

1. はじめに

Learning Analytics (LA) とは、教育分野において学習者や指導者の活動などのデータを分析し、有益な情報の取得及び利用を目的としたデータサイエンスである。教育現場でのデジタル学習環境やオンライン教育システムの普及により、膨大なユーザの学習活動データの収集が可能になっている。そのデータを分析することで、ユーザの学習パターンから成績を予測し、ユーザへのサポートに活かすことができる。

デジタル学習プラットフォームで収集されたデータから学生の成績を予測した研究例として、Okubo らの研究がある [1]。この研究では大学の講義時間内における学習活動から特徴量を作成し、時系列情報を考慮できる深層学習モデルである Recurrent Neural Network (RNN) によって成績の評価レベルを予測している。しかし、この手法では自宅学習など講義時間外の学習活動の影響を考慮できず、また RNN ではどのような活動が予測に影響を与えているかといった分析は困難である。そこで本研究では、多数の特徴量からの特徴量選択とアテンションを用いた深層学習手法である Transformer [2] を用いたモデルによって、学生の学習活動データからの成績予測を行う。

2. 関連研究

2.1 特徴量選択

機械学習における特徴量選択手法には大別して、統計的手法によって特徴量の評価を行って選択する Filter Method、特徴量を入れ替えながらモデルの学習を行って最適な特徴量を選択する Wrapper Method、機械学習における重要度や重みなどの値で選択を行う Embedded Method がある。Filter Method には特徴量の分散のサイズや目的変数との相関係数を利用した選択、Embedded Method は決定木モデルの feature importance を利用した選択などがある。

2.2 Transformer

Transformer は、自然言語処理分野で発表された、再帰や畳み込みを用いずに時系列情報を考慮できるようにしたエンコーダ-デコーダモデルである。Transformer のエンコーダ構造として、まず入力シーケンスの各要素に位置情報を付加する Positional Encoding がある。ブロックは、複数のヘッドそれぞれで Self-Attention を計算する Multi-Head Attention、シーケンスごとに位置情報を整える Feed-Forward Network、残差結合と正規化を行う Add & Norm 層の 3 種類のサブレイヤーで構成される。Transformer では Self-Attention によって時系列情報などの入力シーケンスの関係性を考慮できる。Self-Attention は入力シーケンスごとに作成した 3 つのベクトル Query, Key, Value の内積によって求められる。

3. 提案手法

本研究では、講義時間外の学習活動を含めた多数の特徴量から、特徴量選択によって予測モデルの学習に有効な特徴量を選択し、モデルの学習に用いることを提案する。学習活動を表現する多数の特徴量を作成し使用することで、モデルはより詳細な学習活動のパターンを把握できると考えられる。しかし、特徴量全てを使用すると計算コストが膨大になるため、予測に有効な特徴量を選択することで効率的なモデルの学習を行う。提案手法による成績予測の処理を図 1 に示す。

3.1 特徴量選択

特徴量選択手法には、Filter Method として特徴量の分散の大きさ、成績との Mutual information, Embedded Method として Lasso 回帰の重み, LightGBM の feature importance の計 4 種類の手法を用いる。なお Wrapper Method については、選択するための学習自体に膨大な計算コストがかかるため今回は使用しない。

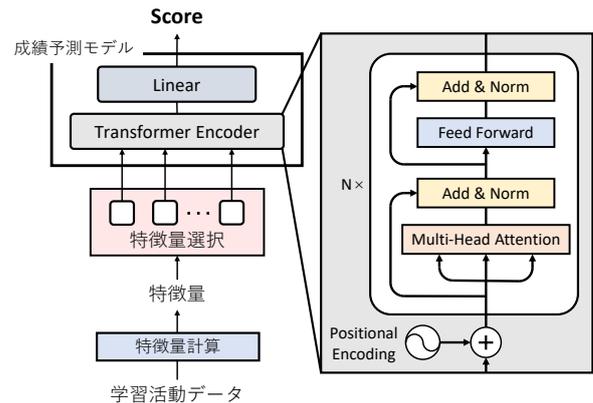


図 1: 提案手法による予測までの処理

3.2 Transformer を用いた予測モデル

予測モデルには、Transformer のエンコーダ構造を利用したモデルを用いる。これにより、入力同士の関係性を考慮でき、また入力について Self-Attention による分析が可能となる。特徴量選択により選択された特徴ベクトルを入力として Transformer のエンコーダによって潜在的な特徴を学習し、全結合層によって成績を出力する。

