

眼球運動を用いた解答戦略有無の判定 — 数学問題を対象として —

宮崎 公彦[†] 青柳 宏紀[‡] 岡田 一洗[‡] 工藤 雅士[‡] 山名 早人[§]

[†] 早稲田大学基幹理工学部 〒169-8555 東京都新宿区大久保 3-4-1

[‡] 早稲田大学大学院基幹理工学研究科 〒169-8555 東京都新宿区大久保 3-4-1

[§] 早稲田大学理工学術院 〒169-8555 東京都新宿区大久保 3-4-1

E-mail: [†] [‡] [§] {loup, aoyagih, k-okada, kudoma34, yamana}@yama.info.waseda.ac.jp

あらまし 近年、オンライン学習が急速に普及し、学習者に合わせた最適な学習環境を提供することが求められている。これらの適応的な学習環境の提供には、学習者の状態把握が不可欠であり、その手法の一つとして眼球運動の利用がある。眼球運動は認知的負荷推定に有効であることが知られており、眼球運動は問題難易度の推定などに用いられている。本稿では、アイトラッカーによる眼球運動データおよびペンタブレットによる手書き入力データを用いて、数学問題解答中における解答者の状態変化を捉え、解答戦略の有無を判定する手法を提案する。眼球運動データから 20 の特徴量を抽出し、Random Forest および XGBoost を用いて、学習者の 3 状態（明確な解法に基づく解答、積極的解法探索、非積極的解法探索）の自動分類を行った。数学問題 10 問を用いた実証実験（被験者数 14 名）の結果、53.6%の正解率を得た。加えて、「解答戦略有り」と「解答戦略無し」の 2 値分類では、70.2%の正解率を得た。

キーワード データ分類, 眼球運動, e-Learning, 数学, 手書き

1. はじめに

近年、オンラインでの学習が急速に普及している。特に 2019 年から現在まで続くコロナウイルスの蔓延による影響は大きく、総務省による令和 3 年の情報通信白書¹では、2020 年 12 月の調査で、高校生の 29.2%、大学生の実に 87.7%が学校での授業の一部ないしは全てをオンラインで受けたと回答している。このように、パソコンやタブレット、スマートフォン等を用いてインターネットを介して学ぶ学習形態を e-Learning と呼ぶ。

従来から行われてきた学校などでの授業では、教師が各学習者の状態を把握して個別に配慮をすることは難しい場合が多い。一方で、インターネットを介して学習を行うことで、学習者の学習進捗や問題の正誤情報などをリアルタイムに取得することができ、学習者の学習状況を簡単に把握することができるようになった。このような背景のもと、学習者の解答過程から得られるデータをテラーメイド教育に活かすことを目的とした研究が盛んに行われている。

学習状況を把握する方法の一つとして、学習者の解答過程から解答戦略を明らかにする手法[1]が提案されている。しかし、問題難易度が上がると一般に解答過程のパターン数が増え、解答戦略の判定が難しくなるという課題がある。そこで本研究では、解答難易度に依存せずに解答戦略を持って解いているかを明らかにする方法を目指す。具体的には、眼球運動を用いて

学習者の状態を把握する。

人間がものを見るときに眼球運動は、固視とサッケードに分けられる。固視はある 1 点を一定時間見つめることであり、空間的にも時間的にも近い注視点の集合のことを指す。一方で、ある固視点から次の固視点へ移る際の素早い目の動きのことをサッケードと呼ぶ。認知負荷や注意力を眼球運動から推定する際に、固視やサッケードの回数や継続時間が特徴量として多く用いられる。また、その他の眼球運動としてまばたきや瞳孔直径の変化も同様に負荷の推定に有効であることが知られている。本研究では、眼球運動に表れる生理反応から学習者の状態を把握し、数学問題の解答過程における学習者の解答戦略の有無を明らかにする。

本稿は、以下の構成をとる。2 節で関連研究を示す。3 節にて、本研究で行う解答戦略有無の判定の提案手法を説明し、4 節で評価実験の説明と実験の結果を示す。5 節で評価を行い、6 節でまとめる。

2. 関連研究

2.1 解答戦略に関する研究

浅井ら[1]は、2016 年に、数学の多項式展開問題の手書きデータを用い、多項式展開に用いた解答戦略の分類を行った。解答戦略が同じである場合、「式変形が類似する」というアイデアに基づき、解答ごとの類似度を比較し、類似度の高い解答をまとめ上げることで、解答戦略の分類を行っている。類似度の計算にあつ

¹<https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r03/html/nd122230.html>

ては、筆記ストロークをサンプリングし、式変形を単位として、各ストローク間の類似度算出を行い、最終的に解答全体としての類似度を求めている。式変形においては途中の変形が省略されることがあるため、DPマッチングを用いた比較を行い、式変形の省略に対応している。そして、解答全体としての類似度を用いたクラスタリングにより、解答戦略別に解答を分類した。

