

ウェアラブルデバイスを用いた英単語4択課題における 確信度および主観的難易度の推定

森 朝春^{1,a)} 長谷川 達人^{1,b)}

受付日 2019年10月16日, 採録日 2020年4月13日

概要: 本稿では, e-learning 学習時の英単語 4 択課題を対象に, 確信度と主観的難易度を推定する手法を開発する. 眼球運動や心拍は知覚や心理状態と関連があることから, 提案手法では 2 種類のウェアラブルデバイス, 3 点式眼電位センサ搭載眼鏡型デバイス (JINS MEME) と胸部装着型心電図モニタリングデバイス (MyBeat) で観測されたセンサデータを用いることで確信度と主観的難易度の推定を行う. 提案手法の有効性を評価するために, 被験者 8 名に対して 2 種類のウェアラブルデバイスを着用した状態で英単語 4 択課題の実施を依頼しデータ収集実験を行った. 計測したセンサデータから 694 次元の特徴抽出を行い, 貪欲法にて特徴選択の後, 機械学習手法の 1 つである SVM を用いて, 確信度と主観的難易度の推定を行った. 分析の結果, 自身のデータを学習データとして使用できることを想定した場合, 確信度を 93.5%, 主観的難易度を 91.8% の精度で推定できることを明らかにした. さらに, 本稿の提案手法を用いることによって, 学習の効率や復習の見逃し率が大幅に改善できることを明らかにした.

キーワード: 機械学習, 学習支援システム, 主観的難易度推定, 確信度推定

Wearable Sensor-Based Estimation of Confidence and Subjective Difficulty for Efficient Vocabulary Learning

TOMO HARU MORI^{1,a)} TATSU HITO HASEGAWA^{1,b)}

Received: October 16, 2019, Accepted: April 13, 2020

Abstract: In this study, we develop a system that estimates the degree of confidence and subjective difficulty of four-choice questions of English vocabulary on e-learning platforms. Heart rate and eye movement differ depending on the learner's mental status and perception. We estimate them using two wearable devices: JINS MEME eyeglass-type device and MyBeat chest-mounted electrocardiogram monitoring device. To evaluate the effectiveness of the proposed method, we conducted an experiment assuming an e-learning environment for eight healthy males. Extracting total 694 features, our method estimates the degree of confidence and subjective difficulty using support vector machine with feature selection by greedy method. The result showed that our method could estimate the learner's confidence and subjective difficulty in answering with 93.5% and 90.8% accuracies, respectively when each learner's own annotated data is used in the training phase. In addition, it is confirmed that our method can greatly improve overall learning efficiency and the ratio of missed reviews.

Keywords: machine learning, learning support system, subjective difficulty estimation, degree of confidence estimation

1. はじめに

近年の教育分野の IT 化にともない, 大学や大手予備校などで, 学習者が Web 上で自学自習を行う e-learning が広く利用されている. e-learning の利点は, 学習を行う際

¹ 福井大学大学院工学研究科情報・メディア工学専攻
Department of Information Science, Graduate School of Engineering, University of Fukui, Fukui 910-8507, Japan

a) ji180260@u-fukui.ac.jp

b) t-hase@u-fukui.ac.jp

に時間や場所の制約が少ない点、また人的コストが削減できる点などがあげられる。e-learning の欠点として、学習者の状態に応じて動的に課題を変更することが難しいという点がある。従来の対面授業では、講師が学習者の様子を観察しながら講義を進めていくため、学習者のしぐさや表情を取得でき、学習者に直接尋ねることもできる。それにより、学習者の理解度やモチベーションをある程度把握して、講義に変化を加えることができる。それに対して、e-learning では、対面授業のように講師が 1 人 1 人の様子を観察するわけではなく、コンピュータ自体が学習者の様子を認識することもできないため、対面学習のように、柔軟な出題を行うことが難しい。そのため知的学習支援システムの実現に向けて、学習者の状態を認識する技術の実現が望まれる。

e-learning を用いた自主学習では、計算や暗記など、繰り返し訓練が必要となる勉強を行うことが多い。暗記学習に着目して考えると、長期的な記憶の定着には、記憶学習を反復して行うことが重要である。心理学者であるヘルマン・エビングハウスの忘却曲線 [1] によれば、学習から一定時間経過時に再び記憶するまでにかかる負担がどれだけ軽減できるかを示す節約率という度合いは、1 時間経過時で 58% である。また、1 日経過時では 33%、1 カ月経過後では 21% まで減少する。すなわち、1 時間後に復習する場合は、初めて学習したときの半分以下の負担で記憶できるが、1 カ月経つと約 80% の負担となり、初めて学習したときと同程度の負担になる。したがって、未記憶な課題は短時間で復習を行う方が記憶にかかる負担が少ない。

復習時にはすべての問題を再学習するよりも、まだ習得できていない、すなわち、「未定着な学習内容」のみを復習する方が効率的である。しかし、現実では未定着な学習内容のみを反復して学習できていないことが多い。なぜなら、学習者は自身の未定着な学習内容を必ずしも問題終了後まで覚えていないからである。学習者が 100 問の英単語課題を行った場合を想定すると、解答終了後に各問題の確信度を学習者が完全に記憶し手動で仕分けを行うことは現実的ではない。また、従来の英単語 4 択解答アプリを用いた場合、学習者は数問～数十問の英単語に 4 択で解答し不正解だった問題を復習することが一般的である。しかし、4 択の場合はまったく分からなくても 4 分の 1 の確率で偶然正解してしまったり、解答が分からなくても選択肢によっては推測して解答できたりする。推測によって正解した場合、4 択問題のときだけ答えが分かる状態になってしまい、実用的な語彙力が備わっているとはいえない。以上より、4 択課題における正解不正解の情報は必ずしも学習者の記憶度合いと一致しないため、不正解の問題のみ復習を行った場合、未定着な学習内容が見逃されている可能性がある。特に e-learning では、解答の入力と採点を容易にする都合上、選択問題が比較的多いため、この問題の解決は重要で

ある。

学習効率が重要な一方で、学習者が継続的に学習を行うためのモチベーションを維持することも重要である。特に e-learning では学習進捗は学習者の自主性に委ねられているため、学習者のモチベーションを低下させない工夫が必要である。もし学習中に学習者にとって難しい問題ばかりが出題されると、その範囲の学習を行うことが億劫に感じてしまい、学習自体を敬遠してしまう可能性がある。したがって、学習者が問題に対して感じる主観的難易度を認識し、学習者ごとに課題難易度を動的に変更することで、学習者のモチベーションを維持できる可能性がある。主観的難易度とは、学習者が学習内容に対して主観的に感じる「簡単」、「難しい」という評価である。すなわち、とある学習内容 X が一般的に難しいか否かを示す情報ではなく、とある学習者が X を難しいと感じるか否かを示す情報である。これは問題自体の難易度とも相関すると思われるが、学習者によって異なるものであり、解答の正否から得られるものでもない。前述したとおり、従来の e-learning では、システムが学習者の様子や印象を観測する機構が備わっていないため、学習者の主観的難易度が観測できず、出題難易度の動的な調整ができなかった。

人の主観的な印象を評価するために、生体情報がしばしば利用されている。そして近年、生体情報を簡易計測可能なウェアラブルデバイスが普及しつつある。現代のスマートフォンのように、将来このようなウェアラブルデバイスが広く普及することを考慮すると、デバイスを利用した学習支援環境を構築することには価値があると考えられる。そこで本研究では、PC 上で英単語 4 択課題を行うという特定の e-learning 学習環境下において、3 点式眼電位センサ搭載眼鏡型デバイス (JINS MEME) と胸部装着型心電図モニタリングデバイス (MyBeat) を用いた確信度、および主観的難易度の推定手法の開発を行う。各々の手法を開発することにより、復習に漏れのない効率的な学習を提供する学習支援システムの実現や継続的な学習支援が可能になると考える。本稿では、先行研究 [2] をふまえて、提案手法と評価実験方法について改めて詳細に説明するとともに、推定に使用する特徴量や作成モデルなど、より詳細な分析内容について述べる。また提案手法が学習効率にどの程度の影響を及ぼすかについての議論も行う。

2. 関連研究

2.1 確信度と主観的難易度

本研究では解答した課題の正否状態を分別することなく、課題解答時に学習者が感じている解答の正しさの程度を確信度と定義する。メタ判断記憶は、正否状態によって確信度判断と feeling of knowing (FOK) 判断の 2 種類に分岐するが、今回は正否で情報を分けることはせずと同様のものとして扱う。張の研究 [3] では、多岐選択問題に対

表 1 確信度と主観的難易度に関する関連研究

Table 1 Related studies on estimation of confidence and subjective difficulty.

