

# 「ひきこもり」防止を目指した学生生活支援システムの提案

北井 悠一朗<sup>1</sup> 水本 旭洋<sup>1</sup> 長谷川 凌佑<sup>1</sup> 内山 彰<sup>1</sup> 石丸 大貴<sup>2</sup> 堀田 牧<sup>2</sup> 足立 浩祥<sup>3</sup>  
工藤 喬<sup>3</sup> 池田 学<sup>2</sup> 東野 輝夫<sup>1</sup>

**概要:** ひきこもりとされる人々は全国に 54.1 万人存在すると推定されている。近年、大学においても、ひきこもり状態に陥ることで授業に出席しなくなり、単位を取得できず留年や退学に至るケースが増加しており、大きな問題となっている。本研究では、ひきこもり状態に至る前に予兆を検知し、医者による改善指導が行えるシステムの実現に向けて、(1) 睡眠時間の乱れ、(2) 授業出席意識の低下、(3) 訪問地点数の減少といったひきこもりの予兆を検知するためのスマートフォンを用いた学生生活支援システムを構築する。睡眠時間の乱れの推定については、スマートフォンのセンサから得た特徴量を用いた機械学習による推定を行う。授業出席意識の低下と訪問地点数の減少は、位置推定を行うことにより推定が可能である。本研究では、学校以外の詳細地点を特定する必要はないため、GPS による絶対位置の特定は行わず、Wi-Fi アクセスポイントの MAC アドレスおよび RSSI(受信信号強度) を用いてクラスタリングを行い、訪問地点数を推定する。実際に手法の妥当性を検証するため実験をした結果、滞在している地点数の推定が可能であること、また、自宅の帰宅時間の推定において、誤差 9.6 分での推定ができていることを確認し、予兆推定の見通しを得た。睡眠推定では Balanced Random Forest を使った睡眠の有無の予測モデルを構築し、F 値 0.86 の結果を得たことから、睡眠の時間を推定できる可能性を示した。

## 1. はじめに

内閣府が 2015 年に実施した「若者の生活状況に関する調査報告書」[1]によると、“自分の趣味に関する用事するときだけ外出し、仕事や勉強など外でコアな人間関係を持たない”という準ひきこもり群を含めた広義のひきこもり群とされる人々は 15 歳から 39 歳において全国に 54.1 万人存在すると推定されている。また同調査で、ひきこもりには長期化する傾向があることが確認された。これを踏まえ、2018 年に内閣府が 40 歳以上の世代を対象として実施した「生活状況に関する調査」[2]においても、広義のひきこもり群とされる人々は推定 61.3 万人存在するとされており、若年層から中高年層まで、ひきこもりが社会問題となっている。

近年、大学において、このようなひきこもり状態に陥ることで授業に出席しなくなり、単位を取得できず留年や退学に至るケースが増加しており、大きな問題となっている。このようなひきこもりの問題に対し、多くの大学では臨床

心理士や精神科医による心理カウンセリングなどのメンタルヘルス対策を行っているが、多くの場合患者本人が症状を自覚してから受診をするため、問題が判明した時には既に深刻化している場合が多い。ひきこもりを未然に防ぐため、出来るだけ早い段階での予兆の検知を行い、患者自身に改善を促すような仕組みが必要である。

しかし、ひきこもりの予兆検知の課題として、大学生がひきこもりまたは準ひきこもりに至る根本的な原因は様々ではないということが挙げられる。例えば、授業に参加しなくなったことでひきこもりに至ったケースをみても、根本的な要因として、友人や先生との対人関係の悪化や、アルバイトなどで忙しく時間通りに起床できなくなったこと、あるいは、授業に対する熱意が低く部活やサークルに参加はするものの授業には出席しないことなど多岐にわたる。そこで、本研究では大阪大学保健センターや医学系研究科精神医学教室と協力する事で、ひきこもり症状に対する予兆を定義し、中長期的な生活データ収集からその予兆の検知を目指す。今回検知を目指す医学的知見から定義したひきこもりの予兆は、(1) 睡眠時間の乱れ、(2) 授業出席意識の低下、(3) 訪問地点数の減少の 3 点である。睡眠時間の乱れは、睡眠時間が不規則になることを意味しており、それによって昼夜の逆転を引き起こしたりすることで、大学への登校やアルバイトへの出勤など規則的な習慣を行う

<sup>1</sup> 大阪大学大学院情報科学研究科  
Graduate School of Information Science and Technology, Osaka University

<sup>2</sup> 大阪大学大学院医学系研究科  
Department of Psychiatry, Osaka University Medical School

<sup>3</sup> 大阪大学キャンパスライフ健康支援センター  
Health and Counseling Center, Osaka University

ことが難しくなる。授業出席率の低下は、睡眠の乱れなどによる要因以外にも、そもそも勉強に対する熱意を失うことによって授業出席意識が低下することで引き起こされることもある。例えば、サークルや部活にのみ参加し、授業に出席しないといた原因はこれに該当する。訪問地点数とは普段から足を運ぶ場所の数を示しており、この減少は、外出の頻度の低下と活動範囲の縮小を意味する。これら3つの要因はいずれも長期化することで、授業に欠席したり、外出そのものができなくなることにつながり、ひきこもりや準ひきこもりの状態に陥ると考えられる。よって本研究では、これら3点を予兆と定義し検出を目指す。

他方、携帯電話やIoTの発展に伴い、スマートフォンやウェアラブルデバイスのセンサの情報を利用することで、個人の生活における様々な行動を分析・推定することが可能となってきた。このようなライフログは様々なシステムで活用されている。例えば位置情報や歩数の推定から、1日の運動の様子や消費したカロリーを日記として残せるアプリケーションや、ウェアラブルデバイスによって収集した心拍などのセンサデータから睡眠を推定することで、日々の睡眠の様子と利用者の要望に合わせた適切な睡眠時間の提案などを行うサービスが存在している。近年では、ライフログを身体健康維持に利用するだけでなく、メンタルヘルスの分野でも活用しようという動きが活発であり、物事に対する熱意や鬱、不安度などの人間の心理状態をライフログから推定しようとする研究が多数実施されている [3-5]。スマートフォンとスマートウォッチを併用した取り組みでは、収集したセンサデータから睡眠やストレスを推定し、鬱の指標となるアンケートの結果との相関を見ることで、どのような活動が鬱の原因となるかを判断する取り組みがなされている [3]。しかし、一般に広く利用されるようなサービスを考えた場合、普及率がスマートフォンに比べ低いスマートウォッチを併用するシステムは好ましくはない [6]。さらに、現在のライフログを利用した健康管理手法では、使用者が能動的に情報を入力し、それまでの蓄積したログとの統計的な差異などの情報から自身の健康や生活改善の指標を得るものが多く、受動的に専門医などのフィードバックを受けられるような仕組みや、生活改善のアドバイスを受けられるような仕組みはいまだ確立されていない。

