

令和 4 年 6 月 3 日現在

機関番号：17102

研究種目：挑戦的研究（萌芽）

研究期間：2018～2021

課題番号：18K18656

研究課題名（和文）中学生の振り返り文に基づく学習行動・学習能力の推定と活用

研究課題名（英文）Estimation of Learning Activity and Learning Performance of Junior High School Student

研究代表者

峯 恒憲（MINE, Tsunenori）

九州大学・システム情報科学研究所・准教授

研究者番号：30243851

交付決定額（研究期間全体）：（直接経費） 4,400,000円

研究成果の概要（和文）：本研究は、学習者（中学生）の学習行動・学習能力の向上を目的として、学習者が書いた振り返り文章から学習行動や学習状況に関連する特徴的な記述表現を自動抽出し、抽出した特徴表現を学習者の学習行動・学習能力改善に向けたアドバイス支援に活用する仕組み作りに関する研究を行った。具体的には学習塾から提供を受けた生徒の振り返り文を利用し、生徒の通う中学校の定期試験の成績の上位・下位を判定するモデルの構築を行った。振り返り文から目的語や動詞、否定・丁寧表現などの特徴を抽出し、重み付けを行い、様々な機械学習器との組み合わせを考慮することで、重み付けしない場合に比べて分類精度が最大30%以上、向上することを確認した。

研究成果の学術的意義や社会的意義

本研究は、中学生の振り返り文を分析し、成績推定を行うことが可能となることを示した初めての研究である。研究の結果、振り返り文には成績推定に有用な特徴があること、各特徴と成績との間に異なる関連の強さがあること、この強さを考慮した重み付けを利用する場合と利用しない場合とでは、成績推定精度に最大30%以上の差が生じること、時期によって成績推定精度が変化し、最高の推定精度を得る機械学習器も一定しない（時期によって変化する、つまりデータの特徴が時期によって異なる）ことなどを確認した。これらから学習状況や学習行動の解釈、学習特性改善のためのアドバイス生成への応用も期待される。

研究成果の概要（英文）：This study aims to create a system for automatically extracting characteristic expressions related to student learning behavior and learning status from their written reflections, and to utilize the extracted expressions to generate advice for improving the student learning behavior and learning ability. Specifically, we constructed a model that judges the top and bottom grades in regular examinations at a student's junior high school using the student written reflections provided by a cram school. We extracted features such as object, verb, negative, and polite expressions from the reflective sentences, weighted them, and combined them with various machine learning tools to improve the classification accuracy by up to 30% or more compared to the case where no weighting was applied.

研究分野：人工知能応用

キーワード：振り返り文 テキストマイニング 機械学習 成績推定 アドバイス支援 特徴抽出

1. 研究開始当初の背景

小中学生などの早期学習者の学習行動、学習能力の向上が求められている。小中学校では、授業内容の振り返り文を書かせることが行われているが、その文章は、主に、教師が読み、児童や生徒の現状把握を行うために利用され、学習行動や学習能力の向上のためのデータとしての利用は皆無であった。研究代表者は、大学生の振り返り文分析については経験があったが、中学生の振り返り文分析については、十分な経験がなかった。

そこで、本申請にあたって事前に中学生の振り返り文章の解析に取り組み、本申請課題の適切さについて調査した。その結果、中学生の成績下位の生徒たちには、読解力や文章記述力に問題があることを確認した。これらは、振り返り文中に、文法や意味的な誤り表現として現れる。一方、文法的、意味的誤りを自動的に推定することは、未だに困難な課題の一つであるが、これまでの経験から、中学生の記述に特化し、その記述方法を機械的に学習することで、推定可能と判断した。また、中間や期末テストの前とその後、特に夏期・冬期休暇の時期との間に成績推定精度に顕著な差が現れることを確認した。大学生の場合は、推定精度の差の要因分析により、授業内容の難易度と関連があることを確認しているが、中学生の場合も、この時期的な推定精度の差の要因分析を行うことで、生徒たちの学習行動や学習能力改善に有効な特徴表現が抽出可能と判断した。

以上から、本申請研究においては、(1) 文法的・意味的誤り表現の推定、(2) 成績推定精度の変動の要因分析、という2つの困難な課題に焦点を当て、これまでの知見や技術を活用し、データ提供元との情報交換を通じてのポジティブフィードバックループを回しながら、中学生たちの学習行動・学習能力改善に向けた仕組みの実現に向け取り組んでいくことが可能と判断した。本研究で実施する研究内容の概要図を図1に示す。

