

# 対人交流を考慮したスマートフォンログによるストレス推定

霧生和樹<sup>†1</sup> 上西康平<sup>†1</sup> 山本直樹<sup>†2</sup> 濱谷尚志<sup>†2</sup> 落合桂一<sup>†2</sup> 深澤佑介<sup>†2</sup>  
木本勝敏<sup>†2</sup> 沖村宰<sup>†3</sup> 寺澤悠理<sup>†3</sup> 前田貴記<sup>†3</sup> 太田順<sup>†1</sup>

**概要**：本研究ではスマートフォンのログを用いて、ストレスを推定する手法の確立を目指した。ストレスと関連のある特徴量とスマートフォンのログを結び付けることで特徴量の算出を行い、ストレスの指標となる値を正解データとして機械学習を行う。対人交流のようにログから直接算出できないものについては、その指標を正解データとした推定モデルを別に構築し、特徴量の算出を行った。提案手法の有効性を実験により得たデータを用いて評価した。

**キーワード**：ストレス、スマートフォン、LF/HF、対人交流

## Stress Estimation by Smartphone Log Considering Interpersonal Exchanges

KAZUKI KIRIU<sup>†1</sup> KOHEI KAMINISHI<sup>†1</sup> NAOKI YAMAMOTO<sup>†2</sup>  
TAKASHI HAMATANI<sup>†2</sup> KEIICHI OCHIAI<sup>†2</sup> YUSUKE FUKAZAWA<sup>†2</sup>  
MASATOSHI KIMOTO<sup>†2</sup> TSUKASA OKIMURA<sup>†3</sup> YURI TERASAWA<sup>†3</sup>  
TAKAKI MAEDA<sup>†3</sup> JUN OTA<sup>†1</sup>

**Keywords**: Stress, Smartphone, LF/HF, Interpersonal exchanges

### 1. 序論

#### 1.1 研究背景

近年メンタルヘルスケアに関心が高まっており、平成 27 年に義務化されたストレスチェックは、個々人にストレスを把握させることでセルフケアを促し、メンタル不調を未然に防ぐ目的がある[1]。ストレスの評価にはアンケートが用いられることが多いが、定期的に行うのは手間がかかる、自身の評価基準を維持することが難しい、無自覚のストレスを評価できない、得点操作の可能性などの問題点がある[2]。すなわち、アンケートを用いて個々人が日常的に正確なストレスを把握し、セルフケアに繋げることは難しい。一方で、客観的な評価方法として生体情報を用いた手法も存在する。例えば、血液中のアルギナーゼ I [3]や心拍変動[4]などを解析することによってストレスを評価することができる。アンケートより定量的な評価ができ有効である一方で、特別なデバイスを装着する必要があったり、定期的な測定が必要であったりと手間がかかりこれまで無関心であった人にセルフケアへの気づきを与えることは難しい。

そこで、WHO はスマートフォンを用いてセルフケアを行うことが有効であると述べている[5]。現在スマートフォンは保有率も高く頻繁に持ち運ぶものとなっており、また、スマートフォンへのアクセスと使用者の精神状態に関連が

指摘されている[6]ことから、スマートフォンから得られる複数のセンサや使用情報のログは使用者の精神状態を推定するのに有効であるとされている。

以上より、スマートフォンのログを用いてストレスを推定することができれば、目視できなかったストレスを簡便かつ客観的に把握することが可能となり、スマートフォンの保有率の高さを考えると、多くの人に気づきを与えることができ、ストレスのセルフケアに対して有用であるといえる。

#### 1.2 ストレスと関連のある行動

スマートフォンのログからストレスを推定するモデルを構築するには機械学習を用いる。この際スマートフォンのログから特徴量を算出するが、ストレスと関連のある行動や項目と結び付けることで、推定に有用な特徴量が得られるだけでなく、推定根拠の特徴量が得られたときに行動と紐づけられ、フィードバックにおいても有用であると考えられる。そこで、表 1 に従来研究の知見に基づいて作成したストレスと関連のある行動や項目のリストを示す。浦川らはストレス対応能力の低い人ほど外出が少なく、趣味が少ないことを述べている[7]。また、照度や気圧といった取り囲む環境もストレスと関連しうる[8][9]。対人交流については、浦川らによって関連が指摘されているだけでなく、ストレスチェックの質問項目中にも相談ができる相手につ

†1 東京大学  
The University of Tokyo  
†2 (株)NTT DOCOMO  
NTT DOCOMO, Inc.

