

情報機器への中毒性をもたらす使用特徴の検出

皆川 智[†] 藤波 香織^{††}

[†] 東京農工大学大学院工学府情報工学専攻 東京都小金井市中町 2-24-16

^{††} 東京農工大学大学院工学研究院先端情報科学部門 東京都小金井市中町 2-24-16

E-mail: [†]sl171726y@st.go.tuat.ac.jp, ^{††}fujinami@cc.tuat.ac.jp

あらまし モバイルアプリや SNS の普及により、人々がデジタルコンテンツに触れる機会は劇的に増加している。その一方で、過度使用による機器への依存や中毒が問題視されている。この問題への対策として、システムによってユーザに介入を行い、ユーザに過度の機器使用を控えさせる研究が行われている。しかし、これらのシステムが介入を行うタイミングや基準はあらかじめ決められた固定値を使用したものやユーザ自身の設定を必要とするものが多く、ユーザにとって適切な介入が行われていない懸念点がある。そこで、本研究ではスマートフォンユーザの使用データを用いた、情報機器の使用状況からユーザの中毒度合いを知るための中毒性のある使用特徴の検出を目的とする。まず、情報機器使用における中毒性評価指標を用いてユーザごとに中毒度を連続値のスコアで定義した。次に、データ収集実験によって得られたデータから計算・選択された特徴量をもとに回帰アルゴリズムを用いて中毒度スコア推定モデルを構築し、スコアの推定に大きく貢献する特徴量を得た。さらに、本研究で構築したモデルを使用した結果、ユーザを 83.3% の精度で中毒性に応じた 2 クラス分類が可能となった。

キーワード ヒューマンプロブ、情報機器、スマートフォン、中毒、回帰分析

Finding features toward recognizing dependence on information devices

Satoru MINAGAWA[†] and Kaori FUJINAMI^{††}

[†] Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology
2-24-16 Naka-cho, Koganei, Tokyo, Japan

^{††} Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology
2-24-16 Naka-cho, Koganei, Tokyo, Japan

E-mail: [†]sl171726y@st.go.tuat.ac.jp, ^{††}fujinami@cc.tuat.ac.jp

Abstract Due to the development of mobile applications and SNS, opportunities for people to touch digital contents are dramatically increasing. On the other hand, dependence and addiction to devices due to excessive use are regarded as problems. As a countermeasure against this problem, research is being conducted to prevent users from using excessive devices by intervening with users through the system. However, these systems intervene in many cases that use fixed values or users' own settings, and the user is not properly interventioned. Therefore, in this research, We aim to find addictive usage features for knowing the user's degree of poisoning from the use situation of information devices by using the usage data of smartphone users.

Key words Human-probe, information devices, smartphone, dependence, Regression

1. はじめに

近年、モバイルアプリや SNS の普及により、人々がデジタルコンテンツに触れる機会は劇的に増加している。その一方で、過度使用による機器への依存や中毒が問題視されている。例えば、情報機器として現代では日常生活に幅広く浸透しているスマートフォンについての調査 [2] では、2016 年時点で、所持者

のうち全体の 2 割以上がスマートフォンに「かなり依存」していると自覚しており、さらに、「やや依存している」と回答した人が全体で 50% を越えていることから、スマートフォンに少なからず依存していると自覚している人が全体で 70% 以上いるという結果が得られている。また、スマートフォンにかなり依存していると回答した人の 21% が、1 日に 7 時間以上スマートフォンに接触していると回答している。このことから、情報機

器に依存している人は使用時間も長くなる傾向にあると考えられる。また、スマートフォンを始めとする情報機器の長時間使用により、身体に悪影響を及ぼす危険性が指摘されている。例えば、スマートフォンなどのデジタルディスプレイから発せられるブルーライトの影響により、目の疲れや痛み、不眠症状が引き起こされる危険性や [3]、機器の過度使用により、指の変形や首への負担が増える危険性が指摘されている [4]。

この問題への対策として、システムによってユーザに介入を行い、ユーザに過度の機器使用を控えさせる研究が行われている。OKEKE らは、Facebook アプリケーションの起動時間に応じてユーザに振動フィードバックを与える介入システム Good Vibrations を開発し、振動がもたらす不快感を利用してユーザのアプリケーション使用時間の削減を試みた [7]。また、Markus らは、スマートフォンユーザが自身で特定のアプリに使用制限をかけることが可能な介入システム AppDetox を開発し、ユーザが自身でルールを設定し自主規制を働きかけることによりアプリケーションの使用時間削減を試みた [8]。これらの介入システムは、導入によりユーザの使用時間の削減に成功している一方で、介入を行うタイミングや基準はあらかじめ決められた固定値を使用したものやユーザ自身の設定を必要とするものが多く、ユーザにとって適切な介入が行われていない可能性がある。

