

令和 4 年 4 月 29 日現在

機関番号：43608
 研究種目：基盤研究(C)（一般）
 研究期間：2018～2021
 課題番号：18K02882
 研究課題名（和文）教学IRと教育ビッグデータを活用した二重機械学習法による学生支援予測モデルの開発

研究課題名（英文）Development of a Predictive Model for Student Support Using Dual Machine Learning Methods with Educational IR and Educational Big Data

研究代表者
 片瀬 拓弥（Katase, Takuya）
 清泉女学院短期大学・その他部局等・教授

研究者番号：70542322
 交付決定額（研究期間全体）：（直接経費） 2,200,000円

研究成果の概要（和文）：本研究は、ラーニング・アナリティクス、性格特性、教学IRデータを活用し、GPAと学校生活満足度を予測するモデルを開発した。開発手法として、クラスター分析、線形重回帰及びニューラルネットワークを採用した。その結果、線形重回帰モデルの予測率は、（GPA、学校生活満足度）=（0.28,0.31）、ニューラルネットワークモデルの予測率は、（GPA、学校生活満足度）=（0.83,0.48）となった。一方、コロナ禍により、平常時データの入手が困難となり、計画変更を余儀なくされた。そこで、学生の質的意見に端を発した対AI信頼感尺度と心理的AIデバイドをアセスメントするテストバッテリーの開発を行った。

研究成果の学術的意義や社会的意義

本研究の意義は、従来のニューラルネットワーク（NN）モデルでブラックボックス化しがちな説明変数において、k-means法のクラスター分析による各クラスター中心からの距離という説明変数を導入した前段処理を活用して「新しいINNモデルの開発手法」を提示したことにある。この方法の利点は、投入変数がいくら増えたとしても、分類されたクラスターは「パターン情報」に集約されることにある。さらに、モデル精度の向上のため、対AI信頼感と心理的AIデバイドを判定する尺度開発を行った。これらの尺度により、AI活用の積極群と消極群を質問紙により判定することが可能となった。

研究成果の概要（英文）：This study utilized Learning Analytics, personality traits, and instructional IR data to develop a model to predict GPA and school life satisfaction. Cluster analysis, linear multiple regression, and neural networks were employed as development methods. As a result, the prediction rate of the linear multiple regression model was (GPA, school life satisfaction) = (0.28,0.31) and that of the neural network model was (GPA, school life satisfaction) = (0.83,0.48).

On the other hand, the corona disaster made it difficult to obtain normal data and forced us to change our plan. Therefore, we developed a scale of trust toward AI that was inspired by students' qualitative opinions and a test battery to assess the psychological AI divide.

研究分野：教育工学

キーワード：ラーニング・アナリティクス 教学IR GPA 学校生活満足度 予測モデル 機械学習 対AI信頼感尺度 心理的AIデバイド

1. 研究開始当初の背景

近年、ICTの発展により教育ビッグデータの利活用が国内外で注目を浴びている。教育ビッグデータに関する分析は、ラーニングアナリティクス(Learning Analytics:以下、LA)と呼ばれ、その必要性が急速に認識されつつある(近藤2016)。一方、大学IR(Institutional Research:大学の意志決定を支援するための情報収集やデータ分析活動)の中でも教学部門に特化した教学IRも急速に普及しつつある(高橋2014)。船守(2014)は、これら2つの研究分野に対し、教学IRで扱う学務・教務・学生調査情報(以下、マクロデータ)とLAで扱う学習履歴(以下、マイクロデータ)の双方を活用した「精緻な教学IR」の必要性を指摘している。一方、マクロ/マイクロデータのような多種多様かつ大量のデータを分析する手法として、機械学習の導入が多角的に検討されている(近藤2016)。機械学習とは、データ解析からパターンを導出し、その結果から新たなデータに対する予測を提供する手法である(橋本2017, 横内・青木2017)。機械学習の活用事例として、オンラインコースの学習ログの分析から、ドロップアウト学生を予測した研究がある(Vihavaineほか2013)。また、近藤・畠中(2016)は、6種類の機械学習法により、大学3年次当初までのドロップアウト学生の40%を予測可能としている。さらに片瀬(2011)は、「教師あり機械学習」のニューラルネット分析を行った結果、専修学校の中退学生を70.4%という精度で予測した。しかし、これらの研究はいずれも中退学生等を予測するモデルであり、総合的な学生支援を行うという観点からは、汎用性に乏しい。