Quotation	Estimation target	ML Method	Target	Using Device	Evaluation Method	Number of Subjects
[13]	Subject difficulty	Nothing	Unknown	Piezoelectric contact sensors, MVpen	User Verification	5
[14]	Confidence	SVM, NB	Input English word spelling	Smart Phone	5fold-CV	21
[15]	Subjective difficulty	SVM	Calculations	Camera	LOSO-CV	14
[16]	Subjective and objective difficulty	Nothing	Word Search	Web platform	Significance Test	725
[17]	High difficulty English words	DNN	English Long Sentence	Eye-tracker	Nothing	5
[18]	Subjective difficulty	SVM	Listening	FreeView	LOSO-CV	10
[19]	Confidence	SVM	English Grammar	Eye-tracker	LODO-CV	11

して、解答選択肢とともに確信度を答えさせることによって、効率的に学習者の情報収集を行う確信度付与法の提案を行っている。彼は混乱区間（正答肢に対する確信度と得点の順序関係が逆転する危険性のある正答確信度区間）という概念を数学的に定義することによって、既存の回答法と比較を行い提案手法の有効性を示している。これによって、多岐選択問題において確信度を推定する価値があることが分かる。次に、本研究では学習者が感じている課題に対する取り掛かりやすさを主観的難易度と定義する。すなわち、谷上 [4] の研究で用いられているような問題ごとの難易度とは異なり、学習者ごとに異なる問題に対する印象のことである。

2.2 学習者の主観と学習支援システム

e-learning に拡張現実やパターン認識技術などの技術的な機能を追加することで学習支援を行うことを目的とした研究 [5], [6], [7], [8], [9], [10] が行われている一方で、学習者自身の理解度や嗜好性などの主観的な内容を活用した学習支援についても研究が行われている [11], [12]。従来のシステムでは、得られる情報から学習者の思考や感情、印象を把握することが難しかった。しかし、なんらかの手段を用いてこれらの主観的印象をコンピュータが認識することができれば、学習者の状況に応じた支援を行うことが可能である。そこで、学習者の課題に対する印象を分析している研究のなかから、今回は特に学習者の確信度と主観的難易度の計測を試みている研究について調査を行った。その概要を表 1 にまとめる。

Hamaguchi ら [13] の研究では、机の下側に圧電式接触センサを取り付けることによって、手書きによって生じる机の振動を測定し、学習者の心理状態を推定する手法を提案している。得られた振動データから学習者が筆記動作の最

中か否かを 82.2%の精度で分類している。そのなかで、課題に費やされた時間の合計に対する思考時間の長さの比を学習者の理解度と定義し、決定木を用いて推定を行ったところ、正解データと順位相関がみられた。この研究では、計算課題に対する主観的難易度を推定できる可能性を示した。一方、MVPen という手書き文字をデジタル化する特殊なペンと、振動を感知するオリジナルの机を使用しなければならないという制約がある。

Maruichi ら [14] の研究では、英単語学習時にスマートフォン上でアルファベット入力を行う際のフリック入力情報を使用して、学習者の確信度を推定する手法を提案している。個人に特化したモデルで 86.8%、汎用的なモデルで 89.1%の推定精度が得られている。この研究では、タッチ間隔などのスマートフォン上で得られるデータから学習者の確信度を推定できる可能性を示した。

Nakamura ら [15] の研究では、学習者の主観的難易度が高いか低いかにについて、顔動作のみを用いて推定を試みている。従来の対面授業と同様に客観的視点から学習者の主観的難易度を推定しようという考えのもとで行われており、個人に特化したモデルで、平均 86.2%、最大 93.6%の推定精度を実現している。この研究では、学習者の顔動作から主観的難易度を推定できる可能性を示した。一方、カメラで撮影した顔画像を用いているため、プライバシー侵害のリスクや、カメラの設置が必要であるという制約がある。

Garain ら [16] の研究では、視線情報と単語の出題頻度から、長文中で学習者個人が難しいと感じた単語の推定を試みている。長文中に出題された単語から、個人が難しいと感じた単語を自動記録することに焦点を当てている。文書を読むだけで、主観的高難易度単語を推定できれば、ユーザはどの単語を学習しなければならないかを簡単に把握することが可能になり、学習補助に有用であると述べている。

Yamada ら [17] の研究では、視点情報を用いて英語多肢選択問題解答時の確信度の推定を行っている。アイトラッカを用いて視線の移動や凝視といった特徴を計測し、個人に特化したモデルで平均 90.1% の推定精度が得られている。この研究では、英文の穴埋め問題に対して行っているため、本研究が対象とする英単語 4 択課題とは形式が異なっている。

繁田ら [18] の研究では、英語のリスニング教材を対象に、眼球運動を用いて学習者の主観的難易度の推定を行っている。この研究は本研究と同様に e-learning 中の学習者の主観的難易度を推定することで、その状況に適した変化を促すことが可能であると述べたうえで、主観的難易度の推定手法を提案している。難しいかやさしいかの 2 値分類で他者で訓練したモデルを用いて、最大 85.8% の精度で推定している。

Adler ら [19] の研究では、マルチタスク中の作業効率を調べるために、課題に対する難易度について主観と客観の 2 つの場合で比較評価を行っている。その結果、客観的な難易度よりも主観的な難易度の方が課題の難しさを研究するためのより良い尺度になる可能性を示している。このことから、問題に対する客観的一律の難易度だけではなく、各個人が課題に感じる主観的難易度を推定できることが有益であることが確認できる。

以上から、学習者の筆記動作や視線、表情などの情報を取得することによって、学習者の心理的な情報を認識することが可能になってきていることが分かる。

2.3 人間の精神状態と生体情報

生体情報と感情や興味などの人間の主観評価の関係について、中山ら [20] の研究で評価を行っている。被験者の興味・関心と瞳孔面積や瞬き率などの間に相関がみられることを示している。また、同論文で知覚情報によって誘発される生体電位の変化について議論を行っている。彼らは立体視を用いて調査を行っているが、知覚情報によって体電位に何らかの変化を示すことが述べられている。

よって、体電位は知覚や記憶といった認識が難しい曖昧な物に対してなんらかの相関があることが実験的に分かっている。

2.4 本研究の位置づけ

本研究の位置づけは、①学習時の印象を、ウェアラブルデバイスを用いて推定する手法を開発すること、②英単語 4 択課題という短時間での解答を求める問題に対する印象推定の実現可否を明らかにすることである。

①に関して、2.2 節で述べたとおり、学習者の主観や印象を学習支援に応用することで学習効率向上を支援する研究が行われている。また、2.3 節で述べたとおり、生理学研究において生体情報と人間の感情や興味などの主観的な印

象についての関連性が示唆されている。以上より、生体情報計測に基づき学習者の主観や印象を推定する手法の開発を行う。生体情報計測は一般に大型の医療機器などを用いて行われるが、本研究では実運用とのトレードオフを考慮し、ウェアラブルデバイスを用いた簡易生体計測データから学習者の主観的印象を推定する手法の開発を行う点が独自性である。ウェアラブルデバイスのみを用いて推定を実現する利点はいくつかある。1 つは、現代のスマートフォンと同様に、将来ウェアラブルデバイスが普及し、誰でもつねにデバイスを着用することが当然な社会になると考えると、提案手法により手軽に効率の良い学習支援を提供できるようになる点である。そのほか、環境設置型のセンサーやカメラとは異なり、ウェアラブルデバイスは利用場所を選ばないという利点もある。また、関連研究 [15], [16] のように顔画像を撮影しないため、プライバシーを気にする学習者も利用できる点で、カメラを用いたシステムとは立ち位置が異なる。