そこで、本研究ではスマートフォンユーザの使用データを用いた、情報機器の使用状況からユーザの中毒度合いを知るための中毒性のある使用特徴の検出を目的とする。本研究の基本方針は以下の通りである。まず、スマートフォン中毒状態の定義付けを行う。今回は情報機器使用における既存の中毒性評価指標を使用して、ユーザごとに中毒度を連続値のスコアとして定義した。また、これらのスコアによってユーザを中毒性のある使用をするユーザとしないユーザの 2 クラスに分類することが可能になる。次に、検出対象となるデータセットを得るためのデータ収集を行う。データ収集のためにスマートフォン使用時のログデータを記録するためのロガーアプリを実装し、被験者に配布する形でデータ収集を行った。このとき、被験者には先述した中毒度のスコアを事前に算出しておき、正解データとした。次に、得られたデータから特徴量を算出し、特徴量からデータ収集の対象となる被験者の中毒度スコアを推定するための推定モデルの作成を行った。今回は線形回帰モデルを使用し、計算された特徴量を説明変数、被験者の中毒度スコアを目的変数として線形回帰式を算出した。さらに、収集したデータをテストデータとし、得られた線形回帰式を用いて算出された中毒スコアの予測値が所属するクラスが正解データが所属するクラスと同一のクラスに正しく分類されるか検証を行った。

本論文の構成を以下に述べる。まず、2. 章でスマートフォンの使用データから問題のある使用の検出を行った関連研究を紹介する。次に、3. 章において提案手法の設計方針について述べ、4. 章でデータ収集システムの実装について述べる。5. 章では収集データから中毒度スコアの推定モデルを作成・評価し、6. 章で全体の考察を行う。

2. 関連研究

スマートフォンの使用データを用いて問題のある使用を検出する研究が行われている。Shin らはスマートフォン使用時の様々なログデータを使用して、起動時間や起動回数をはじめとする様々なイベントの使用特徴から問題のある使用の検出を行った [5]。また、Ding は情報機器への中毒に関してアプリケーションの使用に焦点を当て、スマートフォン使用時のログデータから中毒性をもたらすアプリケーションの種類や、アプリケーション中毒を引き起こすユーザの心理的な要因の特定を行なった [6]。これらの関連研究はいずれもスマートフォン使用時のデータを収集し、中毒性のある使用、または問題のある使用の検出を行なっているが、1 日のライフサイクルの中でスマートフォンを使用する時間帯ごとの特徴が考慮されていない。総務省が 2016 年に行なったインターネット上のメディア利用実態に関する調査では、1 日の中で時間帯別に情報機器の使用率が異なるという結果が示されている [9]。したがって、情報機器を使用する時間帯を使用特徴とし、中毒性をもたらす要因として考慮することは有効ではないかと考えた。本研究では、関連研究の手法に加え、ユーザがスマートフォンを使用する時間帯を考慮した使用特徴を考えることにより、より細かい粒度での中毒性がある使用の検出を試みる。

3. システム設計方針

3.1 スマートフォン中毒の定義

スマートフォンの使用データから中毒性のある使用を検出するためには、中毒状態の定量的に表現可能な定義付けを行う必要がある。情報機器使用時の中毒性を定量化するためのフレームワークに関する研究はこれまでに様々なものが行われてきた [10] [11] [12]。これらの中毒性判定フレームワークはいずれも情報機器への中毒性をもたらす潜在的な予測因子の分析をもとに作成され、アンケート形式で中毒性に関する質問をリッカート尺度で回答することにより、回答点数を中毒度スコアとして提供するフレームワークである。本研究では、これらの評価指標の中でも最新かつ、過去に作成されたフレームワークの内容を参考にして作成された Smartphone Addiction Proneness Scale (以下、SAPS とする) [12] を使用してスマートフォン中毒状態を定義した。SAPS は 15 の質問項目からなるアンケート形式の指標であり、各質問を 5 段階リッカート尺度で回答することにより、中毒度スコアを連続値で取得可能である。さらに、得られた中毒度スコアが 42 点を超えた場合、そのユーザは中毒性のある使用をしていると判断される。したがって、SAPS を実施してもらうことにより、ユーザの中毒性のある/なしの 2 クラスへの分類が可能となる、SAPS のアンケート内容を表 1 に示す。SAPS はアンケートの内部信頼性を表す指標である Cronbach の α 係数が 0.88 と高い信頼性を得ている指標であることから、本研究における中毒性の定義に有効であると考えた。

