

幾何証明問題の解答に対する自動部分点付与の検討

岡田 一洸[†] 田中 聖也[‡] 工藤 雅士[‡] 三浦 将人[‡] 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 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: [†] [‡] [§] {k-okada, seiya, kudoma34, masato, yamana}@yama.info.waseda.ac.jp

あらまし 近年、次世代に必要な能力や資質として論理的思考力の育成が重要視されている。論理的思考力を評価するためには、記述式の問題が有効である。しかしながら、記述式の問題の採点は、ある部分点に対して様々なパターンの表記方法を考慮する必要があるため、採点の自動化が難しいという問題がある。そこで本研究では、学習者の「物事の因果関係を整理し、正解までの道筋を立てて考える能力」である論理的思考力を評価するために中等教育レベルの幾何の証明問題を対象として、解答に部分点を自動的に付与することで採点者を支援するシステムの構築を目指す。証明問題の解答に対して自動的に部分点を付与する初期段階の研究として、データ収集によって得られた図形への書き込みに関するオンライン手書きデータの情報と証明の記述のデータを基に、「完答することのできた解答」か「部分点を付与すべき解答」かの自動分類を行った。ケアレスミスをした解答を、本来であれば完答することのできた解答であると仮定し、「完答することのできた解答」か「部分点を付与すべき解答」かの自動分類を行い、0.833 の正解率を得た。

キーワード 論理的思考力、オンライン手書きデータ、e-Learning、証明問題

1. はじめに

近年、論理的思考力は、文部科学省が掲げている「生きる力」¹の構成要素のひとつとなっており、初等中等教育においても論理的思考力の育成が重要視されている。また、日本が目指すべき未来社会の姿である Society5.0 に必要なリテラシーのひとつとしても論理的思考力は着目されており、企業が学生に求める資質や能力としても重要視されている²。

論理的思考力とは、物事の因果関係を整理し、正解までの道筋を立てて考える能力である。中学校学習指導要領³によると、幾何学について「図形の性質や関係を論理的に考察し表現する力を養う」とあるように、幾何学は論理的思考力の育成を目的としている。論理的思考力は、未来社会に求められる能力であり、学習者がどの程度の論理的思考力を有しているかを測る必要がある。論理的思考力を数学で測るためには、解答過程を評価する必要がある。数学の解答を評価するためには、「正答」と「誤答」の2値の採点ではなく、部分的に解答が間違っている、正しい部分の解答に対して部分点の付与を行う必要がある。幾何証明問題のような記述式の問題の採点は、正解の基準となる部分点に対して、様々な表記方法を考慮する必要があるため、採点の自動化を行うことは困難である。記述式の問題は未だに手作業で採点されている場合がほとんど

であり、採点者の負担となっている。

一方、近年では、教育分野に ICT を活用することによって学習に関する大量のオンラインデータを取得することが可能となった。学習時に取得できるオンラインデータの例としては、タブレット端末と電子ペンを用いて取得可能なオンライン手書きデータがあり、オフライン手書きデータでは取得することのできないストロークごとの時系列データの取得を可能としている。

そこで、本研究では論理的思考力を問う問題として、与えられた仮定を基に、結論までの道筋を立てて表現する必要がある幾何証明問題を対象に、ICT を活用し、オンライン上で解答データを取得し、解答に対して部分点の付与の自動化を目指し、学習者が間違えた箇所や間違えた理由を明らかにすることを検討する。本研究では、学習者が間違えた箇所やなぜ間違えたのかを明らかにする初期段階の研究として、「正しい解答」と「誤った解答」の2値の分類だけでなく、「部分点の付与の必要がある解答」を含めた3値で解答の分類を行う。データ収集によって得た図形への書き込みに関するオンライン手書きデータと証明の記述のテキストデータを取得し、部分点を考慮した解答の分類を行い、内容が似ている解答ごとにまとめることで、採点者の負担軽減を目指す。

数学における手書きの解答データを用いた研究は、解答過程から抽出した特徴量を基に、クラスタリングし、グループ学習時の目的に応じて、能力が似ている生徒同士のグループや能力が異なる生徒同士のグルー

¹ https://www.mext.go.jp/a_menu/shotou/new-cs/index.htm

² https://www.keidanren.or.jp/policy/2018/113_honbun.html

³ https://www.mext.go.jp/content/1413522_002.pdf

プを自動的に形成する研究[1], 学習者が「勘で解いたか」を判定する研究[2]やヒント参照有無の判定を行った研究[3], 幾何証明問題において手書きの図形と手書きの証明の記述を認識する研究[4]などがある. また, 数学の自由回答形式の自動採点に関する研究は, 自然言語処理と機械学習を組み合わせる成績の予測を行った研究[5]などがある. しかし, 我々が調べた限り, 論理的思考力の評価及び自動採点や採点者支援をすることを目的として, 幾何証明問題のオンライン解答データを用いて正しい解答の評価だけでなく, 部分点を付与するべき誤答についても学習者の解答を評価し, 学習者が間違えた部分や間違えた理由を明らかにした研究は存在しなかった.