†3 慶應義塾大学  
Keio University

表 1 ストレスと関連のある行動や指標となる項目

| カテゴリー | 指標となる行動や項目の例                       | 従来研究 |
|-------|------------------------------------|------|
| 外出    | 外出しているか, 歩いているか<br>訪問場所            | [7]  |
| 環境    | 周囲の明るさかどうか                         | [8]  |
|       | 気圧かどうか                             | [9]  |
| 趣味    | 趣味や遊びを楽しめているか<br>休みの日に気晴らしができていますか | [7]  |
| 対人交流  | スマートフォン上での交流                       | [1]  |
|       | 直接会話をしているか                         | [7]  |
| 睡眠    | 睡眠時間, 起床時刻, 就寝時刻                   | [10] |
| 食事    | 食事数と時刻, 所要時間                       | [11] |

いてなど対人交流に関する項目が複数存在することを踏まえると[1], ストレスと関連が大きい項目といえる。最後に, 睡眠や食事もストレスとの関連が指摘されている[10][11]. これらの項目に基づいて特徴量を設計することが精度の向上とフィードバックにおいて有用であると考えられる。

### 1.3 関連研究

本節では, ストレスを推定する研究に関してまとめる。まず, Apple Watch (Apple 社) は身に着けるだけで心拍変動を取得でき, ストレス計測も可能であるとされている[12]. 生体情報を用いているため精度や客観性を確保しやすい一方で, 日常生活中で用いるためには特別なデバイスが必要であるために, 多くの人にストレスを把握してもらいセルフケアを促すことは難しい。Ben-Zeev らはスマートフォンの音声センサ, GPS, 加速度センサ等から取得したデータを用いて, 主観評価のストレスを推定している[13]. スマートフォンのみでのストレス推定を行っている一方で, 参加者自身が記録した主観的なストレスを正解データとしているため, 1.1 節で述べたように自身の評価基準を維持することの難しさや, 無自覚のストレスを反映できないことがあり, 適切なストレスの評価ができない可能性がある。Yamamoto らはスマートフォンから取得したデータを用いて生体情報である LF/HF (Low Frequency / High Frequency) によって評価されたストレスを推定している[14]. スマートフォンのみで生体情報によって評価されたストレスを推定している。しかし, 会社員を対象としているため業務用のスマートフォンよりデータを取得しており, 私用のスマートフォンを用いた場合の効果は明らかになっていない。また, 特徴量が移動, 環境, 趣味, および対人交流の一部に関してしか算出されていない。

### 1.4 研究目的

前節までの議論を踏まえて, 本研究においては私用のスマートフォンを用いて取得したデータを用いて生体情報によって評価されたストレスを推定することを目的とする。

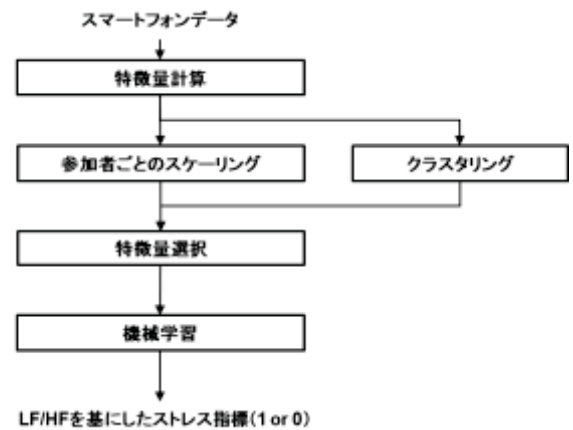


図 1 ストレス推定手法の流れ

手法は[14]をベースとし, 私用のスマートフォンで取得したデータに対して検証を行う。また, ストレスとの関連が大きいとされる対人交流に着目し, 特徴量を設計し, 追加することで精度の向上を図る。

