

様 式 C - 19、F - 19 - 1、Z - 19（共通）

科学研究費助成事業 研究成果報告書



令和 5 年 6 月 13 日現在

機関番号：13801

研究種目：基盤研究(C)（一般）

研究期間：2020～2022

課題番号：20K03092

研究課題名（和文）オンライン教育の受講履歴から学生への総合支援を実現する総合教育システムの研究開発

研究課題名（英文）Research and Development of a Comprehensive Education System that Provides Comprehensive Support to Students Based on their Online Education Course History

研究代表者

永田 正樹（Nagata, Masaki）

静岡大学・情報基盤センター・准教授

研究者番号：30844782

交付決定額（研究期間全体）：（直接経費） 2,100,000 円

研究成果の概要（和文）：本研究課題は、インターネットを用いたオンライン教育に用いる動画教材に含まれる特徴の明確化や履修科目と就職先との関係、希望職種に対する適切な科目履修の提示などを調査および分析した。これら調査には機械・深層学習を用い複数のモデルを開発し、分析結果を定量化および可視化した。モデルから得られた結果をオンライン教育にフィードバックすることで、オンライン教育の品質および学生理解度向上に寄与できる。

研究成果の学術的意義や社会的意義

本研究課題は、オンライン教育で用いる動画教材の品質測定や、学生が将来希望する職種に対して適切な科目履修提案などを過去のデータを基に開発した各モデルにて定量化表示したことに意義がある。オンライン教育は今後も教育現場では必須の教育手法である。本モデルを利用することで、今後作成する動画教材に含めるべき要素の明確化や、学生らへの進路相談の具体的助言として職種に適した必要知識や技術提案などが可能となる。

研究成果の概要（英文）：In this research project, we investigated and analyzed the characteristics of video materials on online education using the Internet, the relationship between courses taken and job placement, and the presentation of appropriate courses for desired occupations. We developed several models using machine and deep learning to quantify and visualize the results of the analysis. By feeding back the results obtained from the models to online education, we can contribute to improving the quality of online education and the level of student understanding.

研究分野：情報ネットワーク、情報セキュリティ、分散処理

キーワード：オンライン教育 動画教材 特徴量 機械・深層学習 テキストマイニング

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

昨今の高度情報化に伴う社会変化に対応すべく学校教育も新たな手法が検討、確立されている。2020 年度小学校から順次導入された新学習指導要領（文部科学省 2017 年）では、情報活用能力を言語能力と同様に位置づけ、教育機関の ICT 環境整備と ICT を活用した学習活動の充実を重要点としている。具体的な手段として、アクティブ・ラーニングの推進やプログラミング教育、情報知識・技術を学ぶ科目「情報」の開設などがある。人生 100 年時代構想会議（内閣官房 2017 年）では、少子高齢化や永年雇用崩壊の実社会において教育および職能技術習得を目的としたリカレント教育を具体的なテーマの 1 つとしている。このような教育手法は ICT をベースとしており、インターネットを用いたオンライン教育の形式と親和性が高い。オンライン教育とは、インターネットや TV 放送などを介して講義コンテンツを配信し、受講者はそれぞれの端末で聴講および参加する教育形態である。

本研究課題はこのオンライン教育に着目した。本研究課題の申請書を提出した 2019 年当時はさまざまなオンライン教育支援システムが存在した。その多くが従前の学習管理システム（LMS：Learning Management System）に対して、動画配信機能を追加したものであった。このようなシステムはインターネットを用いて動画教材を配信することで、遠隔地の受講者への教材提供や受講時刻の制限がないなどのメリットは多数ある。一方で、動画教材そのものの品質や学生らの理解度把握などの効果測定を機能として保持しているものは少ない。そこで本研究ではこの点を課題とし、動画教材に含まれる特徴や受講履歴と就職先との関連分析など、総合的な教育支援システム実現を目指した。

2. 研究の目的

申請当時市場のオンライン教育支援システムは、LMS の一般的な機能である受講者および進捗状況管理、課題提出などの機能に動画配信を追加した形態が多数であった。これら機能はオンライン教育実行に関する必要機能は網羅しているが、学生の理解度測定や進路検討などの次フェーズの支援には至っていない。そこで本研究課題では、「オンライン教育を基とした理解度測定および受講科目と進路との関係」を研究課題の核心をなす学術的問いとして設定した。具体的には以下 3 点について検討した。