4. 提案手法による予測実験

提案手法を用いて、4 つの成績予測実験を行う。1 つ目は、特徴量選択手法の有効性評価である。4 種類の特徴量選択手法を用いた際の予測モデルでの予測精度を比較する。2 つ目は、講義時間外の学習活動データの有効性評価である。講義時間内の学習活動データのみから作成した特徴量と、講義時間外を含めた全時間帯の学習活動データから作成した特徴量とで成績予測を行う。3 つ目は Transformer を利用したモデルと、LA 分野の成績予測タスクにおいて用いられている従来手法のモデルとの比較である。従来手法として、LSTM, MLP, 重回帰分析を用いる。4 つ目は成績の早期予測である。一連の講義の進行途中を想定し、使用する特徴量を制限した成績予測を行う。

4.1 データセット

実験には、九州大学で行われた情報系科目の一連の講義における学習活動データから作成した特徴量を使用する。講義は 8 回行われ、各回で異なる講義資料が電子教材として使用された。また、全講義終了後には 100 点満点のテストが実施された。本実験では、テストの点数を講義における成績として扱う。受講した学生 1300 人のうち、学習データとして 1,200 人分、評価データとして 100 人分のデータを使用する。

学習活動データは、M2B 学習支援システム [3] によって収集された。M2B 学習支援システムは九州大学で利用されているデジタル学習プラットフォームで、学生の電子教材の操作ログなどの学習活動データを記録している。操作ログには、学生が取ったアクションの種類、アクション対象のページ番号、タイムスタンプなどが記録されている。アクションの種類には、ページ送りの「NEXT」や戻りの「PREV」、文字にハイライトを追加する「ADD MARKER」など計 16 種類がある。また、電子教材の使用は講義時間外も可能なため、自宅学習など講義時間外の操作ログも収集される。

特徴量として、講義時間外を含めた全時間帯の学習活動データから各ページにおける各アクション回数と各ページの閲覧時間の特徴量を作成し、また講義時間内の学習活動データのみから 5 分ごとの各アクション回数の特徴量を作成した。これらの特徴量は合計で 9,555 次元となっており、各次元で正規化を行った。

4.2 特徴量選択手法の有効性評価

特徴量選択手法として分散の大きさ, Mutual information, Lasso 回帰, feature importance を用いて各手法の比較を行う. 9,555 次元のうち 128 次元を選択し, 選択した特徴ベクトルを用いて予測モデルの学習と評価を行う.

200 epoch の学習を 5 回ずつ行ったときの, 評価データに対するテスト点数予測値の平均平方二乗誤差 (RMSE) の平均を表 1 に示す. 4 つの手法の中で, Lasso 回帰によって選択された特徴量を用いた場合が最も RMSE が小さかった. このことから, Lasso 回帰が最も予測に有効な特徴量を選択できているといえる.

表 1: 各特徴量選択手法による予測値の RMSE

特徴量選択手法	RMSE
分散の大きさ	9.20±0.43
Mutual information	9.40±0.67
Lasso 回帰	8.90±0.45
feature importance	9.10±0.38

また, 各特徴量選択手法でどのような特徴量が選択されたか調査した. 図 2 に選択された各アクションに関する特徴量数のグラフを示す. 最も予測精度が低かった Mutual information では, ページ送りである「NEXT」のアクションに関する特徴量が集中的に選択されており, 16 種類あるうちの 6 種類のアクションからしか特徴量が選択されていなかった. それに対し最も精度の高かった Lasso 回帰は, 他手法では選択されなかった 6 種類を含めた 14 種類のアクションから特徴量が選択されていた. 予測の際に様々なアクションの特徴量を用いたことで学習活動パターンをより細かく把握でき, 精度が高くなったのではないかと考えられる.