3.2 推定モデル構築方法

推定モデルの構築は、正解値として中毒度スコアが付与されたログデータ群を教師データとして使用した学習により行う。

表 1 SAPS アンケート内容

| 質問番号 | 内容 |
|------|--|
| 1 | スマートフォンの過度使用により、学校の成績が低下したことがある。 |
| 2 | スマートフォンを使っている時間は、家族や友人と一緒に過ごす時間よりも楽しい。 |
| 3 | スマートフォンを使うことを許可されていない時間が苦痛である。 |
| 4 | スマートフォンの使用時間を削減しようとしても、失敗する。 |
| 5 | スマートフォンを使用しているがために、元々やろうとしていたこと(勉強、研究、サークル、部活動など)に支障をきたすことがある。 |
| 6 | スマートフォンを使うことができない間、自分の世界を失ったように感じる。 |
| 7 | スマートフォンがないと落ち着かない、または不安になる。 |
| 8 | 私はスマートフォンの使用時間をコントロールできていると思う。 |
| 9 | 私の過度のスマートフォンの使用について、周りの人は頻繁にコメントする。 |
| 10 | スマートフォンがなくても別に気にならない。 |
| 11 | スマートフォンの使用をやめなければいけないと思いつつも、使い続けてしまう。 |
| 12 | 家族や友人が、「自分はスマートフォンを使いすぎている」と不満や文句を言う。 |
| 13 | スマートフォンの使用は、自分の勉強や研究から気を散らせることはない。 |
| 14 | スマートフォンが使えないとパニックになる。 |
| 15 | スマートフォンの使用に多くの時間を費やすことが習慣になっている。 |

本研究では収集されたログデータをもとに中毒度スコアという連続値の予測を行うため、推定器の作成には線形回帰モデルを利用する。線形回帰モデルは一般的に式 (1) の形で表される。

$$Y = \sum_{i=0}^{n-1} \alpha_i X_i + \beta \quad (1)$$

目的変数 Y は SAPS によって算出されたユーザごとの中毒度スコアを指す。説明変数 X_i は収集されたログデータより計算された特徴量である。 α_i および β は収集データを用いた N 次元の学習により変動する値であり、収集データの実測値と回帰モデルにより算出された予測値の誤差が小さくなるように決定される。 α_i は説明変数にかかる係数であり、説明変数が目的変数に与える影響の大きさを表す数値である。 β は定数項であり、説明変数の変動に左右されない値である。

4. データ収集システム実装

4.1 データ収集実験概要

本研究ではスマートフォンのログデータをもとに 3.1 節で述べた SAPS による中毒度スコアの推定を行うため、ユーザのスマートフォン使用時のログデータおよび中毒度スコアが必要となる。そのため、ログデータ収集用のアプリケーション DataLogApp を Android 上に実装し、普段から Android 端末を使用しているユーザを対象にデータ収集実験を行った。DataLogApp は各被験者のスマートフォンにインストールされ、計測開始ボタンを押して計測が開始されると、バックグラウンドで常に被験者のスマートフォン使用データを監視し、ログデータを端末の内部ストレージに CSV ファイルとして毎秒記録する。実際に記録するデータの詳細は次章で述べる。

データ収集実験は被験者 10 名 (男性 7 名, 女性 3 名, いずれも 20 代大学生) を対象に各被験者ごとに 10 日間, 合計 100 日間行った。実験は以下の流れで行った。まず、被験者には事前に SAPS アンケートに回答してもらい中毒度スコアを算出しておく。次に、実験参加に関する同意書にサインを行った後、DataLogApp を各自の端末にインストールし、実験開始日に記録開始、実験終了日に記録終了のボタンを押してもらう。被験者が実験期間に行う操作はボタンを押す操作のみであり、被験

者の日常的なスマートフォン使用を阻害することはない。なお、今回は事前の SAPS アンケート調査で「中毒あり」のユーザ 5 人、「中毒なし」のユーザ 5 人を被験者として選定した。また、実験期間に関して、各被験者内で日常生活のサイクルを統一するために、平日 (月～金曜日) と休日 (土、日曜日) との生活リズムの差異を考慮して実験期間は全て平日とした。

4.2 収集データ

4.1 節で述べた、DataLogApp で収集するデータを表 2 に示す。

表 2 DataLogApp で収集するデータ

| 種別 | 命名規則 | 説明 |
|---------|-----------------------|---|
| 通常使用 | Time | ・起動時間 |
| | Session | ・起動回数 (画面が ON になってから OFF になるまでが 1 セッション) |
| | AvgTimePerSession | ・1 回のセッションあたりの平均起動時間 |
| | AvgIntervalPerSession | ・平均起動間隔 (セッションが終了してから次のセッションが開始されるまでの平均時間) |
| バッテリー使用 | AvgBatt | ・平均バッテリーレベル |
| | BattChg | ・充電回数 |

さらに、表 2 の分類に加えて、被験者が閲覧しているアプリケーション名の取得を行い、各アプリケーションを所属するジャンル別に分類した。取得するアプリケーションは、被験者のプライバシーを侵害しないように、事前に実験者が選定したアプリケーションのみを取得対象とした。表 3 にアプリケーションのジャンルによる分類および取得対象のアプリケーションを示す。なお、表 3 に記載されていないアプリケーションは全て not allowed というアプリケーション名で記録される。

