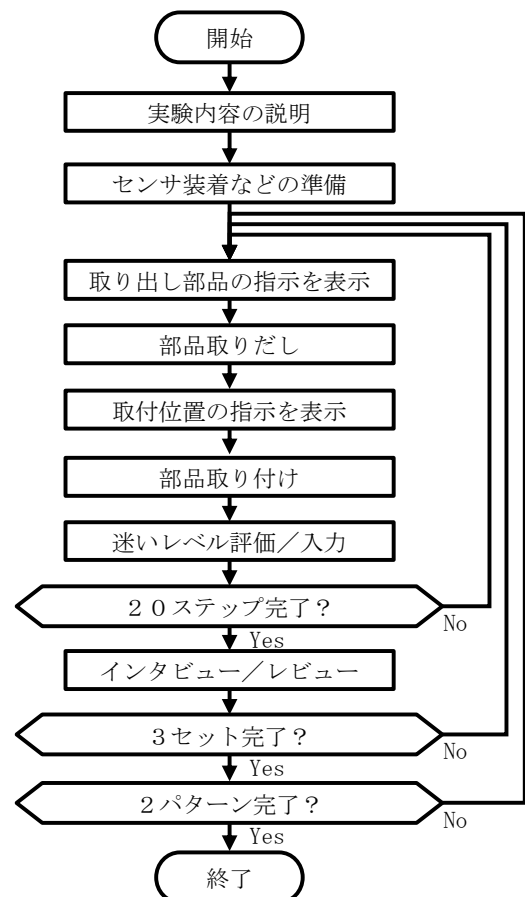


図 2 データ収集実験フロー

個人の認識の違いに起因するノイズ要因を極力削減するため、実験内容の説明を定型文として固定し、その説明者を同一人物で実施することとした。また、実験終了後のインタビューを実施することで、参加者にレベル定義を再確認してもらい、本来の認識と異なる評価結果であれば修正できるように、低ノイズ化に配慮した。

各タスクでは同じ 8 種類のパーツが使用したが、パーツの格納位置を記憶できないよう、部品箱の位置はタスクごとにシャッフルした。

4. EDA データによる迷い認識

4.1 手法

迷いクラスの定義としては、表 1 に示す 5 つの悩みレベルを各クラスに割り当てた。そして迷いレベルの分類パイプラインに関する、さまざまなパラメーターの影響を検証するために、次に述べる 3 つのフェーズで実験を実施した。

まず、第 1 段階の実験は、全参加者の EDA データすべてを統合して 1 つのデータセットとした。データセットから、ウィンドウを用いてサンプリングするが、その条件は 5 秒のウィンドウサイズと 50% のオーバーラップ率とした。ウィンドウサイズを 5 秒とした理由は、EDA 信号の特徴として、立ち上がりと回復を合わせて、3 秒程度の時間が一般的であり、連続した時系列データからウィンドウを切り出す場合の一般的な重複率が 50% であるためである。

第 1 段階における分類器モデルは、4 つの一般的な分類器モデルを用いた。すなわち、Naïve Bayes (NB)、k 近傍法 (kNN)、サポートベクターマシン (SVM)、Random Forests (RF) の 4 つである。なお、ハイパーパラメータは scikit-learn 0.24.2 における既定値を使用した。

検証手法としては、データセット全体に対して層化 10 分割交叉検証を実施した。本報告では、この交差検証の手法を Cross Validation All (CV_ALL) と呼ぶこととする。

CV_ALL の実現方法は、5 段階のレベルごとのサンプル数比率を保ったまま、10 等分したサンプル数の学習データと評価データに分け、10 分の 1 の評価データと残りの 10 分の 9 の学習データという組み合わせで分類器モデルを生成し評価する、という実験を、評価データを 10 等分したデータで順次入れ替えることで 10 回繰り返した。10 回の実験から得られたマクロ平均の F 値の、10 回の平均値が最も高かった分類器モデルを、第 2 段階以降の実験で使用した。