本稿では次の構成をとる. 2 節で関連研究について述べ, 3 節でデータ収集実験について説明し, 4 節で分類実験及び考察を行い, 5 節でまとめる.

2. 関連研究

本節では, 手書きデータを使用した研究や数学の自動採点に関する研究について述べる.

2.1. 数学の手書きデータを用いた研究

吉武ら[1]は, 二次方程式に関する問題のオンライン手書き解答データにおける解答時間及び数式記入時のストローク数, ストロークの時間間隔, 実際に描かれたストロークに対する消しゴムの使用率等を特徴量として抽出し, k-means 法によりクラスタリングを行い, グループ学習時の目的に応じて, 能力が似ている生徒同士のグループや能力が異なる生徒同士のグループを自動的に形成する手法を提案した.

森山ら[2]は, 学習者の論理的思考力の測定の初期段階の研究として, 幾何学図形問題を対象に被験者が「勘で解いたか否か」を分類する手法を提案した. 幾何問題の解答中に書かれたストロークの描画時間, 筆圧, 速度, 座標を用いて特徴量を抽出し SVM を適用することで, 被験者が「勘で解いたか否か」の分類を行い, 0.83 の正解率で分類ができたことを示した.

三浦ら[3]は, 幾何図形問題のオンライン手書き解答データを用いて学習者が「ヒントを見たか」の分類をする手法の提案をした. 従来手法[6]では, 手書き解答時間や解答時に書き込まれた全体のストロークデータの特徴量として抽出していたが, 三浦らは全体の解答時間に対する書き込みが停滞していた割合や幾何図中に限定したストロークデータの特徴量として追加することで F 値が向上することを示した.

幾何の証明問題の手書き解答を用いた例として Jiang らの研究[4]がある. Jiang らは, 手書きの図形と手書きの証明の記述を区別して認識し, 図形の幾何学的な構成要素と証明ステップを対応させる電子ペンを

利用した幾何学定理の証明を支援するシステムを提案した. Jiang らのシステムは, 動的なヒントの提供を可能とし, 図形に書き込まれたストロークと証明の記述を独立して認識するのではなく, ストロークと証明の記述を対応付けて構造化することで, 学習者が図形の幾何学的な構成要素と証明の記述を効率的に入力や消去できるようにした. 手書き図形の認識手法としては, 各ストロークを個別にサンプリングし, ターニングポイントでサブストロークに分割する. なおストローク上の点を P_i , ストロークの両端の点を P_{i-1} , P_{i+1} とすると, 角度 $\angle P_{i-1} P_i P_{i+1}$ が閾値より小さい場合に, 点 P_i をターニングポイントとした. そしてサブストロークがそれぞれ点, 線, 円であるかを判定し, すべてのストロークの判定結果を組み合わせ, 手書き図形の認識を行った.

2.2. 数学の自動採点に関する研究

Erickson[5]らは, 数式と数式以外のテキストの両方を対象とした Bag-of-Words のアプローチを用いたランダムフォレストや XGBoost のツリーベースの機械学習手法と, 自然言語処理と LSTM の深層学習を組み合わせ, 数学の記述式問題の自動採点を行う手法を提案した. データセットとなる解答は, 5 点を満点として教師によって 0 点~5 点のいずれかの整数の点数がつけられている. 学生の記述式問題の解答に対して教師がつけた点数を予測し, 分類した. 各手法により得られる共変量を用いた Rasch モデルで各モデルの性能比較を行った. 提案手法の中ではランダムフォレストを用いたモデルの性能が最もよく, 0.850 の AUC であった. Bag-of-Words のアプローチは, 解答内の構造や単語間の共起性は考慮されていないことに問題がある点やテキストから特徴を抽出する際, 数学用語を含む可能性が高いと仮定したもと, Wikipedia データセットで事前学習された 100 次元の GloVe ベクトルを用いたが, 学生が書いた関数や公式の多くが事前学習された単語埋め込みでは表現することができないことに問題があると示した.