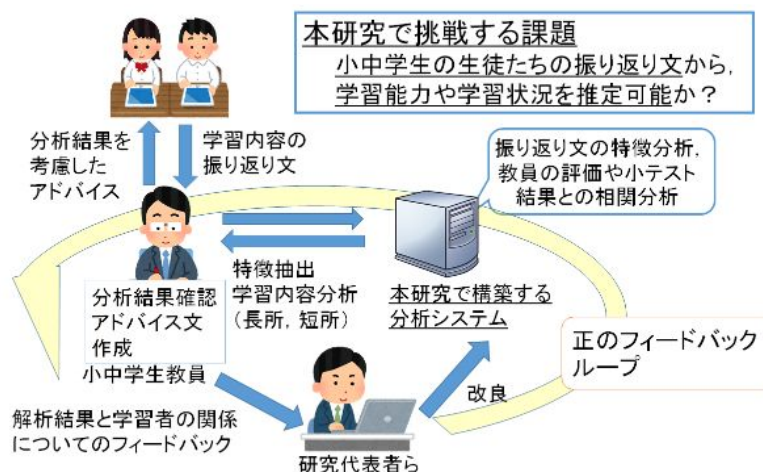


図1. 本研究で実施する研究内容の概要図

2. 研究の目的

本研究は、学習者(中学生)の学習行動・学習能力の向上を目的として、学習者が書いた振り返り文章から学習行動や学習状況に関連する特徴的な記述表現(単語や文末表現、文構造など)を自動抽出し、抽出した特徴表現を学習者の学習行動・学習能力改善に向けたアドバイス支援に活用する仕組み作りを行うことを目的とする。

3. 研究の方法

塾で学ぶ生徒たちの、学習後の振り返り文章データ(年数回のテスト結果含む)に対して、本研究目的の達成に必要な課題(振り返り文章の表層的特徴分析、学習属性の抽出、評価、時期(月)による推定精度の変動を起こす要因についての分析など)に取り組む。特に、テストの点数と相関の高い振り返り文章中の特徴の抽出に注力する。

得られた結果については、随時、データ提供元に報告を行い、生徒の学習状況の変化の有無について聞き取り調査分析・評価を行い、改善策を実施するポジティブフィードバックループを回す(図1)。

4. 研究成果

以下では、主に次の3つの論文の内容を記す。特に、成績推定結果の報告内容に関しては、3

番目の修士論文に詳しくまとめている。

- 細谷，峯：中学生の振り返り文の特徴を利用した成績推定，電子情報通信学会人工知能と知識処理研究会報告，AI2019-8, 39-44, 2019
- 細谷，峯：中学生の授業振り返り文からの成績推定に向けた特徴分析，電子情報通信学会人工知能と知識処理研究会報告，AI2020-14, 1-6, 2021
- 細谷，中学生の振り返り文の特徴分析と成績推定，九州大学修士論文，大学院システム情報科学府情報知能工学専攻社会情報システム工学専攻，2021年2月

(1) 研究で利用したデータ

本研究で扱う中学生の振り返り文は，以下に示す 3 つの質問文への回答として学習後に中学生自らが記述したものである。

質問 1：今日の範囲で分かった事，今日の学習で新しく覚えたこと，理解できたことを教えてください。
質問 2：今日の範囲で分からなかった事，今日の学習で分からなかったことや，質問したいことを教えてください。
質問 3：次回やりたいこと，今分からないことを教えてください。

利用したデータの内容を以下に示す。

生徒数	28 人
日付	2018/4 ~ 2018/11
科目	数学
学年	中学 3 年生
試験結果	6 月，9 月，11 月の定期試験の結果

生徒(下位:id1 と上位：id2)の振り返り文の例

id1・ 数学	質問 1	質問 2	質問 3
04/02	確立を復習する事ができた	なし	次回の英語では新出単語を覚えたり見直したりする
04/16	今日の学習で数学の多項式の乗法を予習する事ができた	特 に あ り ま せ ん	次回は英語のこれから重要になってくる単語を一つ一つ丁寧に覚えたいと思います
04/23	今日の授業で多項式の乗法を復習する事ができた	あ り ま せ ん	次回は英単語を中心に頑張っていきたいです
id2・ 数学	質問 1	質問 2	質問 3
04/03	数学：y を x の式で表す事が難しくて悩んだ		数学がにがてになっているからがんばりたい
04/10	分配法則を使った式を解いた & 比例の問題で $y=ax$ の式を書くところが苦手でわからない		問題を解くときに数字を見のがして失敗する事があるからきをつけてがんばりたい
04/17	因数分解の乗法をあまり間違えずに解けた		式の見間違いで失敗することをなくしていきたい

