

教育 DX におけるデータ活用と個別最適化支援の実現

Data-driven education DX: personalized optimization and support

大政 悠暉^{*1}, 加島 智子^{*1}
Yuki OHMASA^{*1}, Tomoko KASHIMA^{*1}

^{*1} 近畿大学工学部

^{*1} Faculty of Engineering, Kindai University
Email: 2433850038d@hiro.kindai.ac.jp

あらまし：本研究では、教育 DX の進展を背景に、学生個人に適した最適化支援の可能性を探る。ポルトガル中等教育の生徒データを用いて成績を予測する。また、予測データに基づき支援優先度を評価し、教員の負担を軽減しつつリソースを最適活用する支援方法を提案する。

キーワード：教育 DX, 成績予測, 遺伝的アルゴリズム, 機械学習, 個別最適化

1. はじめに

近年、教育現場における DX の進展により、データを活用した教育改善への期待が高まっている⁽¹⁾。特に、学生個人に適した個別支援の提供は教育の質的向上に重要である。しかし、現在の教育現場では学生データが十分に活用されておらず、教員の負担増加や教員が学生指導に割ける時間などのリソースの最適活用が課題となっている。

従来研究ではデータ分析や成績不良者の早期検出、個別支援の提供などが主流である。本研究では、成績予測と遺伝的アルゴリズム(GA)を組み合わせ、リソースを考慮しつつ支援計画そのものを最適化する。

本研究では、学生データを活用し、成績予測モデルで学生の成績を予測する。また、支援優先度を評価し、教員の最適な行動を提案することで、教育現場におけるデータ活用の可能性を提示する。本研究は、教育 DX による個別最適化とリソース活用の改善を目指し、教育の質的向上に寄与するものである。

2. 教育 DX の活用の提案

本研究では、成績予測と最適化技術を組み合わせ、学生データを基に支援計画を策定するシステムを提案する。本研究の支援計画は(1)データ収集・使用、(2)成績予測、(3)個別可視化、(4)最適な支援計画の提案の順に進行する。

2.1 使用データ

本研究では、教育データ利活用ロードマップにおける多様なデータ活用方針を踏まえ、ポルトガル中等教育学校の公開学生データ⁽²⁾を採用する。データセットは幅広い属性を含み、多面的な支援計画に適している。本データセットには、性別や年齢などの基本情報、両親の経歴などの家庭環境、学習時間や課外活動などの学習・生活習慣、家族関係や自由時間、健康状態などの社会・健康、1~3 学期の成績データが含まれ、合計 33 項目存在する。

2.2 成績予測

第 3 学期の成績 (G3) を目的変数とし、それ以外

の属性を説明変数として Random Forest モデルを構築する。予測結果は、成績低下リスクの早期把握や支援の優先順位の算出に活用する。

2.3 現状の可視化

本研究では、学生の現状や支援優先度を明確に示すためにレポートを作成し、データの可視化を通じて効果的な支援計画の立案を支援する。

個人レポート

各学生の情報を提供するもので、学生の基本情報や支援項目とその優先度、成績推移、重要指標のグラフなどを提供する。このレポートにより、各学生の支援分野や優先度が明確になる。

クラスレポート

項目ごとの支援の必要人数や指標間の関係、ヒストグラムなどを提供し、クラス全体の傾向を視覚的に把握して支援方針の立案を支援する。

2.4 教員の行動の最適化

本研究では、教員のリソースを効率的に活用し、学生への支援効果を最大化するために、遺伝的アルゴリズム (GA) を用いて支援計画を最適化する。

モデル化

実際の支援行動につなげやすく、教員の支援で比較的改善が見込まれる指標を重要指標とし、勉強時間、平日/休日の飲酒量、健康状態、欠席日数、家族関係の 6 つを選定した。これらを 0~10 の範囲に正規化し、支援対象とする。GA の適応度関数は、成績予測値と支援優先度の平均向上量を組み合わせた式 1 とし、これを最大化するように最適化を行う。

$$Fitness = \frac{\alpha}{N} \sum_{i=1}^N \Delta G3_i + \frac{\beta}{N} \sum_{i=1}^N \Delta Priority_i \quad (1)$$

ここで、 N は学生数、 $\Delta G3_i$ と $\Delta Priority_i$ は学生 i の予測成績と支援優先度の向上値、係数 α と β はそれぞれの重要度を調整するパラメータである。支援優先度は「Study Support」「Life Guidance」「Interview」「Healthcare」の 4 項目で構成し、これらを 0~5 に