3. データ収集

本節では, 解答の自動分類にて扱う幾何証明問題のオンライン手書きデータ及び論述データを収集するための実験方法について述べる.

3.1. データ収集用アプリケーション

html 及び JavaScript を用いて幾何証明問題の解答データをオンライン上で収集できる Web アプリケーションを開発し, 解答データの収集を行った. 解答の入力には iPad Pro と Apple Pencil を用いた. 作成した Web アプリケーションの解答画面を図 1 に示す.

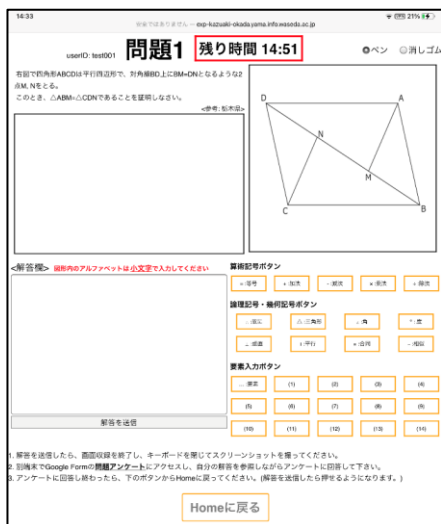


図 1 解答画面

画面上部には、被験者を識別する番号、問題番号、解答の残り時間、手書きの際に「ペン」と「消しゴム」を切り替えるためのボタンの表示がある。「ペン」と「消しゴム」の切り替えについては、「ペン」が選択されている場合は、線を描画することができ、「消しゴム」が選択されている場合は、実際の消しゴムを使用するように画面上に描画した線を消すことができる。解答者は、問題文下部の黒枠内と図形周囲の黒枠内を Apple Pencil を用いて、「ペン」や「消しゴム」機能を使用し、自由に描画することができる。問題文下部の黒枠はメモ欄である。図形に書き込まれた線は、ストロークごとに情報がサーバに送信される。証明問題の解答となる論述については iPad 上のキーボードを使用し、解答欄内に記述してもらい、証明問題の解答となる論述のテキストデータの取得を行った。解答欄右側に配置されているボタンを使用することで、論述する際に使用する記号の入力を行うことができる。論述入力時、キーボードのキーや記号入力ボタンが押下されるごとに、押下されたキーの情報、押下された記号入力ボタンの内容と押下時刻の情報がサーバに送信される。「解答を送信」ボタンを押下することで、解答終了時刻と論述内容がサーバに送信される。

3.2. 収集データの内容と形式

本アプリケーションでは、1つの問題に対して3種類のJSONファイルと1つの幾何証明問題の論述データのテキストファイルが出力される。

1つ目のJSONファイルは、解答時間に関する情報が出力され、以下の2項目の情報を持つ。

- ・ 解答開始時刻
- ・ 解答終了時刻

2つ目のJSONファイルは、図形への書き込みのスト

ローク情報が出力され、以下の6項目の情報を持つ。ストロークとは、Apple Pencil が画面に触れてから離れるまでの一筆のことを指す。

- ・ 1ストローク内の点の順番
- ・ タッチイベント
 1. タッチ開始
 2. タッチ中
- ・ x座標
- ・ y座標
- ・ 描画された時刻
- ・ 筆圧

3つ目のJSONファイルは、解答記述時に入力された文字の情報が出力され、以下の5項目の情報を持つ。

- ・ テキストエリア内のカーソルの位置
- ・ テキストエリアに入力されている総文字数
- ・ 文字や記号が入力された時刻
- ・ 入力された文字
- ・ 入力された文字のキーコード

3.3. 実験に使用した問題

本研究では、三角形の合同や相似を示す問題を5題用意した。問題1、問題2、問題4については、問題文に辺や角度の関係が明示されている問題を作成した。問題3、問題5は、主題を証明するために三角形の合同や相似に自ら気づき、証明する必要がある問題を作成した。実験に使用した問題の詳細は、付録に示す。

3.4. データ収集実験の実施

幾何証明問題のオンライン解答データを収集するために、20名の学生(男性13名、女性7名、大学1年から修士2年)を対象に実験を行った。5題の幾何証明問題の解答してもらった。