- (1) 理解度測定を可能とするオンライン教育支援システムへの機能追加
オンライン教育の理解度には動画教材の品質が関連すると考えられるため、動画の特徴に着目し、どのような特徴が教育に対して良効果をもたらすかを把握するための分析を実施した。関連して動画配信に関する視聴者動向や動画作成手法などについて調査した。「3. 研究の方法」の「(1) 動画特徴量抽出システム」に該当する。
- (2) 受講科目およびその成績と希望進路先との関係を調査
学生が卒業までに受講した複数科目と就職先の産業分類の関連分析を実施した。ある産業分類に就職した学生らの履修科目の傾向や、企業が人材に求める技術ニーズなどを複合的に分析した。「3. 研究の方法」の「(2) 受講科目およびその成績と希望進路先との関係分析」に該当する。
- (3) 希望進路からどのような科目を受講すればよいのかを学生へ指導
(2)の発展的内容として、受講科目と就職先産業分類との関係を基とした、希望進路に就くための適切な科目履修を分析した。また、産業分類に関連する知識や技術の文言と科目シラバスに含まれる文言との関連を分析し、産業分類と科目との関係を調査した。「3. 研究の方法」の「(3) 希望進路に即した適切な履修モデルの提供」に該当する。

3. 研究の方法

(1) 動画特徴量抽出システム〔①〕、〔②〕

理解度測定の 1 手法として動画教材に含まれる特徴を調査した。動画教材にはパワーポイントを使用したものや教員の黒板板書、化学実験風景などを録画・撮影したものが多く、これら動画中に現れる文字や図表、講師の仕草、発言などを動画教材の特徴量と捉え分析した。研究方法としてこれら特徴量を動画教材中から抽出する機械学習モデル「教育動画特徴量分析モデル（以下、「分析モデル」）」を開発した。分析の手順として、はじめに YouTube などの市場から動画教材として高品質・高人気動画（A）を抽出する。次にこれら動画に含まれる特徴量をラベルとして機械・深層学習をし、この特徴量を判別する分析モデルを開発する。次に分析したい動画（B）に対して分析モデルで特徴量を測定し、（A）と（B）の差を把握する。最後に、得られた特徴差を補った動画（C）を作成する。つまり、動画（C）の品質向上が目的となる。

本研究課題では動画教材に含まれる文字や人物などの物体、講師の発話・表情・仕草などを特徴量とした。特徴量の抽出は分析モデルに対して分析対象の動画教材を入力することで実施する。文字や人物の抽出では動画中のこれらの特徴量に対して図 1 のように、文字を緑色枠で「PC_character」、人物（男）をピンク色枠で「man」のように矩形枠表示およびラベリングする。講師発話では、動画中で講師が発した言葉をテキストデータ化し、それをテキストマイニングす

ることで分析した。手順は、まず分析の対象となる動画教材の発話内容のテキストデータを取得する。次にテキストデータを形態素解析し、単語に分割する。最後に分割した単語の出現頻度を特徴量として抽出する。出現頻度が多い単語はフォントサイズが大きくなる。つまりこれらの単語は動画教材で繰り返し用いて強調説明していると推測される。表情では、人の8つの基本表情である「笑顔」、「悲しみ」、「嫌悪」、「侮蔑」、「怒り」、「恐怖」、「驚き」、「中立」から「笑顔」と「驚き」の表情を抽出した。この理由は先行研究から、学生からの評価が高い教員には「笑顔」と「驚き」の表情が多く見られたためである。一方、低評価の教員には「怒り」や「嫌悪」が多くみられる特徴があった。このことから講師表情の「笑顔」、「驚き」が多い動画教材は高品質教材と考えられる。仕草では、動画教材中の講師の手の動きに着目した。黒板やホワイトボードに講師の手首が停止しているときは指している箇所を説明しているときで、ここは動画中で重要な説明をしている箇所といえる。このため、動画中の講師手首が静止している区間を抽出するプログラムを開発した。手首停止を判断するために講師の首と左右手首の3部位の座標を用いた。これら3部位の位置変化を時間経過により算出し、停止区間を判断した。図2は講師手首の停止区間をグラフ化したものとプログラムで抽出した動画のキャプチャ画像である。

