

腕の動きと顔画像による英語問題の確信度推定 Confidence Estimation in Answering English Questions Using Arm Movements and Facial Images

田中 勢也¹⁾ バーゴ アンドリュー¹⁾ 黄瀬 浩一¹⁾
Seiya Tanaka Andrew W. Vargo kouichi Kise

1 はじめに

復習は自身の学習の理解度を測り定着させる際に、非常に重要な作業である。この時、学習者は全ての内容を見直すのではなく、理解が不十分な問題を重点的に復習することが必要となる。復習すべき問題を探し出す時、判断材料としてよく用いられるのが、解答の正誤である。問題を解き、間違っただけを復習することで、正確な知識を身に付ける学習方法が一般的である。

近年、復習すべき問題の判断材料として解答の正誤に加え、新たに解答に対する確信の有無を導入することが試みられている。図 1 に、問題を区別する指標として確信度を導入したときの対応を示す。確信度の導入により、確信がないにもかかわらず偶然正解した問題や、確信をもっていただけ不正解だった問題を発見できるため、効果的な復習が可能となる。

確信度を測る手法として、確信度を学習者自身が記録する手法が挙げられる [1]。しかしこれは学習者にとって余分な工程が加わるため、集中力を阻害する可能性がある。そのため学習者自身が確信度を記録する手間を省略するため、確信度を自動で推定する研究が行われている。

本研究では手軽かつ広範囲な確信度推定の実現を目的に、加速度センサから計測される腕の動作と、Web カメラより計測される顔の動作を用いた確信度推定手法の 2 つについて検討する [2,3]。実験の結果、ユーザ依存において腕の動きから得られる情報が筆記問題の確信度推定に有効であること、顔画像から得られる視線方向の情報が英単語多肢選択問題の確信度推定に有効であることを確認した。

なお、本研究は大阪府立大学大学院工学研究科倫理委員会の承諾を得ていることを付記しておく。

2 関連研究

確信度推定に関する既存手法はこれまでにいくつか提案されている。

Yamada らは、アイトラッカを用いて視線情報を取得し、確信度を推定する手法を提案した [4]。この手法により、Yamada らは英語の多肢選択問題解答した際の学習者の視線から確信の有無を推定し、90%の精度で識別に成功した。アイトラッカは高精度で利用者の視線情報を検知することが可能であるが、現在は学術研究や法人などで利用されるケースが多く日常生活で使われることは少ない。

森らは、眼鏡型デバイスと胸部装着型心電図モニタリングデバイスを用いて頭部動作や眼電位と心拍などのデータを取得し、定着度および主観的難易度を推定した [5]。この手法により、森らは英単語 4 択課題にお

1) 大阪公立大学大学院 情報学研究科

Graduate School of Informatics, Osaka Metropolitan University

| | 正解 | 不正解 |
|------|---------|--------|
| 確信あり | ○ 内容が定着 | ✗ 勘違い |
| 確信なし | ✗ 曖昧な知識 | ✗ 勉強不足 |

○ :要復習

✗ :要復習かつ要注意

図 1 確信度を考慮した正解と不正解

る定着度について最大 81%の推定精度を実現した。森らの手法ではウェアラブルデバイスを用いたことから手軽な利用が期待されるが、これらのデバイスの利用人口は現状少ないため新たに導入が必要となる。

3 提案手法

従来の確信度推定手法には、用いられるセンサが通常環境では利用されていないという問題点がある。本節ではこの問題を克服するため、加速度センサを用いた腕の動きによる確信度推定手法と、Web カメラを用いた顔画像による確信度推定手法の 2 つを提案する。

加速度センサはスマートウォッチのような日常生活で利用されるデバイスに搭載されている。また、web カメラは一般的な PC の多くに標準搭載されている。したがってどちらのデバイスも導入の容易さにおいて既存手法より優れている。以上の点から提案手法の有効性が確認されれば、既存手法より手軽で広範囲な環境での確信度推定の実現が期待される。

以下、腕の動きによる手法と顔画像による手法について並行して述べる。腕の動きによる手法では 1 種類の手法を、顔画像による手法では手法 A と手法 B の 2 種類の手法を提案する。

3.1 データの取得

それぞれのセンサを用いて学習者の振る舞いをデータとして取得する。この時腕の動きによる手法と顔画像による手法の両方において、問題解答用のアプリケーションを用いて 1 問ごとの解答開始時刻、解答終了時刻を記録する。以下、それぞれの手法のデータの取得方法について述べる。