これに対し、本研究ではひきこもり予兆を検知するため、大学生の多くが所持しているスマートフォンのみを用いてライフログを収集し、分析を行える情報収集基盤を構築する。予兆の検知のため、スマートフォンのセンサデータから、睡眠時間の推定と訪問地点の推定を行う。睡眠時間の推定により睡眠時間の乱れを、訪問地点推定により授業への出席と訪問地点数の変化をそれぞれ分析する。睡眠時間推定については、スマートフォンのセンサから得た特徴量を用いた機械学習による推定を行う。

地点推定に関して、本研究において目標となるのは生活する上での訪問地点数の推定であり、訪問地点の地図上における絶対位置を特定する必要はない。そのためGPSによる絶対位置の特定は行わず、Wi-Fi アクセスポイントのMAC アドレスおよびRSSI(受信信号強度)を用いてクラスタリングを行い、訪問地点数を推定する。これによってGPS収集による精神的圧迫感を避けることも期待できる。さらに、各地点に対して、習慣的なデータから意味付けを行うことで、訪問地点の属性の推定も行う。例えば、ある大学生が毎週月曜日の日中に同じMACアドレスを持つ地点を訪れていた場合、その地点を教室だとみなし、学生が授業に出席しているものとする。また、夜間に長時間同じ場所に滞在していればそこを家だとみなす。もしその習慣データから逸脱したデータが観測された場合、授業の欠席や、生活リズムの崩れを推定することが可能となる。

実際に手法の妥当性を検証するため、実験を実施した。その結果、地点数推定実験では、滞在している地点数の推定が可能であること、地点の意味付けを行ったことで、自宅の帰宅時間の推定において、誤差9.6分での推定ができていることを確認した。大学の授業時間が90分であることを考慮すると、本手法が目標としている授業の出席を検知できる見通しを得た。睡眠推定ではBalanced Random Forestを使った睡眠の有無を推定するモデルを構築し、全ての特微量と6つのみの特微量を使用した場合のモデルでそれぞれF値0.86の結果を得た。また、特微量を減らした場合の考察も行った、最適な特微量を用いてモデルを構築した場合、特微量が3つ以上の時F値0.84であることを確認し、本手法がオプトインによるデータ収集で、特微量が抜け落ちた場合も正確な推定ができる可能性を示した。

本稿の構成は以下の通りである。第2章ではライフログを用いた心理状態の推定の手法について既存の関連研究を述べる。また、位置推定や睡眠時間の推定の手法についても既存手法を紹介し、それらに対する本研究の位置付けを明確にする。第3章では本研究が目的とする学生生活支援システムの構成について述べる。第4章では兆候検出のための推定手法について説明を行う。第5章では本手法における地点数・睡眠推定の評価実験の結果を示す。最後に第6章にて本研究のまとめと今後の課題について述べる。

## 2. 関連研究

### 2.1 ライフログを用いた心理状態の推定手法

近年、日々の活動の記録であるライフログを用いた心理状態の推定手法についての研究が活発に行われている [3-5]。例えば、鬱病の可能性のある学生を検出するために、幸福度を推定する手法が提案されている [4]。ライフログの取得に用いる電話やSMSの履歴、画面のオンオフやGPSといったスマートフォンから得られるデータに加えて、ストレスの推定などに用いられる心拍やEDAなどのスマート

ウォッチから得られるデータを1ヶ月間収集しており、幸福度の推定を72%の精度で達成している。また、スマートフォンのみを用いた不安度予測を行なった研究では、スマートフォンのセンサ情報や使用履歴と不安度のアンケートの相関から、どのような操作が不安度を高めているかの予測を行なっている。個人に対して、数日間のデータから予測モデルを作成し、次の日の不安度を予測する手法で、最大7日間のデータを利用することで、8割程度の精度で不安度の変動を判定することに成功している [5]。

本研究では、既存手法とは異なり、ひきこもりの予兆を定義することで、それらと取得データの比較を行う。また、広く使用されるシステムを想定し、普及率の低いウェアラブルデバイスを用いず、スマートフォンのみでの推定を行う。

## 2.2 睡眠時間の推定

モバイル端末を用いた睡眠推定の研究は広く行われてきた [7-10]。モバイル端末の種類として、スマートフォンや、スマートウォッチなどのウェアラブル端末が考えられる。スマートフォンを用いた例では、周りの環境の騒音レベルや画面のオンオフの状況、加速度と組み合わせることで、入眠時刻と起床時刻を推定している。また、得た結果の精度を高めるため、各個人の睡眠のリズムから推定される睡眠時間によって、結果を補正することで、入眠時刻を  $42 \pm 48$  分、起床時刻を  $42 \pm 57$  分の絶対誤差で推定している [8]。また、スマートフォンの内蔵マイクのみを使用して、体の動き、いびきなどの睡眠の質と密接に関わるイベントを検出し、睡眠の質を定量的に評価するシステムを構築した研究も存在している [10]。一方、スマートウォッチを使用した睡眠推定法では、心拍等から睡眠のリズム、質まで測定する研究もなされている [7]。一般に、スマートウォッチなど侵襲型の端末の方が精度が高いとされているが、一方でその侵襲性による睡眠の質の低下が報告されている [11]。そのため、本研究では本来の目的である学生生活の支援に則り、睡眠の質を低下させないように、スマートフォンのみを用いた睡眠の推定を行う。

## 2.3 モバイル端末を用いた位置推定

位置推定の手法として、Wi-Fi アクセスポイントのデータを用いて、Point-of-Interest を推定する取り組みがなされている [12, 13]。GPS と Wi-Fi アクセスポイントの対応のデータベースを作成した研究では、Wi-Fi アクセスによって位置履歴を継続的に取得できるシステムを構築することにより、Wi-Fi アクセスポイントのデータのみでの位置特定を行っている。これによって、GPS では誤差の出やすい屋内環境においても高精度に位置を推定している [12]。

本研究において重要なのは生活する上での訪問地点の推定であり、訪問地点の地図上における絶対位置を特定する

必要はない。また、授業参加の推定を行う上で、屋内にある複数の教室を区別して推定を行いたい。よって、GPS での直接の位置情報取得は行わず、Wi-Fi アクセスポイントによる、訪問地点の推定を行う。これにより使用者の精神的な圧迫感の軽減も期待できる。

## 2.4 本研究の位置付け

本研究では、大学生を対象としたひきこもり予兆検知を目的としたデータ収集基盤を構築することを目的とし、その過程において睡眠時間の推定と、訪問地点の推定を行う。このようなライフログを用いての心理状態推定の研究は多数存在するが、ひきこもりの傾向を対象にしている研究は少ない。そのため、本研究ではひきこもりの心理的傾向を分析し、検知することでひきこもり状態に陥ることを防止することを目標としている。また、今後データ収集を行う過程で、医学的なアンケートを踏まえた心理傾向とライフログの相関を調査する。

## 3. システムの設計

### 3.1 学生生活支援システムの概要

従来の学生と医者間の2者関係では、学生側からの来院があってカウンセリングなどの処置を行うという形態が一般的であり、医者側からの早期での介入というのは難しくほとんど行われてこなかった。このため、学生自身が自らの不調を自覚し、受診するまで問題が表面化しないため、受診時にはすでに問題が深刻化してしまっており、治療が難しくなってしまうケースが珍しくなかった。この問題を解決するために本研究では図3のような情報側の介入を含む3者間でのシステム構築を目標としている。提案