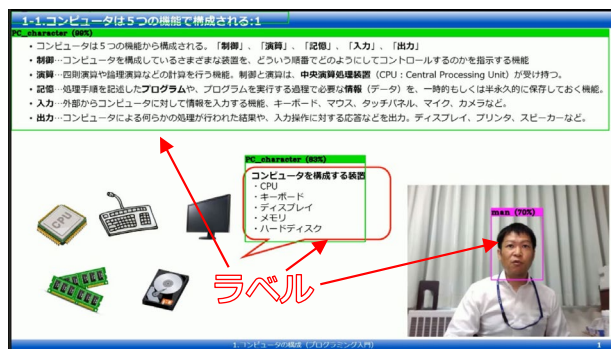


図1 特徴量の矩形表示、ラベリング

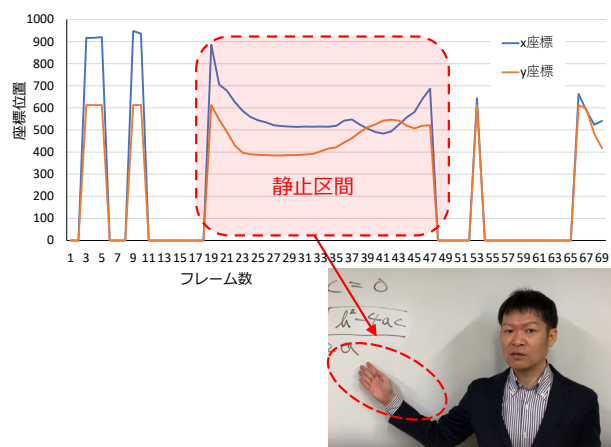


図2 手首の静止区間

(2) 受講科目およびその成績と希望進路先との関係分析[③]

学生らの履修科目や就職先のデータを基にこれらの関連を示す「職種判定モデル（以下、判定モデル）」を開発した。用いた学生データは研究代表者が所属する大学であり、データには履修科目、成績、GPA、就職先、就職先産業分類などが含まれる。平成24年から令和2年までの8年間のデータを用い、個人が特定できないよう事前処理した。判定モデルは過去8年間の学生データを教師データに用い機械学習することで作成した。判定モデルは1年次から4年次の履修科目を入力することで、ある職種に適する確率を示すものである。通常、就職する職種と履修科目には関連がある。たとえば、農業・林業に就職した学生は農学関連の科目を多く履修し、情報通信業に就職した学生は情報学関連の科目を多く履修する。つまり、学生らの履修科目および就職先の産業分類を分析することで、ある産業分類に就職する学生の履修科目の傾向が分かる。このことから判定モデルの機械学習は過去の学生らの履修データと就職データを教師データとして用いた。履修データは全学性が履修する必修科目を除いた選択科目45個を説明変数とし、就職データとして就職産業分類を1個の目的変数とした。45個の履修データは卒業までに履修する選択科目の総数であり、実際の履修傾向に即するよう1、2年次に15科目、3年次に10科目、4年次に5科目とした。内訳は、10個は全就職産業分類での共通科目とし、25個は各就職産業分類特有の科目、残り10個は完全ランダムとした。たとえば、説明変数の科目には「理論と情報特論」、「情報処理」、「情報ネットワーク社会」、「ソフトウェア工学」、「プログラミング応用」などであり、目的変数の就職産業分類は「情報通信業」である。このような教師データを多数用意し、機械学習を実施した。学習後に出力されたモデルに対して、複数の科目を入力すると、それら科目と各産業分類との関連度を示す確率値が算出される。つまり、この確率値は履修した科目群に対してどの産業分類に就職しやすいかの目安となる。なお、科目や就職産業分類のデータは文字列のためそのままでは学習過程が複雑となる。このため、図3に示すようにこれら文字列を数値に変換し、学習効率の向上を図った。

(3) 希望進路に即した適切な履修モデルの提供[④]

ある産業分類に就職を希望する際にどのような科目を履修すればよいのかを明確にするために、公開されている大学の科目シラバスのテキストデータとWeb上で得られるテキストデータを収集し、それらの関係を分析した。分析手法には「共起ネットワーク」と「BERT」を用いた。両者共通の分析手順として、たとえば、シラバス検索システム上で「情報通信」という単語で検索し、検索で得られた複数科目のシラバスのテキストデータを全て収集する。次に、Web上の検索サイトで「情報通信 スキル」などと検索し、テキストデータを収集する。共起ネットワー