評価では、ベースライン手法として、解答時間を特徴量としてクラスタリングする手法との比較を行っている。33名の被験者に電子ペンを用いて4問の多項式展開問題を出題し分類を行った結果、難易度が高い問題に対しては、提案手法とベースライン手法の間に大きな性能差がないことを明らかにしている。一方で、問題難易度が低く答案への書き込みが少ない問題に対しては、提案手法がベースラインに比較して高い分類性能を示すことを明らかにした。すなわち、ストロークを用いた解答戦略分類では、問題難易度がその精度に影響を及ぼすことが示された。

2.2 眼球運動に関する研究

Shojaeizadehら[2]は、2019年に、数学問題を解いている間の眼球運動データを、Random ForestおよびSVMを用いることで分類し、被験者に対するタスク負荷の予測を行った。48人の被験者に対して、10問の数学問題を解く課題を課し、これらの被験者を2つのグループに分けた。時間的な制約がある場合に認知的負荷が高くなることを利用し、1つのグループには1問につき5分の時間制限を設けた。他方のグループは、時間制限を設けずに実験を行った。時間制限を設けたグループはタスク負荷が高いという仮定を置き、この2つのグループの分類を行った結果、69.6%の分類精度を達成している。また、サッケード時と固視時における瞳孔拡張比が分類にもっとも影響を与えることを明らかにした。

Madsenら[3]は、2021年に、オンラインでの教育ビデオ視聴中の眼球運動から、ビデオへの注意の程度を評価した。目の動きの撮影にはアイトラッカーではなくウェブカメラを用いている。ウェブカメラはアイトラッカーに比べてサンプリング周波数および空間分解能が低いという欠点がある。このため、「ビデオ視聴時に注意を払っている」、すなわちビデオの視聴に集中している場合、視聴者の眼球運動が類似していることに着目した。結果、眼球運動とビデオ視聴後のテストの成績に強い相関があることを示している。

3. 提案手法

3.1 問題の解答方法

本稿では、数学問題を解いているときの解答者の状

態（解答戦略）を眼球運動データから明らかにすることを目的とする。特に、従来のストロークを使った解答戦略分類が問題難易度に影響を受けるという問題[1]を解決できる「問題の難易度によらない解答戦略の自動分類」を目指す。

眼球運動データを取得するため、トビー・テクノロジー社のアイトラッカーである Tobii Pro Nano を使用した。アイトラッカーが適切に計測を行うにはアイトラッカーと、ある程度の距離（Tobii Pro Nano の場合には 45–85cm）を保つ必要があり、iPad 等のタブレットを机の上において問題を解くような方法では計測ができない。そこで、手元を見る必要がなく PC 画面を見ながら文字を書くことができるペンタブレットを用いて解答を行うこととした。ペンタブレットは、ワコム社の Wacom Intuos Pro Medium²を使用した。実験のセットアップを図1に示す。

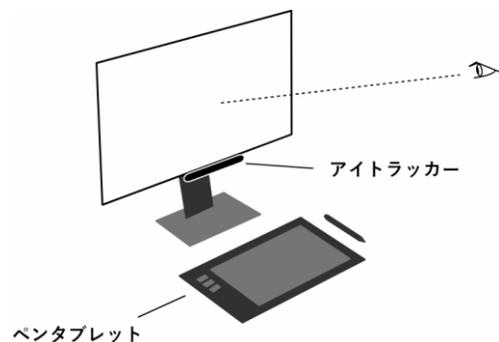


図1 実験のセットアップ

3.2 解答者の状態把握

3.2.1 解答者の状態分類

数学問題を解いているときの解答者の状態は、本人のみが知り得る情報である。そこで、被験者には、全問題の解答が終了した後、各問題での自分の状態を記録するラベル付けを行ってもらい、あらかじめ被験者には表1に示す6状態を提示し、この中から最も適切だと思われる状態を選択するように指示する。

表1 6状態の定義

状態	定義
明確な解法に基づく解答	解法の正否に関わらず、特定の解法を想定して解答を行っている期間
積極的な解法の探索	特定の解法の思考実験または積極的な解法探索
非積極的な解法の探索	解法の想定が不能もしくは厳しい状態
問題文の読み取り見直し	一番初めに問題文を読むとき 計算の見直し、検算
その他	上記5状態のいずれにも当てはまらない場合

「明確な解法に基づく解答」は、例えば点と直線の

² <https://estore.wacom.jp/ja-JP/category/pen-tablets/wacom-intuos-pro/medium.html>

距離を用いて最短距離を算出するなど、明確な解法に基づいて解答を行っている期間を指す。一方で「積極的な解法の探索」は、とりあえずの式変形や、作図など解法は分からないが積極的に解答を行っている期間を指す。また「非積極的な解法の探索」は、お手上げ状態や諦め状態にある期間を指す。