表 3 DataLogApp で取得するアプリケーション名およびジャンル

| ジャンル名 | 命名規則 | 取得するアプリケーション名 |
|-------|---------|---|
| メール | mail | 端末にデフォルトで搭載されているメールアプリ + Gmail |
| ブラウザ | browser | 端末にデフォルトで搭載されているブラウザアプリ + Google Chrome |
| SNS | social | Twitter + Facebook + LINE + Instagram |
| 動画 | movie | YouTube + Amazon Video + Netflix + Hulu + abemaTV |
| 全て | all | 上記のアプリケーション全て + not allowed |

また、1 日の生活サイクルの中で、スマートフォンを使用する時間帯による使用特徴が中毒度スコアに与える影響を調査するために、1 日を図 1 に示すように 3, 6, 12, 24 時間ごとに分割し、それぞれの時間帯において 4.3 節で述べる特徴量計算を行う。

4.3 特徴量計算

4.2 節で収集したデータを用いて推定モデル作成のための特徴量計算を行う。今回は、図 1 に示した各時間帯において、表 3 に示した各ジャンルに対して表 2 に示した使用情報のうち通常使用に該当するものを組み合わせたものを特徴量として計算する。また、各特徴量は [時間帯]_[ジャンル名]_[使用情報] のように命名する。例えば、「3h_3_social_Session」は午前 6:00 から午前 9:00 の間に SNS を起動した回数を表す。したがって、合計 15 種類の各時間帯において、使用される特徴量はジャンル 5 種類 × 通常使用情報 4 種類 + バッテリー使用情報 2 種類の合計 22 個となる。

5. シミュレーション評価実験

5.1 推定モデル作成

本章では、4. 章で収集したデータより計算された特徴量を使

| | | | | | | | | | | | | | | | | | | | | | | | | |
|-------|---|---|------|---|---|------|---|---|------|----|----|-------|----|----|------|----|----|------|----|----|------|----|----|----|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 3h_1 | | | 3h_2 | | | 3h_3 | | | 3h_4 | | | 3h_5 | | | 3h_6 | | | 3h_7 | | | 3h_8 | | | |
| 6h_1 | | | | | | 6h_2 | | | | | | 6h_3 | | | | | | 6h_4 | | | | | | |
| 12h_1 | | | | | | | | | | | | 12h_2 | | | | | | | | | | | | |
| 24h_1 | | | | | | | | | | | | | | | | | | | | | | | | |

図 1 時間帯の分類方法および命名規則

用し、データマイニングツールの Weka [18] による線形回帰モデル構築アルゴリズムを用いた中毒度スコア推定モデルの構築を行い、オフライン環境下で精度評価実験を行う。推定精度評価指標には以下の 2 つの項目を用いる。

- 全データ群の正解値と予測値との相関係数
- 正解値と予測値の二乗平均平方根誤差

さらに、推定モデル構築のための回帰アルゴリズムは複数存在し、それぞれが異なる推定モデルを構築するため、構築される推定モデルの違いによる精度比較が必要となる。したがって、本研究では連続値の推定を対象とする以下に示す 4 種類の推定モデル構築アルゴリズムを使用し、精度の比較を行う。

- 線形回帰 [13]
- 最小二乗メジアン回帰 [14]
- SMO 回帰 [15]
- M5P [16]

まず、以上の 4 つのアルゴリズムを使用して推定精度の比較を行う。線形回帰（以下、LReg とする）は、最も一般的な線形回帰モデル構築アルゴリズムであり、各説明変数における係数から算出した全ての推定値と正解値との間の二乗誤差が最小となるように回帰モデルを決定する手法である。最小二乗メジアン回帰（以下、LMS とする）は、各説明変数における係数から算出した全ての推定値と正解値との間の二乗誤差の中央値が最小となるように回帰モデルを決定する手法である。SMO 回帰（以下、SMOreg とする）は、SVM(Support Vector Machine) を回帰分析に拡張した手法である。M5P は、説明変数を取る値によって分岐判定を行う木構造を作成し、末端の各ノードにおいて LReg による回帰式を作成する手法である。

回帰モデルへの入力とは 3.2 節で説明した通り、収集データから計算された特徴量および SAPS のアンケート結果による中毒度スコアの正解値となる。また、出力として線形回帰式が得られる。実験は、収集された 100 日分のデータのうち、不備があった 23 日分のデータを除外した合計 77 日分のデータを対象に、15 種類の各時間帯ごとに計算された 22 個の特徴量を使用して推定モデルの構築を行い、10 分割交差検証によって相関係数と二乗平均平方根誤差（以下、RMSE とする）を算出した。表 4 および表 5 にその結果を示す。

