

1. はじめに

教育環境のデジタル化に伴い、学生の学習行動データの大規模な収集が可能となった。これらの学習行動を分析し、成績が良い学習パターンとの相違を発見することで成績不良者の学習行動の改善につながる事が期待されている。文献 [1] では、Attention ベースの機械学習モデルである Transformer[2] を用いて成績予測を行い、Attention の可視化により重要な講義や操作を分析している。そこで本研究では、Transformer による成績予測、重要な教材の特定に加え、学習行動の改善を図る。具体的には学習した Self-Attention の Attention weight を解析することで、学習者の行動パターンの傾向を捉え、学習者にフィードバックすることを目指す。

2. 提案手法

本研究では、分類型の Transformer による成績予測を行う。またデータ拡張を行い精度向上を図る。さらに、Attention weight の可視化により、成績ごとの傾向調査を行い、傾向の違いから成績不振の生徒の改善を図る。

2.1. データセットと特徴量作成

実験には、九州大学で行われた講義を受講した学生の学習行動データを用いる。学習行動データとは、電子教材を閲覧した際の操作ログである。操作ログとして収集できる情報は、教材を開く、閉じる、ページ送り、戻り、メモ、マーカーなど 21 種類あり、操作した時間も含まれる。また、各学生の成績も含まれており、成績は A から F の 5 段階である。本実験では、2020 年から 2022 年の 3 年分、165 人の学習行動データを使用して成績予測を行う。入力する特徴量には 12 回の講義資料それぞれに対する各操作の頻度をベクトルとして用いる。

2.2. データ拡張

2.1 節で述べたように、データセットの人数が 165 人とデータ数が少ない。データ数が少ないと汎化性能が低くなるため、データ拡張を行い汎化性能の向上を図る。

データ拡張には Mixup や CutMix を使用する。Mixup の新しいサンプルは式 (1) で算出する。ここで、 $\mathbf{X}$, $\mathbf{y}$ は入力と正解ラベル、 λ は混合比である。混合比 λ はベータ分布を用いてランダムに決定する。

$$\mathbf{X} = \lambda \mathbf{X}_1 + (1 - \lambda) \mathbf{X}_2, \mathbf{y} = \lambda \mathbf{y}_1 + (1 - \lambda) \mathbf{y}_2 \quad (1)$$

また、CutMix は式 (2) で算出する。ここで、 $\mathbf{M}$ はバイナリ行列である。

$$\mathbf{X} = \mathbf{M} \odot \mathbf{X}_1 + (1 - \mathbf{M}) \odot \mathbf{X}_2, \mathbf{y} = \lambda \mathbf{y}_1 + (1 - \lambda) \mathbf{y}_2 \quad (2)$$

2.3. Attention weight の可視化

成績を正しく予測できた学生について、成績ごとの Attention weight の平均を (3) 式で算出する。ここで、 Q , K は Self-Attention の Query と Key、 d_k はその次元数であり、 n は分類に成功したそれぞれの成績の人数である。

$$Attention(Q, K) = \frac{1}{n} \sum_i softmax(\frac{Q_i K_i^T}{\sqrt{d_k}}) \quad (3)$$

3. 精度比較実験

本実験では、Transformer のエンコーダを用いる。ヘッド数 1、埋め込み次元数 128 のエンコーダを 4 層積層する。評価手法として 3 年のうちの 1 年を評価データ、残りを訓練データとして交差検証を行い、分類精度の平均と標準偏差を求める。また、訓練データに対してデータ拡張を適用し、モデルの訓練を行う。ここで、Mixup の適用率は 20%、CutMix の適用率は 40% とし、両手法を用いる場合は適用率を 10% とする。

データ拡張を用いて訓練した Transformer の分類精度を表 1 に示す。データ拡張後の成績分類の精度は Mixup を適用した際に 2.52pt 向上し、データ拡張の有効性を確認することができる。

表 1: Transformer の分類精度

Mixup	CutMix	精度 [%]
-	-	43.28 $\pm$ 8.44
✓	-	45.80 $\pm$ 7.91
-	✓	43.24 $\pm$ 5.87
✓	✓	44.54 $\pm$ 7.89

4. Attention weight の分析

Self-Attention の Attention weight を用いて成績の良い生徒と悪い生徒の分析を行う。ここで、Attention weight の値が大きい教材は成績を予測する上で重要度が高いデータとなる。

4.1. 成績の傾向分析

成績ごとに予測が成功した学生の Attention weight の平均を求め、傾向の分析を行う。2020 年、2021 年のデータで訓練・評価を行った際に正しく予測できた学生の成績ごとの Attention weight の平均を図 1 に示す。図 1 より、成績の良い学生は 12 番目の教材を注視し、成績の悪い学生は 7 番目から 12 番目の教材のように広範囲を注視している傾向があることが分かる。

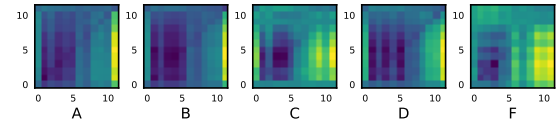


図 1: 成績ごとの Attention weight の平均

4.2. 成績不振の学生の改善

4.1 節で得られた傾向より、成績不振の各学生について Attention weight が大きい教材の操作回数を増加させて、擬似入力データを生成する。擬似データで再度予測を行い、分類結果が向上すれば、その操作が有効であると判断できる。12 番目の教材を除いて、Attention weight が大きい 6 つの教材について、操作回数を 5 倍に増やし、再度予測を行った。

各手法の再分類結果を表 2 に示す。CutMix が最も分類結果の向上する人数が多く、成績 D の学生の 50% の改善が確認できた。また、分類結果が A, B になった 2 人の学生の Attention weight を図 2 に示す。分類結果が向上した学生の Attention weight は、特定の教材を注視しているため成績が不振であったことが分かる。

表 2: 成績 D の学生の再分類精度 [%]

手法	A	B	C	D	F
データ拡張なし	21.05	21.05	0.00	57.90	0.00
Mixup	15.79	26.31	0.00	57.90	0.00
CutMix	15.00	35.00	0.00	50.00	0.00
Mixup&CutMix	20.00	20.00	0.00	60.00	0.00

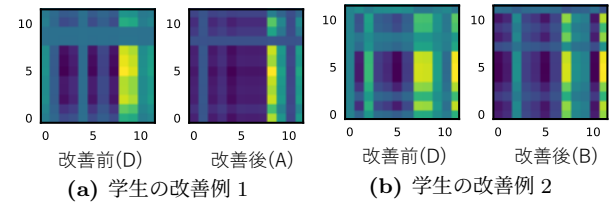


図 2: 成績が改善した学生の Attention weight の比較

5. おわりに

本研究では、Transformer による成績分類を行った。Mixup を行うことで成績分類精度が 2.52pt 高い結果となり、汎化性能の向上が確認できた。さらに、Attention weight の可視化により、成績が良い生徒、悪い生徒の傾向の違いを発見でき、それを活かした学習の改善方法を示した。今後は、成績予測に適した入力特徴について検討する。

参考文献

- [1] A. Satake *et al.*, “Performance Prediction and Importance Analysis Using Transformer”, ICCE, 2021
- [2] A. Vaswani *et al.*, “Attention is all you need”, NIPS, 2017