実験前に実験マニュアルを確認してもらった。マニュアルに記載されている事項を以下に示す。

1. 本実験は、5題の幾何証明問題に解答してもらうものである。
2. 問題の制限時間は、各問題の難易度によって異なる。問題1は15分、問題2と問題4は30分、問題3と問題5は45分である。ただし、問題が解き終わった時点で解答を終了してよい。
3. 図形への書き込みやメモはApple Pencilを用いて行い、論述部分についてはiPad上のキーボードを使用して解答を行う。
4. 画面上の「ペン」を選択すると図形へ線を書き込むことやメモ欄に文字を書くことができ、「消しゴム」を選択すると書いた線を消すことができる。
5. 記号を入力する場合は、記号入力ボタンを使用すること。

3.5. データ収集実験の結果

3.5.1. 完答している解答の条件

問題文に合同や相似を示すべき三角形が明示されている問題である問題 1, 問題 2, 問題 4 においては表 1 に示す 2 項目を満たした解答を「完答している解答」と定義した.

表 1 問題 1, 問題 2, 問題 4 における「完答している解答の条件」

項目	条件
1	合同条件や相似条件を示すための辺や角を全て仮定から導出することができている.
2	合同条件や相似条件が記述できている.

主題を証明するためには, 三角形の合同を証明する必要があることに自ら気づく必要がある問題である問題 3, 問題 5 においては表 2 に示す 3 項目を満たすものを「完答している解答」の条件とした. 主題を証明するために, 第一段階として三角形の合同を証明する必要がある部分を「補題」として定義する. すなわち, 問題 3 と問題 5 は, 補題を証明したうえで主題の証明を行う必要がある問題である.

表 2 問題 3, 問題 5 における「完答している解答」の条件

項目	条件
1	補題の合同条件を示すための辺や角を全て仮定から導出することができている.
2	補題の合同条件や相似条件が記述できている.
3	補題の証明ができており, 主題を証明するために必要な辺や角が全て導出できている.

「完答している解答」の条件の一部の項目を満たしていない解答を「部分点を付与すべき解答」と定義し, 全ての項目を満たしていない解答を「完全に誤った解答」と定義する.

3.5.2. 解答内容の分布と誤答の内容

各問題における「完答した解答数」, 「部分点を付与すべき解答数」, 「完全に誤った解答数」を表 3 に示す. 20 名の被験者を対象にデータ収集を行ったが, VPN の接続不良や問題解答中にブラウザをリロードしてしまったといった理由により, データの取得を行うことができなかった解答があった.

表 3 解答内容の分布

	完答数	部分点を付与すべき解答数	完全な誤答数
問題 1	15	4	0
問題 2	10	8	0
問題 3	9	8	2
問題 4	12	6	0
問題 5	8	6	1

問題 1 から問題 5 における誤答箇所のパターンをそれぞれ表 4,

表 5, 表 6, 表 7, 表 8 に示す.

表 4 問題 1 における誤答箇所のパターン

内容	解答数
$\angle ABD = \angle CDB$ を示す際に, 「 $AB \parallel CD$ 」であることの記述を忘れてしまった.	2
「 $BM = DN$ 」のところを「 $BM = BN$ 」としてしまった.	1
合同条件を「2 組の角とその間の角」と記述してしまった.	1

表 5 問題 2 における誤答箇所のパターン

内容	解答数
合同条件の要素のひとつである $\angle EAB = \angle FAM$ の導出ができなかった.	5
$\triangle ABE \equiv \triangle AMF$ を $\triangle ABE = \triangle AMF$ と記述してしまった.	1
「錯角」という文言を記述し忘れた.	1
対応する頂点の順番を間違えた.	1

表 6 問題 3 における誤答箇所のパターン

内容	解答数
直角二等辺三角形の直角の部分の証明方法が思いつかなかった.	3
補題の合同条件の記載をしなかった.	1
$\angle QRC = \angle PRA$ を $\angle QRC = \angle PRC$ と書いてしまった.	1
$\angle PRQ = 90^\circ$ と書いたつもりであったが, $\angle RQP = 90^\circ$ と書いてしまった.	1
$RA = RC$ を示す際に $\angle RAQ = \angle RCQ$ であることを述べなかった.	1
$\angle RAP = \angle RCQ$ を示す際に, それぞれの角が等しい理由を述べなかった.	1

