

## マルチモーダルデータによるプログラミング作業時の困惑状態推定

倉田 寛大† 辻 愛里†† 藤波 香織††

† 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

†† 東京農工大学 大学院 工学研究院 先端情報科学部門

## 1 はじめに

IT人材の不足が問題となっている昨今、人手を割かずシステムによってさまざまな熟練度のプログラマを多角的に支援することには意義がある。我々はこれまで、相槌機能を実装したインタラクティブな人形を開発し、この人形に対して作業について困惑している事柄を声に出して説明させることで作業者の情報整理を支援することを目指してきた[1]。一方でこの既報では、作業者（プログラマ）が自発的に困惑状態を報告し、それを契機として支援するシステムであったために、支援すべきタイミングを逃す事例が見られた。本研究では、プログラマの視線や姿勢などのマルチモーダルデータを用いた困惑状態推定モデル作成し、困惑検出時にシステム契機で支援を実行することを最終目標とする。本稿ではデータ収集と1センサによる推定を報告する。

## 2 データ収集

## 2.1 概要

困惑状態推定モデルの作成のため、プログラミング作業中のマルチモーダルデータ収集を実施した。被験者は、情報系の学部及び大学院に所属する20代の10名（男性：5名、女性：5名）である。実験では被験者に各センサを装着した状態（図1）で、プログラミングタスクを与えた。タスク中、被験者にはデータ収集後の機械学習の際のラベル付けのため、1分間隔で表示されるポップアップ（図2）から、現在の困惑状態を5段階で回答させる。なおタスク中のWebブラウザを用いた調査は可能とした。

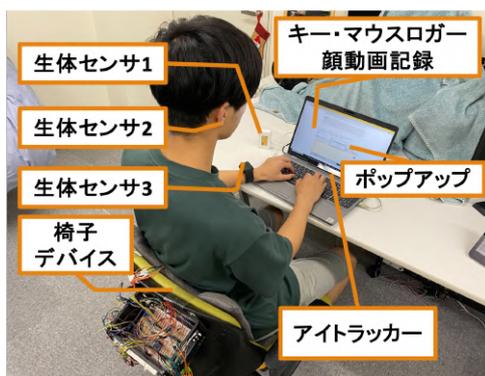


図1: 実験環境

Confusion estimation during programming based on multimodal sensing

† Kandai KURATA †† Airi TSUJI †† Kaori FUJINAMI

† Department of Bio-Functions and Systems Science, Tokyo University of Agriculture and Technology

†† Division of Advanced Information Technology and Computer Science, Tokyo University of Agriculture and Technology

図2: 困惑状態回答用ポップアップ  
(左から困惑レベル1,2,3,4,5)

表1: 5段階困惑レベル

レベル	例
1	迷っていない
2	文法がわからないなど、何とかひとりで解決できる・弱
3	文法がわからないなど、何とかひとりで解決できる・強
4	アルゴリズムがわからないなど、誰かに話して整理したい・弱
5	アルゴリズムがわからないなど、誰かに話して整理したい・強

困惑状態については、実験開始前に表1に示す回答例を被験者に提示することで、被験者間で5段階の基準に揺れが生じないように注意した。

また適切な問題難易度の設定のため、被験者には事前にpaiza[2]を用いてプログラミングスキルの測定を実施させ、実験タスク用の問題も同サイトから各被験者の適正レベルの問題を選定した。なお個々の問題における難易度以外の差による影響を軽減するため、同じ問題を複数の被験者に与えることのないよう設計した。また被験者の疲労を考慮し、実験時間を2時間程度と定めたため、タスク数は被験者によって1から3問と前後した。タスク間には最低5分間の休憩を設けて、安静時のマルチモーダルデータの収集を行った。このデータは、キーボード・マウスおよびアイトラッカーから得られたデータを除外した後、困惑なし状態のデータとみなして取り扱う。

## 2.2 収集データ

各センサによる収集データの詳細を表2に示す。

表2: 収集データ一覧

使用デバイス	収集データ	接触/非接触
生体センサ1 エモコアイ [3]	脈拍 (没入度・眠気度など)	非接触
生体センサ2 心拍センサ [4]	脈拍	接触
生体センサ3 E4 リストバンド [5]	脈拍, 皮膚抵抗, 加速度	接触
椅子デバイス [6]	29箇所での圧力	接触
キーボード・マウス	キー・マウスログ 表情, 瞳孔	接触
Web カメラ	表情, 瞳孔	非接触
アイトラッカー TOBII PRO ナノ [7]	視線遷移, 瞳孔径	非接触

長時間の実験に際して、装着型センサの装着感が被験者のパフォーマンスに与える悪影響を考慮して、接触型デバイスは最低限のものとした。椅子デバイスに関しては小林ら [6] によって開発された、座面および背もたれに合計 29 個の圧力センサを設置し、それらを Raspberry Pi と接続することで圧力データが取得可能なデバイスを使用した。また、各センサでサンプリング周波数が異なるが、一つのセンサデータから一つの分類器を作成し、センサと同数の分類器の出力の多数決を最終出力とするため、データ数の統一のためのアップサンプリングなどは実施しない。