3.2.2 ストロークに基づくラベル付け

表 1 の 6 状態に示される被験者の状態は、1 つの問題解答中に変化する。しかし、被験者が、状態変化点を手掛かりなしに正確に指定することは困難である。そこで、時系列順に並べたストロークデータをもとに状態の変化点を指定する手法を提案する。なお、本稿ではペンがペンタブレットに触れて筆記を開始してから、次にペンがペンタブレットから離れるまでを 1 ストロークとする。

まず、ストロークデータを分割する。具体的には、時系列順に並べたストロークデータについて、1 秒以上間隔が空いた時から次に 1 秒以上間隔が空くまでをひとまとまりとして分割し、このストローク群を以降フレームと呼ぶ。次に、フレームを手掛かりに、ある状態下にあったと考えられる部分を、被験者に指定してもらい（開始フレーム番号と終了フレーム番号）、状態のラベル付け（状態ラベルと呼ぶ）を行う。この作業を状態が遷移するごとに問題解答の最初から最後まで繰り返すことで、問題解答の時系列データを各状態で切り分ける。このようにしてできた、表 1 の 6 状態のうち 1 状態のみを持つ連続フレーム群を、セクションと呼ぶ。

フレーム：連続するストローク間の間隔が 1 秒下である連続ストロークの集合

セクション：同一の状態（表 1 で定義）を持つ連続するフレームの集合

フレームとセクションの関係を図 2 に示す。

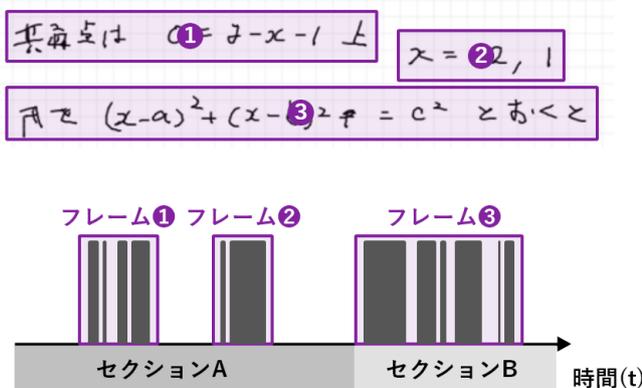


図 2 ストロークデータのフレーム分割とセクション分け

3.3 特徴量の抽出

3.3.1 使用する特徴量

3 状態の分類に使用する特徴量を表 2 に示す。表 2 内の記号の定義は以下の通りである。

- i : 欠損値を除いたサンプリングデータを時系列順に並べたときのナンバリング
- x_i : i 番目の注視点の x 座標
- y_i : i 番目の注視点の y 座標
- $p_{l,i}, p_{r,i}$: i 番目の左目, 右目の瞳孔直径
- t_i : i 番目のデータを取得したときの UNIX 時間
- w, h : 問題を表示するモニターの pixel 数
($w=1920, h=1080$)

表 2 特徴量

特徴量	定義
視線移動量 平均 標準偏差 最大	同一セクション内での単位時間当たりの視線の移動量 (px) $\frac{\sqrt{((x_{i+1}-x_i)*w)^2 + ((y_{i+1}-y_i)*h)^2}}{t_{i+1}-t_i}$
瞳孔サイズ 平均 標準偏差 最大	同一セクション内での左目瞳孔直径 (mm)
瞳孔サイズ変化割合 平均 標準偏差 最大	同一セクション内での単位時間当たりの左目瞳孔直径の変化割合 $\frac{ p_{l,i+1}-p_{l,i} }{t_{i+1}-t_i}$
左右瞳孔サイズ差 平均 標準偏差 最大	同一セクション内での左右の目の瞳孔サイズの差 $\frac{ p_{l,i}-p_{r,i} }{\min(p_{l,i}, p_{r,i})}$
固視回数 継続時間 平均 標準偏差 最大	同一セクション内での固視 (30deg/sec 以内の視線移動量の期間) の回数および継続時間 (ミリ秒)
まばたき回数 継続時間 平均 標準偏差 最大	同一セクション内でのまばたきの回数および継続時間 (ミリ秒)

表 2 に示した同一セクション内の特徴量の集合 X に対し、平均 \bar{X} 、標準偏差 s 、最大 X_{max} を求める。

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (1) \quad s = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2} \quad (2) \quad X_{max} = \max X \quad (3)$$

i : 表 2 の各特徴量から異常値と外れ値 (3.3.2 節参照) を除いたものを時系列順に並べたときのナンバリング

n : 同一セクション内における表 2 の各特徴量から異常値と外れ値を除いた際のデータ数

X_i : 表 2 の各特徴量から異常値と外れ値を除いた i 番目のデータ

特徴量を算出するにあたり、まず、左目右目の注視点および左目右目の瞳孔直径の4つの値のいずれかが欠損している場合は、その時刻に計測された4つのデータ全てが欠損しているとして扱う。