②に関して、大学や大手予備校が導入している英単語学習支援システムが 4 択を採用していることから、本研究でも 4 択課題を採用した。関連研究 [16], [17], [18] では英語長文や英語短文穴埋め問題、リスニング問題を対象としており、どちらも単語のみではなく文章で構成された課題を対象としている。課題が文章の場合、文章の読解を行うために文章の前後を視線が行き来したり、選択肢と文章を視線が行き来したりする。そのため、英単語 4 択課題と比べて、凝視や移動といった視線から得られる情報が多いと考えられる。英単語 4 択課題は、学習者に提示される課題が英単語一語のみであるため、文章で構成された他の課題と異なり何度も問題を見る必要はない。また選択肢も 4 つだけであるため、解答に要する時間も比較的少なくなる。したがって、少ない視線の情報から確信度や主観的難易度の推定可能性を議論するという点も本研究の独自性である。

以上をふまえ、本稿の主要な貢献は以下のとおりである。

1. ウェアラブルデバイスを用いた英単語 4 択問題における確信度および主観的難易度の推定手法を新たに開発する点：筆者らの調査した範囲ではウェアラブルデバイスを用いてこれを実現する手法は見つからなかった。
2. 計測実験を行い、提案手法による確信度および主観的難易度の推定精度の検証を行った点：考察の結果、特に確信度推定に対する実現可能性を示した。
3. 重要因子の分析を行った点：貪欲法を用いた特徴選択に加え、各特徴量と確信度および主観的難易度に関する分析を行い、推定に有用な特徴量を明らかにした。
4. 学習効果に関する検討：提案手法を用いることで学習効果が具体的にどの程度改善されるのか考察を行った。

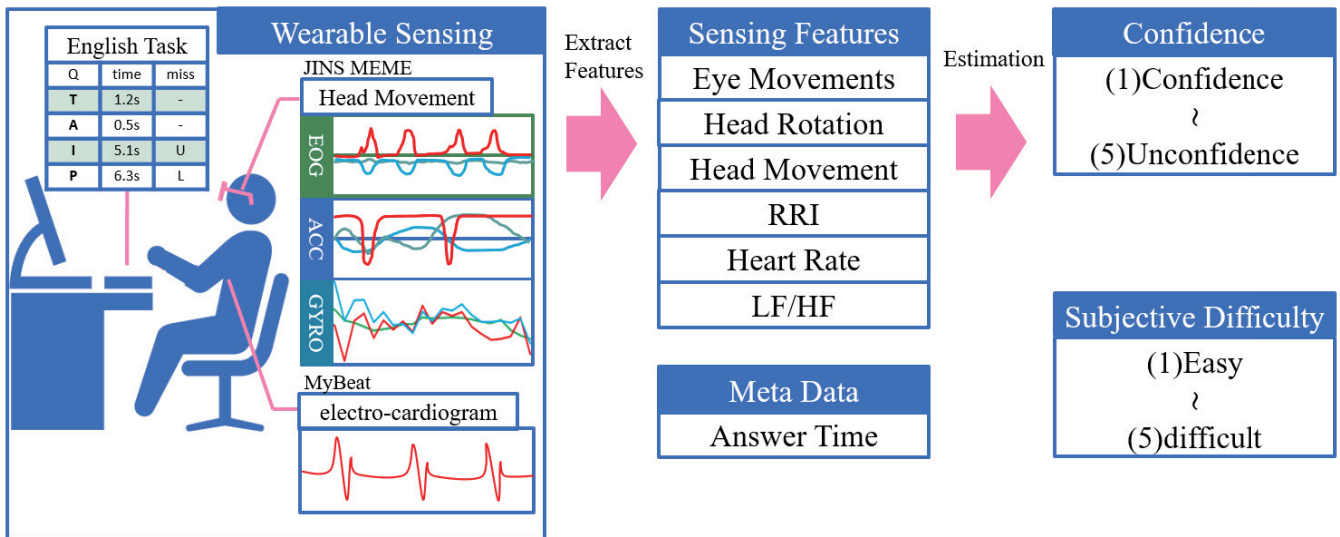


図 1 学習時の確信度と主観的難易度推定における提案手法の概要

Fig. 1 Outline of our proposed method to detect the degree of confidence and subjective difficulty of learning contents.

3. 提案手法

3.1 提案手法概要

本研究では、3点式眼電位センサ搭載眼鏡型デバイスの JINS MEME^{*1}と胸部装着型心電図モニタリングデバイス MyBeat^{*2}の2つのウェアラブルデバイスを用いて e-learning 学習時の AoH (Acceleration of Head) および EOG (Electrooculogram), HR (Heart rate) などの生体情報を簡易計測し、分析することで、学習者の英単語4択課題に対する確信度と主観的難易度を推定する手法を開発する。提案手法の概要を図1に示す。e-learning 上で英単語4択課題学習中の学習者から3点式眼電位センサ搭載眼鏡型デバイス JINS MEME を用いて学習時の AoH, EOG を記録し、胸部装着型心電図モニタリングデバイス MyBeat を用いて HR を記録する。JINS MEME は、眼球の動きによって生じる電位の変化である EOG を簡易計測することが可能なデバイスである。関連研究 [15], [20] で示されているように、眼の動きは心的変化と相関があるため、確信度および主観的難易度の推定に有効であると考えられる。MyBeat は、心臓が血液を送り出す際の ECG (Electrocardiogram) を簡易計測することが可能なデバイスである。HR を計測するウェアラブルデバイスは腕時計型が一般的であるが、今回は HR に限らず ECG の情報の有効性を評価するために、試験的に ECG を簡易計測できる MyBeat を用いて分析を行う。得られたデータから推定に有用だと考えられる特徴量の抽出を行い、機械学習を用いて、学習者の確信度と主観的難易度を推定する。

本研究では、確信度と主観的難易度を5段階で定義する。

主観的難易度は簡単から難しいを段階的に(1)から(5)で定義した。確信度は以下のように定義した。

- (1) 完璧に知っていて解答した。
- (2) 4 択の選択肢から推測したうえで自信がある。
- (3) 4 択の選択肢から推測したうえでたぶんあっている程度に自信がある。
- (4) 4 択の選択肢から推測したがあまり自信がない。
- (5) 完全に分からない。

3.2 特徴量抽出

今回使用する3点式眼電位センサ搭載眼鏡型デバイス JINS MEME は、3点式眼電位センサと6つの加速度センサを搭載しており、AoH および EOG を簡易計測することが可能である。実際取得を行うデータ区分としては、眼電位 ($EOG_R, EOG_L, EOG_H, EOG_V$), 加速度 (ACC_X, ACC_Y, ACC_Z), 角加速度 ($GYRO_X, GYRO_Y, GYRO_Z$) となっている。取得の際、サンプリング周波数は 100 Hz, 加速度センサの計測範囲は $\pm 2g$, 角加速度の計測範囲は $\pm 250 dps$ とした。予備実験として筆者が計測を行った結果、感度 16,384 [LSB/g] で加速度センサの値は $[-16618, 7867]$, 眼電位センサの値は $[-1855, 1877] \mu V$ の範囲で計測された。

MyBeat は、脈波センサを搭載しており ECG を簡易計測することが可能である。実際取得を行うデータ区分としては、RRI (R-R interval), LF/HF (High Frequency/Low Frequency), HR を計測することが可能である。取得の際、ECG のサンプリング周波数は、1,000 Hz とした。予備実験として筆者が計測を行った結果、RRI の値は $[554, 815] ms$, LF/HF の値は $[-2.083, 0.993]$, HR の値は $[74, 108] bpm$ の範囲で計測された。

計測した時系列データは、図2のように設問ごとにフ

^{*1} JINS MEME, <https://jins-meme.com/ja/>

^{*2} MyBeat, <https://www.uniontool-mybeat.com/>

表 2 特徴ベクトルを計算するための関数リスト
Table 2 Function list for calculating feature vector.