(2) データ分析

本研究で用いたデータでは，6 月に退塾する生徒と 7 月に入塾する生徒が多く，6 月と 7 月の間で生徒の入れ替わりが特に多いため，分析の期間を，期間 A：4 月～6 月と期間 B：7 月～11 月の 2 つに分けた。また，テストの点数の中央値以下の生徒に成績下位，点数が中央値より大きい生徒に成績上位のラベルを付けてグループ分けを行なった。分析に当たって，3 つの仮説を立て，その検証を行った。

[仮説 1] 上位の生徒と下位の生徒の間で目的語・動詞ペアの割合，目的語の語彙の共通率，丁寧語の割合，振り返り文の類似度に違いがある。

[仮説 2] 下位の生徒と上位の生徒の間で否定語の割合，振り返り文末記入の割合に違いがある。また質問 1 および 3 と質問 2 の傾向が異なる。

[仮説 3] 振り返り文の単語数，目的語の単語数と点数との間に正の相関がある。

成績推定では，機械学習器で成績推定モデルを構築した。機械学習器として Support Vector Machine (SVM)，Naive-Bayse (NB)，Multilayer perceptron (MLP) を用いた。推定結果の評価には 5 分割交差検定を実施した。成績推定の精度に影響が出るため，訓練用のデータにテスト用データ

の生徒のコメントが入らないようにした。成績推定で利用する素性として、8つの素性（振り返り文のベクトル、振り返り文の類似度(BoW)、目的語のベクトル、目的語の類似度(BoW)、目的語・動詞ペアの有無、目的語の単語数、否定語の有無、丁寧語の有無)を利用した。

分析では、各素性の有用性を確認するため、目的語・動詞ペアの割合、目的語の語彙の共通率、丁寧語の割合、振り返り文の類似度、振り返り文末記入の割合、否定語を含む振り返り文の割合について調べるとともに、各素性と点数の間の相関を算出した。その結果、平均的に強い相関がある素性は無かったが、単語数、丁寧語はともに点数と平均的に弱い相関があることがわかった。

各仮説の検証結果を以下に示す。

仮説1：3つの素性に関する仮説1は成り立っていた。目的語動詞ペアの有無、丁寧語、振り返り文の類似度は上位か下位かを区別するのに利用できる特徴と言えると考えられる。

仮説2：仮説2は一部の時期を除き成り立たなかった。

仮説3：各素性と点数の間に強い相関が見られるものは無かったため、仮説3は成り立たなかった。ただし、単語数と丁寧語について、平均的に弱い相関が見られ、他の素性よりも相関が強いいため、単語数と丁寧語は組み合わせると成績推定に利用できると推定される。

以上、仮説1がいくつかの素性で成り立ち、かつ仮説3が成り立たないことから、成績の分類がしやすくなる素性は存在するが、点数との相関ではなく点数からつけたラベルとの相関があることがわかった。

(3) 成績推定

成績推定に有効な素性を発見し特徴を分析すること、成績推定に有効または無効である振り返り文の特徴を発見することを目的とし、実験を行った。その際、次の4種類の手法(手法1.1, 1.2, 2.1, 2.2)を試し、有効性を比較した。

手法1.1 8つの素性と3種類の機械学習器で成績推定：8つの素性を1種類～3種類組み合わせ、成績推定の入力として92種類の素性の組み合わせを用いる手法。推定を行うテストの月までに記述された、連続する2週間分の振り返り文を各素性に交換する。1週ずつ利用する振り返り文をずらし、4月から6月までの12週から全11回分の推定結果を得る。

手法1.2 手法1.1で精度が高い素性について複数の質問間の素性を利用した成績推定：質問1から3の振り返り文全てを結合し、有用な素性を組み合わせると成績推定を行う手法。

手法2.1 投票法を用いた成績推定：手法1.1において各テストに対して最も成績推定精度が高かった質問、素性、機械学習器の組み合わせ3通りに対して、投票法を用いる手法。投票法の概念図を図2に示す。