2. 研究の目的

上記の学術的背景から、以下2つの課題を解決する必要があると考えた。

- (1)マクロ/マイクロデータの双方を活用した「精緻な教学IR」の必要性
 - (2)機械学習による支援ニーズにマッチした汎用性の高い学生支援予測モデルの開発
- つまり、本研究は、上記2つの課題を解決するための方法論を探究・開発・評価することにある。

3. 研究の方法

図1は、本研究における予測モデルの開発工程図である。図1のA~F部は、それぞれ、A部が「LAデータとしての活用するリメディアル教材のLMS学習ログ」、B部が「性格特性を決定するために活用する性格検査(調査データ)」、C部が「k-means法によるクラスター分析(教師なし機械学習)」、D部が「教務学生部等から提供を受ける教学IRデータ」、E部が目的変数である「1年春学期末のGPAと学校生活満足度(hyper-QU)」を示している。そして、F部が「教師あり機械学習による予測モデルの開発」を示す。また、モデル開発に必要なデータは以下とした。

- 目的変数 : 1年春学期末のGPA(GPA)、学校満足度の偏差値(SAT)
- LAデータ : ログイン回数、総学習時間、平均学習時間、クリア率
- 性格データ : 外向性、調和性、誠実性、開放性、神経症傾向
- 教学データ : 日本語プレースメントテスト、出身高校偏差値

(1)教師なし機械学習(k-means法によるクラスター分析)

図1のC部「k-means法によるクラスター分析(教師なし機械学習)」について述べる。k-means法は、「教師なし機械学習」のクラスタリング問題を解くためのアルゴリズムである。片瀬(2017)は、LAデータの学習ログから得られる「学習スタイル」のクラスター数を3つ、また、性格検査から得られる「性格特性」のクラスター数を4つとしており、本研究でも同様とする。このクラスター分析により、次段階の「教師あり機械学習」へ投入する情報量が削減されるとともに、投入前の説明変数について、パターン情報という有意味な解釈を与えることができる。学習スタ

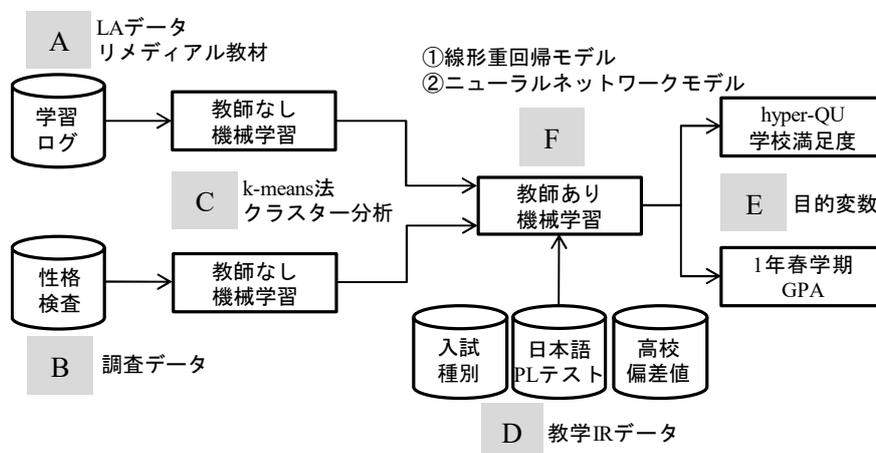


図1 予測モデルの開発工程図

イルと性格特性のパターン情報は、**図1**のF部の「教師あり機械学習」において、2つのケースで分析する。

(ケース1) 各クラスター情報をダミー変数とする場合

(ケース2) 各クラスター中心からの距離変数を用いる場合

(2) 教師あり機械学習 (ニューラルネットワークモデル)

