

## 成績予測システムの構築と予測に用いる特徴量とデータ量の検証

Construction of a Grade Prediction System and  
Validation of Features and Data Volume for Predictions駒谷 優斗<sup>1)</sup>望月 久稔<sup>1)</sup>

Yuto KOMATANI Hisatoshi MOCHIZUKI

## 1 はじめに

ICT の発展により、教育の場面で LMS などの技術が用いられるようになった。これらの技術を教育や学習の場面で利用するメリットの 1 つは学習に関するログが大量に残ることである。それらのデータを収集して分析し、教育・学習を改善することをラーニングアナリティクス (Learning Analytics) といい、オンライン授業を実施する学校が増えている中で、注目されている [1]。その中に成績予測があり、受講生の成績を予測できれば、教員が成績不良の学生を早期発見し指導することなどが可能となる。そこで本研究では、Python の Web フレームワークである Flask [2] を用いて、機械学習による成績予測システムを構築し、学生と教員それぞれに対して予測成績や特徴量などの情報をフィードバックする。特徴量は、大学生の授業を対象に Moodle などを実施した小テストのログから抽出する。ランダムフォレストを用いて成績を予測し、用いる特徴量やデータ量を検証する。

## 2 成績予測システムと特徴量

Web ブラウザ上で機械学習による成績予測を実行するシステムの構築と、成績予測に用いる特徴量について述べる。

## 2.1 成績予測システム

Flask を用いて Web ブラウザ上に成績予測システムを構築する。Google フォームまたは Moodle を用いて実施した授業の小テストのログを入力データとし、受講生の最終成績を予測する。システムにおけるモデルの構築と、成績の予測それぞれについて述べる。

まず、過去の小テストのログをブラウザ上で選択し、機械学習モデルを構築する。本稿ではモデルにランダムフォレストを用いた。小テストのログから得点や受験回数などの特徴量を抽出し、グリッドサーチを用いてその授業に適したランダムフォレストのモデルを構築する。

次に、未知の成績を予測する際は、モデルの構築時と同様に小テストのログデータを選択することで、特徴量を抽出し構築したモデルを用いて成績を予測する。予測結果ページでは以下の 4 つの情報が表示される。

- 各学生の予測成績
- 全体の小テストの得点率推移
- 各学生の小テストの得点率推移
- 各小テストにおけるヒストグラムと箱ひげ図

更に、各学生の予測された成績は csv や Excel, ods, pdf 形式で教員がダウンロードできる。得点率推移や箱ひげ図などもダウンロードできる。

また、データベースを用いた認証機能により、教員だけでなく受講生が自身の予測成績や予測に用いられた小

表 1 実験に用いたデータ

実験名	構築用データ	予測用データ
W	2021 年度	2022 年度
X	2021 年度	2023 年度
Y	2022 年度	2023 年度
Z	2021, 2022 年度	2023 年度

テストの情報などをブラウザ上で確認できる。

## 2.2 特徴量の定義

予測システムで用いる特徴量を小テストのログデータから 4 つ定義する。

- 得点率の平均 (初回受験点数)
- 得点率の平均 (最高点数)
- 得点率の上がり幅を加算した得点率の平均
- 受験回数の合計

小テストは複数回受験可能であるため、学生によっては何度も受験する場合がある。そこで、受験回数と得点率の上がり幅を定義する。

得点率の上がり幅を加算した得点率の平均は、複数の定義方法を用いて検証した [3]。本稿ではその中で最も精度の高かった定義を用いる。

## 3 データ量と特徴量の評価

2.2 節の特徴量を用いて、授業の成績を予測する。実験データには 2021 年度から 2023 年度までに実施した学部 1 年生対象の情報リテラシーに関する授業の Moodle のログを用いる。

昨年度までのデータを用いて今年度の未知の成績を予測するため、モデルの構築に用いたデータの次年度以降を予測用として用いた。実験に用いた年度の組み合わせを表 1 に示す。以降ではそれぞれの実験を、表 1 に示した W, X, Y, Z を用いて表す。

## 3.1 データ量

授業のどの段階のデータであれば成績を予測できるのかを検証するため、小テストの 1 回目から最後まで全段階のデータを用いて予測し評価する。それぞれの実験における  $R^2$  上位 5 位までの構築用データと予測用データの量と  $R^2$  を表 2 に示す。データ量は小テストのログデータ数を表し、最大で 10 または 11 である。例えば、実験 W で構築用が 5、予測用が 6 の場合、2021 年度の初回から 5 回分の小テストのログデータを用いて構築し、2022 年度の初回から 6 回分の小テストのログデータを用いて予測、検証したことを表す。また、以降ではログデータの組み合わせを (構築用データ量, 予測用データ量) として (5, 6) などと表す。特徴量はそれぞれのデータの中で最も精度の高かった組み合わせを用いた。

表 2 より、構築用と予測用のデータ量が近い組み合わせ

1) 大阪教育大学 Osaka Kyoiku University

表 2  $R^2$  上位 5 件のデータ量の組み合わせ

順位	W			X			Y			Z		
	構築用	予測用	$R^2$	構築用	予測用	$R^2$	構築用	予測用	$R^2$	構築用	予測用	$R^2$
1	7	7	0.816	7	7	0.644	9	11	0.505	9	10	0.675
2	8	7	0.813	10	10	0.631	4	5	0.474	8	8	0.658
3	10	10	0.794	8	8	0.631	4	6	0.469	10	10	0.657
4	11	10	0.790	9	10	0.628	8	11	0.463	10	11	0.655
5	10	9	0.789	9	9	0.615	6	8	0.457	7	8	0.645