第 2 段階では、検証方法を 2 つ追加し、比較検証を実施した。追加した検証方法は、個別 10 分割交叉検証と、一人抜き交差検証の 2 つである。本報告では、個別 10 分割交叉検証を、Individual Cross Validation (CV_INDIV)、一人抜き交差検証を Leave-One-Person-Out Cross Validation (LOPO_CV) と呼ぶこととする。

CV_INDIV の実現方法は、CV_ALL と同じであるが、対象とするデータセットが異なる。すなわち、参加者の EDA データを 1 名分ずつに分け 16 個のデータセットとし、各データセットに対して CV_ALL と同等の実験を実施した。これにより、計 16 回分の実験結果を取得した。

また、LOPO_CV の実現方法は、参加者の EDA データを 1 名分ずつに分け、そのうち 15 名分のデータを統合して学習データとし、残り 1 名分のデータを評価データとする、という組み合わせを 1 パターンとし、評価データとする 1 名分のデータを、16 名分順次入れ替えることで、16 パターンの組み合わせのデータに対して、CV_ALL と同等の実験を実施し、計 16 パターン分の実験結果を取得した。

第 3 段階では、時系列 EDA データからある長さの時間範囲で切り出す処理における、ウィンドウのサイズを、5 秒の整数倍で、5 秒から 30 秒までの 6 通りに変化させ、得ら

れた 6 通りのデータセットを用いて、ウィンドウサイズによる認識精度の違いを検証した。

ウィンドウサイズの範囲として 30 秒を上限とした理由は、ウィンドウサイズを長くするほど得られるサンプル数が減り、十分な学習が見込めなくなる懸念が生じ、またウィンドウサイズが 5 秒の場合においても発生頻度が低かったレベル 4 やレベル 5 のサンプルデータ数が、30 秒よりも長いウィンドウサイズでは 0 になってしまい、識別器のクラス数が不足してしまう場合が多く発生するようになったためである。

また、迷いレベルごとに異なるサンプル数に起因する、分類精度への影響を除去するために、アンダーサンプリングとオーバーサンプリングを行った場合の分類精度も追加実験として実施し、F 値を比較確認した。

さらに、レベル 1 から 4 までの 4 つの迷いレベルを 1 つのレベルとみなしたデータセットを別途作成し、迷いレベルを認識する代わりに、迷いの有無を認識する 2 レベル(バイナリ)分類の実験も実施した。さらに、5 レベルと 2 レベルの中間とみなせる、3 レベル分類および 4 レベル分類の実験も実施し、分類精度を確認した。この時のレベル定義を、表 2 に示す。

表 1 で定義していた、5 段階の迷いレベルを左端の列「定義した迷いレベル」に記載している。この迷いレベル 0 から 4 までの 5 段階を、例えば、左端から 2 番目の列「2 レベル分類」に示すように、レベル 0 とレベル 1 の 2 つに割り振ることで、2 段階の迷いレベルとみなす再定義を行う。同様に、3 列目「3 レベル分類」では、定義した迷いレベル 1 と 2 を合わせて、レベル 1 と再定義し、定義した迷いレベル 3 と 4 を合わせて、レベル 2 と再定義した。

表 2 迷いレベルの再定義

定義した 迷いレベル	再定義した迷いレベル		
	2 レベル 分類	3 レベル 分類	4 レベル 分類
0	0	0	0
1	1	1	1
2	1	1	2
3	1	2	3
4	1	2	3

4.2 結果と議論

図 3 に、第 1 段階の実験で作成したデータセットのサンプル数の、迷いレベルごとの内訳を示す。サンプリングは、5 秒のウィンドウサイズと 50% のオーバーラップ率という条件である。図 3 に示すように、迷いレベルごとのサンプル数には大きな開きがあるレベルが存在し、特にレベル 1 以下のレベルのサンプル数とレベル 2 以上のレベルのサンプル数を比べると、少なくとも 2.3 倍以上であることが分かった。

まずは、サンプル数の違いはそのままとし、模擬組立実験で得られたデータそのものに対する悩みレベル分類実験を実施した。一方で、サンプル数の違いによるレベル間の認識精度の違いが生じる可能性も考えられるため、この違いに起因する認識精度への影響を、オーバーサンプリングおよびアンダーサンプリングで吸収したデータセットを作成し、比較する実験を追加で実施したので後述する。