## 2. 提案手法

[14]をもとにした手法全体の流れを図 1 に示す。スマートフォンの各データを取得したら, ストレスと関連のある行動をもとに特徴量を計算する。その特徴量を用いてクラスタリングを実施することで, 行動が類似している参加者同士でクラスタを作成し, クラスタごとに機械学習を実施する。機械学習を実施する前に, 参加者ごとに平均 0, 分散が 1 となるようにスケールリングを行う。スケールリング後の特徴量から推定に有効と考えられる特徴量をクラスタごとに選択し, 日ごとの LF/HF が各参加者の平均をより大きい場合 1, 以下の場合 0 と設定したものを正解データとした機械学習を行う。以下の項では, [14]と手法の異なる特徴量計算と機械学習の項目について述べる。

### 2.1 特徴量計算

表 1 に示したストレスと関連のある行動や項目のリストとを照らし合わせて特徴量の設計を行うことを 1.2 節で述べた。[14]で作成されている特徴量をリストと照らし合わせると, 位置情報と加速度センサ, Activity を用いた特徴量 (63 個) は移動や行動を示す指標であり, 外出に関する項目の特徴量となる。次に気圧と照度を用いた特徴量 (18 個) は参加者を取り囲む環境を示す指標であり, 環境に関する項目の特徴量となる。次に, イヤホンジャックの使用, 画面 ON/OFF, 全アプリ起動回数, 電池, ストレージ, 曜日を使用した特徴量 (29 個) はスマートフォンを通じて音楽を聴くなど息抜きの指標であり, 趣味に関する項目の特徴量となる。最後に電話と SMS の起動回数を使用した特徴量 (8 個) はスマートフォン上での対人交流の指標となり, 対人交流の特徴量となる。[14]においては, 他に業務アプリの起動回数が特徴量として用いられているが, 私用のスマ

スマートフォンを対象とする本研究にそぐわないため除外する。以上の計 118 個の特徴量に、本研究の対象である対人交流に関する特徴量を追加する。

対人交流については、スマートフォン上での対人交流と対面での対人交流の 2 種類がある。前者については各アプリの起動履歴を参照することで特徴量化できる。一方、対面の対人交流についてはスマートフォンにより取得するデータから直接特徴量として算出することができない。そこで、あらかじめスマートフォンのデータと対面の対人交流の指標を用いて機械学習を行い、推定するモデルを作成し、その出力を特徴量として組み入れることで対処する。以上の議論より、対人交流に関する特徴量として以下の 120 の特徴量を追加する。

- アプリ起動履歴を使用した特徴量 (24 個)  
アプリの起動履歴を用いて各携帯会社のキャリアメール、PC メール (Yahoo Mail と Gmail の合算)、LINE, Twitter, Facebook, Instagram それぞれの起動回数を 0 時～6 時 (Q1), 6 時～12 時 (Q2), 12 時～18 時 (Q3), 18 時～24 時 (Q4) の各区間において集計する。
- 対人交流 (対面) に関する特徴量 (96 個)  
対人交流 (対面) に関しては、スマートフォンのログから直接算出することができないため、別に推定するモデルを構築し、その出力を用いることを述べた。本研究においては、1 時間ごとにスマートフォンのデータから 96 の特徴量を算出し、その 1 時間における会話時間が閾値以上を 1、閾値未満を 0 としたものを正解データとして学習を行った。閾値としては 10 分, 20 分, 30 分, 40 分を設定し、計 4 つの推定モデルを構築した。この 4 つの推定モデルから出力された該当日の 0 時～23 時の各 1 時間における推定値である 1 または 0 を特徴量 (4×24 個) とする。