次に、各特徴量の算出方法を説明する。視線移動量および瞳孔サイズ変化割合を求めるにあたり、データが欠損していた場合、時間的に連続したデータ間の移動量や変化割合を正確に求めることができない。そこで、アイトラッカーのサンプリング周波数が60Hzであることから、2フレームに満たない32ミリ秒を閾値とし、連続するデータの時間間隔が32ミリ秒以上の場合は欠損があったとして、当該データ間の特徴量は算出しないものとする。また、隣接するデータの時間間隔が75ミリ秒以上425ミリ秒[4]未満のデータ欠損をまばたき、32ミリ秒以上75ミリ秒未満および425ミリ秒以上の場合を理由不明のデータ欠損として扱う。

固視の検出には、目の移動速度により分類を行うI-VT Fixation Filter[5]を用いる。固視とサッケードの閾値は30deg/sec[5]とした。I-VT Fixation Filterでは、閾値以上の視線移動速度を持つ期間をサッケード、それ以外を固視と判定するが、固視と判定された期間の中には固視継続時間が極端に短いものも存在する。通常、固視は最低でも200ミリ秒[6]継続されることが知られているため、固視と判定されたものについて、その期間が200ミリ秒以上のもののみを正式に固視とみなす。

3.3.2 異常値と外れ値の除外

表2で得られた各被験者の特徴量に対して、問題ごとに、原因が分かっている異常な値である異常値と、原因不明の異常な値である外れ値の除外を行う。

まず、まばたきによる異常値の除外を行う。データ欠損期間が75ミリ秒以上425ミリ秒未満の場合まばたきであると判定する。この時、まばたきによるデータ欠損期間の前後数サンプリングデータは通常よりも瞳孔サイズが小さく検出される。したがって、75ミリ秒から425ミリ秒のデータ欠損期間に隣接する前後50ミリ秒[7]以内のデータを削除する。

次に外れ値の除外を行う。表2で得られた各被験者の特徴量から被験者ごとに異常な瞳孔サイズ変化割合、瞳孔サイズ、左右瞳孔差を削除するため、これら3つのデータについて第1四分位数Q1、第3四分位数Q3、四分位範囲IQRを計算し、Q1およびQ3からIQRの1.5倍以上離れているデータを外れ値として削除する。

3.3.3 特徴量の正規化

被験者間の差異の影響を抑えるため、表2で得られた各特徴量について、被験者ごとにZスコアを求めて標準化を行う。

$$z_i = \frac{X_i - \bar{X}}{\sigma} \quad (4)$$

i : 同一被験者のすべての問題における全セクションのナンバリング（順不同）

X_i : 同一被験者のセクション*i*における特徴量

σ : 同一被験者のすべての問題における全セクションの特徴量の標準偏差

3.4 解答戦略有無の判定手法

セクション分けされたデータには、眼球運動データとストロークデータおよびその間の解答者の状態ラベルが含まれる。眼球運動データから表2で示した特徴量を抽出し、20個の特徴量からRFE (Recursive Feature Elimination) による特徴量選択を行う。選択された特徴量を使用しRandom Forest及びXGBoostを用いて解答戦略の有無の自動判定を行う。なお、分類対象となる状態ラベルは、「明確な解法に基づく解答」「積極的な解法の探索」「非積極的な解法の探索」の3つであり、この中で「明確な解法に基づく解答」の状態ラベルを持つセクションを「解答戦略有り」、その他を「解答戦略無し」とする。本稿は、解答者の解答戦略の有無を判定することを目的としているため、解答戦略の有無に関係のないラベルである「問題文の読み取り」「見直し」「その他」の分類は行わない。

4. 評価実験

4.1 データ収集

4.1.1 データ収集用アプリケーション

数学問題解答中の眼球運動データおよびストロークデータを取得するため、Webアプリケーションを開発してデータ取得を行った。解答画面を図3に示す。

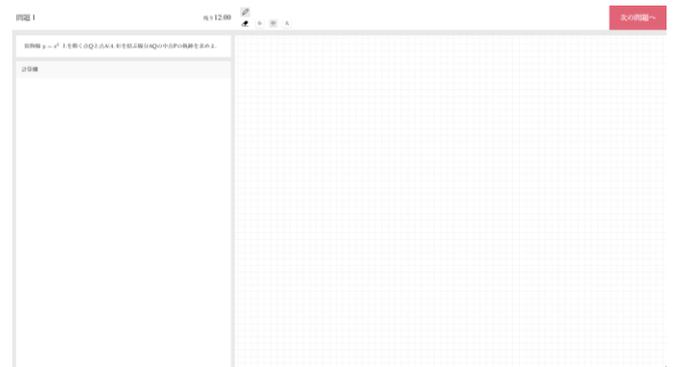


図3 解答画面

解答画面の左上には問題、左下には計算用紙として使える計算欄がある。実際の解答は右側にあるマスの目のある解答欄に書く。右上には次の問題へ移るボタンがあり、このボタンを押すもしくは12分経過で次の問題へ遷移する。計算欄と解答欄にはペンタブレットを用いて線を描くことができるほか、消しゴムツールを選択することで描いた線を消すこともできる。