特徴量抽出関数	説明	数式	種類 (T: Time, F: Freq.)
nor(s)	Min-max normalization	$s_i = \frac{s_i - \min(s)}{\max(s) - \min(s)}$	T
stn(s)	Standardization	$s_i = \frac{s_i - \bar{s}}{\sigma}$	T
mean(s)	Arithmetic mean	$\bar{s} = \frac{1}{N} \sum_{i=1}^N s_i$	T, nor(T), F
std(s)	Standard deviation	$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (s_i - \bar{s})^2}$	T, nor(T), F
max(s)	Maximum values in array	$\max_i(s_i)$	T, stn(T), F
min(s)	Minimum values in array	$\min_i(s_i)$	T, stn(T), F
rms(s)	Root mean square	$\sqrt{\frac{1}{N} \sum_{i=1}^N (s_i)^2}$	T, nor(T), F
cor(s, v)	Correlation coefficient	$r = \frac{\frac{1}{N} \sum_{i=1}^N (s_i - \bar{s})(v_i - \bar{v})}{\sqrt{\frac{1}{N} \sum_{i=1}^N (s_i - \bar{s})^2 \frac{1}{N} \sum_{i=1}^N (v_i - \bar{v})^2}}$	T, nor(T), stn(T), F
qrang(Q)	quantile range	$Q_{\frac{3}{4}} - Q_{\frac{1}{4}}$	T, nor(T), stn(T), F
maxfreqInd(s)	Frequency component	$\operatorname{argmax}_i(s_i)$	F

N: signal vector length, Q: Quartile, T: Time domain features, F: Frequency domain features.

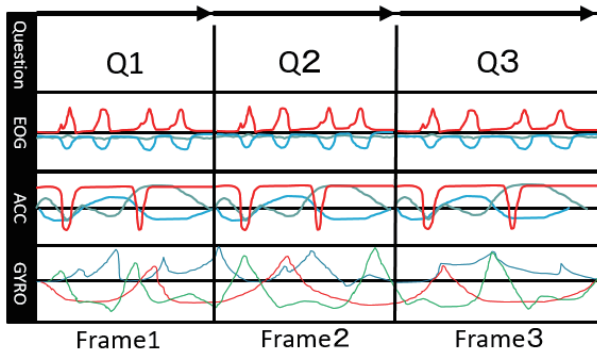


図 2 時系列データのフレーム化
Fig. 2 Sensor data division by each question.

フレーム化を行い、各フレームに対して成分ごとに各次元に対する特徴量の抽出を行った。解答時間によって各フレームの長さが異なるため、設問ごとに計測指標のサンプル数は異なっているが、特徴抽出時に、平均値や標準偏差といった、データ数によって影響を受けない統計量に変換することで、この影響を極力排除している。一般的な眼球運動計測では、測定前にキャリブレーションとして基準電位の計測を行うが、本研究ではキャリブレーションは実施していない。したがって、後述する LOSO-CV の評価は、キャリブレーション未実施時の精度評価であり、後述する LOO-CV の評価は、キャリブレーション実施を想定した際の精度評価を行っている。JINS MEME で取得したデータに関しては、個人差による影響を小さくするため、正規化や標準化を施したデータを作成し、生データ、正規化を施したデータ、標準化を施したデータの 3 つに対して同様の特徴抽出を行った。基本的な特徴量（平均、標準偏差、最大、最小、四分位数）に加えて、各軸間の相関係数や、FFT (Fast Fourier Transform) を施したときの周波数成分（低周波数成分、中周波数成分、高周波数成分）のピーク、それに対応する周波数を特徴量として抽出した。MyBeat で

Algorithm 1 Greedy Feature Selection

```

Require: F
F_adapted ← none
for j = 1 ... 10 do
    F_m = argmax_{1 ≤ i ≤ N-1} {Accuracy(F_adapted ∪ F_i)}
    F_adapted.append(F_m)
    Remove F_m from F
end for
return F_adapted
    
```

取得したデータに関しては、基本的な特徴量（平均、分散、標準偏差、最大、最小）のみを特徴量として抽出した。2 つのセンサから抽出した特徴量は合計で 694 次元となった。今回抽出に使用した関数を表 2 に示す。

3.3 機械学習アルゴリズム

機械学習アルゴリズムには、様々な手法が提案されているが、今回、Python の scikit-learn パッケージを用いて探索的に機械学習アルゴリズムの選定と特徴選択、パラメータチューニングを行った。その結果、特徴選択+SVM (Support Vector Machine [21]) のデフォルトパラメータが最も高い推定精度を示した。以降は上記のアルゴリズムを用いた結果で議論を行う。特徴選択には、各特徴量を追加して、推定精度の評価を行い、そのなかで最も精度が良いものを追加していく貪欲法を用いて有用な特徴量を精査した。今回は特徴選択中の推定精度の収束具合から、特徴選択で選択される特徴量の最大数を 10 個とした。

本手法について具体的に説明する。すべての特徴量 F のうち、推定精度が最高となる 1 つの特徴量

$$F_m = \operatorname{argmax}_{1 \leq i \leq N} \text{Accuracy}(F_i)$$

を求め、F から F_m を採用済み特徴群 F_adapted に移動する。F_m を除いた F に対して、

$$F_m = \operatorname{argmax}_{1 \leq i \leq N-1} \{Accuracy(F_{adapted} \cup F_i)\}$$

を求め、同様の手続きを繰り返し、 $|F_{adapted}| = 10$ で終了する。これまでの探索結果より最高の推定精度を達成した際の $F_{adapted}$ を選択済み特徴量として採用する (Algorithm 1 を参照)。

4. 実験概要

4.1 実験概要

提案手法の有効性を検証するため、e-learning を想定した環境でデータ計測実験を行った。実験は研究に同意を得た 22~29 歳の大学生、社会人の健常男性 8 名を対象に行った。被験者の TOEIC スコアの平均は 487.1 点 (標準偏差 160) であった。今回使用したのは、自作の英単語 4 択課題 210 問である。出題課題は、一般的な英語学習に合わせるため TOEIC 問題を読み解くための必須単語から選出した。中條 [22] が TOEIC 必須単語をレベル別に振り分けていたため、そのなかの「TOEIC 語彙 1」より 70 問、「TOEIC 語彙 2」より 70 問、さらに、上記 140 問と重複がないように「Learning through Official TOEIC Test items」[23] から 70 問を選出し、合計 210 問を今回の出題課題とした。出題する単語の正解を除いた選択肢について、他の単語の解答をランダムに表示した場合、正しい答えが分からない際の解答の容易さに影響を及ぼす可能性がある。そこで今回はまったく無関係の単語ではなく、接頭語や接尾語が等しいものや発音が類似しているものを選出することで選択肢から解答を予想しにくくしている。

実験で使用したアプリケーションの画面例を図 3 に示す。上部に問題となる英単語が出題され、その下にある 4 つの日本語選択肢から正解となる項目を選択する。図 3 から分かるように問題に関する設問文は表示されないため、視線が設問文に向かう可能性を除外している。解答後に図 4 と図 5 を表示し、今回推定対象となる正解データを取得する。学習者は全問解答終了時にすべての確信度を覚えていることは容易ではないが、各問題解答直後であれば、

Question	transform
Choice	変形させる ← transpose
	転動する ← transfer
	変身する ← transform
	送り届ける ← transmit

図 3 英単語課題解答画面

Fig. 3 Example of four-choice vocabulary tasks.