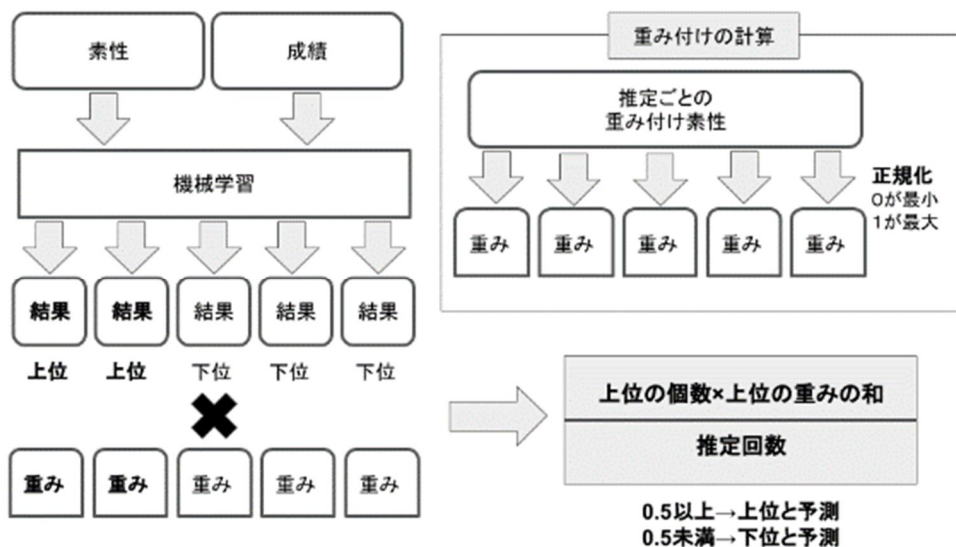


図2 投票法の概念図

手法2.2 票の重みづけを行なった投票法を用いた成績推定：2週ごとの成績推定結果にその週の生徒全体の振り返り文の特徴を重みとして利用し投票法を行う手法。10個の重み付けの素性について値が大きいほど重みを大きくする場合と、値が小さいほど重みを大きくする場合の2通りがあり、全20個の重み付けの素性を1種類～2種類組み合わせると210通りの重みの計算を試す。

手法 1.1 についての考察

従来の研究で用いられている振り返り文ベクトルを利用した成績推定よりも本手法が高い精度を得た。すなわち、単純に、振り返り文から得られる単語ベクトルを使うのではなく、振り返り文や目的語の類似度など、さまざまな素性を考慮することが、高い成績推定精度を得るために求められることが確認された。

手法 1.2 についての考察

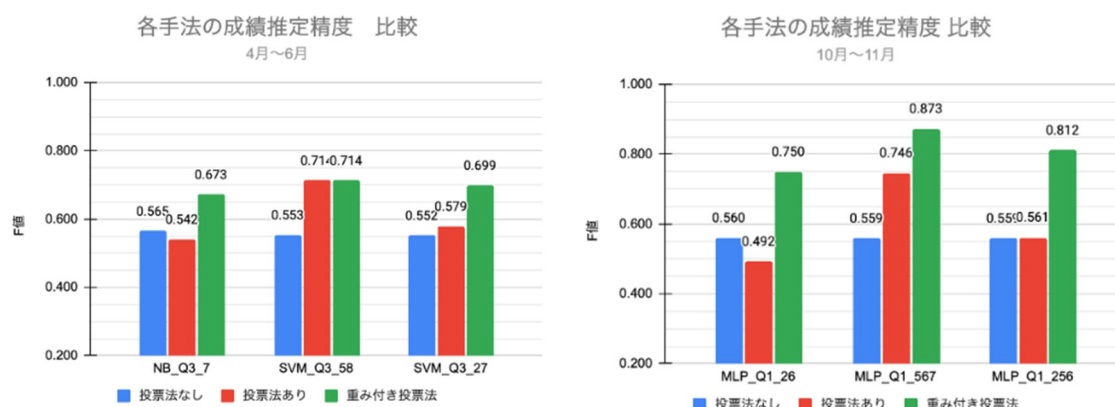
この手法では成績推定精度は向上しなかったが、異なる質問から得られる別々の素性を組み合わせた実験を行うとより高い精度が得られる可能性がある。

手法 2.1 についての考察

手法 1.1 の結果と比較すると、投票法を用いた方が成績推定精度が最大約 2 割向上することを確認した。月ごとの結果から、テストがある月（6 月、9 月、11 月）よりもその月から離れた月（4 月、7 月、10 月）の振り返り文を利用した方が、成績推定が高い傾向あることが分かった。