使用する問題は、数学IIの図形と方程式の分野か

ら 10 問選出した。5 段階評価 (1-5) の難易度のうち 2 から 4 までを均等に出題する。以下に使用した問題の一例を示す。

問 1: 放物線 $y = x^2$ 上を動く点 Q と点 A(4, 0) を結ぶ線分 AQ の中点 P の軌跡を求めよ。

4.1.2 収集データの内容と形式

本アプリケーションは、問題解答時に視線データとストロークデータの 2 種類のデータを JSON 形式で出力する。視線データは問題が画面に表示された時に自動的に取得が開始され、次の問題へ移るまで継続して計測される。一方、ストロークデータは、計算欄と解答欄に記入があった場合にデータが記録される。記録されるデータの種類と説明を表 3、表 4 に示す。

表 3 視線データ

データ名	定義
left_gaze_point	左目の注視点の x, y 座標
right_gaze_point	右目の注視点の x, y 座標
left_pupil_diameter	左目の瞳孔サイズ (mm)
right_pupil_diameter	右目の瞳孔サイズ (mm)
validity	上記 4 つのデータの有効性 ³
time	UNIX 時間 (秒)

表 4 ストロークデータ

データ名	定義
canvas_id	計算欄か解答欄のどちらに書き込まれたかを表す番号
type	PointerEvent ⁴ インターフェイスにより記録されるイベント名
tool	使用したツール名 (ペンもしくは消しゴム)
x	マウスポインタの x 座標 (px)
y	マウスポインタの y 座標 (px)
pressure	ペンの筆記圧力 (0 - 1, 8192 段階)
eraser_width	消しゴムツールの大きさ (直径 50px, 100px, 200px)
time	UNIX 時間 (ミリ秒)

注視点座標は、画面の左上を(0, 0)、右下を(1, 1)とする座標系である。一方、マウスポインタの座標は、計算欄もしくは解答欄の左上(0, 0)からの距離を px で表す座標系である。

4.1.3 データ収集実験の実施

大学生 14 名 (男性 10 名, 女性 4 名) を対象に、数学問題の解答実験を行った。実際の実験手順を以下に示す。

1. ペンタブレットの操作方法の説明
2. アイトラッカーの説明および Tobii の Eye Tracker Manager を用いたキャリブレーション

³ 有効性とはデータが欠損しているか否かを意味する

3. 操作に慣れるための 10 分間のチュートリアル問題 (データ取得対象外)
4. 数学問題 10 問の解答
5. ラベル付け

アイトラッカーは目に赤外線を照射することで眼球運動を計測するため、PC 画面以外の天井や手元を見ると正確な計測ができない。そこで、被験者には、問題解答中には PC 画面を可能な限り見つめるように指示をした。また、実際の解答実験に移る前に、ペンタブレットでの筆記に慣れるために、10 分間のチュートリアル問題を解いてもらった。チュートリアル問題の間に、PC 画面のみを見る練習も行うように被験者に指示をした。

4.2 データ収集実験の結果

各問題の難易度、全被験者の問題ごとの得点率、平均解答時間および平均ストローク回数を表 5 に示す。解答の一部が記述されているなど、半分以上の部分点が与えられる解答は半分得点できたとして得点率を算出している。また、図 4 にストローク回数と解答時間および問題の正誤の散布図を示す。

表 5 各問題の難易度、得点率、平均解答時間および平均ストローク回数

問題番号	1	2	3	4	5	6	7	8	9	10
難易度	3	2	3	3	4	2	4	4	2	2
得点率 (%)	27	81	31	31	17	14	32	18	14	18
解答時間 (秒)	平均	375	310	479	373	457	450	289	358	388
	標準偏差	184	172	175	152	217	164	220	149	230
ストローク回数	平均	234	286	285	258	309	286	176	223	230
	標準偏差	161	165	180	141	193	142	146	124	158

表 5 から、問 2 を除いて全体的に得点率が芳しくなく、多くの場合完答まで至っていないことが分かる。全体的には、得点率が低いほど平均解答時間が長くなるほか、平均解答時間が長くなると平均ストローク回数も増える傾向にある。

図 4 より、解答時間に対してストローク回数が多くなるほど正答する確率が高いことがわかる。一方で、部分点のみの解答はその多くが短い解答時間となっている。これは、解答の一部を出した時点で完答をあきらめて次の問題へ進んでしまうためだと考えられる。

⁴ https://developer.mozilla.org/en-US/docs/Web/API/PointerEvent#pointer_event_types