### 3 困惑状態推定モデルの作成と性能評価

2.2 節で示した収集データのうち、本稿ではエモコアイによって得られたデータによる結果のみ報告する。

#### 3.1 使用データ

エモコアイは 24GHz ドップラー方式を用いて計測する脈の情報から、没入度（集中）、眠気度（退屈、怠惰）、活性度（緊張、元気）、消耗度（疲労）といった人の感情推定データを取得できるデバイスであり、これら 4 項目とこれらをもとに計算されるワークスコア、さらに脈拍、短期間脈拍の計 7 項目のデータが 1Hz で記録される。実験後、得られたデータを被験者からの困惑状態の回答を基に 5 クラスに分類した。なお、被験者からの困惑レベルの回答は最短で 1 分間隔であるが、回答が得られてから前回の回答直後までのデータを同じ困惑レベルとみなしてラベル付けを行った (図 3)。困惑レベル別データ数を表 3 に示す。各困惑レベルで十分な数のデータが収集されたことから、エモコアイから得られた 7 項目を特徴量として用いる。

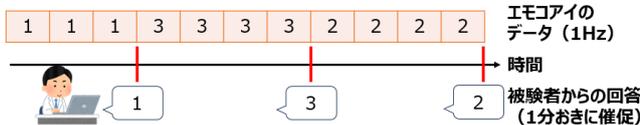


図 3: ラベル付けの概要

表 3: 困惑レベル別データ数

困惑レベル	1	2	3	4	5
データ数 [個]	39633	10374	6209	4920	6559

#### 3.2 機械学習による困惑状態の推定

分類器として Support Vector Machine (以下, SVM) と Random Forests (以下, RF) を用い, 3.1 節で示した特徴量群を用いた場合について F 値を算出した。なお, 分類器の実装は Python のライブラリ scikit-learn (バージョン 0.24.2) を用いて行い, SVM のカーネル (kernel) を linear, poly, rbf, sigmoid の 4 パターン, 正規化パラメータ (c) を 0.1, 1, 10, 100 の 4 パターンでグリッドサーチを行い, 最も F 値の高いパラメータでテストを行った。RF においても, 弱分類器の数 (n estimators) を 10, 100, 500, 1000 の 4 パターンで変化させ最も F 値の高いパラメータを採用した。これら分類器を用いて 5 段階の困惑レベルについて層化 10 分割交差検証を行った結果, SVM の F 値が 0.44, RF の F 値が 0.751 となった, SVM と比べて RF の方が高い F 値を示す結果となった。最後に RF による分類器の混同行列を図 4 に, 各レベルの評価指標を表 4 に示す。

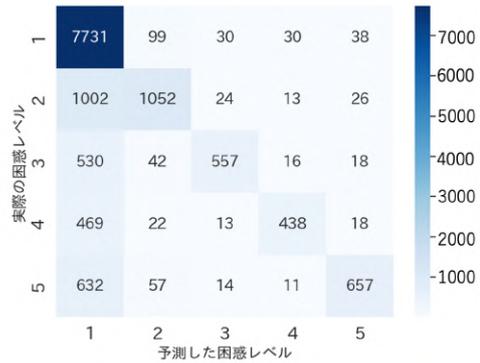


図 4: 混同行列 (RF)

表 4: 各レベルの評価指標 (RF)

レベル	precision	recall	F 値
1	0.746	0.975	0.845
2	0.827	0.497	0.621
3	0.873	0.479	0.619
4	0.862	0.456	0.596
5	0.868	0.479	0.617

### 4 考察

3.2 節に示した結果から, 0.751 の F 値で 5 段階の困惑状態を分類可能であることが明らかになった。SVM による分類器の結果と比較すると高い F 値であるが, 図 4 から誤分類のほとんどが困惑レベル 1 に分類されていることが確認できる。表 4 の通り, レベル 2 以上における recall が低いことから, ユーザーが支援を必要としているにも関わらず, モデルは困惑なしと分類してしまう可能性が高いことがわかる。これに関しては現状, スライディングウィンドウによる特徴量の計算を実施せず, 得られた生データを特徴量としているため, ウィンドウごとの平均や分散などを新たに特徴量とすることで, 改善が望めると考えられる。

### 5 おわりに

本研究では, プログラマの作業中のマルチモーダルデータから困惑状態推定モデルを作成し, 困惑状態の推定を目指した。1つのセンサデータを基に分類器を作成したところ, 0.751 の F 値で 5 段階の困惑状態を分類可能であったが, レベル 2 以上は再現率が 0.5 以下であり, 支援契機を逸失しがちである。今後は, 分類器の学習に用いる特徴量の検討を進め, パラメータチューニングを行うことで性能の向上に努める。センサモダリティごとの分類モデルを作成後, それらの結果の多数決によって出力を決定する方針で実装を進める。

### 参考文献

- [1] 倉田 他. 人形との相互作用によるプログラミング時の思考整理支援. HAI シンポジウム 2023.
- [2] paiza. IT エンジニア向け国内最大の転職・就職・学習プラットフォーム. <https://paiza.jp>. (最終アクセス 2024 年 1 月 7 日)
- [3] FUJITSU. 感情推定センサー エモコアイ ® 評価キット販売開始のお知らせ. <https://www.fcl.fujitsu.com/resources/news/press-releases/2023/20230331.html>. (最終アクセス 2024 年 1 月 9 日)
- [4] SWITCH SCIENCE. 心拍センサ. <https://www.switch-science.com/products/1135>. (最終アクセス 2024 年 1 月 9 日)
- [5] empatica. E4 wristband. <https://e4.empatica.com/e4-wristband>. (最終アクセス 2024 年 1 月 9 日)
- [6] 小林 他. ながら運動促進システム実現のための基礎検討. 第 20 回情報科学技術フォーラム (FIT2021) .
- [7] tobii. TOBII PRO ナノ. <https://www.tobii.com/ja/products/eye-trackers/screen-based/tobii-pro-nano>. (最終アクセス 2024 年 1 月 9 日)