本報告では、予測精度の高い「ニューラルネットワークモデル」についてのみ述べる。ニューラルネットワークモデルでは、全変数をモデルに投入した。解析ツールは、MATLAB 2018を使用した。ネットワークの学習にはバイズ正則化法を用い、全データの50%をモデル開発用、25%をモデルの過適合防止用、残り25%を完全独立したモデル検証用として使用する。また、隠れニューロン数は10個とし、数十回の学習試行の中でモデル検証用の決定係数が最大となる予測モデルを選択する。

4. 研究成果

(1) 研究開始当初からの目的の成果

対象者は、パイロット校にX年度～(X+3)年度に入学した1年生、計331名とした。教務部から提供された各種個人情報データは、個人情報が特定できないように匿名化処理を行った。また、欠損データについては、ペアワイズ方式による分析とした。**図2**(a)～(d)に「ニューラルネットワークの予測モデル」を示す。

(a) ケース1のGPA予測モデル開発

(b) ケース2のGPA予測モデル開発

(c) ケース1のSAT予測モデル開発

(d) ケース2のSAT予測モデル開発

図2(a)(b)によれば、GPA予測モデルにおいて、決定係数は(ケース1, ケース2) = (0.69, 0.83)となり、ケース2の方が高い。この結果は、ケース2のクラスター情報の方が多いためと考えられる。線形重回帰モデルと比較すると約55%の予測精度の向上が見られた。一方、**図2**(c)(d)によれば、SAT予測モデルにおいて、決定係数は(ケース1, ケース2) = (0.27, 0.48)となり、ケース2の方が高くなっているが、線形重回帰モデルと比較すると約17%の予測精度の向上に留まった。予測精度があまり向上しなかった要因について、今後の検討が必要である。