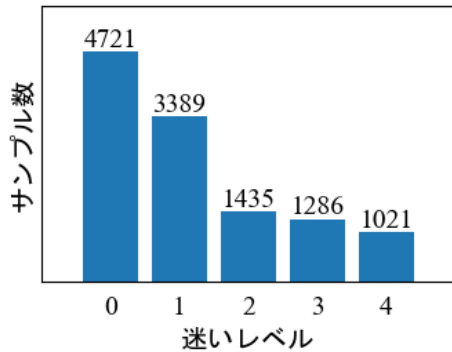


図 3 サンプルデータ数 (CV_ALL)

図 4 は第 1 段階の実験結果を示しており、4 つの分類器モデルの F 値を比較した結果である。

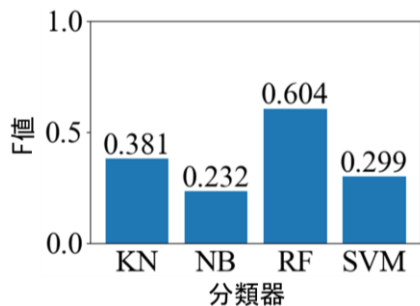


図 4 分類器モデルごとの F 値 (CV_ALL)

RF がマクロ平均の F 値が 0.604 であり、4 つの識別器モデルのうちで最も高い値を示した。したがって、第 2 段階以降の識別器モデルは RF を適用した。

第 2 段階の実験結果として、3 つの検証手法のマクロ平均の F 値を比較した。その結果を図 5 に示す。

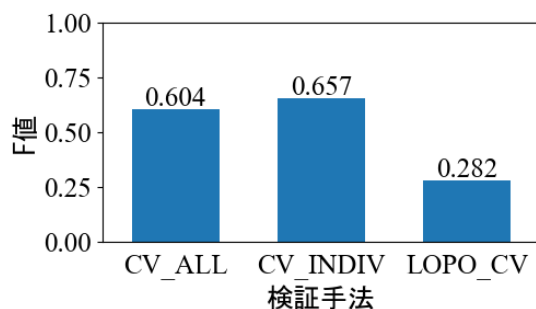


図 5 検証手法ごとの F 値

CV_ALL がマクロ平均の F 値が 0.604 であり、それを上回る 0.657 という結果となったのが CV_INDIV であった。一方、LOPO_CV は 0.282 であり、CV_INDIV から LOPO_CV の間で、0.375 の劣化が見られた。

CV_ALL のデータセットに比べて CV_INDIV のデータセットは個人別のデータセットであるという特徴を持つため、他の参加者の特徴が混じらないデータセットであると考えられる。一方、LOPO_CV のデータセットは、同一人物のデータが学習データと評価データの両方に含まれることがないデータセットであるという特徴を持つ。

参加者ごとに異なる要因、すなわち、個人に依存した要因としては、EDA 信号そのものと、知覚した悩みレベルを 5 段階に評価する際の解釈の 2 つが考えられる。[4]によると同じ負荷条件における EDA データが個人ごとに異なる傾向を持つことが示唆されている。また、悩みレベルの判定基準を表 1 に示した主観的な表現で定義して説明しているため、参加者個人の受け取り方に依存して解釈が変化する可能性もあると考えられ、特にレベル数が多いほど、隣接するレベル間の違いに個人差が大きくなり、共通の指標としにくくなると推察される。

図 6 は RF によって得られた混同行列であり、3 つの上位レベル (2, 3, 4) の区別がうまくいったことを示唆しているが、下位のレベル 1 はしばしばレベル 0 (「迷いなし」) に誤って分類されていた。その理由の 1 つは、レベル間の境界の曖昧さであると考えられる。

レベル0	2737	648	182	78	68
レベル1	1015	1398	153	182	133
レベル2	395	278	618	56	29
レベル3	136	229	45	761	99
レベル4	48	167	25	74	693
	✓レベル0	✓レベル1	✓レベル2	✓レベル3	✓レベル4