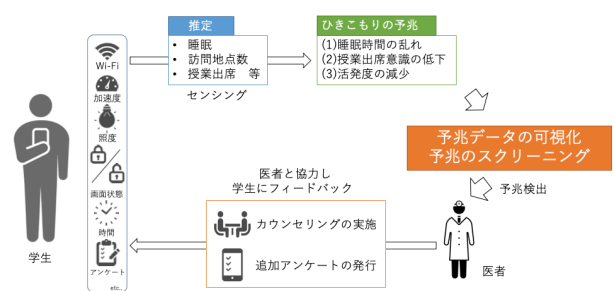


図3 学生生活支援システム概要

するシステムでは、学生からスマートフォンで収集した生活データの分析を行う。そして分析結果から、睡眠の状態や授業の出席状況などの生活の状況を推定し、それらを組み合わせることでひきこもりの予兆検知を行う。一定期間収集する中で、ひきこもりの兆候などが見られる学生に対しては生活状況の改善を提案するとともに医者側に分析結果を通知し、面談などによるメンタルケアを実施する。これにより従来では難しかった学生の生活リズムの変化をよ

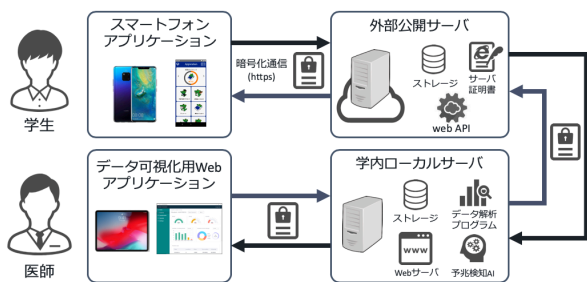


図 4 学生生活支援システム構成

り早い段階から知ることができ、医者側は問題を抱えている可能性のある学生に対して早期介入を行うことができる。

### 3.2 学生生活支援システムの構成

このシステムは医学の正しい知見をもとに兆候を判断し、適切なアドバイスを行う必要がある。そのため、本研究は大阪大学保健センターや医学系研究科精神医学教室と協力しシステムの構築を目指している。本研究で想定するシステムの構成を図4に示す。本システムでは大きく分けて4つの実装を行う。

#### 3.2.1 スマートフォンアプリケーション

スマートフォン上のセンサから得られる情報や、Wi-Fiの情報収集し送信を行う。また、生活状況に関する簡単なアンケートの回答の送信も行う。使用者のプライバシーに配慮し、情報収集は全てオプトイン/アウト形式で行い、それぞれのデータについて収集の有無を利用者側で選択できる機能を実装する。送信内容はアプリ上で暗号をかけたのちさらにhttpsを利用した暗号化通信を行う。これによりデータは二重の暗号をかけられた状態で外部公開サーバへと送られる。また、収集データはそれのみで個人を特定できるような形式は取らず、医者側のみで個人との紐付けを行えるようにする。アプリはデータを送信するだけではなく、睡眠予測を用いた睡眠日誌の提供や、出席管理のリマインダなどにも活用し、ひきこもりの兆候が見られた場合は医者からのフィードバックが届くようにする。

#### 3.2.2 外部公開サーバ

アプリによって暗号化されたデータの一時的な保存を行う。外部からアクセスできるサーバに収集データを直接置くことを回避することでセキュリティの担保を図る。学内ローカルサーバから送信要求があった場合保存されたデータを送信し、外部公開サーバ内からは削除する。睡眠日誌のデータや医者からのフィードバックなどが行われる場合も同様に、学内サーバから送られてくる暗号化されたデータを一時保存し、スマートフォンアプリに送信する。

#### 3.2.3 学内ローカルサーバ

アプリから学外サーバを経由して送られてきたデータを復号し、データの分析や予兆検知などの処理を行う。送られてきた生のデータの保存を行うが、学内からのアクセス

のみに対応しており、医者などの限られた人間しかアクセスできないようにすることでセキュリティに配慮している。分析情報や医者からのフィードバックはhttps通信を用いて、学外サーバを経由しスマートフォンアプリに送信する。この時、スマートフォンアプリ側で独自に発行される公開鍵を用いて暗号化することによりデータに二重の暗号を施す。

#### 3.2.4 データ可視化用 Web アプリケーション

医者側の閲覧用に主に利用されるアプリケーションである。分析により得た生活状況の情報やアンケートの回答などが閲覧できるようになっている。閲覧情報が誰のものであるかは医者のみが紐付けできるようにする（対応表などはネットから隔離する）ことで、万が一暗号化されていないデータが流出しても匿名性が担保される。生活状況から問題等が見受けられる学生に対して追加のアンケートや、カウンセリングの提案などのフィードバックもこれを通して行う。

このようなシステムのもと実際にサービスを運用し、学生生活支援を行うことが本研究の目標である。

### 3.3 収集データ基盤の設計

ひきこもりの予兆を検知するため、ライフログの収集を行う必要がある。本研究では、学生に対して広く適応できるシステムの構築を目指しており、ライフログの収集を学生のほとんどが所有しているスマートフォンのみを用いて行う。ひきこもりの予兆検知実現に向けた課題として、学生がひきこもりまたは準ひきこもりに至る根本的な原因は一樣ではないということが挙げられる。例えば、授業に出席できない原因として、友人や先生との対人関係の悪化、アルバイトなどが忙しく時間通りに起床することが難しいこと、授業に対する熱意が低く部活やサークルに参加はするものの授業には出席しないことなどが考えられる。そこで、本研究では大阪大学保健センターや医学系研究科精神医学教室と協力し、ひきこもり症状に対する予兆を定義し、中長期的な生活データ収集からその予兆の検知を目指す。今回検知を目指す医学的知見から定義したひきこもりの予兆は、(1) 睡眠時間の乱れ、(2) 授業出席意識の低下、(3) 訪問地点数の減少の3点である。睡眠時間の乱れは、昼夜の逆転による生活リズムの乱れを引き起こし、日中の活動に悪影響を及ぼす。そのため、規則的な活動を行うことが難しくなる。授業参加意識の低下は、睡眠の乱れや、部活やサークルなどに熱意が移ってしまうことなど、様々な要因から生じる。長期化することにより大学に行く意味を見い出せなくなりそのままひきこもりや準ひきこもりになってしまうことが推測できる。訪問地点数の減少は外出の頻度や活動範囲の縮小を示しており、これの常態化は準ひきこもりの定義そのものと合致する。