3.1.1 腕の動きによる手法

学習者はタブレットに英単語を筆記する形で問題を解答する。このとき、加速度センサを学習者の利き腕に取り付け、腕の動きを計測する。データとして 3 軸加速度、3 軸角速度、クォータニオンを取得する。

3.1.2 顔画像による手法

学習者は PC 上で英文法の多肢選択問題を解答する。このとき、解答時の学習者の振る舞いを PC 内蔵の Web カメラより録画し、動画データに OpenFace [6]

表 1 視線に関するデータ (顔画像による手法)

| No. | 取得するデータ |
|-----|-----------------------|
| 1-3 | x, y, z 軸の左目の視線ベクトル |
| 4-6 | x, y, z 軸の右目の視線ベクトル |
| 7 | 上下の視線方向 |
| 8 | 左右の視線方向 |

表 2 姿勢に関するデータ (顔画像による手法)

| No. | 取得するデータ |
|-------|--------------------------|
| 1-3 | x, y, z 軸のカメラに対する頭の位置 |
| 4-6 | x, y, z 軸のカメラに対する頭の回転 |
| 7 | 顔を表現する点分布モデル (PDM) のスケール |
| 8-10 | x, y, z 軸の PDM の回転 |
| 11-12 | x, y 軸の PDM の平行移動 |

の FeatureExtraction を用いる。手法 A では、視線に関するデータ、顔の姿勢に関するデータを取得する。表 1, 表 2 に手法 A で使用するデータを示す。手法 B では、上下の視線方向と左右の視線方向のデータのみを取得する。

3.2 データの加工, 整形

取得したデータを用いて、問題 1 問ごとにデータの加工, 整形を行う。ここで、これまでの研究で確信度推定に解答時間が特に有効であることが判明している。そのため、腕の動きによる手法と顔画像による手法 A について、問題解答用アプリケーションより取得した解答時間を特徴量として追加する。以下、それぞれの手法のデータの加工, 整形方法について述べる。

3.2.1 腕の動きによる手法

取得したデータから特徴量を抽出する。解答時間を含めて合計 33 種類の特徴量が抽出される。表 3 に抽出する特徴量を示す。

3.2.2 顔画像による手法 A

取得したデータからそれぞれ基本統計量 (平均, 最大, 最小, 中央値, 標準偏差) を算出して特徴量として抽出する。視線に関するデータから解答時間を含めて合計 41 種類の特徴量が抽出される。また、姿勢に関するデータから解答時間を含めて合計 61 種類の特徴量が抽出される。

3.2.3 顔画像による手法 B

取得した視線方向のデータを 2 次元座標にプロットして、問題一問につき 1 枚の画像を作成する。これにより、問題解答中の視線方向の変化を描画する。

3.3 識別

3.3.1 腕の動きによる手法と顔画像による手法 A

抽出した特徴量を用いて確信の有無を判別する識別器を作成する。Support Vector Machine (SVM) を使用し正解問題から確信なしを検出する識別器と不正解問題から確信ありを検出する識別器の 2 つを作成する。これにより偶然正解した問題と誤った知識で解いた問題を検出する。またパラメータは特徴数を n , 学習データの分散を v とし、 $C=1.0$, $\gamma=\frac{1}{n \cdot v}$ とする。

3.3.2 顔画像による手法 B

作成した画像から確信の有無を判別する識別器を作成する。Convolutional Neural Network (CNN) を使用し、問題

表 3 抽出する特徴量 (腕の動きによる手法)

| No. | 特徴量 |
|-------|---------------------|
| 1 | 解答時間 |
| 2-4 | x, y, z 軸加速度の平均 |
| 5-7 | x, y, z 軸角速度の平均 |
| 8-11 | クォータニオンの平均 |
| 12-14 | x, y, z 軸加速度の振幅平均 |
| 15-17 | x, y, z 軸加速度の位相平均 |
| 18-20 | x, y, z 軸加速度の分散 |
| 21-23 | x, y, z 軸角速度の分散 |
| 24-27 | クォータニオンの分散 |
| 28-30 | x, y, z 軸加速度の振幅分散 |
| 31-33 | x, y, z 軸加速度の位相分散 |

データを確信あり, 確信無しの 2 クラスに分類する。この時 CNN のモデルは線形モデルを利用し, 2 つの CNN ブロックと各ブロックの後の max-pooling をベースのネットワークとした。バッチサイズを 8, カーネルサイズを 3×3 , 活性化関数を ReLU, 出力層の活性化関数をシグモイド関数, 最適化関数を RMSprop, 入力画像サイズを 128×128 とした。