図 6 RF の混同行列 (CV_ALL)

以上から、本報告の条件においては、他人の悩みレベルを推測するには不十分であるが、同じ人物であれば、悩みレベルを推測することも可能であると考えられる。

第 3 段階の実験結果として、ウィンドウサイズによる認識精度の違いをマクロ平均の F 値を用いて比較した。その結果を図 7 に示す。CV_ALL と CV_INDIV においては、ウィンドウサイズ 5 秒が最も高い F 値となり、それぞれ 0.604 と 0.657 であった。ウィンドウサイズが 10 秒、15 秒と大きくなるほど F 値は減少するが、15 秒以上はほぼ同じ値で横ばいであった。これの理由としては、より大きなウィンドウに含まれるノイズデータの量がより多くなることが原因である可能性が考えられる。また、他に考えられる理由としては、ウィンドウサイズを大きくするほど、限られた数の EDA データから得られる学習データの数が減少するため、

大きなウィンドウでの学習が不足する可能性である。一方、LOPO_CV はウィンドウサイズ 5 秒が最も低い F 値の 0.282 であり、ウィンドウサイズが 10 秒、15 秒と大きくなるほど F 値は増加し、20 秒以上で 0.370 と横ばいとなった。

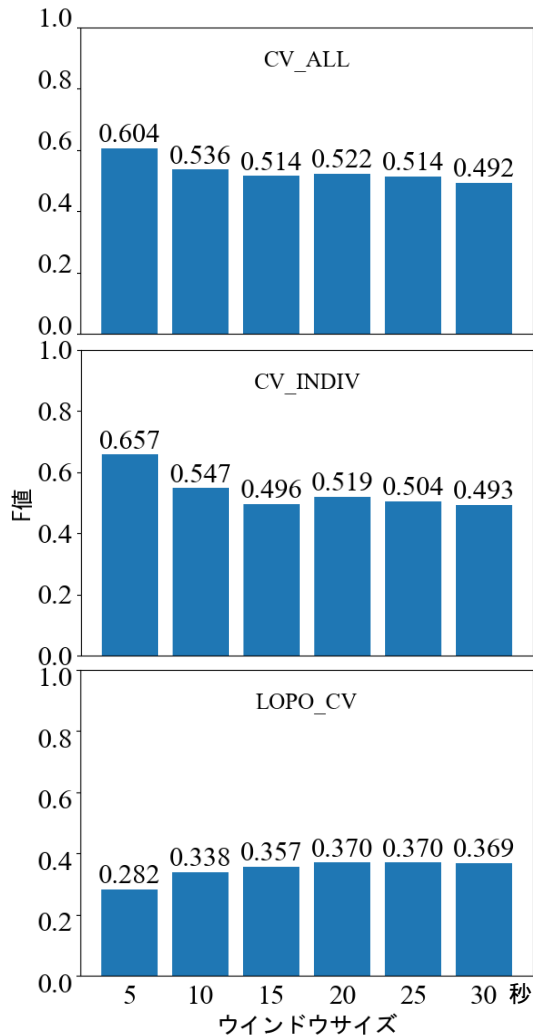


図 7 ウィンドウサイズと F 値の関係

以上から、個人に依存しない迷いレベルを推定するためには、20 秒以上のウィンドウサイズが有利であると考えられる。また、個人の悩みレベルを推定するためには、5 秒のウィンドウサイズがよいと考えられる。

4.3 追加実験と議論

サンプル数の不均衡に起因する F 値への影響がどの程度か確認するために、オーバーサンプリングおよびアンダーサンプリングによりサンプル数をそろえて学習を実施した追加実験の結果を、図 8 に示す。

CV_ALL/CV_INDIV/LOPO_CV において、ウィンドウサイズ 5 秒の場合、オーバーサンプリングにより、それぞれ-0.017/-0.038/0.017 の増減が見られ、アンダーサンプリングにより、それぞれ-0.032/-0.119/0.013 の増減が見られた。同様に、ウィンドウサイズ 20 秒の場合、オーバーサンプリングにより、それぞれ 0.004/-0.025/0.008 の増減が見られ、アンダーサンプリングにより、それぞれ-0.035/-0.093/-0.008 の減少が見られた。