手法 2.2 についての考察

手法 1.1、手法 2.1 の結果と比較すると重み付けを行った投票法を用いることで、さらに成績推定精度を向上させることができることが分かった。最も成績推定精度が高い 10 月～11 月の推定では、振り返り文の文字数・単語数の標準偏差（バラツキ）が多いほど成績推定の精度が高く、重みを大きくすることで投票法の成績推定精度が向上した。成績推定の回ごとに重み付け素性と各推定での手法 1.1 の推定精度の相関を計算すると、相関係数 0.45 と他の素性と比べて大きい正の相関があった。これが高い成績推定精度を得た要因と考えられる。4 月から 6 月と、10 月から 11 月の期間での成績推定結果を、以下に記す。



成績推定についてのまとめ

4 種の手法（手法 1.1 から手法 2.2）と様々な成績推定精度を向上させるための提案を行った。本実験の成績推定精度が高かった素性から成績下位の生徒の振り返り文の特徴として、文章に丁寧語が含まれる割合が低く、振り返り文の類似度、目的語の類似度が低いという結果を得た。成績上位の生徒同士では学習した内容についての記述が似ることから、目的語の類似度が高くなると考えられる。

(4) その他の成果について

2020 年 2 月から、Covid-19 の影響を受け、研究期間が 2 年延長となった。その間に、新たな共同研究先からデータ提供を受けられるようになり、研究の発展につながった。以下に、(3)までに述べていない主な 4 つの成果を記す。

Intelligent Tutoring System の利用ログを利用した STEM キャリアの推定
the 2017 ASSISTments Data Mining Competition で 2 位となり、成果の内容を(Makhlouf and Mine, JEDM2020)で発表

塾講師の生徒の学習状況報告文に基づく、生徒の模擬試験成績の推定

成果の一部が Fateen, Ueno and Mine, ICCE2021 が Best Student Paper Award 候補に選出

キーワードに基づく塾講師の生徒の学習状況報告文の自動生成

振り返り文に対する自動フィードバック機能の構築

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件/うち国際共著 2件/うちオープンアクセス 1件）

1. 著者名 Jihed Makhoulouf and Tsunenori Mine	4. 巻 12
2. 論文標題 Analysis of Click-Stream Data to Predict STEM Careers from Student Usage of an Intelligent Tutoring System	5. 発行年 2020年
3. 雑誌名 The Journal of Educational Data Mining	6. 最初と最後の頁 1-18
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1. 著者名 Shaowen Peng, Xianzhong Xie, Tsunenori Mine and Chang Su	4. 巻 LNAI11224
2. 論文標題 Vector representation based model considering randomness of user mobility for predicting potential users	5. 発行年 2018年
3. 雑誌名 PRIMA2018, LNAI11224	6. 最初と最後の頁 70 -- 85
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-03098-8_5	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

〔学会発表〕 計21件（うち招待講演 1件/うち国際学会 12件）

1. 発表者名 Jihed Makhoulouf and Tsunenori Mine
2. 発表標題 Automatic Feedback Models to Students Freely Written Comments
3. 学会等名 28TH INTERNATIONAL CONFERENCE ON COMPUTERS IN EDUCATION（国際学会）
4. 発表年 2020年

1. 発表者名 Jihed Makhoulouf and Tsunenori Mine
2. 発表標題 Prediction Models for Automatic Assessment to Student Free-written Comments
3. 学会等名 12th International Conference on Computer Supported Education (CSE2020)（国際学会）
4. 発表年 2020年

1. 発表者名 Jihed Makhoulouf and Tsunenori Mine
2. 発表標題 Predicting Student Exam Scores Based on Click-stream Level Data of Their Usage of an E-Book System
3. 学会等名 DC&LAK20 at the 10th International Learning Analytics & Knowledge Conference (LAK20) (国際学会)
4. 発表年 2020年

1. 発表者名 Ristu Saptono and Tsunenori Mine
2. 発表標題 Time-based Sampling Methods for Detecting Helpful Reviews
3. 学会等名 The 2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT2'0) (国際学会)
4. 発表年 2020年

1. 発表者名 Shaowen Peng and Tsunenori Mine
2. 発表標題 Mixture-Preference Bayesian Matrix Factorization for implicit feedback datasets
3. 学会等名 The 35th ACM/SIGAPP Symposium On Applied Computing (国際学会)
4. 発表年 2020年