以上の手順をもって計 238 個の特徴量を算出する。以上の特徴量算出の流れを図 2 にまとめる。



図 2 特徴量算出の流れ

## 2.2 使用する学習器

推定には K Nearest Neighbor (KNN) [15], Support Vector Machine (SVM) [16], Random Forest [17], Multilayer Perceptron (MLP) [18], Xgboost [19] によるアンサンブル学習 [20] を使用する。学習用のデータセットを用いて個別に学習させ、ストレスを推定するモデルを作成する。検証時には特徴量を各分類器に入力しそれぞれ出力を行い、結果を多数決により統合することで最終的な出力とする。

## 3. 実験

### 3.1 実験手法

実験参加者は 20 代の大学生 19 人 (男性 16 人, 女性 3 人) であった。参加者の日常生活において、心拍データは心拍センサ “Mybeat WHS-1” (ユニオンツール社) を用いて 9 時～20 時の間取得を行い、スマートフォンのデータはアプリを用いて 24 時間取得を行った。このデータ取得を最大 6 週間行った。本実験は東京大学倫理審査専門委員会の許可 (審査番号: 17-139) のもとに行い、書面にてインフォームドコンセントを行った。

### 3.2 実際に取得したデータと評価方法

3.1 節のデータ取得により 663 日分のデータ (1 : 0 = 305 日 : 358 日) を得た。本研究ではクラスタ数を 4 とし、手法の評価はクラスタごとに 1 人を検証データ、残りを学習用とするのを人数分繰り返す交差検証を行う。精度は Accuracy (全データに対し 1 と 0 を正しく分類した割合), Sensitivity (1 のデータのうち正しく 1 と分類できた割合), Specificity: (0 のデータのうち正しく 0 と分類した割合) の 3 値を用いて評価する。

### 3.3 従来手法との比較

[14] 中の手法の私用のスマートフォンへの適用性を検証するために、[14] に掲載されている精度 (従来モデル) と [14] で用いられた特徴量 (118 個) をのみを用いて本研究におけるデータに適用した推定精度 (従来手法) を比較する。それぞれの精度を表 2 にまとめる。従来モデルと従来手法を比べると、すべての指標において数値が低下している。これは業務用のスマートフォンに対して私用スマートフォ

表 2 従来モデルと従来手法における推定精度の比較

| 手法    | Accuracy | Sensitivity | Specificity |
|-------|----------|-------------|-------------|
| 従来モデル | 0.711    | 0.606       | 0.773       |
| 従来手法  | 0.667    | 0.544       | 0.771       |

表 3 従来手法と提案手法における推定精度の比較

| 手法   | Accuracy | Sensitivity | Specificity |
|------|----------|-------------|-------------|
| 従来手法 | 0.667    | 0.544       | 0.771       |
| 提案手法 | 0.679    | 0.544       | 0.793       |