4 実験

提案手法の有効性を検証するためそれぞれ実験を行った。本節ではまず行った実験について触れ, その後, 結果・考察を述べる。

4.1 実験条件

4.1.1 腕の動きによる手法

実験には日本人大学生 10 名が参加した。このうち男性は 7 名であり女性は 3 名であった。平均年齢 21.4 歳, 年齢の標準偏差は 0.49 であった。実験参加者の利き手の手首に加速度センサを取り付け, サンプル周波数 1000Hz でデータを取得した。実験参加者は英単語筆記問題を 1 時間解答し, 1 問解き終えるごとに 5 段階の確信度アンケートを行った。確信度は 5 に近づくほど確信があり, 1 に近づくほど確信がないと評価するよう指示した。また, 確信ありと確信なしの問題のデータの偏りをなくすため, 問題 1 セット 10 問の内, 確信度アンケートで 4, 5 と解答した問題が 3 問以下の場合には次のセットの問題難易度を下げ, 8 問以上の場合には問題難易度を上げるように調整した。謝礼は 1000 円分の Amazon ギフトカードとした。

4.1.2 顔画像による手法

実験はグループ P とグループ S の 2 回に分け実施した。グループ P の実験には日本人大学生, 大学院生 10 名が参加した。このうち男性は 8 名, 女性は 2 名であり, 平均年齢 22.5 歳, 年齢の標準偏差は 0.67 であった。実験参加者には 1 時間問題を解答するよう指示した。謝礼は 1000 円分の Amazon ギフトカードとした。

グループ S の実験には日本人大学 1 年生 9 名が参加した。このうち男性は 6 名, 女性は 3 名であった。実験参加者には 3 時間の授業の中で, 合計 95 分間問題を解答するよう指示した。

両グループにおいて, 実験参加者は英文法多肢選択問題を解答し, 1 問解き終えるごとに 5 段階の確信度アンケートを行った。確信度は 5 に近づくほど確信があり, 1 に近づくほど確信がないと評価するよう指示した。ま

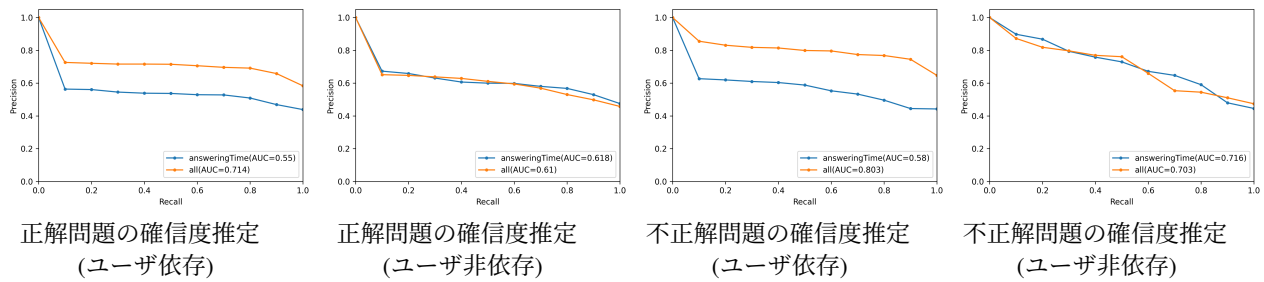


図 2 腕の動きによる確信度推定の Precision-Recall 曲線

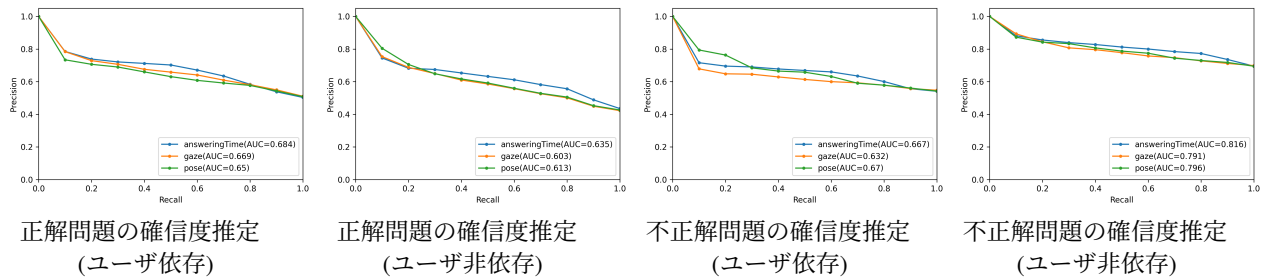


図 3 顔画像による手法 A の確信度推定の Precision-Recall 曲線

た、正誤の問題のデータの偏りをなくすため、問題 1 セット 10 問の内、正解した問題数が 3 問以下の場合は次のセットの問題難易度を下げ、8 問以上の場合は問題難易度を上げるように調整した。