クでは、得られたテキストデータを分析に用いるため形態素毎に分割する。科目シラバスと Web 上のテキストデータが単語まで分割され、単語間の類似性や関係性などから科目と職業に関連する特徴的な単語間関係を分析する。次に、単語中の名詞のみを抜き出し、抜き出した単語から1つの Web サイト分の単語組み合わせである共起を作成し、共起回数をカウントしたデータフレームにする。次に共起カウントを基にした共起ネットワークを作成する。図4は「情報通信」で検索した際の共起ネットワークである。円の大きさが単語の出現回数を表しており、単語同士の距離が単語間の関係性の近さを表している。BERT は深層学習モデルの Transformer を用いた自然言語処理モデルであり、本課題では Attention 機能を用いてテキストデータの注目すべき箇所の可視化を実施した。図5は Attention にて重要単語を可視化した例である。赤色が濃いほど重要な単語となる。これを利用してテキストデータ中の単語をピックアップし、科目、職種で必要な単語を把握する。評価指標には、precision (適合率)、recall (再現率)、f1-score (F 値) などがあり、開発に用いた Python ライブラリ sklearn の classification_report にてそれぞれを求めた。precision と recall がそれぞれ共に高い値であると f1-score も高くなるため、f1-score を総合的な判断値とした。

例1：情報通信業

科目1	科目2	科目3	...	職業分類
理論と情報理論	情報処理	情報ネットワーク社会	...	情報通信業



科目1	科目2	科目3	...	職業分類
400010	1400020	2400030	...	0

例2：農業・林業

科目1	科目2	科目3	...	職業分類
農業環境演習	分子遺伝学	植物バイオサイエンス入門	...	農業・林業



科目1	科目2	科目3	...	職業分類
9400099	4200045	5200055	...	1

図3 データ変換例

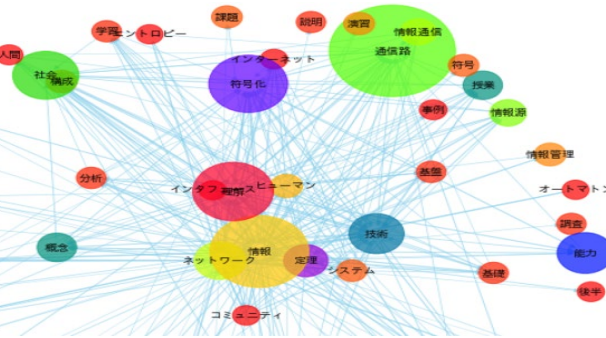


図4 「情報通信」共起ネットワーク

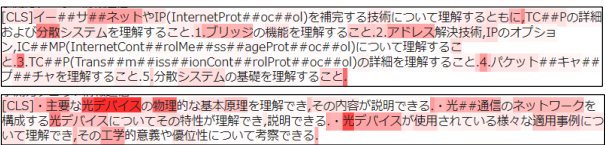


図5 Attention の可視化例

4. 研究成果

(1) 動画特徴量抽出システム

教育動画に含まれるさまざまな特徴を抽出する機械学習モデルを開発した。物体検出では、動画に現れるさまざまな物体を判別し、その中から教育に関連する文字や数式などを抽出した。今回は数学に関するものと限定した。これは教育に関する文字や数式、記号は多数あり、すべてを認識するためには膨大な学習データと計算コストが必要となるため、機械学習を簡単にするために教科を限定した。開発したモデルは四則演算記号や学習に用いた数種類の数式の判別を確認した。本モデルは判別できる対象は少ないが、たとえば音符記号や地図の記号など他教科の特徴を追加学習することで複数教科に対応するモデル開発が可能となる。講師発話では動画教材「整数と小数」の分析結果例として、動画から得られたテキストデータのマイニングから、「カード」や「お金」といった単語の登場頻度が多くまた関連が強いことがわかった。これらは数学固有の単語ではないが、社会生活の具体例を示しながら説明していると示唆され、良質な動画教材を検討する尺度の1つと考えられる。表情では先行研究から高品質教材の特徴として講師表情に「笑顔」「驚き」が多く含まれることから、これら2つの表情を判別するモデルを開発した。人間の表情を判別する既存モデルは存在するが、8つの基本表情全てを学習に用いることでそれぞれの判別精度が低下する点や、さまざまな人種を用いていることで特有人種の表情判別が困難な点などに課題がある。このため、本モデルは表情を2つに限定し、かつ人種を日本人限定とした。この結果、既存モデルと本モデルの比較では「笑顔」の認識率はほぼ同程度であったが、「驚き」の認識率を約30%向上できた。仕草では、動画教材中の講師の手首が静止している区間の開始時間と終了時間を取得でき、この区間内を物体や発話、表情などの各モデルがそれぞれの特徴を重点的に取得する目安となる。動画教材は時間が長いものもある。長時間動画の分析は計算コストが増加する。動画中にある教育重点区間を把握できれば効率よく特徴量分析が可能となる。仕草を検出するモデルは他モデルが分析する動画区間を提示することができ、結果的に合計計算時間の短縮につながった。今回はすべてのモデルで判別対象を限定することで認識精度を向上した。将来的には複数教科や講師人種・言語にかかわらず分析可能なモデル開発を目指す。