正規化した値の合計とする。具体的には重要指標とその平均値との差分を用い、値が平均以下のときの平均との差分によって算出する。また、制約条件は以下の通りである。

1. 支援の合計が教員のリソースを超えない。
2. 改善後の指標が 0~10 の範囲を超えない。

アルゴリズムのフロー

支援計画の最適化は、以下の手順で進める。

1. データの準備
重要項目を正規化し、支援優先度を計算する。
2. 初期解の生成
少しずつステップを割り当て、制約条件を満たす初期解をランダムに生成する。
3. 評価関数の計算
解の適合度を算出し、リソース超過分に応じてペナルティを付与する。
4. 選択・交叉・突然変異
エリート選択とトーナメント選択を用いて、適合度が高い個体を選択する。また、一様交叉を用いて新たな解を生成し、突然変異として各指標を±1の範囲で変化させる。
5. 終了条件
1000世代が経過する、または100世代適合度が改善しない場合に終了とする。

3. 結果

3.1 成績予測結果

本研究で使用するデータセット 395 件のうち 325 件でモデルを作成し、残りの 70 件を予測した結果は表 1 の通りである。結果より、おおむね高精度に成績の予測ができていると考えられる。

表 1 成績予測スコア

項目	結果
Mean Absolute Error (MAE)	1.11
Mean Squared Error (MSE)	3.49
R ² Score	0.83

3.2 個人/クラスレポート

クラスレポートの結果は図 1、個人レポートの結果は図 2 の通りである。クラスレポートより、本クラスは学習支援及び健康支援を必要としている人が多いことが分かる。また、個人レポートの学生は、他と比較して Health の値が低く、ヘルスケアを特に必要としていることが分かる。

【必要支援の状況】

支援内容	人数	必要としているID
学習支援	34	78,371,234,316,393,124,334,15,291,0,113,307,70,384,137,73,25,72,141,158,196,39,386,132,168,193,157,177,255,358,82,18
学業指導	12	390,234,113,307,384,25,46,304,132,5,82,18
面談	6	391,140,25,141,46,108
保健/健康支援	27	55,390,223,316,116,393,155,324,15,391,33,76,384,72,141,93,46,108,39,377,193,157,289,18

図 1 クラスレポート (一部抜粋)

【レポート優先度】 (各項目0~5)

レポート項目	優先度(数値)
Study support	0.49
Life guidance	1.54
Interview	0.00
Healthcare	2.00
合計(割合)	4.03

【グラフ】

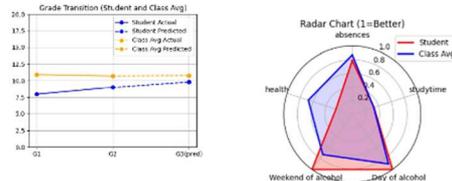


図 2 個人レポート (一部抜粋)

3.3 最適化結果

成績予測を行った 70 名の学生を対象に、ランダムに改善を行った場合と GA を用いた支援計画の立案を行った結果が表 2 の通りである。

表 2 最適化結果

ランダムな改善による提案			
指標	改善前	改善後	差分
G3 の平均予測成績	10.77	10.94	0.17
支援優先度の平均	3.94	1.94	2.00
遺伝的アルゴリズムによる提案			
G3 の平均予測成績	10.77	10.83	0.06
支援優先度の平均	3.94	2.38	1.55

最適化結果より、ランダムな改善と比較して、提案手法は G3 の平均予測成績および支援優先度の平均ともにより上昇しており、効果的な支援計画の立案を行うことができていることが分かる。

4. おわりに

本研究では、学生の多様なデータを用いて学生の成績を予測した。また、レポートにより学生の現状を可視化し、効果的な支援の基盤を構築した。さらに、GA を活用した支援計画の最適化により、成績や支援優先度の向上が確認された。本研究は教育 DX による個別最適化支援の可能性を示すとともに、教育の質的向上や教員の負担軽減に寄与するものである。今後は、実際の教育現場のデータを用いた検証や既に成績の高い学生のさらなる伸長支援、コスト設定の実現化などが課題となる。また、長期的な学習効果の検証やプライバシーへの配慮も含め、教育現場との連携を進める必要がある。

参考文献

- (1) デジタル庁、総務省、文部科学省、経済産業省: “教育データ利活用ロードマップ”, https://www.digital.go.jp/assets/contents/node/information/field_ref_resources/0305c503-27f0-4b2c-b477-156c83fdc852/20220107_news_education_01.pdf (参照 2025.1.27)
- (2) Kaggle: “Student Performance Data Set”, <https://www.kaggle.com/datasets/impapan/student-performance-data-set> (参照 2025.1.27)