これを正確に自己判断することは可能であると考え、被験者全員の条件を統一するため、事前に問題内容の通知は行わなかった。2 種類のウェアラブルデバイスを装着した状態で、PC 上に表示される課題を、マウスクリック操作のみで解答してもらいデータの収集を行った。課題を行う前に、JINS MEME および MyBeat の計測値が正常となるまで待機したうえで実験を行った。取得したデータは、眼電位と頭部動作および RRI, LF/HF, HR である。その後、得られたデータを元に特徴量の抽出を行い、機械学習を用いて、確信度と主観的難易度の推定を行う。

4.2 計測データ

本実験では被験者 8 名 × 210 問の 1,680 データを計測した。このなかで問題解答中に連続で 2 回クリックしてしまい、問題をまったく見ることなく答えてしまった 2 問を正常に計測できなかったデータとして除外した。最終的に 1,678 データを採用した。本実験で得られた確信度と主観的難易度の正解ラベルの分布を表 3 に示す。表中の A-H は、今回データ計測に協力してもらった被験者 8 名の識別子である。この表から確信度においても、主観的難易度においても被験者によって、分布にばらつきがあることが見てとれる。

提案手法による推定を実施する前に、本研究が推定対象とする確信度と主観的難易度が本質的には同じものではないかという疑問が生じる。そこでこれらの混同行列を表 4 に示す。横軸が主観的難易度、縦軸が確信度を示している。表を見て分かるとおり、確信度と主観的難易度の一致度合いは約 68% であり、必ずしも一致するわけではないことを確認した。おおむね確信度 ±1 が主観的難易度となっているが、どちらかといえば主観的難易度を高めにする傾向が

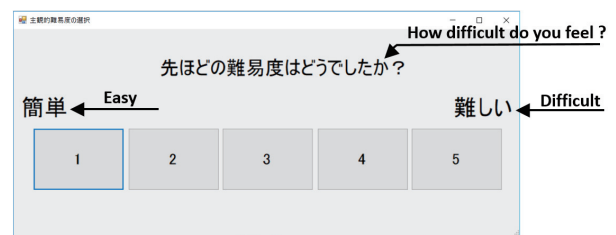


図 4 主観的難易度解答画面

Fig. 4 Screen for assessing degree of subjective difficulty.

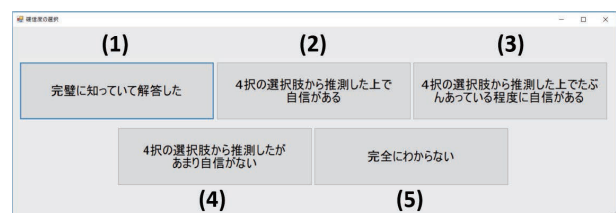


図 5 確信度解答画面

Fig. 5 Screen for assessing degree of confidence.

表 3 得られた主観的難易度と確信度の分布

Table 3 Distribution of data for subjective difficulty and answer confidence.

Subjective difficulty	A	B	C	D	E	F	G	H	Sum
(1)	40	0	150	0	99	124	114	41	568
(2)	78	26	15	18	47	32	44	36	296
(3)	29	36	3	137	18	34	30	48	335
(4)	24	100	24	40	3	12	17	28	248
(5)	39	48	18	15	43	8	4	56	231
Confidence	A	B	C	D	E	F	G	H	Sum
(1)	54	35	144	0	84	121	110	36	584
(2)	58	45	28	106	26	31	36	45	375
(3)	40	35	15	45	37	37	39	31	279
(4)	35	53	16	53	28	18	21	34	258
(5)	23	42	7	6	35	3	3	63	182

表 4 確信度と主観的難易度の混同行列

Table 4 Confusion matrix of confidence and subjective difficulty.

		Degree of confidence					Sum
		(1)	(2)	(3)	(4)	(5)	
Degree of subjective difficulty	(1)	520	36	10	2	0	568
	(2)	48	173	60	13	2	296
	(3)	13	134	141	42	5	335
	(4)	3	30	65	139	11	248
	(5)	0	2	3	62	164	231
Sum		584	375	279	258	182	1678

読み取れる。主観的難易度は、学習者の課題に対する取り掛かりやすさであり、覚えやすいかどうかを示している。そして確信度は、解答した問題が正解であるかどうかの自信度合いを示す。そのため、取り掛かりやすい単語、つまり主観的難易度が高いと感じている単語は、学習者にとって習得しやすい単語であるため、確信度も自ずと高くなる。またその逆もいえるため、今回のような結果が得られた可能性が高い。以上から、確信度と主観的難易度は別物であり、それぞれを推定する価値はあると考える。

今回は確信度と主観的難易度のそれぞれに対して、1, 2, を確信度=高、難易度=易として、3, 4, 5を確信度=低、難易度=難として2値分類問題で推定精度を評価する。

4.3 評価方法

評価は LOSO-CV と被験者別 LOO-CV という2つの手法で実施する。LOSO-CV (Leave-One-Subject-Out Cross-Validation) は、1人分のデータを検証用に、残りを学習用にして精度評価を行い、全員分交差検証する評価手法である。被験者別 LOO-CV (Leave-One-Out Cross-Validation) は、被験者1名のデータの中で、1インスタンスを検証用に、残りを学習用にして精度評価を行い、全インスタンス交差検証する評価手法である。

5. 実験結果と考察

5.1 確信度推定

8人の計測データから特徴量抽出を行い、LOSO-CVと被験者別 LOO-CVで確信度推定の精度評価を行った。得られた結果を表5に示す。表5の各行について説明する。「BASELINE」とは、正解データのなかで最も解答数の多かったものに予測が集中した場合の推定精度を表している。「CONVENTIONAL」とは、従来の4択アプリのように正否情報に基づき推定した場合の精度である。「TIME」とは、従来のe-learningでも取得可能な解答時間のみを特徴量として用いた場合の推定精度である。「EOG+AoH」とは、JINS MEMEより得られたデータ(眼電位、加速度、角加速度)とそのデータに高速フーリエ変換を適用して得られたデータから抽出した特徴量を用いた場合の推定精度を示している。「ECG」とは、MyBeatを用いることで得られたデータから抽出した特徴量を用いた場合の推定精度を示している。残りは、これらの特徴量の組合せごとに推定を行った結果であり、「ALL」は「TIME」、「EOG+AoH」、「ECG」のすべての特徴量を用いた場合を表している。前述したとおり、どの項目も各特徴量群から貪欲法にて特徴選択を行った場合の推定精度である。

各センサ間での精度を見ていくと、今回用いた2種類のウェアラブルデバイスによって得られたデータを用いることによって、最大93.5%の推定精度が得られている。これは「BASELINE」より30.5%、解答時間のみを用いた場合より15.4%推定精度が向上している。したがってウェアラブルデバイスで簡易計測したEOG, AoH, ECGを用いることで確信度推定の精度が向上することを明らかにした。「EOG+AoH」と「ECG」について精度を比較すると、単体で用いた場合でも他の特徴量と併用した場合でもEOGとAoHより得られた特徴量を用いた方が、高い精度が得られた。LOSO-CVと被験者別 LOO-CVの結果に着目する。両評価手法で最も高い推定精度が得られたALLに着目す

表 5 確信度と主観的難易度の推定精度

Table 5 Estimation results of confidence and subjective difficulty (%).