(2) 受講科目およびその成績と希望進路先との関係分析

判定モデルに用いるアルゴリズムは、LinearSVC、K 近傍法、勾配ブースティング (GBDT) をそれぞれ検証した。学習に用いたデータ数は、100、500、1000、2000、5000、10000 で比較した。この内、訓練データは8割、テストデータは2割とした。なお、今回はアルゴリズム比較検証を主としたため、訓練データに対してある産業分類に就く学生らの履修科目がある程度重複する

よう事前加工をした。たとえば、「情報通信業」に就いた学生らは「情報処理」や「プログラミング基礎」などの科目が重複している。今回はこの加工を施したダミーデータにて学習を実施した。評価値は「正解率 (accuracy)」を用いた。これは、テストデータを判定モデルに入力し、予測結果と答えがどれくらい一致しているかを判断する指標である。すべて正解した場合 1.0 となる。比較結果では、LinearSVC は今回用いた訓練データでは反復回数を増やして試行してみたが、いずれも収束しなかったため参考値とした。GBDT は訓練データを増やすことで正解率の向上がみられた。今回の検証ではデータ数 5000 で最も正解率が高かった。K 近傍法では、K 値により多少の変化は見られたが、訓練データ数の多少による変化は乏しかった。判定モデルは、1 年次から 4 年次の履修科目を入力することである職種に適する確率を示すものである。これを用いて、産業分類のみを入力することでその産業分類に就くための履修計画を提示することもできる。今回はダミーデータを用いた検証であったが、一定の結果を得られたため、今後は実データを用いることや訓練データの増加などでモデル精度向上を図る。

(3) 希望進路に即した適切な履修モデルの提供

共起ネットワークでは、科目シラバスおよび Web サイトで共に共起カウント上位 300 データを用いた。科目シラバスの結果 1 例として「情報通信」で検索した場合、静岡大学では「通信路や基盤に関する能力」を「理解」する科目群、A 大学では「ネットワークのプロトコル」を「理解」する科目群など、同一検索単語にもかかわらず異なる単語のつながりや関係性を確認した。Web サイトではテキストデータが多すぎたため単語同士の関係がうまく読み取れなかった。BERT では、f1-score が 90%を超えたモデルを用い、同様に「情報通信」で検索した。科目シラバスでは共起ネットワークでの結果に追加して、「イーサネット」、「ブリッジ」、「光デバイス」などの単語が注目された。Web サイトではある程度の単語が注目されたが、明らかに「情報通信」との関連は薄かった。共起ネットワーク、BERT、両者とも Web サイトの分析が期待通りといかなかった。これは扱うテキストデータが膨大であり、また検索語と関係が薄い単語が多数含まれることなどが原因と考えられる。今後の課題として、適切なデータ量の把握や不要単語や表現の削除などのデータクレンジングを検討している。