1. 発表者名 細谷桃子, 峯 恒憲
2. 発表標題 中学生の授業振り返り文からの成績推定に向けた特徴分析
3. 学会等名 電子情報通信学会 人工知能と知識処理研究会
4. 発表年 2021年

1. 発表者名 加藤博之, 石岡恒憲, 峯 恒憲
2. 発表標題 短答式試験における自動採点のための概念辞書を用いたデータ拡張手法の提案
3. 学会等名 電子情報通信学会 人工知能と知識処理研究会
4. 発表年 2021年

1. 発表者名 細谷桃子, 峯 恒憲
2. 発表標題 中学生の振り返り文の特徴を利用した成績推定
3. 学会等名 電子情報通信学会 人工知能と知識処理研究会
4. 発表年 2019年

1. 発表者名 加藤 博之, 石岡 恒憲, 峯 恒憲
2. 発表標題 短答式試験における自動採点のための品詞別ノイズ付与手法の提案
3. 学会等名 JAWS2019
4. 発表年 2019年

1. 発表者名 古賀詩織, 峯 恒憲, 廣川佐千男
2. 発表標題 W2Vを用いたLSTMの前処理としての特徴選択の有効性
3. 学会等名 電子情報通信学会 人工知能と知識処理研究会
4. 発表年 2019年

1. 発表者名 古賀詩織, 峯 恒憲, 廣川佐千男
2. 発表標題 LSTMの前処理としての特徴選択の有効性
3. 学会等名 第6回自然言語処理シンポジウム, 電子情報通信学会 言語理解とコミュニケーション研究会 (NLC)
4. 発表年 2019年

1. 発表者名 逆瀬川滉大, 峯恒憲, 廣川佐千男
2. 発表標題 双極空間への埋め込みを利用した文書識別
3. 学会等名 人工知能学会 第111回人工知能基本問題研究会(SIG-FPAI)
4. 発表年 2020年

1. 発表者名 Jihed Makhlouf and Tsunenori Mine
2. 発表標題 Predicting Student Exam Scores Based on Click-stream Level Data of Their Usage of an E-Book System
3. 学会等名 DC@LAK20 at the 10th International Learning Analytics & Knowledge Conference (LAK20) (国際学会)
4. 発表年 2020年

1. 発表者名 Shaowen Peng and Tsunenori Mine
2. 発表標題 Mixture-Preference Bayesian Matrix Factorization for implicit feedback datasets
3. 学会等名 The 35th ACM/SIGAPP Symposium On Applied Computing (国際学会)
4. 発表年 2020年

1. 発表者名 Jihed Makhoulouf and Tsunenori Mine
2. 発表標題 Prediction Models for Automatic Assessment to Student Free-written Comments
3. 学会等名 The 12th International Conference on Computer Supported Education (CSEDU2020) (国際学会)
4. 発表年 2020年

1. 発表者名 Jihed Makhoulouf and Tsunenori Mine
2. 発表標題 Investigating Reading Behaviors within Student Reading Sessions to Predict their Performance
3. 学会等名 International Workshop on Predicting Performance Based on the Analysis of Reading Behavior: A Data Challenge 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Tsunenori Mine and Sachio Hirokawa and Takahiko Suzuki
2. 発表標題 Does Crime Activity Report Reveal Regional Characteristics?
3. 学会等名 IMCOM2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Jihed Makhoulouf and Tsunenori Mine
2. 発表標題 Investigating How School-Aggregated Data Can Influence in Predicting STEM Career from Student usage of an Intelligent Tutoring System
3. 学会等名 EDM 2018 Workshop on Scientific Findings from the ASSISTments Longitudinal Data (2018) (国際学会)
4. 発表年 2018年

1. 発表者名 Jihed Makhoulf and Tsunenori Mine
2. 発表標題 Predicting if students will pursue a STEM career using School-Aggregated Data from their usage of an Intelligent Tutoring System
3. 学会等名 EDM2018 (国際学会)
4. 発表年 2018年

1. 発表者名 Tsunenori Mine
2. 発表標題 Artificial Intelligence and Education
3. 学会等名 Rakuten Technology Conference 2018 (招待講演)
4. 発表年 2018年

1. 発表者名 Shaowen Peng, Xianzhong Xie, Tsunenori Mine and Chang Su
2. 発表標題 Vector representation based model considering randomness of user mobility for predicting potential users
3. 学会等名 PRIMA2018
4. 発表年 2018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
---------	---------