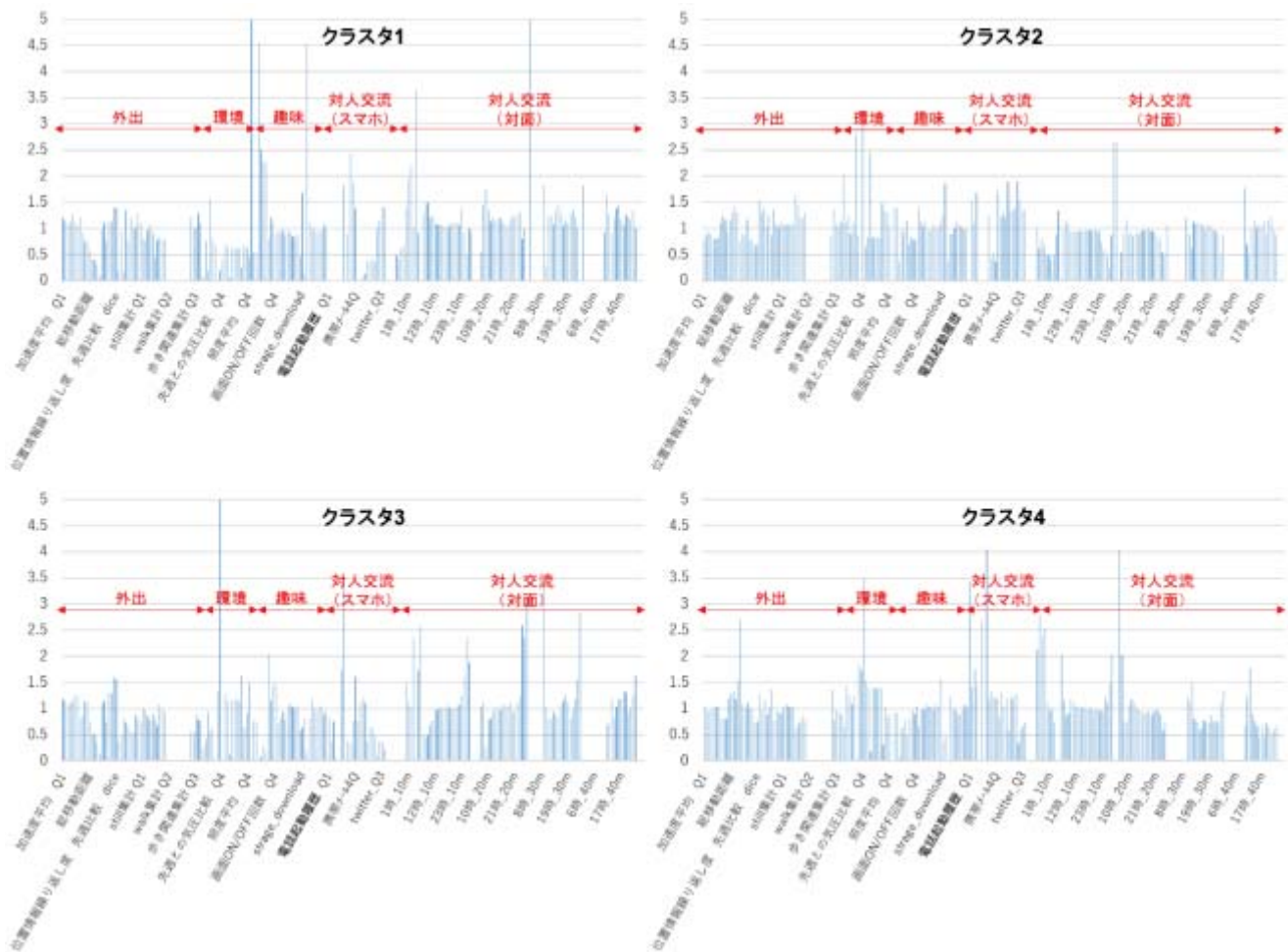


図 3 各クラスタの特徴[横軸：特微量，縦軸：それぞれのクラスタの平均 / 全クラスタの平均]

ンの方が参加者ごとに多様な行動を反映しているからだと考えられる。次に、従来手法と本研究における提案手法における推定結果を表 3 にまとめる。従来手法と提案手法を比較すると Accuracy, Specificity の指標において向上しており、対人交流の特微量を考慮することの有効性を示している。

### 3.4 各クラスタの特徴

各クラスタにおける特徴を図 3 に示す。横軸は 238 個の特微量を示し、縦軸はそれぞれの特微量における各クラスタの平均を全体の平均で割った値を示している。

まず、クラスタ 1 は外出に関する特微量において 1 を下回るものが多く、移動が少ないことが分かる。また、メールでの交流は多い一方で、LINE 等の交流は少ないことが見てとれる。対面での対人交流の特微量は 1 より大きいものが多く、対面での対人交流をよく行っているクラスタであることが分かる。次にクラスタ 2 を見ると、外出に関する特微量の値から移動を多く行っていることが見てとれ、LINE や Twitter を用いて交流を多く行っている傾向がある。クラスタ 3 については、外出に関する特微量において 1 を下回るものが多く一方で、位置情報の繰り返し度の値は大