表 7 問題 4 における誤答箇所のパターン

内容	解答数
$\angle GFE = \angle GHD$ の導出ができなかった.	2
3 組の角が等しいことを述べてしまった.	2
$\triangle GFE \sim \triangle GDH$ を $\triangle GFE \equiv \triangle GDH$ と書き間違えてしまった.	1
$\angle GFE = \angle GDH$ を $\angle GFE = \angle GDA$ と書き間違えてしまった.	1

表 8 問題 5 における誤答箇所のパターン

内容	解答数
△EDM 及び△FDM が直角二等辺三角形であることを示す際の根拠を省略した.	3
△EDM 及び△FDM が直角二等辺三角形であることを証明できなかった.	1
「(6), (9)より」と書くべき箇所を「(6), (6)より」と書いてしまった.	1
錯角と記述するのを忘れてしまった.	1

4. 解答の分類方法と評価

本節では、収集した幾何証明問題の解答データの分類手法と評価結果について述べる.

4.1. 解答の分類手法

データ収集実験によって得ることのできた幾何証明問題のオンライン解答データから特徴量を抽出し、線形 SVM 及び XGBoost を用いて解答の分類を試みる. 線形 SVM と XGBoost はそれぞれ Python のライブラリである scikit-learn⁴と xgboost⁵を用いて実装を行い、線形 SVM 及び XGBoost のハイパーパラメータは、デフォルトの値で学習を行う. 収集した解答のうち、全問題を通じて「完全に誤った解答」は 3 件のみであり、目的変数の分布に偏りが生じ、「完全に誤った解答」を含めた 3 値での分類は困難であったため、本分類では「完答することのできた解答」、「部分点を付与すべき解答」の 2 値で分類を行うこととした. 分類実験は、全問題を用いた分類実験と問題毎の分類実験を行う. 全問題を用いた分類実験においては、問題の内容に依存しない特徴量を抽出することで、様々なパターンの問題に対応することのできる学習器の構築を行う. 問題毎の分類実験においては、問題の内容に依存する特徴量を追加し、解答の分類を行う. 図形への書き込みを行っていなかった解答と VPN の接続不良等の理由でデータの取得が行えていなかった解答はラベル付けから除外をした.

4.2. 全問題での分類実験

4.2.1. 使用する特徴量

三浦らの研究[3]にて使用された特徴量を参考に、問題の内容に依存しないと考えられる特徴量の抽出を行った. 抽出した特徴量を表 9 に示す.

また、Python の機械学習ライブラリの scikit-learn のモジュールである MinMaxScaler を用いて各特徴量の正規化を行った. f_i を特徴量 F の i 番目の特徴量として式 1 を用いて正規化を行い、各特徴量の値を 0~1 の範囲で表現した.

表 9 問題に依存しないと考えられる特徴量

特徴量	定義
描画速度	平均 最大 最小 標準偏差 解答画面で「ペン」選択時の図形への書き込みに関するストロークの描画速度
消去速度	平均 最大 最小 標準偏差 解答画面で「消しゴム」選択時の図形に書き込んだ線の消去速度
筆圧	平均 最大 標準偏差 「ペン」選択時の図形への書き込みに関する Apple Pencil から取得したタッチ圧力
消しゴム使用時間割合	図形に入力を行っていた時間に対する図形に描いた線を消していた時間の割合

$$f = \frac{f_i - \min(F)}{\max(F) - \min(F)} \quad (1)$$

4.2.2. 実験結果

表 9 に示した特徴量を使用し、「完答することのできた解答」と「部分点を付与すべき解答」に分類した場合の各モデルの正解率を表 10 に示す. leave-one-out 交差検証を行い、正解率を算出した. RFE(Recursive Feature Elimination)を用いて各特徴量の重要度を算出し、最も正解率が高くなる特徴量の組み合わせとなる特徴量の選択を行った. 特徴量の順位を表 11 に示す.

表 10 各モデルの正解率

	線形 SVM	XGBoost
特徴量選択前	0.571	0.442
特徴量選択後	0.610	0.494

表 11 特徴量の順位

順位	線形 SVM	XGBoost
1	消去速度_最小	消去速度_平均
2	筆圧_最大	消去速度_最大
3	消しゴム使用時間割合	筆圧_最小
4	消去速度_標準偏差	描画速度_最小
5	筆圧_最小	描画速度_平均
6	描画速度_最小	描画速度_標準偏差
7	消去速度_最大	消去速度_標準偏差
8	描画速度_最大	消去速度_最小
9	描画速度_標準偏差	筆圧_最大
10	消去速度_平均	描画速度_最大
11	筆圧_平均	消しゴム使用時間割合
12	描画速度_平均	筆圧_平均