いずれの時間帯においても M5P が他の 3 種類のアルゴリズムと比較した場合に相関が強くなる傾向にあるという結果が得られた。また、時間帯の分類により、時間帯別に相関が強いものと弱いものがあることから、時間帯の分類は中毒度スコアの予測因子として有効である可能性があると考えられる。しかし、今回各時間帯において使用した 22 個の特徴量を全て使用して推定モデルを構築した場合に最良の解が得られるとは限らない。

表 4 全ての特微量を使用した場合の相関係数

| 時間帯 | LReg | LMS | SMOreg | M5P |
|-------|--------|--------|--------------|--------------|
| 3h_1 | 0.048 | 0.186 | 0.269 | 0.096 |
| 3h_2 | -0.355 | -0.045 | -0.001 | 0.049 |
| 3h_3 | 0.281 | 0.204 | 0.418 | 0.456 |
| 3h_4 | 0.121 | 0.064 | -0.115 | 0.140 |
| 3h_5 | 0.240 | 0.207 | 0.275 | 0.359 |
| 3h_6 | 0.124 | -0.201 | -0.071 | 0.192 |
| 3h_7 | -0.105 | 0.054 | 0.022 | 0.054 |
| 3h_8 | 0.189 | 0.073 | 0.138 | 0.293 |
| 6h_1 | 0.025 | -0.045 | 0.042 | 0.044 |
| 6h_2 | 0.294 | 0.141 | 0.069 | 0.240 |
| 6h_3 | 0.240 | 0.113 | 0.349 | 0.365 |
| 6h_4 | 0.264 | 0.043 | -0.004 | 0.292 |
| 12h_1 | 0.168 | 0.133 | 0.161 | 0.137 |
| 12h_2 | 0.286 | -0.124 | -0.068 | 0.416 |
| 24h_1 | 0.154 | 0.046 | 0.231 | 0.378 |

表 5 全ての特微量を使用した場合の RMSE

| 時間帯 | LReg | LMS | SMOreg | M5P |
|-------|-------|-------|--------|--------------|
| 3h_1 | 3.916 | 6.408 | 9.062 | 3.814 |
| 3h_2 | 4.055 | 205.1 | 5.313 | 3.830 |
| 3h_3 | 4.290 | 8.402 | 3.716 | 3.354 |
| 3h_4 | 4.376 | 5.725 | 4.852 | 4.056 |
| 3h_5 | 4.261 | 5.010 | 4.707 | 3.924 |
| 3h_6 | 4.612 | 12.92 | 15.046 | 4.206 |
| 3h_7 | 8.842 | 4.819 | 5.869 | 4.503 |
| 3h_8 | 7.122 | 11.31 | 5.086 | 3.863 |
| 6h_1 | 4.759 | 10.92 | 4.585 | 3.872 |
| 6h_2 | 4.086 | 5.668 | 4.622 | 3.872 |
| 6h_3 | 7.027 | 6.026 | 4.192 | 3.637 |
| 6h_4 | 4.139 | 6.963 | 4.964 | 3.757 |
| 12h_1 | 9.007 | 7.384 | 6.078 | 4.641 |
| 12h_2 | 3.909 | 5.435 | 5.751 | 3.507 |
| 24h_1 | 5.899 | 6.377 | 4.717 | 4.008 |

また、22 種類全ての特微量を使用した場合、構築された回帰式の係数が多くなり、本研究の目的でもある中毒度スコアの推定に影響を与える特微量の特定が困難である。したがって、次章で特微量選択を行った上で再度精度比較を行うこととした。

5.2 特微量選択

5.1 節より、推定精度向上のための特微量選択を行う。本研究では、以下の 2 つの観点から特微量の選択を行う。

- 特微量の組み合わせ数の変化による推定精度への貢献度を考慮した特微量削減
 - 特微量間の相関関係を考慮した特微量削減
- 特微量の組み合わせ数の変化による推定精度への貢献度を