きく、移動が少なく、決まった移動が多いことが分かる。画面 ON/OFF の値の大きさからスマートフォンへのアクセスが多いことが見てとれる一方で、スマートフォン上での対人交流の特微量においては 1 を下回るものが多く、交流を行わない傾向にある。最後にクラスタ 4 については、まず、最大移動距離において 1 を超えるなど移動が多いクラスタである。また、スマートフォン上での対人交流の特微量において 1 を超えるものが多いことから交流が盛んであることが見てとれる。一方、対面での対人交流については値が 1 を下回り、対面での交流をあまり行っていないことが分かる。

そして、各クラスタにおける精度を図 4 に示す。クラスタごとに精度が異なり、クラスタ 1 や 3 において Accuracy が 0.7 を超えた一方で、クラスタ 2 においては Accuracy が 0.624 しかない。すなわち、今後精度を向上させるにはクラスタ 2 のような制度の低いクラスタに着目し、その特徴に合わせて特微量を再設計、追加する必要がある。

### 3.5 推定に有効とされた特微量

Random Forest を用いて各クラスタにおいて推定に有効とされた上位 10 の特微量を表 3 に示す。

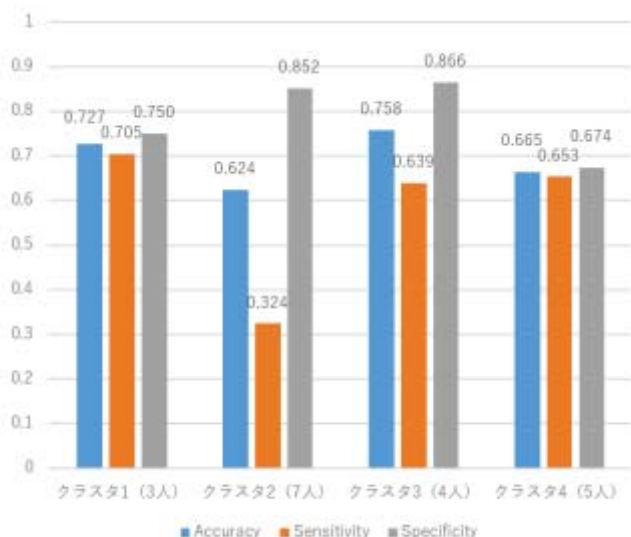


図 4 各クラスターの推定精度

クラスター1は外出に関する特徴量が多く、ストレスと外出の関連がみられる。また、スマートフォン操作やスマートフォン上での対人交流に関する特徴量もあり、ストレスがスマートフォンの操作に反映されているといえる。クラスター2は外出に関する特徴量よりも Battery やアプリや LINE の使用などスマートフォン操作にストレスが反映されやすい傾向にある。クラスター3は、上位は外出に関する特徴量であり、クラスター1, 2同様スマートフォン操作に関する特徴量も上位に入っている。また、照度や気圧の特徴量が上位に入っており、参加者を取り囲む環境とストレスに関連の強いクラスターとなっている。クラスター4は外出に関する特徴量と関連の強いクラスターとなっている。会話の推定値を用いた特徴量も入っており、対面での対人交流との関連も見えてくる。

最後に全体としては、外出に関する特徴量が多く推定に有効であるとされており、従来手法の特徴量の有効性が示されている。一方で、本研究において追加した対人交流に関する特徴量についても推定に有効とされており、[14]においては主に移動とスマートフォン操作においてしか議論がなされていなかったが、新たな観点から議論可能となり対人交流を考慮することの有用性が示されたといえる。