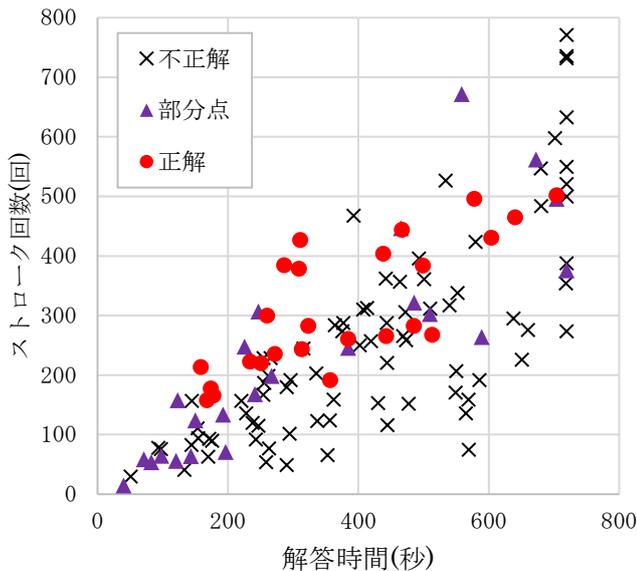


図 4 ストローク回数と解答時間および問題の正誤の散布図

最後に、問題ごとの「明確な解法に基づく解答」「積極的な解法の探索」「非積極的な解法の探索」の3つの状態ラベルのついたセクションの割合（3 ラベル以外のラベルがついたセクションを含めずに計算した割合）と難易度および平均得点率の相関係数を表 6 に示す。

表 6 各状態ラベルのついたセクションの割合と難易度および平均得点率の相関係数

	明確な解法に基づく解答	積極的な解法の探索	非積極的な解法の探索
難易度	0.0438	-0.219	0.219
得点率	0.845	-0.559	-0.788

表 6 より、各状態ラベルと得点率は一定の相関関係が確認されたが、難易度との相関関係は確認されなかった。このことから、被験者が感じる難易度には差がある可能性があることが伺える。

5. 評価

5.1 データの選別

3.3 節で得た各セクションにおける特徴量をもとに、各セクションの3つの状態ラベル「明確な解法に基づく解答」「積極的な解法の探索」「非積極的な解法の探索」の分類を行う。

3.3 節で得た各セクションにおける特徴量の集合には、アイトラッカーから十分にデータが取得できなかったものも含まれているため、各セクションの継続時間 t から、本来得られるべきサンプリング数 ($t \cdot 60\text{Hz}$) を算出し、実際のサンプリング数とその半分に満たないセクションを除外する。その結果、セクション数が著しく少ない被験者が存在したため、セクション数が

2 以下の被験者のデータを評価用のデータから除外した。最終的に 12 人、計 235 個の状態ラベルの付いたセクションデータをもとに評価を行う。

5.2.3 状態ラベルの 3 値分類結果

5.2.1 RFE による特徴量選択

20 個の特徴量から分類精度の向上に寄与する特徴量を適切に選択するために、RFE (Recursive Feature Elimination) による特徴量選択を行う。状態ラベルの付いたセクションデータ 235 個を 5 分割し、分割された 5 個のデータをそれぞれテストデータとし、その他を訓練用データとして 5 回特徴量を選択する。また、各訓練用データで選択された特徴量から、Random Forest および XGBoost を用いて 3 値分類を行い、5 回の分類精度の平均を求める。最終的に、5 回の特徴量選択で得られた特徴量の重要度を足し合わせ、分類に影響を与える特徴量をランキング化する。その結果、Random Forest では 17 個もしくは 20 個すべての特徴量を選択した場合、XGBoost では 10 個もしくは 14 個の特徴量を選択した場合に最も高い 52% の分類精度となった。

5.2.2 3 状態ラベルの 3 値分類結果

5.2.1 節で選択された特徴量をもとに、状態ラベルの付いたセクションデータ 235 個について再度 5 分割交差検証を行った結果を表 7 に示す。

表 7 3 状態ラベルの 3 値分類結果

分類器	特徴量	分類精度
RF	17 個 視線移動量：平均/標準偏差/最大 瞳孔サイズ：平均/標準偏差/最大 瞳孔サイズ変化割合：平均/標準偏差/最大 左右瞳孔サイズ差：最大 固視：回数/継続時間[平均/標準偏差/最大] まばたき：回数/継続時間[平均/標準偏差]	51.5%
RF	20 個全て	50.6%
XGB	10 個 視線移動量：平均/標準偏差/最大 瞳孔サイズ：平均/標準偏差 固視：回数/継続時間[平均/標準偏差/最大] まばたき：回数	53.6%
XGB	14 個 視線移動量：平均/標準偏差/最大 瞳孔サイズ：平均/標準偏差 瞳孔サイズ変化割合：標準偏差/最大 左右瞳孔サイズ差：最大 固視：回数/継続時間[平均/標準偏差/最大] まばたき：回数	52.3%