表 6 各時間帯において選択された特徴量一覧（網掛けされた行は 10 種類以上の時間帯で選択された特徴量名を指す）

| 選択された特徴量名\時間帯 | 3h_1 | 3h_2 | 3h_3 | 3h_4 | 3h_5 | 3h_6 | 3h_7 | 3h_8 | 6h_1 | 6h_2 | 6h_3 | 6h_4 | 12h_1 | 12h_2 | 24h_1 |
|-------------------------------|------|------|------|------|------|------|------|------|------|------|------|------|-------|-------|-------|
| BattChg | | | | | | | | | | | | | | | ○ |
| AvgBatt | | ○ | ○ | ○ | | | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | |
| Time_all | | | | | | | | | | | | | | | |
| Session_all | ○ | | ○ | ○ | ○ | ○ | | | | ○ | ○ | | ○ | | ○ |
| AvgTimePerSession_all | | | | | ○ | ○ | | | | | ○ | | ○ | ○ | |
| AvgIntervalPerSession_all | | | | | | | | | | | | | | ○ | ○ |
| Time_mail | ○ | | | | ○ | ○ | ○ | | | | ○ | ○ | ○ | ○ | |
| Session_mail | | | | | | | | ○ | ○ | | ○ | ○ | | ○ | ○ |
| AvgTimePerSession_mail | ○ | ○ | | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | | |
| AvgIntervalPerSession_mail | | | | ○ | | | | | | | | | | | |
| Time_browser | | | | | ○ | ○ | ○ | | | | ○ | ○ | | ○ | |
| Session_browser | ○ | | | | ○ | ○ | | | | | ○ | ○ | | | ○ |
| AvgTimePerSession_browser | ○ | ○ | | | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| AvgIntervalPerSession_browser | | | | | ○ | ○ | | | | | ○ | | | | |
| Time_social | ○ | | | | ○ | ○ | | | | | ○ | | ○ | | |
| Session_social | ○ | | | | | | | ○ | ○ | | | | | | |
| AvgTimePerSession_social | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| AvgIntervalPerSession_social | | | | | | | | | | | | | | | |
| Time_movie | | | | | | | ○ | | | | ○ | | | ○ | |
| Session_movie | ○ | | | ○ | | | | ○ | ○ | | ○ | | ○ | | ○ |
| AvgTimePerSession_movie | ○ | | | | | | | | | | | | ○ | ○ | ○ |
| AvgIntervalPerSession_movie | ○ | | | | | | | | | | | | | | |
| 選択された特徴量数 | 11 | 4 | 3 | 6 | 10 | 10 | 7 | 7 | 7 | 3 | 14 | 6 | 9 | 9 | 9 |

考慮した特徴量削減では、まず、中毒度スコア推定に関する特徴量を、貢献度が高い順番にランク付け可能な属性選択アルゴリズムを使用してランク付けした。今回は Weka を使用して既存の属性選択アルゴリズムである CfsSubsetEval [17] と GreedyStepwise を用いた。CfsSubsetEval は、各特徴量の個々の予測能力および特徴量間の相関関係を考慮して特徴量選択を行う手法である。また、選択された各特徴量は目的変数と強い相関を持ち、かつ、各特徴量間の相関が弱いものが望ましいとされる。次に、ランク付けされた特徴量のうち、貢献度が高い順に特徴量を追加した時の 10 分割交差検証による推定精度の変化を検証する。これを全ての時間帯、4 種類の回帰アルゴリズムに対して行い、各時間帯において最も推定精度が高くなる回帰アルゴリズムおよび特徴量数を算出した。

また、特徴量間の相関関係を考慮した特徴量削減では、回帰分析においてある特徴量が他の特徴量と強い相関を持っている場合に発生する多重共線性を考慮し、各特徴量間で強い相関を持っているもの同士は相関係数が低い方を排除するという形で特徴量の削減を行った。今回は、0.9 以上の相関を持っている特徴量の組み合わせを強い相関を持つ組み合わせとした。

以上の 2 種類の手法で特徴量削減を行った結果を表 6 に示す。表 6 は行が特徴量名、列が時間帯を表し、各時間帯において選択された特徴量のセルを○で表す。なお、回帰アルゴリズムに関しては、3h_1 のみ SMOreg、残りの時間帯は全て M5P が最も高精度という結果が得られた。

5.3 選択された特徴量を用いた推定結果

本章では、5.2 節で選択した特徴量および回帰アルゴリズムを使用して再度推定精度の比較実験を行う。表 7 にその結果を示す。

表 7 より、本研究で選択した特徴量および回帰アルゴリズム

表 7 各時間帯において選択された特徴量および回帰モデルを使用した際の相関係数 (左) および RMSE (右)

| 時間帯 | 相関係数 | 回帰モデル | 時間帯 | RMSE | 回帰モデル |
|-------|---------------|------------|-------|---------------|------------|
| 3h_1 | 0.186 | SMOreg | 3h_1 | 3.813 | SMOreg |
| 3h_2 | -0.069 | M5P | 3h_2 | 3.874 | M5P |
| 3h_3 | 0.514 | M5P | 3h_3 | 3.226 | M5P |
| 3h_4 | 0.248 | M5P | 3h_4 | 3.736 | M5P |
| 3h_5 | 0.269 | M5P | 3h_5 | 3.634 | M5P |
| 3h_6 | 0.192 | M5P | 3h_6 | 4.207 | M5P |
| 3h_7 | 0.055 | M5P | 3h_7 | 4.504 | M5P |
| 3h_8 | 0.254 | M5P | 3h_8 | 3.723 | M5P |
| 6h_1 | 0.102 | M5P | 6h_1 | 3.793 | M5P |
| 6h_2 | 0.361 | M5P | 6h_2 | 3.530 | M5P |
| 6h_3 | 0.299 | M5P | 6h_3 | 3.636 | M5P |
| 6h_4 | 0.306 | M5P | 6h_4 | 3.766 | M5P |
| 12h_1 | 0.365 | M5P | 12h_1 | 3.509 | M5P |
| 12h_2 | 0.417 | M5P | 12h_2 | 3.507 | M5P |
| 24h_1 | 0.5211 | M5P | 24h_1 | 3.2257 | M5P |