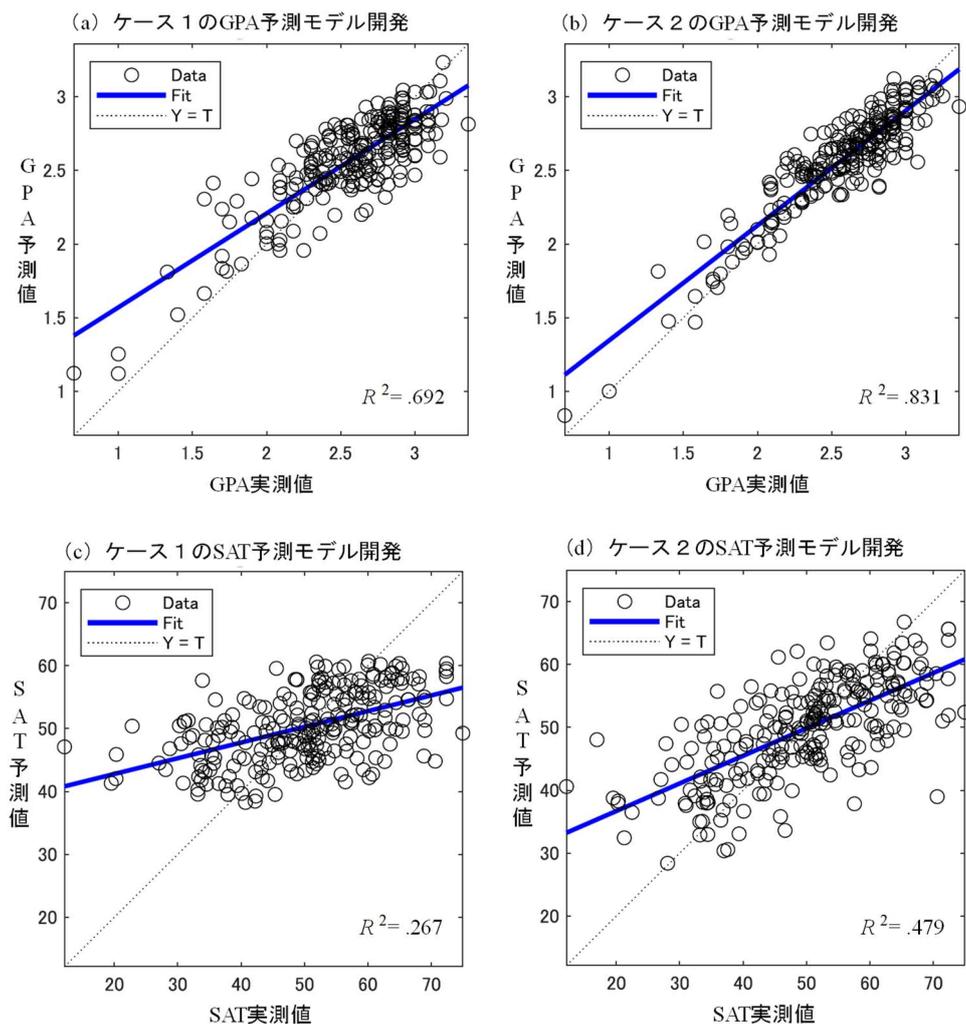


図2 ニューラルネットワークモデルによるGPA及びSAT予測モデル

本研究では、LA データ、性格特性、教学 IR データを活用し、1 年春学期末の GPA と学校生活満足度を予測するモデルを開発した。本研究で開発された手法では、直接投入する説明変数の数を削減するため、「k-means 法のクラスター分析による各クラスター中心からの距離」という新しい方法論を提示した。この方法の利点は、投入変数がいくら増えたとしても、分類されたクラスターは「パターン情報」に集約されることである。さらに、ニューラルネットワークモデルでは、一般的に投入する説明変数が多くなるほど予測精度は高くなる。しかし、その一方で計算過程はブラックボックスになりがちであるが、本研究では、説明変数の意味解釈が容易なクラスター分析（教師なし機械学習）を前段処理に活用している。つまり、意味解釈をブラックボックス化せず、予測精度を向上させる新しいモデル開発方法を提示することができたと考えている。

(2) 新型コロナ禍による研究計画変更後の成果

最終年度は、新型コロナの蔓延により、研究計画の変更と期間延長を余儀なくされた。一方、当初の研究計画により取得した「学生の質的記録」の分析結果に端を発して「対 AI 信頼感尺度」と「心理的 AI デバイドをアセスメントするテストバッテリー」の開発を行った。その開発理由として、AI 予測を就職活動支援に活用することに積極的な学生と消極的な学生が存在したためである。この結果は、「AI 予測を活用すること自体に別要因が存在する可能性」を示唆するという新たな仮説を生んだ。その影響を考慮しないとモデル精度の向上は十分に図れないと考えたからである。以下に、最終年度の研究成果を挙げる。

①全国 Web 調査により、対 AI 信頼感尺度を作成した。妥当性の検討は、対人信頼感尺度と TIPI-J を用いた。さらに、当該尺度を用い、性別・年代別のコホート分析を行った結果、対 AI 信頼感について、性別・年代別に違いがみられた。

②全国 Web 調査により、心理的 AI デバイドをアセスメントするテストバッテリーを作成した。テストバッテリーは、対 AI 信頼感尺度と最新電子機器使用態度尺度とした。さらに、このテストバッテリーを用いて、性別・年代別のコホート分析を行った結果、年代別の心理的 AI デバイドの実態を把握できた。

以上の尺度を開発したことにより、今後、AI 活用の「積極群」と「消極群」を簡単な質問紙により判定することが可能となった。この尺度の活用により、よりモデル精度の向上を図ることができるのではないかと考えている。