各予兆を検知するため図5のようなフローに基づきデー

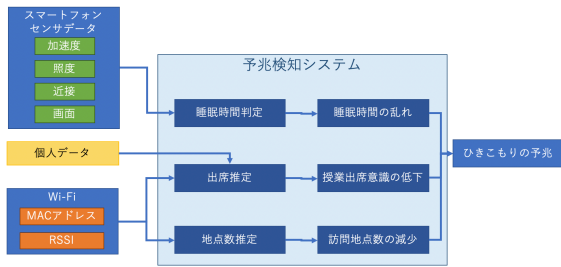


図 5 データ収集による予兆検知の概要

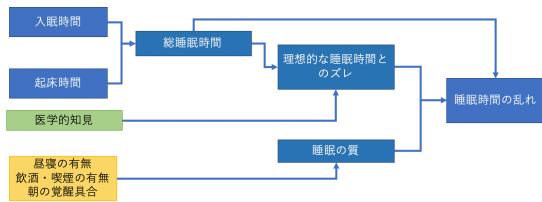


図 6 睡眠時間推定のフロー図

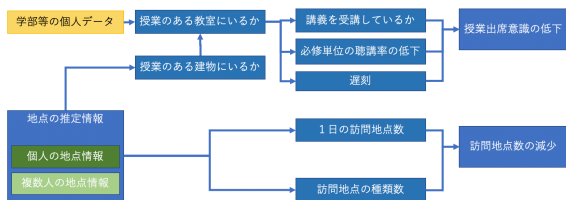


図 7 出席推定と地点数推定のフロー図

データの収集を行う。これは予兆検知のために必要なデータの流についてまとめた図となっており、それぞれの予兆をスマートフォンから取ることのできるデータとアンケートなどの個人データを組み合わせることにより推定を行う。睡眠推定・出席推定・地点数推定に関して、図 6, 7 の流れにより推定を行う。これは、予兆検知のために推定が必要な情報をどのようなデータの組み合わせによって推定するかを表している。睡眠の場合はスマートフォンが、画面がオフで動きがなく、周りが暗い状態になっている時間を睡眠時間と考えることができる。これはスマートフォンの加速度センサや照度センサから予測できる。また、カバンや、ポケットなどでも同様の状態が考えられるが、近接センサを使用すれば判別が可能であると考えられる。そこで、これらの要素が大きく睡眠に関わっていると仮定し、特徴量として機械学習からモデルを構築することで睡眠の推定を図る。授業出席意識の低下と訪問地点数の減少は位置情報により推定を行うことが可能である。GPS を利用することで詳細な位置情報を取得することができるが、一方でユーザーに与える精神的な圧迫度が強いと考えられ、長期的な継続利用を考慮する場合、ユーザーの負担を減らすため、GPS を用いない位置推定手法が必要である。そこで、本研究では、WiFi の MAC アドレスと RSSI を用いて、訪問地点数、自宅にいるか、大学にいるか、講義を受けているか

表 1 使用 Wi-Fi データ

データ項目	単位
MAC アドレス	任意の名称
受信信号強度	dBm

程度の推定を行うことで、精神的な圧迫度を低くする。また、詳細な情報が必要と考えられる場合においては簡単なアンケートに回答してもらうことで確認する。それぞれの手法に関して、第 4 章にて詳しく説明し、第 5 章において、検証、評価を行う。

また、実際のデータ収集では、推定に関して医学的な知見が必要になる場合が考えられる。例えば、睡眠の乱れを検知する際、理想的な睡眠時間とのズレを判別する必要があるが、これには医学的知見が必須である。また、睡眠の質を調査する際は、個人的な精神状態をスマートフォンだけで検知することは難しく、医学的指標に基づいたアンケートとの相関により判別を行う。

## 4. 推定手法

### 4.1 地点数推定手法

本手法では Wi-Fi データを利用して地点数推定を行う。収集するデータの一覧は表 1 に示す。ここではクラスタリング方法を説明したのち、長期間での予測を想定とした特徴量の再生成について説明する。

#### 4.1.1 クラスタリング

データの分析には凝集型階層的クラスタリングを用いた。凝集型階層的クラスタリング手法は、それぞれ独立したクラスタを、クラスタ間の距離が最も近いもので併合していく。併合された新クラスタと残りのクラスタ間での距離を再計算し、新クラスタを含めたクラスタの中から最も距離の近いものを併合する。同様にクラスタの併合と、距離の再計算を繰り返し、全体のクラスタが 1 つになった時点で終了する。クラスタ併合の過程は木構造で表すことができ、各クラスタが併合された時のクラスタ間の距離を基準に最終的なクラスタ数を決定することができる。クラスタリングのプリプロセッシングとして、取得した各データを信号強度が強い順にソートし、上位 10 種の MAC アドレスのリストを作成した。そのため、クラスタリングのための特徴量は各時間での取得した MAC アドレスのリストのみである。特徴量同士の類似度  $s(C_1, C_2)$  は次式で定義した。  $C_n$  は各特徴量を示す。

$$s(C_1, C_2) = \frac{1}{\min(|C_1|, |C_2|)} \sum_{x_1 \in C_1} \sum_{x_2 \in C_2} d(x_1, x_2)$$

$$d(x_1, x_2) = \begin{cases} 1 & (x_1 = x_2) \\ 0 & (x_1 \neq x_2) \end{cases}$$

この類似度の定義の元、各クラスタを最終的に 1 つのクラスタになるまで併合する。併合の優先度は取得した時間が

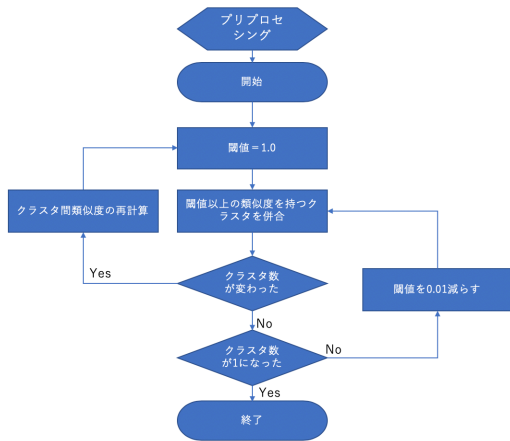


図 8 クラスタリングのアルゴリズム

近いリストの類似度が高くなると仮定し、類似度が同じものが複数ある場合は取得時間が近いものから優先して併合することとした。続いて、このクラスタリングのクラスタ凝集のためのアルゴリズムを図 8 に示す。まず、閾値を完全な一致を示す 1.0 に設定する。その後閾値以上の類似度をもつクラスタ同士を併合する。この時その閾値に対して併合が起こらなかった場合閾値を下げる事で、併合を進める。今回の手法では閾値の減少幅は 0.01 に設定した。併合が起きた場合、新しいクラスタ間での類似度の再計算を行う。その後類似度の閾値を 1.0 にリセットし、再度クラスタ同士の併合を行う。このアルゴリズムにより、生成された階層的クラスタに対して特定の閾値を与えることでその閾値でのクラスタリング結果を得ることができる。

#### 4.1.2 特徴量の再生成

クラスタリングを行うことにより一定の期間滞在した場所などの判別を行うことができる。一方で長期間での運用を考えた場合、移動等の短い間隔で検出されるであろうクラスタが残りに残るのは、地点の属性を考えることに対して悪影響を与える可能性がある。そこで、一定時間以上の滞在が確認できた地点を主要な訪問地点であると考え、その地点を示すクラスタを形成する MAC アドレスから新たな特徴量を生成する。主要な訪問地点の抽出の例を図 9 に示す。この図では、下の 2 つのグラフが 1 日のクラスタリング結果を示しており、20 分以上滞在した地点であるクラスタをそれぞれ四角で囲っている。2 つのグラフの四角で囲まれたクラスタが、上の 1 週間のクラスタリング結果を表すグラフの四角で囲われているクラスタと対応しており、下の 2 つのグラフでの短時間のクラスタの変化は上のグラフには反映されていない。このように、移動や、短時間の滞在を示すクラスタを除去し、主要な訪問地点のみを示すクラスタのみで再構成を行うことが、ここでの目的である。そのため、長時間の滞在を示すクラスタがもつ特徴量から新たに特徴量の生成を行う。新たな特徴量を生成する場合、このクラスタを形成している MAC アドレスを全