RFE による特徴量選択を行った結果、線形 SVM における正解率が最も高くなる特徴量の組み合わせは、上位 1 個の特徴量を使用したときであり、正解率は 0.610 であった. XGBoost における正解率が最も高く

⁴ <https://scikit-learn.org/stable/modules/svm.html>

⁵ <https://xgboost.readthedocs.io/en/latest/python/index.html>

なる特徴量の組み合わせは、上位 6 個の特徴量を選択したときであり、正解率は、0.494 であった。特徴量選択前と特徴量選択後と比較すると、線形 SVM を用いた正解率は 0.039 高くなり、XGBoost を用いた正解率は 0.052 高くなった。しかし、表 10 に示した結果のように、特徴量選択の有無や分類に使用した学習器に関わらず、得られた正解率自体は低かった。問題の内容に依存しないと考えられる特徴量を抽出したが、分類ができていないことが分かる。

4.3. 問題毎の分類実験

4.3.1. 使用する特徴量

三浦らの研究[3]にて使用された特徴量を参考に、問題に依存すると考えられる特徴量を抽出した。本実験では、「完答した解答」と「部分点を付与すべき解答」の思考過程の違いや傾向による解答の分類を試み、思考過程が表れる図形への書き込みから特徴量を抽出し、テキストの内容を特徴量として用いずに分類実験を行った。抽出した特徴量を表 12 に示す。表 9 に示した問題に依存しないと考えられる特徴量と表 12 に示した問題に依存すると考えられる特徴量を使用し、問題ごとに「完答することができた解答」と「部分点を付与すべき解答」の分類を行った。7 人の被験者が図形への書き込みを行っていない問題 1 は、分類の対象とする問題から除外した。

表 12 問題に依存すると考えられる特徴量

特徴量		定義
経過時間		解答画面に問題が表示された時間から解答を送信するまでに経過した時間
端点移動距離	平均	図形に書き込まれた n 番目のストロークの終点と n+1 番目のストロークの始点の距離
	最大標準偏差	
ストローク長	最大	図形に書き込まれたストロークの長さ
	最小	
ストローク描画本数		解答画面で「ペン」モード選択時にストロークが図形に描画された本数
最終文字数		解答送信時に記述されていた文字数
等式出現回数		解答送信時に記述されていた「＝」で結ばれた数式の数

4.3.2. 実験結果

表 9 と表 12 に示した特徴量を合わせた全ての特徴量を使用し、線形 SVM 及び XGBoost の学習モデルの構築を行った。問題ごとに leave-one-out 交差検証で検証を行う。問題 2 から問題 5 における各モデルの正解率を表 13 に示す。問題 5 は、線形 SVM も XGBoost も同じ正解率であったが、問題 2, 3, 4 については XGBoost に比べ、線形 SVM の方が正解率は高かった。

しかしながら、いずれの結果も正解率自体は低く、分類はできていないと考えられる。

表 13 モデルごとの正解率

	線形 SVM	XGBoost
問題 2	0.722	0.556
問題 3	0.500	0.438
問題 4	0.647	0.647
問題 5	0.571	0.571

RFE (Recursive Feature Elimination)を用いて各特徴量の重要度を算出し、最も正解率が高くなる特徴量の組み合わせとなる特徴量の選択を行った。問題 2 から問題 5 の特徴量選択後の結果を表 14 に示す。

表 14 特徴量選択後の正解率

	線形 SVM	XGBoost
問題 2	0.722	0.889
問題 3	0.563	0.563
問題 4	0.647	0.706
問題 5	0.786	0.571

XGBoost を用いた問題 2 の分類を除き、特徴量選択後の結果も正解率自体は低く、分類はできていないと考えられる。問題 2 の XGBoost を用いた解答の分類において RFE による特徴量選択を行った結果、上位 1 件の特徴量を選択した際の正解率が最も高くなり、特徴量選択前と比べ、正解率は 0.333 向上した。特徴量選択によって XGBoost では 0.889 の正解率を得たが、「描画速度_最大」の特徴量のみで分類を行った場合が最も高い正解率となっており、データを増やした場合に、分類精度が下がる可能性が高い。

4.4. 問題毎の分類実験(ケアレスミス「完答した解答」としたときの分類)