最大の減少幅は、CV_INDIV のウィンドウサイズ 5 秒のアンダーサンプリングにおける-0.119 の減少であった。最大の増加幅は、LOPO_CV のウィンドウサイズ 15 秒のオーバーサンプリングにおける 0.018 の増加であった。

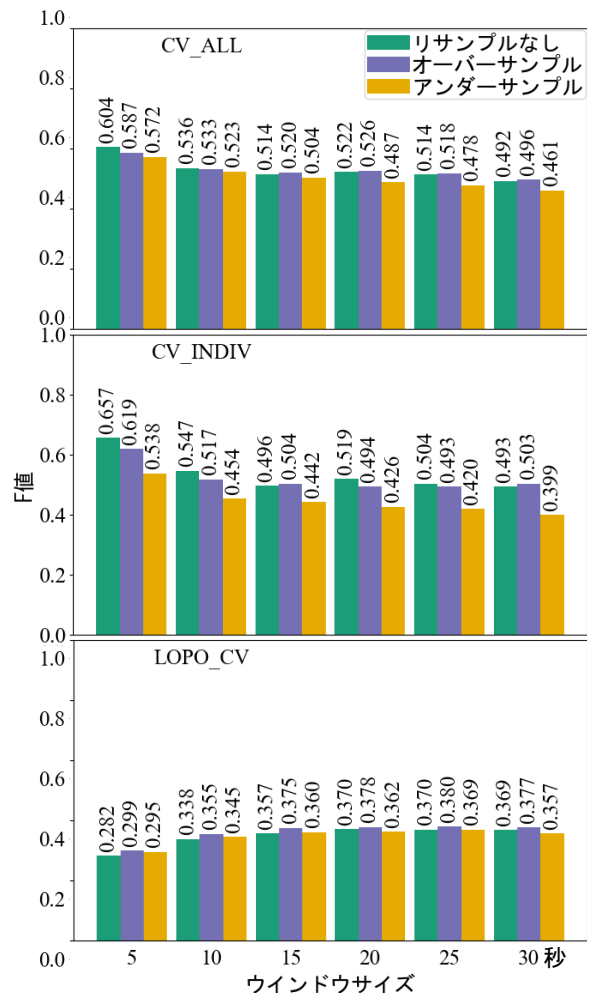


図 8 リサンプリングと F 値の関係

以上より、リサンプリングによる F 値の大幅な改善は見られなかったといえる。学習データのサンプル数が不均衡であっても、各レベルのデータ数としては十分確保できていたのではないかと考えられる。

2 レベル分類および 3 レベル、4 レベル分類のデータセットを用いた追加実験の結果を、図 9 に示す。

2 レベル分類は、CV_ALL/CV_INDIV/LOPO_CV で最大 0.719/0.740/0.627 の F 値となり、3 レベル分類で最大 0.601/0.659/0.281、4 レベルで最大 0.597/0.676/0.330 と比べて、高い F 値であった。一方、2 レベル分類以外の、3/4/5 レベル分類間で差分としては、大きな違いは見られず、3 レベル分類および 4 レベル分類の最大の F 値は、5 レベル分類の最大の F 値である 0.604/0.657/0.370 と同等もしくはそれ以下の値であり、クラス数を減らしても明らかな改善は見られなかった。

以上より、迷いレベル 0 とその他の迷いレベルとに分けた 2 レベル分類では、比較的精度よくクラス分類ができて一方、迷いレベル 1~4 の分類に関しては精度が不足しているといえる。これは、迷いレベル 1 から 4 の境界が