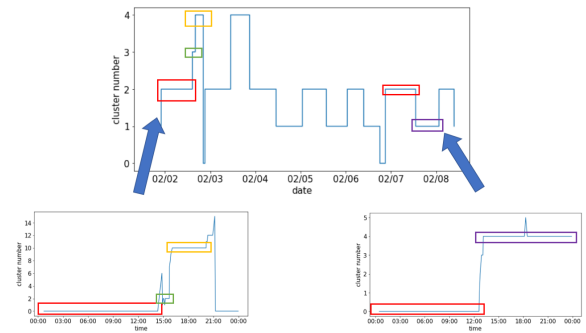


図 9 主要な訪問地点の抽出

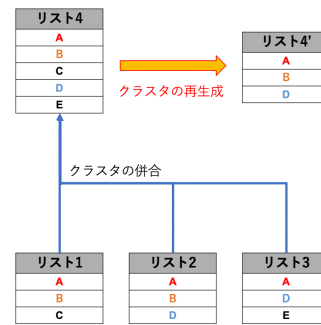


図 10 クラスタ再生成の一例

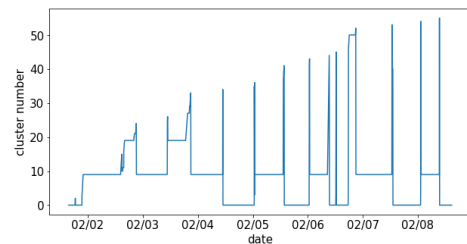


図 11 生データをクラスタリング

て適応すると、ほとんどの地点では確認されていない MAC アドレスを含んでしまい、誤検出を引き起こす可能性がある。そこで、再生成時に、図 10 のようにクラスタの生成に寄与が高かった MAC アドレスのみを抽出する。抽出の基準はクラスタ形成を行っていた元データのうちの 2/3 以上に含まれていることとした。これにより細かなノイズのようなクラスタを除去し、検知したい主要なクラスタのみを比較の対象とすることができる。この操作によって生成されたデータ群に対して、もう 1 度同じ手法でのクラスタリングをかけることによって、長期間での主要な訪問地点のみを検出できる。また、ある期間で主要なデータとみなされたが、さらに長い期間ではその傾向が見られなかったクラスタに関しては、再クラスタリングにおいて、その他のクラスタとして一律に番号 0 を割り振った。実際に取得したデータによる比較を図 11, 12 に示す。縦軸はクラスタ番号を示しており、各クラスタは横軸の時刻に訪問していた地点を表す。連続して同じクラスタが測定されてい

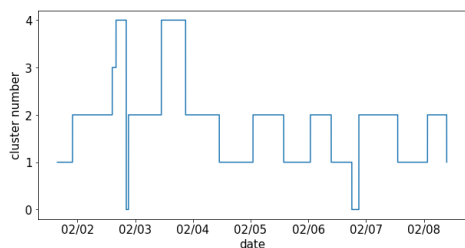


図 12 再生成し再度クラスタリング

表 2 睡眠推定における特徴量

特徴量	内容
slots	時間 単位は hour
acc_x	加速度の変化量
acc_y	
acc_z	
mag_x	地磁気の変化量
mag_y	
mag_z	
gyro_x	角速度の変化量
gyro_y	
gyro_z	
batteryStatus	充電中か否か
batteryLevel	充電の度合 (0~100)
illuminance	照度
proxi	近接センサの値 (0~5)
pressure	気圧
screen	画面の ON/OFF

る時間は、ある地点にその時間だけ滞在していることを示す。取得データをそのままクラスタリングした図 9 では 1 週間の間に 50 種類以上のクラスタが検出されていることがわかる。一方で、ノイズのようなクラスタは、移動や短時間の滞在などの地点を表しており、同じ番号のクラスタも検知されていることが少ないことからわかるように、主要な地点を示してはいないと推定できる。図 12 は再生成した特徴量に対して再度クラスタリングを適応したグラフである。図 11 のグラフ上で長時間の滞在が推定されるクラスタをうまく抽出できていること、また、短時間のノイズのようなクラスタが除去できていることがわかる。今回は第 5 章で示す実験で収集した 1 週間のデータであったが、期間が長期間になればなるほど有用な処理であると考えられる。

## 4.2 睡眠推定手法

モバイル端末を用いた睡眠推定については、これまでいくつかの研究が行われている。これらの研究では、モバイル端末から取得したセンサデータを用いて、睡眠時間、または、睡眠状態を推定するモデルを構築している。本研究においては、大半の大学生が所持しているスマートフォンにより取得できる特徴量のみを用いて、機械学習により推定モデルを構築することとする。

### 4.2.1 特徴量

表 2 は、推定モデルの構築に用いる特徴量の候補を表している。slots は時間帯を表しており、値の範囲は 0~23 で、AM, PM を区別した時刻である。acc\_n, mag\_n, gyro\_n は、それぞれ直前に取得した値との差分を特徴量とした。batteryStatus は充電中か否かを 2 値によって表す。proxi は近接センサの値であり、最も離れている時に 0 を、最も近づいている時に 5 を表す。その他の特徴量に関してはセンサごとの計測値を用いる。既存手法のいくつかは、マイクから取得できる特徴量を用いて精度が高いモデルを構築しているが、プライバシーの問題を考慮し本研究からは除外することとした。

### 4.2.2 特徴量の学習方法

[9] などの既存研究においては、回帰モデルにより睡眠時間を推定するモデルを構築しているが、医師はカウンセリングにおいて、眠っている間の時間帯が塗りつぶされた睡眠日誌を確認して診察を行うため、睡眠時間を直接推定する回帰モデルは提案システムの目的に適さない。そのため、任意の時間におけるセンサ値から睡眠中か覚醒中か判定する分類モデルを構築する。また、収集するデータは、覚醒時の方が睡眠時よりサンプル数が多く、実際に使用する場合においても覚醒時のデータが多くなることが予想される。このようなサンプル数に偏りがある不均衡データでは、そのままモデルを構築すると、サンプル数が少ないラベルの分類精度が低いモデルが生成される。そのため、決定木ごとにサンプル数を調整することで、不均衡なデータに対応可能な Balanced Random Forest によりモデルを構築する。モデルの検証においては Leave-One-Group-Out Cross-Validation を用いた。