XGBoost を用いて 10 個の特徴量で分類したときに、最も高い分類精度 53.6% を得た。この時の混同行列を表 8 に示す。

表 8 の混同行列から分かる通り、「明確な解法に基づく解答」と「積極的な解法の探索」に比べて「非積極的な解法の探索」は分類精度が低くなった。また、「非積極的な解法の探索」は「積極的な解法の探索」に分類される傾向にある。

表 8 XGBoost と特徴量 10 個を用いて 3 値分類を行った場合の混同行列

		予測値		
		明確	積極	非積極
真値	明確	42	38	6
	積極	22	69	14
	非積極	9	20	15

5.3 解答戦略の有無の判定

次に、解答戦略の有無を判定する。「明確な解法に基づく解答」を解答戦略有り、その他 2 つの状態ラベルを解答戦略無しとし、「明確な解法に基づく解答」とその他 2 つの状態ラベルについて 2 値分類を行い、解答戦略の有無の判定を行う。5.2.1 節と同様に、「明確な解法に基づく解答」とその他の 2 つのラベルの 2 値分類について、RFE による特徴量選択を行い、選択された特徴量をもとに再度 5 分割検証を行った。いずれも分類器は Random Forest を用いた。その結果を表 9 および表 10 に示す。

表 9 「明確な解法に基づく解答」とその他の 2 つの状態ラベルの 2 値分類結果

分類器	特徴量	分類精度
RF	16 個 視線移動量：標準偏差/最大 瞳孔サイズ：平均/標準偏差 瞳孔サイズ変化割合：平均/標準偏差/最大 左右瞳孔サイズ差：平均/標準偏差 固視：回数/継続時間[平均/標準偏差/最大] まばたき：回数/継続時間[平均/標準偏差]	70.2%

表 10 「明確な解法に基づく解答」とその他の 2 つの状態ラベルの 2 値分類における混同行列

		予測値	
		明確	積極・非積極
真値	明確	43	43
	積極・非積極	27	122

表 9 より、「明確な解法に基づく解答」とその他の 2 つの状態ラベルの分類精度は 70.2%であった。表 8 および表 10 より、「明確な解法に基づく解答」と「積極的な解法の探索」、「積極的な解法の探索」と「非積極的な解法の探索」の 2 つの分類精度が悪い。一方で、「明確な解法に基づく解答」と「非積極的な解法の探索」の分類精度はこれら 2 つと比較して良い結果になっている。

5.4 被験者・問題・難易度ごとの分類結果

5.2 節および 5.3 節の結果は、被験者、問題に依存することなく学習した結果である。そこで、最後に、被

験者ごと、問題ごとに学習した場合の分類精度を求めることで、現在の特徴量が被験者の差異を吸収できない特徴量であるのか、それとも問題の差異を吸収できない特徴量であるのかを判断する。

まず、5.2 節で行った RFE による特徴量の選択を、被験者、問題および難易度ごとに「明確な解法に基づく解答」「積極的な解法の探索」「非積極的な解法の探索」の状態ラベルのついた 3 種類のセクションデータについて 3 値分類を行い、最高の分類精度を得たときの特徴量の選択数および分類精度の平均を表 11 に示す。また、表 11 に示した特徴量を用いて、再度被験者ごとおよび問題ごとに leave one out 法で分類を行ったときの分類精度を表 12 に示す。いずれも分類器は Random Forest を用いた。なお、被験者ごとの分類精度を求めるにあたり、状態ラベルが 3 つ揃っていない被験者 2 名のデータは評価対象から除いた。同様に、状態ラベルが 3 つ揃っていない問題番号 2 のデータも評価対象から除いた。

表 12 より、被験者ごとに分類した場合には、10 名中 9 名において 5.2 節で得た 3 値分類の分類精度を上回った。一方で、問題ごとに分類した場合は被験者ごとに分類した場合よりも分類精度の向上は見られなかった。また、難易度ごとの分類においても、難易度 2 の分類精度が 10%程度低いものの明確な差は無かった。以上より、本研究で使用した特徴量は被験者間の差異を吸収できていない可能性があることが分かった。

6. まとめ

本稿では、解答中の眼球運動データを時間的に区切りその間の解答者の状態をラベル付けすることで、解答戦略を持って解いているかどうかの判定を行った。

大学生 14 人、幾何問題 10 問の解答データから特徴量を抽出し、RFE による特徴量選択を行った。その後、Random Forest および XGBoost を用いて「明確な解法に基づく解答」「積極的な解法の探索」「非積極的な解法の探索」の 3 つのラベルについて 3 値分類を行った結果、XGBoost を使用した際に 53.6%の正解率を得た。加えて、「明確な解法に基づく解答」を解答戦略有り、「積極的な解法の探索」と「非積極的な解法の探索」を解答戦略無しとして、解答戦略の有無の判定である 2 値分類を行ったところ 70.2%の正解率を得た。また、被験者ごとおよび問題ごとに分類を行った結果、被験者ごとに分類を行った場合により高い正解率を得たことから、今回使用した特徴量は被験者ごとの差異を吸収できていない可能性がある。一方、解答戦略の分類正解率が問題難易度に依存するかについて検証した結果、問題難易度が高い方が解答戦略の分類正解率が上がる傾向が見られたものの、問題数が少ないことから