不明瞭であることと、本報告で使用した 11 種類の特徴量では十分に特徴をとらえていない可能性が考えられる。

今後は、心拍数や体温などの他の生体信号を特徴量に導入することで、推定精度と汎化性能の改善を試みる。

謝辞

本研究は JSPS 科研費 JP21K11992 の助成を受けた。

参考文献

- [1] Mario Heinz, Sebastian Buttner, and Carsten Rocker, Exploring users' eye movements when using projection-based Assembly assistive systems, In Proceedings of International Conference on Human-Computer Interaction, pages 259–272, Springer, 2020,
- [2] Pierre Fite-Georgel, Is there a reality in industrial augmented reality? In Proceedings of 2011 10th IEEE international symposium on mixed and augmented reality, pages 201–210, IEEE, 2011,
- [3] Markus Funk, Andreas Bachler, Liane Bachler, Thomas Kosch, Thomas Heidenreich, and Albrecht Schmidt, Working with augmented reality? a long-term analysis of in-situ instructions at the assembly workplace, In Proceedings of the 10th International Conference on pervasive Technologies Related to Assistive Environments, pages 222–229, 2017,
- [4] J, Shukla, M, BarrEDA-Angeles, J, Oliver, and G, C, Nandi, "Feature Extraction and Selection for Emotion Recognition from Electrodermal Activity", *IEEE Trans, Affective Computing*, 12(4), pp, 857-869 (2019)

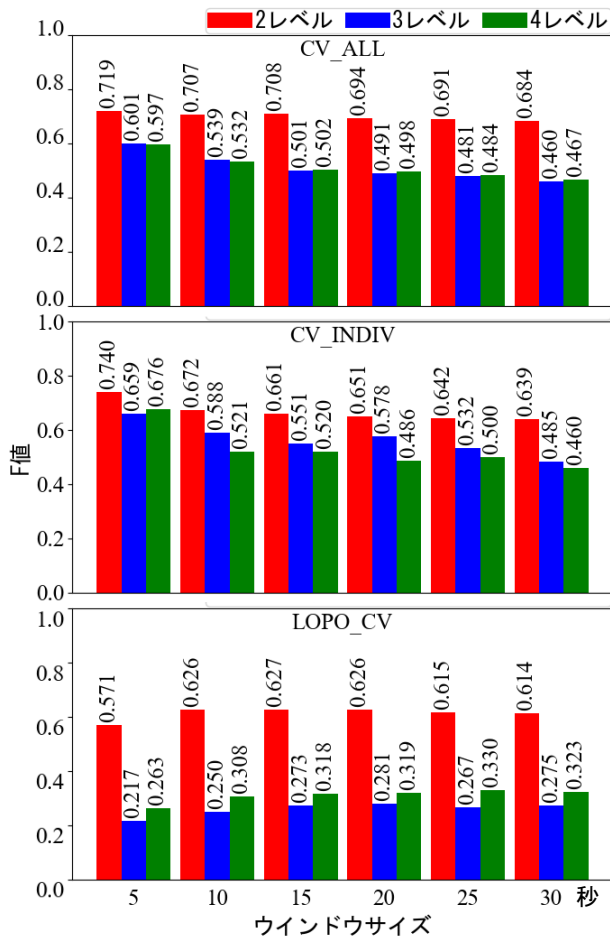


図 9 2レベル～4レベル分類のF値

リストバンド型センサーから得られる情報としては、EDA データ以外に心拍や体温などの情報があるため、それらの情報と組み合わせた特徴量により、悩み認識精度の向上やロバスト性向上を検討する予定である。

また、EDA データのフィルター処理におけるパラメータをチューニングすることと特徴量の工夫や取捨選択することにより個人依存分の特徴を隠蔽し非依存分の特徴のみを切り出す検討を行い、精度向上を検討する予定である。

5. おわりに

EDA データに基づいた組立作業中の迷いレベルを認識するための予備調査を行った。EDA データから 11 種類の特徴量を抽出し、5 段階の迷いレベルでアノテーションした計 16 名分のデータセットを作成した。推定対象者の EDA データを含むデータセットを用いる場合は、ウィンドウサイズ 5 秒の学習データで構築した RF 分類器が、最も高い 0.657 のマクロ平均の F 値を示し、また、推定対象者の EDA データを含まないデータセットを用いる場合は、ウィンドウサイズ 20 秒の学習データで構築した RF 分類器が、最も高い 0.370 の F 値を示した。これにより、推定対象者を含むデータセットが利用することで、EDA データから組立作業の迷いレベルを推定可能な見込みを得た。