Features	Degree of confidence		Degree of subjective difficulty	
	LOSO-CV	LOO-CV	LOSO-CV	LOO-CV
BASELINE	63.0	63.0	74.3	74.3
CONVENTIONAL	69.6	69.6	66.0	66.0
TIME	77.3	78.1	69.2	82.4
ECG	70.7	73.2	68.0	80.2
EOG+AoH	75.9	93.0	72.8	88.3
TIME+EOG+AoH	79.3	92.6	74.7	89.8
TIME+ECG	78.2	80.8	74.0	84.3
EOG+AoH+ECG	75.9	92.9	72.9	88.1
ALL	80.8	93.5	74.6	90.8

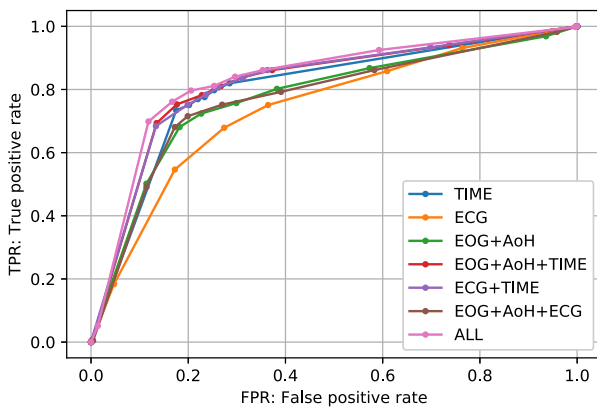


図 6 LOSO-CV 使用した確信度推定モデルの ROC 曲線

Fig. 6 Receiver operating characteristic curve of confidence estimation model with LOSO-CV.

ると、他者のデータで学習したモデル (LOSO-CV) では 80.8%だったが、個人に特化させたモデル (LOO-CV) では 93.5%と 12.7%の精度向上が確認できる。したがって、提案手法は個人に特化させることによって大幅に推定精度を向上できる可能性がある。

LOSO-CV にて作成した推定モデルに関して、ROC 曲線 (Receiver operating characteristic curve) の描画を行った。threshold は、0 から 1 まで 0.1 刻みとした。その結果を図 6 に示す。この図から分かる通り、検証したすべての推定モデルの中で「ALL」を用いて作成したモデルが最も良いモデルであることが確認できる。作成した「ALL」のモデルに対して最適なカットポイントを求めたところ (TPR=0.761, FPR=0.167) となった。また他のモデルについても、解答時間だけを用いた推定モデルより、センサから得られた特徴量を含めた推定モデルの方が良い性能を示している。

5.2 特徴量に関する分析

提案手法では、大社ら [16] や山田ら [17] のように、観測したセンサデータから人間が意味を理解できる情報 (瞬き回数など) への変換を行わず、センサデータから有用と思

表 6 LOSO-CV で選択された特徴量

Table 6 Selected features in LOSO-CV.

Index	Contents
(1)	Response Time
(2)	Minimum value of $GYRO_Y$ applied standardization
(3)	Max value of high frequency component in EOG_V
(4)	Correlation coefficient of the absolute values of between EOG_L and EOG_H
(5)	Correlation coefficient of frequency between ACC_X and ACC_Z
(6)	Max value of high frequency component in EOG_H
(7)	Correlation coefficient of frequency between EOG_H and EOG_V
(8)	Correlation coefficient of the absolute values between ACC_X and ACC_Z
(9)	Max value of high frequency component in EOG_R
(10)	Median of EOG_R applied standardization

われる特徴量を多数抽出し、特徴選択を行うという手法を採用した。ここでは、どの特徴が有用に働いたかという点について議論を行うとともに、各特徴量に対して有意差がみられるかについての分析を行う。

まず汎用的なモデル (LOSO-CV) で有用に働いた特徴量について考える。今回最も高い精度が得られた際に貪欲法にて選択された特徴量を表 6 に示す。この表をみても有用な特徴量として解答時間が最初に選択されている。別途分析した結果、表 5 の「TIME+EOG+AoH」、「TIME+ECG」でも、一番初めに解答時間の特徴量が選択されることを確認した。Kelley ら [24] は、確信度は想起に要した時間の長さから間接的に推測されると論じている。容易に想起された解答に対して、確信度は高くなり、想起されるまでに時間を要した解答に対しては確信度は低くなるため、解答時間が確信度推定に対して有効にはたらいと考えられる。ここから解答時間の長さは確信度や主観的難易度に影響を与えると考えられる。実際に解答時間を特徴量に含めるほうが含めない場合と比較して精度の向上が確認できる。他の特徴量についてみると、EOG と AoH の周波数成分から抽

出した特徴量が半分を占めるという結果が得られた。ここから、周波数成分には、各被験者に対して共通する特徴が含まれている可能性があることが示唆される。また EOG から抽出した特徴量が主に選択されているため、EOG の変化を取得することによって、確信度に対する何らかの変化を得られていることが示唆される。本研究では、EOG を視線方向などの特徴に落とし込まなかった。しかし、取得した EOG 各成分の組合せによって視線の動きを疑似的に取得することは可能である。周波数成分に変換を行うことでノイズによる影響を下げつつ視線的な情報を取得できたのではないかと考えられる。学習者が解答選択を行う際の思考動作、具体的には首を傾げる動作や目線の揺らぎ動作を EOG の周波数成分の相関係数や AoH によって疑似的に取得することで、今回の推定に有効に働いたと考えられる。つまり表 6 の (2) の角加速度 Y の最小値は、思考動作中の悩みのサインとなる首を傾げる動作による情報を取得できており、(4) や (7) では、目線の上下などの情報を取得できていると考える。

一方で、被験者別に LOO-CV で評価を行った際には、被験者ごとに有効にはたらい特徴量の組合せが異なる結果となった。したがって、事前にデータを取得することで各個人に特化させることができれば、より高い精度での推定が可能である。また、有用にはたらく特徴から学習者をグループに分類することができれば、学習方法の傾向分析などの研究に役立つ可能性がある。

次に今回抽出した特徴量の有意差について議論していく。全特徴量に対して確信度が高い場合と低い場合で Welch の t 検定を行い有意差の有無を確認した。全特徴量 694 個の中、有意水準 0.01 以下で有意差が認められた特徴量は 579 個みられた。ウェアラブルデバイスによって得られたデータから抽出した特徴量は、確信度を分類するのに有用な要素を含んでいることが確認できる。

5.3 主観的難易度推定

確信度と同様に、得られたデータから特徴量の抽出を行い、8 人のデータで LOSO-CV、および被験者別 LOO-CV により主観的難易度推定の精度評価を行った。得られた結果を表 5 に示す。各行の意味は 5.1 節で説明したとおりである。

LOSO-CV に着目すると、最大で 74.6% の精度が得られているが、「BASELINE」が 74.3% であることをふまえると、他者のデータを用いて学習したモデルでは、主観的難易度の推定はできないという結果となった。一方、被験者別 LOO-CV の結果では推定精度が向上しているため、個人差の影響を強く受けているものと考えられる。

被験者別 LOO-CV についてみると、最大 90.8% の精度が得られている。「BASELINE」よりも 16.5% 上昇、解答時間のみを用いた場合よりも 8.4% の精度の向上が確認で

きる。したがって、ウェアラブルデバイスで簡易計測した EOG, AoH, ECG を用いることで得られたデータが主観的難易度の推定に有用であることが確認できる。「ECG」に関しては、解答時間と併用して用いた場合、「BASELINE」と比較して 10% 上昇、解答時間のみを用いた場合よりも 1.9% の精度の向上が確認できる。ここから、ECG から得られるデータが主観的難易度の推定に有用であることが確認できる。

主観的難易度では、他人のデータを用いた学習モデルで推定を行うことはかなわなかった。しかし、個人のデータを用いて学習を行うことで高い推定精度を得られているため、センシングの有用性は確認できている。したがって、個人適応などの技術を応用することで、他者のデータで構築した学習モデルでも高精度に主観的難易度が予測できるよう改善案を模索する必要がある。

5.4 確信度と主観的難易度の違い

関連研究では、長文問題やリスニング問題など「解答に時間を要する課題」に対して、確信度や主観的難易度を「それぞれ単体」で推定する研究が行われていた。一方、本研究では英単語 4 択問題という「解答時間が比較的小さい問題」を対象に、「確信度と主観的難易度両方」の推定に取り組んでいる。評価実験の結果、従来のアプリでも取得できる時間情報に加えて、ウェアラブルデバイスによるセンシングを応用することによって、解答時間が比較的小さいような問題に対しても、確信度や主観的難易度を推定できる可能性を示した。

確信度と主観的難易度両方の推定実験を行ったことで、いくつか興味深い結果を確認した。1 つは、表 4 で示したように、確信度と主観的難易度は似ているものの必ずしも一致するわけではないということである。もう 1 つは、他者のデータを用いてモデルを学習した場合、確信度はある程度推定できるが、主観的難易度はまったくできなかったという点である。これは確信度と主観的難易度はある程度の関連を持っているが、似て非なるものであるという裏付けにもなると考えている。したがって、実際の学習支援システムに応用する際に、確信度は復習問題の見逃し防止のために、主観的難易度はモチベーションの低下を軽減するための出題難易度の自動調整に、別々に応用することができる可能性がある。

6. 実運用とその効果

6.1 自己データの収集と実運用

提案手法では、確信度と主観的難易度両方において、LOO-CV による評価で高い推定精度を確認している。一般的な機械学習の応用アプリケーションでは、自身の正解ラベル付きデータを取得することが困難であることから、被験者別 LOO-CV での評価結果ではなく、LOSO-CV によ

表 7 従来手法での混同行列

Table 7 Confusion matrix of conventional method.