問題 2 において分類を行うことができなかった原因として、誤答データのラベル付けを行う際に、「BM=DN のところを BM=BN としてしまった」という間違いや「錯角であることの記述を忘れてしまった」というようなケアレスミスによる間違いと「合同条件の要素のひとつである $\angle EAB = \angle FAM$ の導出ができなかった」というような導出過程が分かっていたことによる間違いを「部分点を付与すべき解答」として同一のラベル付けを行ってしまったことに原因があるのではないかと考えた。ケアレスミスは注意していれば防ぐことのできた間違いであり、論理は正しい解答である。そこで表 5 に示した問題 2 の間違いにおける、「 $\triangle ABE \equiv \triangle AMF$ を $\triangle ABE = \triangle AMF$ と記述してしまった。」、「“錯角” という文言を記述し忘れた。」、「対応する頂点の順番を間違えた。」といった間違いを、「完答した解答」

として扱い、「合同条件の要素のひとつである $\angle EAB = \angle FAM$ の導出ができなかった。」と解答した5つの解答を「部分点を付与すべき解答」としてラベル付けを再度行い、「完答した解答」と「部分点を付与すべき解答」の分類を行った。RFEによる特徴量選択前後のSVMとXGBoostの分類結果を表15に示す。SVMは、上位8件の特徴量が選択され、XGBoostは上位2件の特徴量が選択された。上位8件までの特徴量の順位を表16に示す。

表 15 正解率

	線形 SVM	XGBoost
特徴量選択前	0.667	0.778
特徴量選択後	0.833	0.833

表 16 上位 8 位までの特徴量の順位

順位	線形 SVM	XGBoost
1	経過時間	最終文字数
2	筆圧_最小	描画速度_最大
3	端点移動距離_標準偏差	経過時間
4	等式出現回数	ストローク長_最大
5	消去速度_標準偏差	消去速度_標準偏差
6	最終文字数	等式出現回数
7	描画速度_標準偏差	消しゴム使用時間割合
8	描画速度_最大	筆圧_平均

表14に示した特徴量選択後の問題2のSVMの正解率よりも「ケアレスミスの解答」を「完答した解答」と扱った場合のSVMの正解率が0.111向上している。XGBoostを用いた分類においても正解率0.833を得ており、幾何証明問題における誤答の分類を行う場合、ケアレスミスをした解答を論理が不十分な解答と区別して分類を必要がある可能性が考えられる。

5. おわりに

本稿では、解答者の解答過程を評価するために、Web上で幾何証明問題のオンライン解答データを収集し、解答の部分正解の判定を行い、自動的分類する手法を提案した。本手法を用いて分類を行った結果、「完答できた解答」、「部分点を付与すべき解答」の2値分類における最も良い正解率は0.889であったが、RFEによって選択された特徴量は1つだけであったため、データを増やした場合に、分類精度が下がる可能性が高い。また、ケアレスミスをした解答を「完答できた解答」として扱い、「完答できた解答」か「部分点を付与すべき解答」かの分類を行ったところ、SVMとXGBoostの双方のモデルで0.833の正解率を得た。誤答を含めて解答の分類を行う場合、ケアレスミスによる間違いなのか、解法が分からないことによる間違いなのかを判定する必要があることが考えられる。

今後の課題としては、正解のデータの収集だけでなく、より多くの誤答のデータの収集を行うことや解答の部分正解の判定において有効な特徴量を特定する必要がある。また、幾何問題において解答者が間違えた箇所がケアレスミスによる間違いであるかどうかの自動判定を行うシステムの開発を行うことが挙げられる。

謝辞

本研究の一部は科学研究費補助金(20H0492)によるものである。

参考文献

- [1] D. Yoshitake, B. Flanagan, H. Ogata. “Supporting Group Learning Using Pen Stroke Data Analytics,” Proc. of the 28th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education, pp.1-6, 2020.
- [2] 森山優姫菜, 山名早人. “シーケンス情報を用いて幾何学の手書き解答データを解答パターンに分類する手法,” 早稲田大学大学院基幹理工学研究科情報理工・情報通信専攻修士論文, 2018.
- [3] 三浦将人, 村上統馬, 中山祐貴, 山名早人, “オンライン手書き解答データを用いた解答者のヒント参照有無の判定 ～ 幾何学問題を対象として ～”, 電子情報通信学会, 信学技法, vol.119, no.393, pp.27-32, 2020.
- [4] Y. Jiang, F. Tian, H. Wang, X. Zhang, X. Wang, G. Dai. “Intelligent understanding of handwritten geometry theorem proving,” Proc. of the 15th international conference on Intelligent user interfaces, pp.119-128. 2010.
- [5] J. A. Ericson, A. F. Botelho, S. McAteer, A. Varatharaj, N. T. Heffernan. “The Automated Grading of Student Open Responses in Mathematics,” Proc. of the 10th International Conference on Learning Analytics & Knowledge, pp.615-624, 2020.
- [6] 村上統馬, 山名早人. “オンライン手書きデータを用いた幾何学図形問題に対する答案の自動分類と解答者の理解度推定,” 早稲田大学基幹理工学部情報理工学科卒業論文, 2018