#### 4. 結論と今後の展望

本研究においては、私用のスマートフォンを用いて取得したデータを用いて生体情報によって評価されたストレスを推定することを目的とした。対人交流とストレスの関係に着目し、従来手法の118個の特徴量に対人交流に関する特徴量として120個の特徴量を追加して精度の向上を図った。参加者実験において手法を評価し、従来手法より Accuracy, Specificity が向上し、対人交流の特徴量を組み入

表 4 Random Forest によって有効とされた各クラスターにおける上位 10 の特徴量

| クラスター | 順位 | 特徴量                    |
|-------|----|------------------------|
| 1     | 1  | 位置情報繰り返し度 先週比較 Jaccard |
|       | 2  | 画面 ON/OFF 回数 Q3        |
|       | 3  | 総移動距離                  |
|       | 4  | 新しく訪れた場所               |
|       | 5  | 最大移動距離                 |
|       | 6  | 位置情報繰り返し度 先週比較 Dice    |
|       | 7  | Twitter Q4             |
|       | 8  | foot 集計 1 日            |
|       | 9  | ストレージ (写真)             |
|       | 10 | Battery Q3             |
| 2     | 1  | 全アプリ起動履歴 Q1            |
|       | 2  | 加速度分散 Q2               |
|       | 3  | 加速度分散 Q3               |
|       | 4  | 総移動距離                  |
|       | 5  | Battery Q3             |
|       | 6  | LINE Q3                |
|       | 7  | ストレージ (ダウンロード)         |
|       | 8  | Battery 1 日            |
|       | 9  | Battery Q4             |
|       | 10 | ストレージ (カメラ)            |
| 3     | 1  | tilt 集計 Q4             |
|       | 2  | 最大移動距離                 |
|       | 3  | still 集計 Q2            |
|       | 4  | 気圧平均 Q2                |
|       | 5  | 照度平均 Q3                |
|       | 6  | 画面 ON/OFF 回数 Q2        |
|       | 7  | 画面 ON/OFF 回数 Q3        |
|       | 8  | 全アプリ起動履歴 Q4            |
|       | 9  | 気圧平均 Q3                |
|       | 10 | 全アプリ起動履歴 Q3            |
| 4     | 1  | still 集計 Q3            |
|       | 2  | foot 集計 Q2             |
|       | 3  | 加速度分散 Q4               |
|       | 4  | 歩き関連集計 1 日             |
|       | 5  | 加速度分散 Q3               |
|       | 6  | 17 時における推定 (20 分を閾値)   |
|       | 7  | 加速度平均 Q4               |
|       | 8  | still 集計 1 日           |
|       | 9  | 気圧平均 Q3                |
|       | 10 | ストレージ (カメラ)            |

れる有効性が示された。また、推定に有効とされた特徴量においても対人交流に関する特徴量が上位に入っており、このことからストレスの推定に対人交流を考慮することの有用性が示されたといえる。