4.2 検証方法

4.2.1 確信度ラベルの取り扱い

実験参加者によって確信度の分布には偏りが見られた。そのため、確信度のラベルの中央値を確信の有無の閾値として設定し、中央値より上のラベルを確信あり、中央値より下のラベルを確信なしと定めた。腕の動きによる手法と顔画像による手法 A では、中央値のラベル自身は確信あり、確信なしのうちサンプルの少ないクラスとした。顔画像による手法 B では、中央値のラベル自身はどちらのクラスにも含めず、データセットから除外した。

4.2.2 交差検証

取得したデータセットを提案手法を用いて識別する際に交差検証を行い精度を算出した。腕の動きによる手法と顔画像による手法 A では、ユーザ依存、ユーザ非依存の両方を検証した。一方、顔画像による手法 B では、ユーザ非依存のみについて検証を行った。ここでユーザ依存とは、実験参加者自身の問題解答時の振る舞いを学習データとして識別器を作成することを指す。一方ユーザ非依存とは、推定する対象の実験参加者以外の参加者について問題解答時の振る舞いを学習データとして受け取り、識別器を作成することを指す。ユーザ依存では、10-fold cross validation で交差検証を行った。また、ユーザ非依存では leave one out cross validation で交差検証を行った。この時、腕の動きによる手法と顔画像による手法 A では、SMOTE を用いて確信ありの問題数と確信なしの問題数が同数になるようにサンプル数を調整した。

4.2.3 評価方法

提案手法の有効性について、それぞれ評価を行った。動きによる手法と顔画像による手法 A では、正答・

誤答の識別器ごとに PR(Precision-Recall) 曲線を描画し、AUC(Area Under the Curve) から評価した。そして解答時間のみを特徴量としたときの識別器の PR 曲線をベースラインとして、提案手法の PR 曲線と比較し評価した。一方顔画像による手法 B では、正答率の学習曲線を描画し、最終エポックの正答率とチャンスレートと比較することで評価を行った。ここでチャンスレートは全ての問題数のうち確信ありのクラスに属する問題数の割合とした。

5 結果・考察

5.1 腕の動きによる手法

本項では腕の動きによる手法の確信度推定の実験結果とその考察について述べる。図 2 に問題正解時、不正解時のユーザ依存、ユーザ非依存における確信度推定の PR 曲線をそれぞれ示す。ユーザ依存では提案手法の PR 曲線の AUC が問題正解時、問題不正解時ともにベースラインを上回っている。したがって提案手法の有効性が確認され、腕の動きの情報が確信度推定に寄与していることがわかる。しかし、ユーザ非依存では提案手法とベースラインの PR 曲線に差は見られなかった。

以下、実験結果について考察する。ユーザ依存での確信度推定の結果から、学習者にはそれぞれ確信がある時と確信がない時に腕の動きに異なる振る舞いの傾向がある可能性が考えられる。一方、ユーザ非依存の確信度推定で提案手法の有効性が確認できなかったことから、学習者の腕の動きの振る舞いの癖は一人ひとりで異なり、提案手法では全ての学習者に共通する確信の有無により生じる癖を見つけ出すことができなかったことが考察される。

5.2 顔画像による手法 A

本項では顔画像による手法 A の確信度推定の実験結果とその考察について述べる。図 3 に問題正解時、不正解時のユーザ依存、ユーザ非依存における確信度推定の PR 曲線をそれぞれ示す。

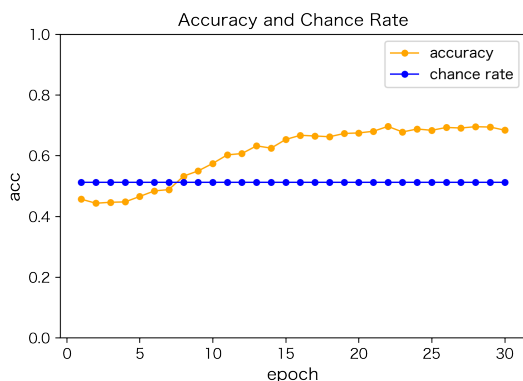


図 4 顔画像による手法 B の確信度推定の学習曲線

問題正解時、不正解時にかかわらずユーザ依存、ユーザ非依存におけるいずれの場合においてもベースラインである解答時間の PR 曲線の AUC と提案手法の AUC の間に大きな差は見られなかった。したがって、提案手法では顔画像データからの確信度の推定への有効性は見られなかった。