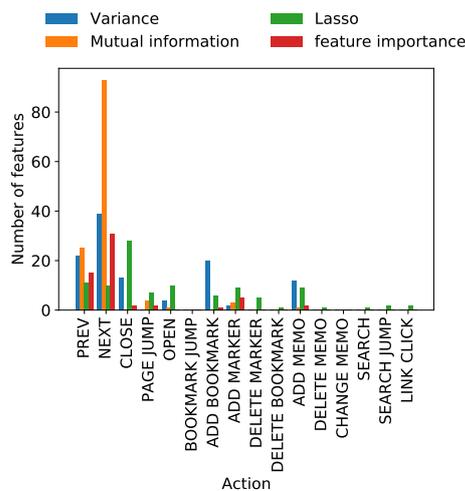


図 2: 各手法で選択されたアクションに関する特徴量

4.3 講義時間外の学習活動データの有効性評価

講義時間外を含めた特徴量と, 講義時間内のみの特徴量による予測との比較を行う. 各講義内で行った 16 種類のアクションの回数を講義時間内のみの特徴量とし, 計 128 次元を Transformer モデルに入力する. 講義時間外を含めた特徴量としては, 4.2 節と同じく 9555 次元の特徴量から特徴量選択によって 128 次元の特徴ベクトルを選択する. 選択手法には Lasso 回帰を用いる.

5 回ずつ学習を行ったときの, 予測値の RMSE の平均を表 2 に示す. 講義時間内のみの場合と比較して, 講義時間外を含めた特徴量を用いた場合に RMSE が低くなった. 結果から, 講義時間外の学習活動データが予測精度向上に有効であると考えられる.

表 2: 各特徴量選択手法による予測値の RMSE

特徴量	RMSE
講義時間内のみ	9.10±0.69
講義時間内+講義時間外	8.90±0.45

4.4 従来手法との比較

成績予測タスクにおいて主に用いられる LSTM, MLP, 重回帰分析を用いた予測との比較を行う. これらのモデルの特徴量には, 16 種類のアクションを各講義内で行った回数, 128 次元の特徴ベクトルを使用する.

5 回ずつ学習を行ったときの, 予測値の RMSE の平均を表 3 に示す. 従来手法と比較して Transformer を利用したモデルの RMSE が最も低く, 高精度な予測ができることが分かった.

表 3: 各モデルによる予測値の RMSE

予測モデル	RMSE
LSTM	9.34±0.52
MLP	9.37±0.53
重回帰分析	10.02±0.52
Transformer	8.90±0.45

4.5 成績の早期予測

予測結果を教育現場にフィードバックすることを考えたとき, より早期の段階で成績を予測したり, その時々での予測結果を示すことで, 適切な支援に活かせることが望ましい. そこで, 一連の講義の進行途中を想定した, 成績の早期予測実験を行う. 作成した特徴量 9,555 次元のうち, 特定の講義回までに使用された講義資料に関する特徴量から特徴量選択を行い, Transformer モデルの学習および評価に使用する. 例えば, 4 回目の講義時点での予測では 1~4 回目で使用された講義資料に関する特徴量から選択を行う.

各特徴量選択手法を用いて学習を行ったときの 1~8 回目までの RMSE の推移を図 3 に示す. いずれの手法を用いた場合でも概ね講義が進むにつれて RMSE が小さくなった. また, Lasso 回帰は全体を通じて RMSE が最も小さいか僅差で 2 番目に小さく, 早期予測においても高い精度であることが分かった.

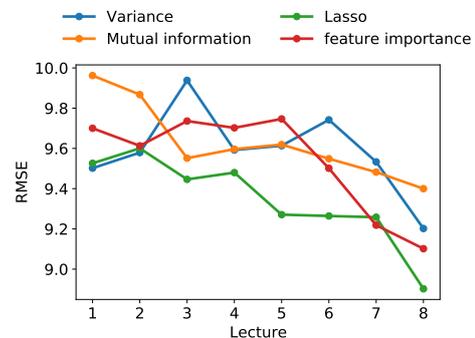


図 3: 各特徴量選択手法での RMSE の推移

5. おわりに

本研究では, 特徴量選択と Transformer を利用したモデルによるテスト点数の予測を行った. 特徴量選択手法としては Lasso 回帰が最も精度が高くなった. また, 成績予測タスクにおける従来手法と比較して, Transformer を利用したモデルは高精度の予測結果が得られた. 今後は, 作成する特徴量の改善とより効果的な特徴量選択を検討する.

参考文献

- [1] F.Okubo, *et al.*, “A Neural Network Approach for Students’Performance Prediction”, LAK17, 2017
- [2] A.Vaswani, *et al.*, “Attention is all you need”, NIPS, 2017
- [3] H.Ogata, *et al.*, “M2B System: A Digital Learning Platform for Traditional Classrooms in University”, LAK17, 2017

研究業績

- [1] A.satake, *et al.*, “Performance prediction and importance analysis using Transformer”, ICCE, 2021 (他 1 件)