引用・参考文献

- 近藤伸彦 (2016) 大学におけるビッグデータ・アナリティクスと教学 IR. 大手前大学 CELL 教育論集, Vol. 6, pp. 11-18
- 高橋哲也, 星野聡孝, 溝上慎一 (2014) 『学生調査と e ポートフォリオならびに成績情報の分析について —大阪府立大学の教学 IR 実践から』, 京都大学高等教育研究, 20, pp. 1-15.
- 船守美穂 (2014) 『デジタル技術は高等教育のマス化問題を救えるか? —MOOCs, 教育のビッグデータ, 教学 IR の模索』, 情報知識学会誌, 24(4), pp. 424-436.
- 橋本泰一 (2017) データ分析のための機械学習入門. SB クリエイティブ, 東京
- 横内大介, 青木義充 (2017) イメージでつかむ機械学習入門. 技術評論社, 東京
- Vihavainen, A., Luukkainen, M., Kurhila, J. (2013) Using Students Programming Behavior to Predict Success in an Introductory Mathematics Course. Proceedings of the 6th International Conference on Educational Data Mining, pp. 300-303
- 近藤伸彦, 畠中利治 (2016) 学士課程における大規模データに基づく学修状態のモデル化. 教育システム情報学会誌, 33(2) : 94-103
- 片瀬拓弥 (2011) ニューラルネットによる専修学校の中途退学者早期予測方法の開発. 2010 年度科学研究費実績報告書 (研究課題 22910004)
- 片瀬拓弥 (2017) 『教学 IR のための学生支援モデルの試作—学習スタイルと性格特性のクラスターリングを活用して—』, 日本教育工学会研究会報告集 リフレクション活動の支援/インストラクショナルデザイン/一般 (JSET17-4), pp. 159-166.

5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 6件）

1. 著者名 片瀬拓弥	4. 巻 40
2. 論文標題 Web調査結果の信頼性を向上させる方法の検討	5. 発行年 2022年
3. 雑誌名 清泉女学院短期大学 研究紀要	6. 最初と最後の頁 87-96
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 -
1. 著者名 片瀬拓弥	4. 巻 4
2. 論文標題 心理的AIデバイドをアセスメントするテストバッテリーの作成	5. 発行年 2021年
3. 雑誌名 日本教育工学会報告集	6. 最初と最後の頁 67-72
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 無
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 -
1. 著者名 片瀬拓弥	4. 巻 3
2. 論文標題 人工知能（AI）に対する信頼感尺度の作成と信頼性・妥当性の検討	5. 発行年 2021年
3. 雑誌名 日本教育工学会報告集	6. 最初と最後の頁 172-179
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 無
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 -
1. 著者名 片瀬 拓弥	4. 巻 38
2. 論文標題 AI予測とテキストマイニング分析による短大生の就職活動支援方法の研究	5. 発行年 2020年
3. 雑誌名 清泉女学院短期大学 研究紀要	6. 最初と最後の頁 12-21
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 -

1. 著者名 片瀬拓弥	4. 巻 JSET19-3
2. 論文標題 短大生の就職活動終了時期の予測モデルの開発	5. 発行年 2019年
3. 雑誌名 日本教育工学会研究会 報告集	6. 最初と最後の頁 27-34
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 無
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 -

1. 著者名 片瀬拓弥	4. 巻 37
2. 論文標題 ラーニング・アナリティクス、性格特性、教学IRデータを活用した GPAと学校生活満足度の予測モデルの開発	5. 発行年 2018年
3. 雑誌名 清泉女学院短期大学 研究紀要	6. 最初と最後の頁 1-10
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 -

〔学会発表〕 計1件 (うち招待講演 0件 / うち国際学会 0件)

1. 発表者名 片瀬拓弥
2. 発表標題 短大生の就職活動終了時期の予測モデルの試作
3. 学会等名 日本教育工学会 第35回大会 (名古屋大学)
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

清泉女学院リポジトリ
<https://seisen-jc.repo.nii.ac.jp/>

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
---------	---------