## 5. 評価実験

### 5.1 地点推定

#### 5.1.1 実験概要

ここでは Wi-Fi ログを用いて地点数を推定するための実験環境について述べる。本研究では、位置情報を詳細に取得することなく地点数を推定することで、精神的な圧迫度を避ける目的がある。また、実際に授業出席の推定を想定して、ある地点に滞在している時間を推定する実験を行った。実験用の情報収集デバイスとして、LG Nexus5 を使用し、表 1 のデータを一定時間間隔で取得できるアプリケーションをインストールし被験者に配布した。クラスタリングにおける最終的な閾値の設定に関しては図 13 のようなエルボー図に閾値とクラスタ数の関係をまとめ、比較的クラスタ数が一定の数に収束するとみられる閾値を選択する。本実験では地点数推定実験で得られたエルボー図を基準とし、全ての実験において閾値を 0.65 に設定した。評価のため、地点数の推定と地点属性を利用した滞在時間推定のそれぞれについて個別で実験を行った。地点数推定実

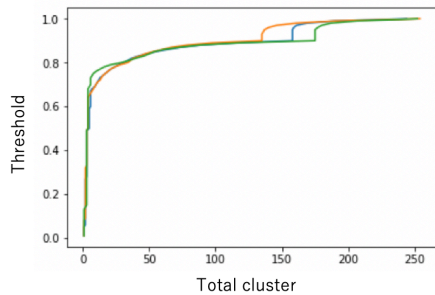


図 13 閾値の減少に伴うクラスタ数の変化

験は取得地点数が現実の状態に則しているかを判定するために行った。実験方法は 22~24 歳の男性被験者 6 名を 2 グループに分け、各グループを異なる地点に滞在させた。移動のパターンについては表 3 に示す。データの取得間隔を 10 秒とし、現在の地点を出発してから次の地点を出発するまでの時間間隔を 15 分とした。評価基準は、この滞在していた地点数を正しく検出できたかどうかとした。データ収集間隔が Android 端末の処理仕様により、ズレが生じる場合があったため、設定した間隔でデータ収集を行えない期間が存在した。

地点属性を利用した滞在時間推定実験では、22~27 歳の男性被験者 9 名に対し 1 週間の間、配布したスマートフォンを各自の所有のものと同様に持ち歩くものとし、データの取得間隔を 5 分に設定して実験を行った。真値として配布のスマートフォンで GPS 情報も同様の時間間隔で取得したが、地点数推定実験と同様に設定時間間隔にズレが生じる場合があったため、真値との誤差が 5 分でない場合が存在した。属性推定についてはデータ収集期間に日中最も長い時間を過ごした場所を自宅として設定を行った。評価として、自宅の属性推定から、推定した帰宅時間と実際の時間との誤差を検証した。基準として、帰宅時刻としたのはある地点の滞在している時間よりも到着したと考えられる時刻の方が有用であると考えたためである。本研究はひきこもりの予兆検知を目的としており、予兆として考えられる授業出席意識の低下を検出する場合は、通常授業の出席は何分間教室にいたかよりも、遅刻の有無やいつから来ていたかなどが評価となる。そのため今回は滞在時間ではなく、到着時刻を評価基準とした。

### 5.1.2 地点数推定実験

実際に各グループが収集したデータのクラスタリング結果を図 14, 15 に示す。縦軸が検出された各クラスタの番号、横軸が時間を表し、グラフの変化しない長さから、ある地点での滞在時間が推定できる。

グラフの各色は参加者一人あたりのデータを表しており、グループ A の結果を図 14 に、グループ B の結果を図 15 にそれぞれまとめている。参加者によって取得データが異なるためクラスタ数にばらつきはあるものの、15 分程度の滞在とみられるクラスタはどちらも 3 つ程度存在すること

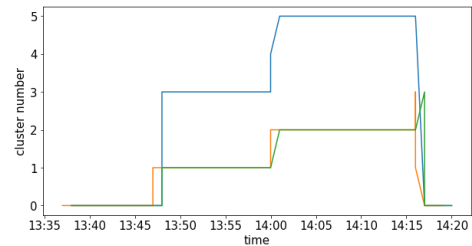


図 14 グループ A のクラスタリング

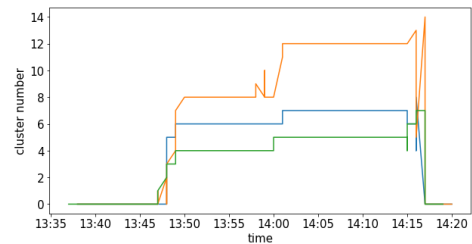


図 15 グループ B のクラスタリング

が確認できる。これは開始地点も含めた地点数と合致している。グループ B が A に比べて多くの地点数を検出しているのは、A が建物の上下の移動をエレベーターで行ったのに対して B では隣の建物への徒歩での移動だったため、徒歩での移動中に多くの MAC アドレスを検出したためだと推測できる。この結果から、本手法によって地点数推定の検証ができたと言える。

### 5.1.3 時間推定実験

本実験では、各被験者の 1 週間分のデータから自宅を推定し、帰宅時刻推定の性能を評価した。帰宅時間は自宅とみなせるクラスタが最初に観測された時間、真値は GPS のデータから自宅の半径 150m に最初に到達した時間とする。クラスタリングを行ない、真値の GPS データと比較した結果、全体での平均誤差は 9.6 分となった。各参加者における平均誤差を表 4 に示す。使用した端末での測定では、設定した取得時間間隔に対して、実際の取得時間に誤差が生じることがあった、それにより、各被験者の間で誤差のばらつきが生まれたと考えられる。しかしながら、どの参加者においても 9.6 分の誤差と大きく乖離することはなかった。この結果は、授業出席意識の低下を判断する際、授業時間が 90 分であることを考慮すると、大きな遅刻や欠席に対して対応ができることを示している。

## 5.2 睡眠推定実験

### 5.2.1 実験概要

本実験ではスマートフォンから収集した特徴量から睡眠推定モデルを生成し評価を行なった。収集した特徴量は 4.2.1 項で示したものである。22~26 歳の男性 7 名の被験者に対して、被験者所有のスマートフォンにデータ収集用のアプリをインストールしてもらい、普段の生活中的データを収集した。データは 1 分間隔で収集し、それぞれの取



表 3 実験グループ移動パターン

	～13:45	～14:00	～14:15	それ以降
A	建物 P 5 階	建物 P 3 階	建物 P 1 階	建物 P 5 階
B	同上	建物 Q 1 階 P 側	建物 Q 1 階 P 逆側	同上

表 4 帰宅時間との誤差

参加者	平均誤差 (分)
A	7.5
B	15.5
C	7.2
D	11.1
E	7.5
F	7.0
G	6.7
H	6.3
I	15.0

表 5 全ての特微量を使用した場合の推定精度

	Precision	Recall	F1-score	support
Wake	0.92	0.88	0.90	73613
Sleep	0.68	0.78	0.73	24056
micro avg	0.86	0.86	0.86	97669

表 6 特微量の重要度

	Importance
slots	0.373514
illuminance	0.156453
batteryStatus	0.104039
batteryLevel	0.098374
pressure	0.058113
screen	0.045141
acc_z	0.029621
mag_x	0.023478
mag_z	0.023193
mag_y	0.020047
gyro_x	0.017428
acc_y	0.015831
gyro_y	0.013926
proxi	0.008515
gyro_z	0.007104
acc_x	0.005224