本研究課題進行中に新型コロナウイルスの大流行に見舞われた。通学の規制により多くの教育機関ではオンライン教育が実施された。そこで本研究課題の補助的な研究として、オンライン教育の学習効果を調査、分析した〔5〕。学生および教員に複数項目のアンケートを実施した。項目は学生側には、パソコン利用度、スマートフォン利用度、興味、わかりやすさ、理解度など、教員側には、動画教材の作成回数、動画作成時のリメイク回数、動画編集方法、動画作成で苦労した点などである。これらを変数とし重回帰分析を実施した。調査の結果、リメイク状況は学生の理解度にマイナスの影響を与え、編集状況はプラスの影響を与えることが確認された。これに関連して、情報リテラシーの高い教員は学生の理解度にプラスの影響を与え、情報リテラシーの低い教員は理解度にマイナスの影響を与えるという結果も明らかになった。さらに、情報リテラシーの高い学生はより高い理解度を示しており、これらの結果からオンライン教育における学習効果を最大化するためには、学生および教員相互の情報リテラシーが不可欠であることがわかった。

本研究課題におけるコロナ禍の影響は大きなものであった。研究計画立案時にオンライン教育はリカレント教育や遠方留学生などの理由が多数であったが、コロナ禍により一般社会へ多く普及した。普及速度も著しく、情報系、教育系だけでなくさまざまな研究会で日々議論が重ねられた。このような状況下において、本研究課題は当初計画の内容だけでなく関連した補助調査や分析など多岐にわたった。これらは研究を進める上での多くの知見となり、最終的な研究目的に近づくべく現在も研究を進行中である。今回の研究課題では、すべてで期待する精度は得られなかったが、オンライン教育に関する動画教材や受講科目と進路との関連性、希望職種に対する科目や取得技術の提案など一定の成果を得られた。今後も新たなデータ追加やアルゴリズムの検証を実施し、更なる精度向上を目指していきたい。

<引用文献>

- ① 永田 正樹、阿部 祐輔、教育動画に含まれる特徴の教育作用効果分析、日本教育工学会研究会 (学習支援システム/一般)、2020
- ② 横山 要、原崎 佑一郎、塩崎 雅基、阿部 祐輔、永田 正樹、教育動画に含まれる特徴量を可視化する Web アプリケーションの研究開発、情報処理学会 第 83 回全国大会、2021
- ③ 永田 正樹、阿部 祐輔、井口 敦史、機械学習を用いた希望進路に即した履修科目の分析、日本教育工学会研究会 (教育の情報化/一般)、2022
- ④ 山本 開世、向島 響、塩崎 雅基、阿部 祐輔、永田 正樹、シラバスと Web サイトの文言を用いた職能スキルの分析、情報処理学会 コンピュータと教育研究会 169 回研究発表会、2023
- ⑤ Masaki Nagata, Masaki Takeda, Information Literacy and Acceptance in Online Learning by Analysis Focusing on Learning Effectiveness: What is Universal Design for Online Learning?, 10th The Asia-Pacific Library and Information Education and Practice Conference, 2021, 208-218

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計13件（うち招待講演 1件 / うち国際学会 2件）

1. 発表者名 永田正樹，阿部祐輔，井口敦史
2. 発表標題 機械学習を用いた希望進路に即した履修科目の分析
3. 学会等名 日本教育工学会研究会（教育の情報化 / 一般）
4. 発表年 2022年

1. 発表者名 山本開世，向島響，塩崎雅基，阿部祐輔，永田正樹
2. 発表標題 シラバスとWebサイトの文言を用いた職能スキルの分析
3. 学会等名 情報処理学会 コンピュータと教育研究会 169回研究発表会
4. 発表年 2023年

1. 発表者名 Masaki Nagata, Masaki Takeda
2. 発表標題 Information Literacy and Acceptance in Online Learning by Analysis Focusing on Learning Effectiveness: What is Universal Design for Online Learning?
3. 学会等名 10th The Asia-Pacific Library and Information Education and Practice Conference（国際学会）
4. 発表年 2021年

1. 発表者名 永田正樹，山崎國弘，磯部千裕，長谷川孝博
2. 発表標題 視聴者維持率を基にした動画発信効果および評価手法についての考察
3. 学会等名 第25回学術情報処理研究集会
4. 発表年 2021年

1. 発表者名 永田 正樹、阿部 祐輔
2. 発表標題 教育動画に含まれる特徴の教育作用効果分析
3. 学会等名 日本教育工学会
4. 発表年 2020年

1. 発表者名 横山 要、原崎 佑一郎、塩崎 雅基、阿部 祐輔、永田 正樹
2. 発表標題 教育動画に含まれる特徴量を可視化するWebアプリケーションの研究開発
3. 学会等名 情報処理学会
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織			
	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------