付録

データ収集実験に使用した問題を以下に示す。

右図で四角形 ABCD は平行四辺形で、対角線 BD 上に $BM=DN$ となるような 2 点 M, N をとる。
このとき、 $\triangle ABM \cong \triangle CDN$ であることを証明しなさい。

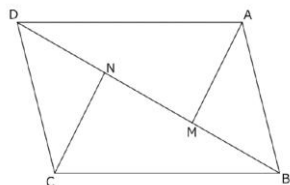


図 A 三角形の合同を証明する問題 (問題 1)

右図で $\triangle ABC$ は $AB=AC$ の二等辺三角形で、点 D は辺 BC 上にあり、 $AD=BD$ である。
頂点 A を通り、辺 BC に平行な直線上に $AB=AM$ となる点 M をとり、線分 BM が線分 AD, AC と交わる点をそれぞれ E, F とする。
このとき、 $\triangle ABE \cong \triangle AMF$ であることを証明しなさい。

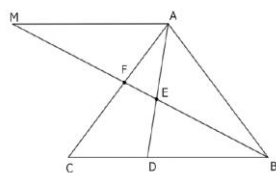


図 B 三角形の合同を証明する問題 (問題 2)

$\triangle ABC$ は $\angle BAC=90^\circ$ の直角二等辺三角形である。点 P は点 A, B いずれにも一致しない辺 AB 上、点 Q は点 A, C いずれにも一致しない辺 AC 上にあり、 $AP=CQ$ とする。 $\angle BAC$ の二等分線と辺 BC との交点を点 R とし、点 P と点 Q、点 P と点 R、点 Q と点 R をそれぞれ結んだ。
このとき $\triangle RQP$ は直角二等辺三角形であることを証明しなさい。

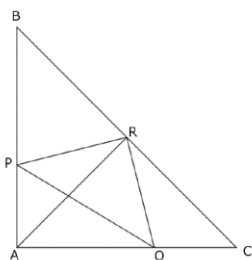


図 C 補題として三角形の合同を証明する必要がある問題 (問題 3)

四角形 ABCD は平行四辺形であり、 $\angle BAD$ の二等分線と BC の交点を E、 $\angle CDA$ の二等分線と BC の交点を F、 $\angle BAD$ の二等分線と $\angle CDA$ の二等分線の交点を G とし、 $\angle BAD$ の二等分線と DC を延長した線の交点を H とする。

このとき、 $\triangle GFE \sim \triangle GDH$ を証明しなさい。

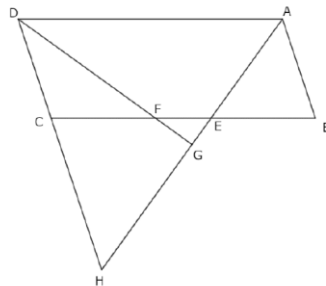


図 D 三角形の相似を証明する問題 (問題 4)

右図 $\triangle ABC$ において、辺 BC の中点を点 M とする。辺 AB, AC 上にそれぞれ点 D, E をとると $MD=ME$ 、 $\angle DME=90^\circ$ となった。点 B を通り辺 AC に平行な直線と直線 EM の交点を点 F とし、点 D を通り辺 AB に垂直な直線と、点 E を通り辺 AC に垂直な直線の交点を点 G とする。
このとき、 $\angle EDF=90^\circ$ であることを証明しなさい。

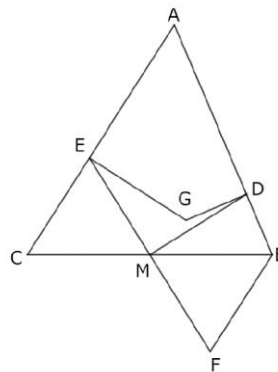


図 E 補題として三角形の合同を証明する必要がある問題 (問題 5)