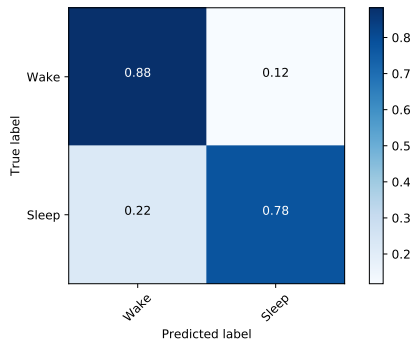


図 16 全ての特微量を使用した場合の混同行列

得データに対して、真値として一日に一回起床時に就寝時間と起床時間の記録をとった。分に関しては 15 分間隔で入力する (00, 15, 30, 45)。これは主観評価であり、睡眠の真値としては誤差があるものであるが、実際に医学のカウンセリングで睡眠日誌を入力する際は主観評価での入力となるため、ひきこもりを対象としたシステムを考えた場合主観評価の方が、追加装置を用いた正確なものよりも適しているものと考えた。7 名の被験者から収集した、合計 92 日間のデータから生成した特微量から Balanced Random Forest を用い学習を行った。深さを 10, 決定木の数を 32 に設定を行い、Leave-One-Group-Out Cross-Validation によって検証することで生成したモデルの性能を評価した。

### 5.2.2 全ての特微量を使用した場合

全ての特微量を使用した場合の Confusion matrix を図 16 に、モデルの推定精度を表 5 に示す。実際に起きていた時間を 9 割近い精度で推定でき、実際寝ていた時間の 8 割近い精度で推定できていることが分かる。しかしながら、構築したモデルでは、実際に寝ていた時間のうち 2 割程度を誤って起きてしていると推定しており、適合率も低くなっている。これでは、睡眠時間を過少評価してしまうことになるため、より誤りが少ないモデルを構築する必要がある。

全体の特微量を使用したモデルの構築に対して各特微量

の重要度を示した表が表 6 である。重要度の高さから、時間・明るさ・バッテリーの状況が大きく影響していることがわかる。一方で近接センサの値や、9 軸センサの値はモデル構築においてあまり影響していなかった。

### 5.2.3 提供される特微量による影響

提案システムでは、オプトイン/アウトによりデータを収集することを想定しているため、ユーザーの許可により、推定モデルに入力できる特微量の種類が異なることが予想される。そこで、提供される特微量による差を確認するために、特微量の組み合わせごとに睡眠推定モデルを構築し、推定精度への影響を確認する。しかしながら、全ての特微量を使用した場合の各特微量の重要度 (表 6) から睡眠の推定に寄与しない特微量があることは明らかであるため、比較的重要度が高い 6 つの特微量 (slots screen) について、全組合せの推定モデルを構築し確認を行った。

図 17 は、組合せで選択する特微量の数ごとに、最高の特微量の組合せと最低の特微量の組合せの推定精度を表している。

表 7 は、重要度上位 6 位までの特微量を全て用いて構築したモデルの精度を表している。F 値に関しては、全ての特微量を使用したモデルと同様に 0.86 という結果が得ら

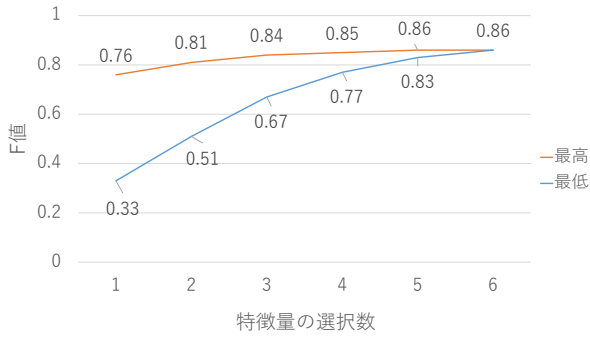


図 17 提供される特徴量の数による推定精度への影響

表 7 重要度上位 6 位までの特徴量を使用した場合の推定精度

	Precision	Recall	F1-score	support
Wake	0.93	0.88	0.90	73613
Sleep	0.68	0.80	0.73	24056
micro avg	0.86	0.86	0.86	97669

れているが、睡眠時の Recall が全ての特徴量を使用したモデルより高くなっている。

また、特徴量をさらに減らした場合には、特徴量 3 つを用いたモデルでは F 値が最高になるモデルの場合、0.84 と、全体の特徴量を用いた場合と変わらない精度を示した。このモデルは、slots, illuminance, batteryStatus という重要度上位 3 位までの特徴量を用いて構築されたものである。重要度が最も高い slots(時間帯) は、各特徴量数において F 値が最高になるモデルに常に含まれている特徴量であった。これは、ユーザーが提供するかどうかには依存しない特徴量であるため、ユーザーの選択によって特徴量が欠落しても、0.76 以上の推定精度で推定が行えると言える。しかし、時間帯だけで構築されるモデルは、単に一般的な大学生の睡眠時間が出力されるだけであるため、個々人の睡眠推定を行うためには、その他の特徴量が提供される必要があると考えられる。

#### 5.2.4 15 分毎の推定を行う場合

睡眠日誌は 15 分毎の睡眠を主観的に記録することが多いため、各特徴量について、15 分毎の統計量を新たな特徴量として生成し、15 分毎の睡眠状態を推定するモデルを構築した。統計量としては、数値があるものについては、平均、最大、最小、分散といった基本統計量を、battery status や screen については、15 分のうちに 1 度でも ON/OFF になったかを特徴量とした。また、slots に関しては、5.2.2 項と同様に時間帯を用いた。

図 18 と、表 8 は、新たに生成した合計 57 個の特徴量を用いた場合の混同行列と推定精度を表している。

1 分毎の睡眠状態を推定した 5.2.2 項のモデルと比べて、F 値は 0.86 と同様の結果であるが、覚醒状態の Precision と睡眠状態の Recall が大きく向上し、誤推定も少なく、覚醒/睡眠両方の Recall も 0.85 以上になっていることが分かる。

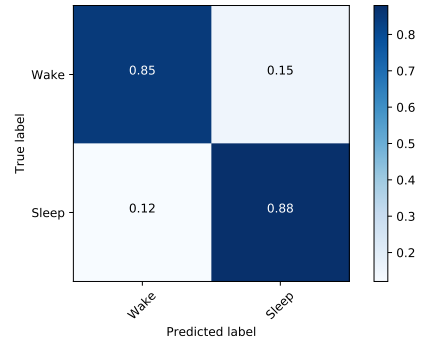


図 18 全ての特徴量を使用した場合の混同行列 (15 分毎)

表 8 全ての特徴量を使用した場合の推定精度 (15 分毎)

	Precision	Recall	F1-score	support
Wake	0.96	0.85	0.90	5308
Sleep	0.65	0.88	0.75	1711
micro avg	0.86	0.86	0.86	7019

表 9 特徴量の重要度 (上位 10 位まで)

	Importance
slots	0.112393
illuminance_avg	0.095324
illuminance_max	0.076125
batteryStatus_off_once	0.063131
batteryStatus_on_once	0.042928
illuminance_var	0.039807
batteryLevel_avg	0.039504
batteryLevel_min	0.032166
gyro_z_var	0.028267
gyro_x_min	0.026868