のうち、24h_1、すなわち 24 時間を 1 つのセグメントとし、て M5P アルゴリズムを使用して推定モデルを作成した場合に収集されたデータと中毒度スコアとの相関が最も強くなるという結果が得られた。このとき、選択された特徴量数は AvgTimePerSession_social、AvgTimePerSession_browser、AvgTimePerSession_movie、Session_movie、Session_browser、BattChg、AvgIntervalPerSession_all、Session_all、Session_mail の 9 個である (表 6)。また、1 日を 3 時間ごとに分割した場合、3h_1 の相関係数が他の時間帯と比較して高くなっていることから 6:00 から 9:00 の使用はスコアとの相関が強くなり、1 日を 6 時間ごとに分割した場合、6h_2、6h_4 の相関係数が他の時間帯と比較して高くなっていることから、6:00 から 12:00、および

18:00 から 24:00 の使用は他の時間帯と比較して中毒スコアとの相関が強くなるという結果となった。

5.4 中毒クラス分類可能性調査

次に、5.3 節で求めた最も高精度で中毒度スコアを予測可能な 24h_1, M5P, 特徴量 9 個を使用して作成された推定モデルの中毒クラス分類可能性調査を行う。3.1 節より、SAPS は中毒度スコアの点数に応じて回答者を「中毒あり」、「中毒なし」の 2 クラス分類が可能であると述べた。そこで、収集データをテストデータとして上記の推定モデルを使用して中毒度スコアを予測した際に、算出されたスコアと正解値のスコアが同じクラスに正しく分類される割合を正解率として算出した結果、正解率は 83.3% となった。

6. 考察

6.1 選択された特徴量に関する考察

本研究では、精度向上のために特徴量の削減を行い、推定に貢献する特徴量を選択した。表 6 の一部の特徴量を見てみると、異なる時間帯でも選択されることが多いものが見受けられた。特に、AvgBatt, AvgTimePerSession_mail, AvgTimePerSession_browser, AvgTimePerSession_social では全 15 種類の時間帯のうち、10 種類以上の時間帯で選択されている。AvgBatt は端末の平均バッテリーレベルであるため、端末の平均バッテリーレベルが中毒度スコア予測への貢献度が高くなる可能性が示唆された。また、AvgTimePerSession は 1 回のセッションあたりの平均使用時間であるため、1 回の起動における長時間の SNS・ブラウザ・メールアプリの使用時間は中毒度スコアへの貢献度が高いという可能性が示唆された。一方で、本研究で行ったデータ収集の被験者は 10 名と少なく、個人のスマートフォン使用に関する習慣が収集データに影響を与えた可能性が考えられる。したがって、収集データに個人差が与える影響の調査および推定モデル作成に用いるデータ数の増加が今後の課題として挙げられる。

6.2 時間帯ごとの分類に関する考察

本研究で特徴量の計算を行う際、基本使用情報、ユーザが使用したアプリのジャンル、ユーザがスマートフォンを使用した時間帯による 3 つの観点から特徴量の計算を行った。このうち、基本使用情報およびアプリケーションのジャンルによる分類は既存研究で行われており、これらの分類の有効性が示唆されている。残りの時間帯による分類に関しては筆者が独自に考案したものであるため、ここでは時間帯による分類の有効性について考察を行う。表 7 において最も推定モデルの精度が高かったのは 24h_1 であり、そこからさらに時間帯ごとに分割していくといずれの時間帯においても精度は下がるという結果になった。その一方で、表 7 からわかる通り、ある一定の時間で 1 日を分割した場合、各時間帯のブロックごとに推定精度にばらつきが見受けられることが読み取れる。したがって、時間帯による分類は、1 日の中でより細かい粒度で中毒度スコアへ影響を与える予測因子を考える上で有効である可能性が示唆された。しかし、現時点では各時間帯における相関係数同士の相対的な比較のみで判断しているため、今後はこれらの相関係