今後は、本研究においては導入できなかった食事や睡眠に関する特徴量を加えることで精度の向上を図り、またセルフケアを促すフィードバックの方法についても検討を行う。

## 参考文献

- [1] 厚生労働省: <https://www.mhlw.go.jp/bunya/roudoukijun/anzeneisei12/pdf/150507-1.pdf>
- [2] 堀 輝, 香月 あすか, 菅 健太郎, 吉村 玲児: “客観的なストレス評価方法について,” 日本職業・災害医学会会誌, Vol. 66, pp. 330-334, 2018.
- [3] Keiki Ogino, Tatsuo Ito, Eri Eguchi, Kenjiro Nagaoka: “Association of Arginase I or Nitric Oxide-Related Factors with Job Strain in Healthy Workers,” *PLOS ONE*, Vol. 12, No. 4, 2017.
- [4] Bruce Pomeranz, Robert J. B. Macaulay, Margaret A. Caudill, Ilan Kutz, Dan Adam, David Gordon, Kenneth M. Kilborn, A. Clifford Barger, Daniel C. Shannon, Richard J. Cohen, Herbert Benson: “Assessment of Autonomic Function in Humans by Heart Rate Spectral Analysis,” *American Journal of Physiology*, Vol. 248, No. 1, H151-H153, 1985.
- [5] Emily Anthes: “Pocket Psychiatry: Mobile Mental-Health Apps Have Exploded onto the Market, but Few Have Been Thoroughly Tested,” *Nature*, Vol. 532, No. 7597, pp. 20-23, 2016.
- [6] Russell B. Clayton, Glenn Leshner, Anthony Almond: “The Extended iSelf: The Impact of iPhone Separation on Cognition, Emotion, and Physiology,” *Computer-Mediated Communication*, Vol. 20, No.2, pp. 119-135, 2015.
- [7] 吉井 清子, 近藤 克則, 平井 寛, 松田 亮三, 斎藤 嘉孝, 村田 千代栄: “ストレス対処能力 SOC(sense of coherence)と社会経済的地位と心身健康,” 公衆衛生, Vol. 69, No. 10, pp. 825-829, 2005.
- [8] 杉本 賢: “照明環境要素の生体への影響に関する研究—照度と生理的負担の関係 (その 2)—,” 照明学会誌, Vol. 65, No. 4, pp.171-175, 1981.
- [9] 佐藤 純: “気象変化による慢性痛悪化のメカニズム,” 日本生気象学会雑誌, Vol. 40, No. 4, pp.219-224, 2003.
- [10] Robert A. Hicks, Eliot R. Garcia: “Level of Stress and Sleep Duration,” *Perceptual and Motor Skills*, Vol. 64, No. 1, pp. 44-46, 1987.
- [11] Susan J. Torres, M.Nutr.Diet, Caryl A. Nowson: “Relationship between Stress, Eating Behavior, and Obesity,” *Nutrition*, Vol. 23, No.11-12, pp.887-894, 2007.
- [12] David Hernando, Surya Roca, Jorge Sancho, Álvaro Alesanco, Raquel Bailón: “Validation of the AppleWatch for Heart Rate Variability Measurements during Relax and Mental Stress in Healthy Subjects,” *Sensors*, Vol. 18, No. 8, 2018.
- [13] Dror Ben-Zeev, Emily A. Scherer, Rui Wang, Haiyi Xie, Andrew T. Campbell: “Next-Generation Psychiatric Assessment: Using Smartphone Sensors to Monitor Behavior and Mental Health,” *Psychiatric Rehabilitation Journal*, Vol. 38, No. 3, pp. 218-226, 2015.
- [14] Naoki Yamamoto, Keiichi Ochiai, Akiya Inagaki, Yusuke Fukazawa, Masatoshi Kimoto, Kazuki Kiriu, Kouhei Kaminishi, Jun Ota, Tsukasa Okimura, Yuri Terasawa, Takaki Maeda: “Physiological Stress Level Estimation Based on Smartphone Logs,” *Proceedings of The 11th International Conference on Mobile Computing and Ubiquitous Networking*, 2018.
- [15] N. S. Altman: “An Introduction to Kernel and Nearest Neighbor Nonparametric Regression,” *The American Statistician*, Vol. 46, No. 3, pp. 175-185, 1992.
- [16] Corinna Cortes, Vladimir Vapnik: “Support-Vector Networks,” *Machine Learning*, Vol. 20, No. 3, pp. 273-297, 1995.
- [17] Leo Breiman: “Random Forests,” *Machine Learning*, Vol.45, No. 1, pp. 5-32, 2001.
- [18] Jürgen Schmidhuber: “Deep Learning in Neural Networks: An Overview,” *ELSEVIER*, Vol. 61, pp. 85-117, 2015.
- [19] Tianqi Chen, Carlos Guestrin: “XGBoost: A Scalable Tree Boosting System,” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794, 2016.
- [20] Lior Rokach: “Ensemble-Based Classifiers,” *Artificial Intelligence Review*, Vol. 33, No. 1-2, pp. 1-39, 2010.