表 10 重要度上位 8 位までの特徴量を使用した場合の推定精度 (15 分毎)

	Precision	Recall	F1-score	support
Wake	0.96	0.85	0.90	5308
Sleep	0.65	0.89	0.75	1711
micro avg	0.86	0.86	0.86	7019

表 9 は、新たに生成した特徴量の重要度上位 10 位までを表している。表 6 では、slots の重要度だけ高くなっていたが、slots と illuminance\_avg との差は小さいことが分かる。また、重要度が 0.03 以上の特徴量は、slots, illuminance, battery status, battery level という、表 6 の上位 4 位までの特徴量から生成される特徴量である。表 10 は、これらの特徴量のみを使用した場合の推定精度を表しており、睡眠状態の誤推定が減ったことで、Recall が 0.89 まで増加している。

しかしながら、睡眠状態の Precision を見ると 0.65 と低く、覚醒状態を睡眠状態と誤推定していることが多いことが分かる。このような誤推定は、昼間や夕方に発生していることがあり、そのタイミングでは、照度が低い状態で充電を行っていた。これは、充電中にスマートフォンの画面をテーブルに向けて置いていたり、画面の上に物を置い

ていたりする場合などが考えられる。そのため、スマートフォンの向きや、照度センサの状態（近接センサなどから物が被さっていないか推定）など、特徴量を追加することで、誤推定を少なくできると考えられる。

## 6. おわりに

本研究では、ひきこもりの予兆検知を目的としたシステムを提案し、それを達成するための基盤の設計を行った。ひきこもりの予兆である睡眠時間の乱れ、授業出席意識の低下、訪問地点数の減少の検知を行うため、本手法は大学生の間で広く普及しているスマートフォンを用いた睡眠推定と地点推定の手法を提案し、それぞれについての妥当性を検証するため、評価実験を実施した。その結果、地点推定では、滞在している地点数の推定が可能であること、また、地点の意味付けを行ったことで自宅の帰宅時間の推定において、誤差 9.6 分での推定ができていることを確認した。また睡眠推定では、スマートフォンで収集した合計 92 日分のセンサデータを特徴量とし、Balanced Random Forest によって睡眠の有無を判定するモデル構築を行った。その結果 F 値 0.86 の結果を得た。また、特徴量が少なくなった場合の考察も行い、最適な特徴量でモデル構築を行った場合、特徴量が 3 つの場合でも F 値 0.84 であることを確認した。特徴量が減少したモデルにおいても大きく性能が下がらないことから、本手法がオプトインによるデータ収集で、特徴量が抜け落ちた場合も正確な推定ができる可能性を示した。一方で本稿では睡眠の有無を 2 値で判定するに留まったため、今後の研究において睡眠時間の推定を行う。また、実験において、修士以上の学生を対象にデータを収集したが、睡眠の乱れ検出を可能にするために、今後は、睡眠に問題がある生徒のデータも収集してモデルを改良する。システム面の課題として、ユーザーがシステムに対して継続的利用を行うような仕組みについて検討する必要がある。また、センサデータ収集にあたり、電池消費が多である場合継続的使用が望めない可能性があるため、電池消費を考慮したデータ収集方法を考える。そして、実際に学生を対象とした情報収集基盤の構築および学生生活支援アプリの開発を行い、学生から情報を収集する実験を行う予定である。この実験では、集めた学生生活データと、医学的な知見に基づいたアンケートによる心理状況の相関などを見ることで、センサデータから抽出できる特徴がどれほど予兆検知に対して有用かを調査する。

謝辞 本研究の一部は、文部科学省 Society 5.0 実現化研究拠点支援事業「ライフデザイン・イノベーション研究拠点」の支援を受けたものである。

## 参考文献

[1] 内閣府. 若者の生活に関する調査報告書 (平成 27 年度). <https://www8.cao.go.jp/youth/kenkyu/>

- hikikomori/h27/pdf-index.html.
- [2] 内閣府. 生活状況に関する調査 (平成 30 年度). <https://www8.cao.go.jp/youth/kenkyu/life/h30/pdf-index.html>.
- [3] Rui Wang, Weichen Wang, Alex daSilva, Jeremy F. Huckins, William M. Kelley, Todd F. Heatherton, and Andrew T. Campbell. Tracking depression dynamics in college students using mobile phone and wearable sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 2, No. 1, pp. 43:1–43:26, March 2018.
- [4] Natasha Jaques, Sara Taylor, Asaph Azaria, Asma Ghandeharioun, Akane Sano, and Rosalind Picard. Predicting students' happiness from physiology, phone, mobility, and behavioral data. In *Proceedings of 2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 222–228, September 2015.
- [5] 伊藤拓, 深澤佑介, 沖村宰, 山下祐一, 前田貴記, 太田順. スマートフォンセンサログとアプリ履歴からの不安度予測モデルの構築. Technical Report 55, 東京大学, (株)NTT DOCOMO, 慶應義塾大学, 国立精神・神経医療研究センター, 慶應義塾大学, 東京大学, March 2017.
- [6] デロイトトーマツ. 世界モバイル利用動向調査 2017. <https://www2.deloitte.com/content/dam/Deloitte/jp/Documents/about-deloitte/news-releases/jp-nr-nr20171214.pdf>.
- [7] Marko Borazio and Kristof Van Laerhoven. Predicting sleeping behaviors in long-term studies with wrist-worn sensor data. In *Proceedings of International Joint Conference on Ambient Intelligence*, pp. 151–156. Springer, 2011.
- [8] Florian Wahl and Oliver Amft. Data and expert models for sleep timing and chronotype estimation from smartphone context data and simulations. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 2, No. 3, p. 139, 2018.
- [9] Zhenyu Chen, Mu Lin, Fanglin Chen, Nicholas D Lane, Giuseppe Cardone, Rui Wang, Tianxing Li, Yiqiang Chen, Tanzeem Choudhury, and Andrew T Campbell. Unobtrusive sleep monitoring using smartphones. In *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare*, pp. 145–152. ICST, 2013.
- [10] Tian Hao, Guoliang Xing, and Gang Zhou. isleep: unobtrusive sleep quality monitoring using smartphones. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, p. 4. ACM, 2013.
- [11] Michael Littner, Clete A Kushida, W McDowell Anderson, Dennis Bailey, Richard B Berry, David G Davila, Max Hirshkowitz, Sheldon Kapen, Milton Kramer, Daniel Loube, et al. Practice parameters for the role of actigraphy in the study of sleep and circadian rhythms: an update for 2002. *Sleep*, Vol. 26, No. 3, pp. 337–341, 2003.
- [12] Jun Rekimoto, Takashi Miyaki, and Takaaki Ishizawa. Lifetag: Wifi-based continuous location logging for life pattern analysis. In *Proceedings of LoCA*, Vol. 2007, pp. 35–49, 2007.
- [13] Katayoun Farrahi and Daniel Gatica-Perez. What did you do today?: discovering daily routines from large-scale mobile data. In *Proceedings of the 16th ACM international conference on Multimedia*, pp. 849–852. ACM, 2008.