数のばらつきに有意差が見られるか検証を行う必要がある。

6.3 中毒クラス分類に関する考察

5.4 節で求めた中毒クラス分類正解率の内訳を表 8 に示す。表 8 より、「中毒なし」クラスに正しく分類された割合が 84.9% であるのに対し、「中毒あり」クラスに正しく分類された割合が 76.9% となったことから、実際には「中毒あり」クラスに所属するユーザの誤分類が多い結果となった。言い換えると、実際には「中毒なし」クラスに所属するユーザが「中毒あり」クラスに誤分類されるケースが多かった。そこで、中毒スコアの正解値を真値、推定モデルによる予測値を測定値として平均誤差を求めたところ、+0.9 となった。したがって、誤分類の割合における差異は、予測値が全体として正解値より高いスコアを取る傾向にあったことが原因だと考えられる。本研究では中毒性のある使用の検出を目的としているため、今後は「中毒あり」クラスへの分類精度の向上が課題となる。具体的には、中毒度スコアの予測値と正解値との絶対誤差がマイナスの値取るように推定モデルを作成する方法などが考えられる。

表 8 中毒クラス分類正解率の内訳

| 正解クラス\予測クラス | 中毒あり | 中毒なし |
|-------------|-------|-------|
| 中毒あり | 76.9% | 23.1% |
| 中毒なし | 15.1% | 84.9% |

7. おわりに

本研究ではスマートフォンユーザの使用データを用いた中毒性のある使用特徴の検出を行った。まず、情報機器使用における中毒性評価指標である SAPS を用いて連続値のスコアという形で中毒度を定義し、データ収集実験によって得られたデータから計算・選択された特徴量をもとに回帰アルゴリズムを使用して推定モデルを作成した。その結果、1 日を 1 つのセグメントとし、SNS・ブラウザ・メールアプリの 1 回のセッションあたりの平均起動時間などを含む選択された 9 種類の特徴量と M5P アルゴリズムを使用した場合に最も高精度で中毒度スコアを予測可能な推定モデルが作成可能という結果が得られた。また、このモデルを使用した場合、83.3% の割合でユーザを正しいクラスへ分類することが可能となった。今後の課題として、推定モデル作成に使用するデータ数の増加や、個人差が推定モデル作成に与える影響の調査、特徴量選択方法の再検討、推定器作成に用いる回帰モデルの追加などが挙げられる。

文 献

- [1] Goldman, J., et al, "Participatory Sensing: A citizen-powered approach to illuminating the patterns that shape our world," Foresight and Governance Project, White Paper, 2009.
- [2] 2016 年スマホ依存に関する調査, https://mmdlabo.jp/investigation/detail_1563.html. (2017/04/21 閲覧)
- [3] スマホから出る「ブルーライト」体内に悪影響 | いしゃまち, <https://www.ishamachi.com/?p=904>. (2017/04/21 閲覧)
- [4] スマホや PC で痛めやすい首の症状, <http://www.xn--68j2bq4hz700azqionwoy6d.net/kubisuma77788.html>. (2017/04/21 閲覧)
- [5] Shin, C., et al., "Automatically detecting problematic use of smartphones.", Proceedings of the 2013 ACM international

- joint conference on Pervasive and ubiquitous computing (pp. 335-344).
- [6] Ding, X., et al., "Beyond Smartphone Overuse: Identifying Addictive Mobile Apps.", Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems.
 - [7] OKEKE, F., et al., "Good Vibrations: Designing and Evaluating a Ubiquitous Intervention for Reducing Digital Consumption.",
 - [8] Markus L., et al., "AppDetox: helping users with mobile app addiction.", Proceedings of the 12th international conference on mobile and ubiquitous multimedia.
 - [9] 平成 28 年 情報通信メディアの利用時間と情報行動に関する調査 | 調査報告書 http://www.soumu.go.jp/main_content/000492877.pdf. (2018/02/17 閲覧)
 - [10] Bianchi, A., et al., "Psychological Predictors of Problem Mobile Phone Use.", Journal of Korean Academy of Nursing, 39(6).
 - [11] Koo, H, Y., "Development of a Cell Phone Addiction Scale for Korean Adolescents.", Journal of Korean Academy of Nursing, 39(6).
 - [12] Kim, D., et al., "Development of Korean smartphone addiction proneness scale for youth.", PLoS One, 9(5), e97920.
 - [13] McCullagh, P. "Generalized linear models." European Journal of Operational Research 16.3 (1984): 285-292.
 - [14] Simpson, D, G. "Introduction to Rousseeuw (1984) Least Median of Squares Regression." Breakthroughs in Statistics. Springer, New York, NY, 1997. 433-461.
 - [15] Shevade, S, K., et al. "Improvements to the SMO algorithm for SVM regression." IEEE transactions on neural networks 11.5 (2000): 1188-1193.
 - [16] Eibe F., et al., "Using Model Trees for Classification." Mach. Learn., Vol. 32, No. 1, pp. 63(2013)76, July 1998.
 - [17] Hall, M, A. "Correlation-based feature selection for machine learning." PhDthesis, The University of Waikato, 1999.
 - [18] Machine Learning Group at University of Waikato. Weka3. Data mining with open source machine learning software in java. <http://www.cs.waikato.ac.nz/ml/weka/>. (2018/02/16 閲覧)