表 3 予測用データの量が 5 以下の  $R^2$  上位 5 件の組み合わせ

順位	W			X			Y			Z		
	構築用	予測用	$R^2$	構築用	予測用	$R^2$	構築用	予測用	$R^2$	構築用	予測用	$R^2$
1	5	5	0.763	5	5	0.493	4	5	0.474	5	5	0.507
2	6	5	0.738	3	4	0.416	3	5	0.436	4	5	0.481
3	4	4	0.693	4	5	0.364	2	4	0.254	3	4	0.393
4	7	5	0.679	6	5	0.328	1	3	0.180	4	4	0.334
5	5	4	0.664	3	5	0.303	1	2	0.116	3	5	0.316

表 4 データ量の組み合わせ上位 50 件における最も精度の高い特徴量の組み合わせの度数分布

W		X		Y		Z	
特徴量	度数	特徴量	度数	特徴量	度数	特徴量	度数
a,d	27	a,d	12	b,d	11	a,c,d	12
b,c,d	7	a,b,d	10	b,c	10	a,b,c,d	8
a,b,d	5	b,d	8	b,c,d	8	a,b,c	5
a,b,c,d	5	a,b,c,d	5	a,b,c,d	5	b,d	4
c,d	2	b,c,d	4	c,d	3	a,b,d	4

せが上位に多いことが分かる。4 つの実験の  $R^2$  上位 20 件の全 80 通りの中で、データ量の差が 1 以下である組み合わせは 49 通りであった。これは授業の内容が毎年同じであるため、同じ授業回までのデータを用いることで精度が高くなったと考える。W, X では最も精度の高い組み合わせが (7, 7) であり、全てのデータを用いる場合よりも高い精度で成績を予測できた。

また、20 位以内で最も少ない予測用データの量は 5 であり、その中でも (5, 5) の組み合わせは、W では 10 位で  $R^2$  が 0.763, X では 26 位で  $R^2$  が 0.493, Z では 17 位で  $R^2$  が 0.505 であり、高い精度で予測できた。よって W, X, Z では約半分までのデータを用いることで、精度を保って成績を予測できることが分かった。

Y は最も精度の高い組み合わせでも  $R^2$  が 0.505 であり、他の実験と比較して予測精度が低い。これは、学生の成績分布に変化があることが原因であると考えられる。2022 年度は他の 2 年度分と比較して最終成績の中央値と平均値が低かった。実験に用いた授業は、年度によって提供する資料や実施するテストの難易度に差がないため、受講生のレベルが下がったと考えられる。よって、2022 年度の評点の分布が変化したことから、2022 年度を用いた Y の予測精度が低かったと考える。

次に、5 回目よりも早い段階で成績を予測できるか評価するために、予測用データの量が 5 以下の  $R^2$  上位 5 件の組み合わせを表 3 に示す。表 3 より、予測用データが 4 または 5 の場合はどの年度も精度を保って予測できた。しかし、予測用データが 3 以下の場合、W 以外では  $R^2$  が 0.3 未満であり、予測できなかった。よって、4 回目または 5 回目以降の小テストのデータを用いること

で、精度を保って予測できた。すなわち、授業の半分以降のデータがあれば、成績を予測できる結果であった。

### 3.2 特徴量

3.1 節の実験における、データ量の組み合わせ  $R^2$  上位 50 件における、最も精度の高い特徴量の組み合わせの度数分布を表 4 に示す。

表 4 より、上位の特徴量の組み合わせに d が多く含まれることが分かる。詳しく調べると、4 つの実験において、上位 11 位までの全ての組み合わせに d が含まれていた。これは、d の受験回数の特徴量に、d 以外の小テストの得点率に関する特徴量を組み合わせることで、受講生の取組状況をより反映できたからであると考えられる。実際に、X で最も精度の高かった (7, 7) のデータ量の組み合わせにおける、d と a,d それぞれを用いた場合の  $R^2$  は、d 単体で約 0.223 であったのに対し、a を組み合わせることで約 0.421 向上し、約 0.644 であった。

### 4 おわりに

成績予測システムを構築し、更にそのシステムに用いる特徴量と、必要なデータ量について検証した。全ての実験において授業の半分程度のデータを用いることで、精度を保って成績を予測できた。また、精度の高い特徴量の組み合わせに d の特徴量が多く含まれた。

以上より、本稿で定義した特徴量を用いることで、授業の途中までのデータで成績を予測できることが分かった。今後の課題として、成績予測システムの UI の改善やフィードバックの拡張、利用者からの評価などが挙げられる。

#### 参考文献

- [1] 緒方広明：ラーニングアナリティクス：教育ビッグデータの分析による教育変革，ネクストコム：情報通信の現在と未来を展望する，KDDI 総合研究所，Vol.45，pp.12-21，2021.
- [2] Welcome to Flask — Flask Documentation (3.0.x)  
入手先：<<https://flask.palletsprojects.com/en/3.0.x/>> (参照 2024-5-27).
- [3] 駒谷優斗，望月久稔：Random Forest による成績予測とモデルに用いる特徴量とパラメータの検証，研究報告教育学習支援情報システム (CLE)，2023-CLE-40，pp.1-6，2023.