Answer \ Actual	Know	Unknown	Precision
Correct	748	299	41.4%
Failed	211	420	66.6%
Recall	78.0%	58.4%	69.6%

表 8 提案手法の混同行列

Table 8 Confusion matrix of our proposed methods.

	Predicted \ Actual	Know	Unknown	Precision
LOSO-CV	Know	752	116	86.6%
	Unknown	207	603	74.4%
	Recall	78.4%	83.9%	80.8%
LOO-CV	Know	885	42	95.5%
	Unknown	74	677	90.1%
	Recall	92.3%	94.2%	93.5%

る評価結果で議論を行うべきである。一方、将来的に本提案手法を実運用することを考えると、設問ごとに学習者に対して正解ラベルを尋ねる機能を実装することで、実運用環境においても自身の正解ラベル付きデータを取得することは難しくない。すなわち、ユーザビリティとのトレードオフを考慮しながら、運用時に学習モデルを学習者自身に適応させていくことが可能である。したがって、LOO-CVの結果は実運用時の最大達成可能精度を意味しており、今回被験者別 LOO-CV で高い精度を確認できたことは、本研究の有用性を示すうえで重要な結果となる。実運用環境において自身の正解データを尋ねることは、学習者の負担になることが懸念されるが、能動的学習手法などを応用することで、少ない負担で精度を上げることも可能であると考えている。

6.2 提案手法による学習効率の変化

本提案手法の確信度推定により、英単語学習がどの程度効率化されるかについて考察する。被験者が記憶済みと解答した課題、つまり 3.1 節で述べた確信度の選択肢で (1)、(2) と選択された課題を復習不要課題と定義する。同様に確信度の選択肢で (3)~(5) と選択された課題を要復習課題と定義する。今回、学習効率について考えるための基準として、従来のアプリで採用されている課題の正否を用いて復習課題を決める場合の混同行列を表 7 に示す。また、提案手法で最も高い推定精度が得られた個人に特化したモデル (LOO-CV + ALL 特徴量) と他人のデータで学習したモデル (LOSO-CV + ALL 特徴量) で、各課題が要復習か否かを推定した結果の混同行列を表 8 に示す。これを基に、従来手法が、どの程度未定着な学習内容を見逃していたのかを明らかにするとともに、それに対する本提案手法の有用性を議論する。それぞれ重要となる数値を抜粋し、

表 9 従来手法と提案手法の比較

Table 9 Compared conventional and proposed methods.

	Conventional	LOSO-CV	LOO-CV
Overall efficiency	69.6%	80.8%	93.5%
Missing the reviews	41.6%	16.1%	5.8%
Excessive reviews	22.0%	21.6%	7.7%

意味付けを述べたものを表 9 に示す。表 9 の各行について説明する。「Overall efficiency」とは、未定着単語か否かの推定精度を示す。「Missing the reviews」とは、学習者が覚えられていない単語のうち正しく認識できていない単語の割合を表している。「Excessive reviews」とは、学習者に要する復習単語のうち、学習者がすでに覚えられている単語の割合を表している。以下の議論の中では、各項目をそれぞれ全体効率、復習見逃し率、過剰復習率と呼称する。計算式を以下に示す。式中の TP, TN, FP, FN は、混同行列の要素数を示している。

$$\text{Overall efficiency} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{Missing the reviews} = 1 - \frac{TN}{FP + TN}$$

$$\text{Excessive reviews} = 1 - \frac{TP}{TP + FN}$$

従来のアプリケーションから得られる学習効率について着目すると、全体効率が 69.6%である。学習者が覚えられていない 719 個の英単語のうち 299 個が正しく認識できておらず、復習見逃し率が 41.6%となっている。さらに、すでに習得済みである英単語 959 個のうち 211 個が不正解となっており、過剰復習率が 22.0%である。不正解問題は、本来、覚えているつもりだったが忘れてしまった英単語である可能性が高いため、未修得単語と判断し復習対象とすべきではあるが、ここでは効率について議論しているため、その議論は後述する。

提案手法における学習効率について着目すると、全体効率が、他者のデータで構築した汎用的なモデル (LOSO-CV) で 80.8%、個人に特化したモデル (被験者別 LOO-CV) を作成することで 93.5%である。学習者が覚えられていない 719 個の英単語のうち、LOSO-CV で 603 個、LOO-CV で 677 個を正しく認識することができており、復習見逃し率が LOO-CV で 16.1%、LOO-CV で 5.8%となっている。余剰復習率も LOO-CV で 21.6%、LOO-CV で 7.7%となっている。従来手法と提案手法の結果を比較すると、提案手法は全体効率を各モデルともに向上させることができており、復習見逃し率も 41.6%から LOO-CV で 16.1%および被験者別 LOO-CV で 5.8%と大幅に軽減することができており、より正確な復習必須単語の選別が可能になったといえる。過剰復習率に関しては、LOSO-CV では大幅な軽減を行うことはかなわなかったが、被験者別 LOO-CV を用いることで 22.0%から 7.7%と 14.3%軽減することができ

ている。以上から、提案手法を用いることでより効率の良い学習を提供することが可能になったといえる。

前述したとおり、記憶済みと考えているにもかかわらず不正解であった場合は、復習対象として扱った方が適切であると思われる。よって、解答不正解の場合は、すべて復習を行い、正解の場合は、提案手法で推定を行い未記憶と判定された場合に復習を行うことで、より効率的な学習支援システムを提供することが可能である。なお、パラメータを調整することによって再現率適合率のバランスを変え、無駄な復習を行いたくない人や未記憶単語を確実に復習したい人に向けたチューニングを施すことも可能である。

7. 終わりに

本研究では、英単語 4 択課題における確信度および主観的難易度の推定手法の開発に取り組んだ。e-learning 学習中の被験者から 2 つのウェアラブルデバイス JINS MEME と MyBeat を用いてデータを取得し、そこから抽出した特徴量を使用して機械学習を行うことにより、確信度と主観的難易度の推定を行った。提案手法を評価した結果、確信度は LOSO-CV で最大 80.8%、LOO-CV で平均最大 93.5% の推定精度を得た。ウェアラブルデバイスで簡易計測した EOG, AoH, ECG を用いることで精度が向上することを示した。主に解答時間が確信度を判定するのに役立っていることは今回の検証から明らかであるが、解答時間を用いなかった場合においても、同程度の精度が得られている。使用デバイスごとの推定精度を見る限り、EOG や AoH の方が ECG よりも個人に対して有用な特徴が取得できている可能性があることが分かった。

主観的難易度は、他者のデータで学習モデルを構築する場合、現段階では実用的な推定精度は実現できなかった。自身のデータを学習できることで、平均で最大 90.8% の精度が得られたため、センシングで得られたデータに主観的難易度を判断する何らかの特徴が含まれていることを確認した。確信度の推定では、他人のデータを用いることである程度の推定精度を得ることができたのに対して、主観的難易度では、うまく推定することはできなかった。ここから、主観的難易度の推定を実用的にするには、個人適応などのさらなる手法の改善が必要であるといえる。