表 11 被験者・問題・難易度ごとに分類した場合の特徴量の重要度

被験者番号/問題番号/問題難易度	被験者番号										問題番号										難易度		
	2	3	4	5	6	7	8	10	11	12	1	3	4	5	6	7	8	9	10	2	3	4	
特徴量の選択数	8	6	2	2	10	9	3	2	10	1	3	3	3	4	10	7	7	8	6	6	3	6	
視線移動量	平均		✓			✓	✓			✓				✓			✓	✓	✓				
	標準偏差	✓				✓	✓			✓									✓	✓	✓		
	最大	✓						✓						✓								✓	
瞳孔サイズ	平均								✓				✓					✓					
	標準偏差		✓		✓					✓	✓						✓					✓	
	最大			✓		✓				✓								✓			✓		
瞳孔サイズ変化割合	平均		✓			✓							✓	✓	✓	✓	✓				✓		
	標準偏差						✓										✓	✓				✓	
	最大						✓						✓				✓		✓			✓	
左右瞳孔サイズ差	平均	✓	✓	✓			✓			✓					✓		✓	✓		✓			
	標準偏差		✓										✓				✓						
	最大				✓	✓											✓	✓					
固視	回数	✓					✓			✓		✓					✓						
	継続時間 平均	✓				✓	✓			✓				✓	✓	✓	✓			✓	✓		
	標準偏差	✓					✓			✓		✓			✓	✓		✓	✓	✓	✓		
まばたき	回数		✓			✓	✓			✓		✓			✓	✓					✓		
	継続時間 平均					✓			✓		✓			✓									
	標準偏差	✓						✓						✓	✓	✓	✓	✓			✓		
最大					✓		✓								✓								
分類精度の平均 (%)	77	53	43	68	55	61	17	60	86	60	37	47	73	70	45	63	75	49	34	51	60	58	

(注) 特徴量の重要度 小 大
 選択された特徴量 ✓

表 12 被験者・問題・難易度ごとの分類精度

被験者番号/問題番号/問題難易度	被験者番号										問題番号										難易度		
	2	3	4	5	6	7	8	10	11	12	1	3	4	5	6	7	8	9	10	2	3	4	
分類精度 (%)	79	73	64	68	63	64	33	58	84	88	44	50	64	60	45	65	72	57	44	52	63	60	

十分な検証には至っていない。しかし、被験者ごとの差異よりも問題ごとの差異の方が小さいことを確認した。

今後の課題としては、新しい特徴量の検討が挙げられる。例えば、筆記速度や筆記圧力といった手書きストロークから得られる特徴量や、画面上のしている関心部分を表す AOI (Area of Interest) などである。また、ラベル付けが正確に行われているか不明瞭な点も多く、被験者に正確にラベル付けを行ってもらえるように誘導を行う方法も検討すべきである。

謝 辞

本研究の一部は科学研究費補助金(20H04292)によるものである。

参 考 文 献

[1] H. Asai, H. Yamana, "Detecting Student Frustration based on Handwriting Behavior," Proceedings of the ACM Symposium on User Interface Software and Technology (UIST), pp.77-78, 2013.

[2] M. Shojaeizadeha, S. Djamasi, R. C. Paffenroth, A. C. Trapp, "Detecting task demand via an eye tracking machine learning system," Decision Support Systems Vol.116, pp.91-101, 2019.

[3] J. Madsen, S. U. Julio, P. J. Gucik, R. Steinberg, L. C. Parra, "Synchronized eye movements predict test scores in online video education," Proceedings of the National Academy of Sciences, Vol.118, No.5, pp.1-9, 2021.

[4] O. V. komogortsev, D. V. Gobert, S. Jayarathna, D. H. Kkoh, S. M. Gowda, "Standardization of Automated Analyses of Oculomotor Fixation and Saccadic Behaviors," IEEE Transactions on Biomedical Engineering, Vol.57, No.11, pp.2635 - 2645, 2010.

[5] D. D. Salvucci, J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," Proceedings of the 2000 symposium on Eye tracking research & applications, pp.71-78, 2000.

[6] T. A. Salthouse, C. L. Ellis, "Determinants of Eye-Fixation Duration," The American Journal of Psychology, Vol.93, No.2, pp.207-234, 1980.

[7] M. E. Kret, E. E. Sjak-Shie, "Preprocessing pupil size data: Guidelines and code," Behavior Research Methods, Vol.51, No.3, pp.1336-1342, 2019.