以下、提案手法が有効でなかった原因について考察する。原因の一つとして、OpenFace で取得された時系列データからの特徴量の抽出方法が挙げられる。今回の手法では、1 問分の問題を解き始めてから解き終わるまでの時系列データをまとめて平均や標準偏差といった基本特徴量として抽出していた。しかし実際は、問題を解いている中の実験参加者の動きの変化のより詳細な変化が確信度の推定に寄与する可能性が考えられる。実際、問題を解いている学習者を観察していると、わからない時に首を傾げる動作をしたり俯く動作が確認された。こうした学習者の局所的な振る舞いの変化を特徴量として抽出することができれば、確信度推定の精度改善が実現できると考察される。

5.3 顔画像による手法 B

本項では顔画像による手法 B の確信度推定の実験結果とその考察について述べる。図 4 に顔画像による手法 B の確信度推定の学習曲線を示す。提案手法の最終エポックの正答率は、69.98%となった。またこの時、チャンスレートは 51.18%であった。したがって、正答率と図 4 から提案手法の有効性が確認された。すなわち、Web カメラから取得される視線方向のデータは確信度の推定に寄与する情報であることが明らかとなった。

以下、実験結果について考察する。提案手法で作成された画像を確認した時、確信がある時とない時でプロットの散らばりや密度に差が見られた。これは学習者がわからない問題に出くわしたときに視線が同じ場所にとどまり続けたり、何度も問題を見直すことにより視線の方向があちこちにばらつくといった振る舞いの変化が原因であると考えられる。このような振る舞いの違いが確信度の推定に寄与したことが予測される。さらに精度を向上するためにはより学習者の振る舞いを表現するような描画の手法を検討する必要がある。例えば、視線だけでなく顔そのものの動きも合わせて表現できるような描画手法が実現できれば、より高精度に確信度を推定できると考察される。

6 結論

確信度を考慮した効果的な学習の実現のため、センサデータで観測した学習者の振る舞いから確信度を推定することが試みられている。従来の手法では用いられるセンサが通常環境では用いられないものが多く、導入が困難であるという問題点があった。そこで本研究ではこの問題を克服する為に、加速度センサを用いて腕の動きから確信度を推定する手法と、Web カメラを用いて顔の動きから確信度を推定する手法について提案した。実験の結果、SVM を用いた顔画像による手法とユーザ非依存における腕の動きによる手法では提案手法の有効性は確認できなかった。しかし、ユーザ依存では腕の動きから得られる情報が筆記問題の確信度推定に有効であること、CNN を用いた手法においてユーザ非依存で、顔画像から得られる視線方向の情報が英単語多肢選択問題の確信度推定に有効であることを確認した。

今後の課題としては、より精度の高い確信度推定を実現することがあげられる。具体的には、CNN を用いる手法においてより学習者の振る舞いを表現するような描画の手法を検討することや、学習者の学力や性格といった個人の特性に適応させて確信度を推定することなどが考えられる。

参考文献

- [1] Lazar Stankov, Sabina Kleitman, and Simon A Jackson. Measures of the trait of confidence. In *Measures of personality and social psychological constructs*, pp. 158–189. Elsevier, 2015.
- [2] Seiya Tanaka, Andrew W. Vargo, Motoi Iwata, and Koichi Kise. Confidence estimation via wrist movement. In *Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp/ISWC '21 Adjunct)*, September 2021.
- [3] 田中勢也, Andrew W. Vargo, 黄瀬浩一. 腕の動きと web カメラを用いた英語問題の確信度推定. 電子情報通信学会総合大会, 2022.3.
- [4] Kento Yamada, Olivier Augereau, and Koichi Kise. Estimation of confidence based on eye gaze: an application to multiple-choice questions. In *Proc. UbiComp2017*, pp. 217–220. ACM, September 2017.
- [5] 朝春森, 達人長谷川. 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. 情報処理学会論文誌, Vol. 61, No. 8, pp. 1306–1318, aug 2020.
- [6] Tadas Baltrušaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, pp. 59–66, 2018.

謝辞

本研究の一部は JST CREST (JPMJCR16E1), JST Trilateral AI Research (JPMJCR20G3), JSPS 科研費基盤 (B) (20H04213), JSPS 国際共同研究強化 (B) (20KK0235), 阪大 Society5.0 グランドチャレンジの補助による。