今回は、英単語 4 択課題という限定的な学習内容で推定を行った。今回得られた結果が汎用的に使用できるかを確認するため、今後、別内容の選択式課題において同様の結果が得られるか調査する必要がある。また本提案手法を実際の学習支援システムに導入した際、実際に効率の良い学習を提供することが可能であるかどうか調査していく必要がある。

謝辞 本研究の一部は科学研究費助成事業若手研究 (16K16175, 19K20420) の助成によるものである。ここに謝意を表す。

参考文献

- [1] Ebbinghaus, H.: *Memory: A contribution to experimental psychology*, Dover Publications, reprint edition (1987).
- [2] Mori, T. and Hasegawa, T.: Estimation of degree of retention and subjective difficulty of four-choice English vocabulary questions using a wearable device, pp.0605–0610 (online), DOI: 10.1109/TENCON.2018.8650315 (2018).
- [3] 張 一平: 確信度テスト法と項目反応理論—新たなモデルと実践的応用, University of Tokyo Press (2007).
- [4] 谷上亜紀: 問題の難易度に関する情報が確信度判断および FOK 判断に及ぼす効果, 心理学研究, Vol.73, No.3, pp.243–250 (オンライン), DOI: 10.4992/jjpsy.73.243 (2002).
- [5] de la Guía, E., Camacho, V.L., Orozco-Barbosa, L., Brea Luján, V.M., Penichet, V.M.R. and Lozano Pérez, M.: Introducing IoT and Wearable Technologies into Task-Based Language Learning for Young Children, *IEEE Trans. Learning Technologies*, Vol.9, No.4, pp.366–378 (2016).
- [6] Muñoz-Cristóbal, J.A., Jorrín-Abellán, I.M., Asensio-Pérez, J.I., Martínez-Monés, A., Prieto, L.P. and Dimitriadis, Y.: Supporting Teacher Orchestration in Ubiquitous Learning Environments: A Study in Primary Education, *IEEE Trans. Learning Technologies*, Vol.8, No.1, pp.83–97 (online), DOI: 10.1109/TLT.2014.2370634 (2015).
- [7] Wu, T., Yang, T., Hwang, G. and Chu, H.: Conducting Situated Learning in a Context-Aware Ubiquitous Learning Environment, *Proc. 5th IEEE International Conference on Wireless, Mobile, and Ubiquitous Technology in Education (WMUTE 2008)*, pp.82–86 (online), DOI: 10.1109/WMUTE.2008.9 (2008).
- [8] Chen, C.-M.: Personalized e-learning system with self-regulated learning assisted mechanisms for promoting learning performance, *Expert Systems with Applications*, Vol.36, No.5, pp.8816–8829 (online), DOI: <https://doi.org/10.1016/j.eswa.2008.11.026> (2009).
- [9] 孫 勝国, 甘泉瑞応, Huang, T., He, A., 程 子学: 学習者の学習順序や反応パターンに基づいた学習状態推論法を用いる Web-based 教育支援システム, 情報処理学会論文誌, Vol.46, No.2, pp.327–336 (2005).
- [10] Lee, S.H., Choi, J. and Park, J.: Interactive e-learning system using pattern recognition and augmented reality, *IEEE Trans. Consumer Electronics*, Vol.55, No.2, pp.883–890 (online), DOI: 10.1109/TCE.2009.5174470 (2009).
- [11] 浅井洋樹, 山名早人: オンライン手書き情報を用いた未定着記憶推定システム, 情報処理学会研究報告, コンピュータと教育研究会報告, Vol.2014, No.1, pp.1–6 (2014).
- [12] Mothukuri, U.K., Reddy, B.V., Reddy, P.N., Gutti, S., Mandula, K., Parupalli, R., Murty, C.H.A.S. and Magesh, E.: Improvisation of learning experience using learning analytics in eLearning, *Proc. 2017 5th National Conference on E-Learning E-Learning Technologies (ELELTECH)*, pp.1–6 (online), DOI: 10.1109/ELELTECH.2017.8074995 (2017).
- [13] Hamaguchi, N., Yamamoto, K., Iwai, D. and Sato, K.: Estimation of Subjective Difficulty and Psychological Stress by Ambient Sensing of Desk Panel Vibrations, *SICE Journal of Control, Measurement, and System Integration*, Vol.5, No.1, pp.2–7 (online), DOI: 10.9746/jcmsi.5.2 (2012).
- [14] Maruichi, T., Augereau, O., Iwata, M. and Kise, K.: Keystrokes Tell You How Confident You Are: An Ap-

- plication to Vocabulary Acquisition, *Proc. 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pp.154–157 (online), DOI: 10.1145/3267305.3267609 (2018).
- [15] Nakamura, K., Kakusho, K., Murakami, M. and Minoh, M.: Estimating Learners' Subjective Impressions of the Difficulty of Course Materials in e-Learning Environments, *Proc. Distance Learning and the Internet Conference*, pp.199–206 (2008).
- [16] Garain, U., Pandit, O., Augereau, O., Okoso, A. and Kise, K.: Identification of Reader Specific Difficult Words by Analyzing Eye Gaze and Document Content, *Proc. 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Vol.1, pp.1346–1351 (online), DOI: 10.1109/ICDAR.2017.221 (2017).
- [17] Yamada, K., Kise, K. and Augereau, O.: Estimation of Confidence Based on Eye Gaze: An Application to Multiple-Choice Questions, *Proc. 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proc. 2017 ACM International Symposium on Wearable Computers*, pp.217–220 (online), DOI: 10.1145/3123024.3123138 (2017).
- [18] 繁田亜友子, 小池武士, 濱本和彦, 野須 潔: 英語リスニング電子教材を対象とした眼球運動測定による学習者の主観難易度の推定, *電気学会論文誌 C (電子・情報・システム部門誌)*, Vol.131, No.4, pp.800–807 (オンライン), DOI: 10.1541/ieejieiss.131.800 (2011).
- [19] Adler, R.F., Rubin, D., Mohammad, A.R., Irfan, A., Senadeera, H. and Nguyen, T.: That's Easy! The Effects of Objective and Subjective Task Difficulty when Multitasking, *Procedia Manufacturing*, Vol.3, pp.5467–5474 (online), DOI: <https://doi.org/10.1016/j.promfg.2015.07.684> (2015).
- [20] 中山 実, 清水康敬: 生体情報による学習活動の評価, *日本教育工学雑誌*, Vol.24, No.1, pp.15–23 (オンライン), DOI: 10.15077/jmet.24.1.15 (2000).
- [21] Cortes, C. and Vapnik, V.: Support-Vector Networks, *Machine Learning*, Vol.20, No.3, pp.273–297 (online), DOI: 10.1023/A:1022627411411 (1995).
- [22] 中條清美: 英語初級者向け「TOEIC 語彙 1, 2」の選定とその効果, *日本大学生産工学部研究報告 B 文系/研究報告専門委員会 (編)*, Vol.36, pp.27–42 (2003).
- [23] Service, E.T.: *TOEIC Test learning vocabulary with official question* (in Japanese), The Institute for International Business Communication (2013).
- [24] Kelley, C.M. and Lindsay, D.S.: Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions, *Journal of Memory and Language*, Vol.32, No.1, pp.1–24 (1993).



長谷川 達人 (正会員)

2009年石川工業高等専門学校電子情報工学科卒業。2011年金沢大学工学部情報システム工学科卒業。同年株式会社富士通北陸システムズ入社。2014年東京医療保健大学助手。2015年金沢大学大学院自然科学研究科修了。博士(工学)。2017年より福井大学大学院工学研究科講師。専門はコンテキストウェアネスと教育工学。IEEE, 日本シミュレーション&ゲーミング学会各会員。



森 朝春

2018年福井大学工学部情報・メディア工学科卒業。同年福井大学大学院工学研究科情報・メディア工学専攻入学。2020年福井大学大学院工学研究科情報・メディア工学専攻修了。