

学習履歴を用いたデータ解析におけるデータ欠損の補完の検討

D-15 Investigation of Lacking Data Complementing Method on Learning Analytics

甘利 美季† 小川 賀代†

Miki AMARI† Kayo OGAWA†

† 日本女子大学理学部

† Faculty of Science, Japan Women's University

1. はじめに

新型コロナウイルス感染症拡大の影響により、オンライン授業の導入が急速に進んだ。オンライン授業は場所や時間の自由度が高い反面、学習意欲の維持や意思疎通が困難であることからドロップアウトしてしまう学生が発生するという課題がある。これまでに、学習管理システムに蓄積された学習履歴データからドロップアウトの兆候がある学生を早期発見し支援を行うための方法として、主成分分析 (PCA) 及び自己組織化マップ (SOM) を用いた解析手法について提案してきた。PCA からは時間に関するデータが特に着目すべきであること、SOM からはドロップアウト兆候が概ね授業の中盤で検知できることが明らかになっている^[1]。本研究では、これまで十分に検証ができていないデータ欠損の取り扱い手法についての検討を行う。また、本解析手法の汎用性についても評価を行う。

2. 解析について

2.1 解析手法

本解析では PCA 及び SOM を用いている。PCA とは、多量の説明変数を集約することでより少ない合成変数にする手法である。この結果は主成分得点と主成分負荷量で評価され、変数の関係性や影響力がわかる。また、SOM とは入力層と出力層の 2 層構造からなる、教師なし学習ニューラルネットワークであり、データの位相を保ちながら 2 次元まで圧縮する。この結果は 2 次元の図として出力される。

2.2 解析に用いるデータ概要

本解析手法の汎用性についても評価を行うために、解析手法の提案^[1]時とは異なる学習履歴データを用いる。今回は、筆者らが所属する数物科学科 2 年次の選択必修科目である「光情報処理」の学習履歴を用いた。本講義は全回オンライン形式で実施し、出席確認のための小テスト (全 12 回) を、授業終了後 24 時間以内に受験することを課している。受講者数は 48 名であり、使用学習履歴データは、小テスト各回の点数、解答に要した時間、着手するまでの時間と定期試験の点数の全 37 変数である。

3. 解析結果

3.1 データ欠損に対する取り扱い手法の検討

小テスト未提出者には、解答に要した時間、着手までの時間のデータが存在しない。これまで、欠損データは、平均代入法によって補完して解析を行っていたが、一部、成績上位者が下位者に混ざることがあった。そこで今回は、データ欠損値を他の補完手法によって、上位者と下位者の分離がより明確になるか検討を行った。今回は、表 1 に示す 5 種類の方法について比較検討を行った。パターン①がこれまでの手法である。パターン②～⑤は、未提出者の特性をより強調するために、着手までの制限時間である 24 時間を上回

る値を入力した。各パターンで SOM を 50 回実施したときの結果を表 2 に示す。いずれの方法も、下位者の中に学習ログが似ている上位者が 1 名含まれてしまった。これを許容した分離では、従来方法であるパターン①が、90 % と高い確率で分離することができた。

3.2 解析手法の汎用性に対する評価

本解析手法の汎用性の評価として、新たな学習履歴データを用いて、ドロップアウト兆候が検知可能かどうかの検討を行った。全授業 12 回を 3 回、4 回、6 回ごとに区切り、SOM で 50 回ずつ解析を行った。その結果、3 回ごとに区切ったうちの第 4～6 回で、成績上位者と下位者のグループの分離が最も高い頻度で確認でき、その確率は 74 % であった。図 1 は、第 4～6 回の結果である。このことから、授業中盤付近で学生のドロップアウト兆候が検知可能であるといえる。以上の結果はこれまでに行った解析の結果と概ね一致しており、これまでに行ってきた解析の手法は、他の科目においても有効であると推測される。

4. まとめ

本研究では、これまでに行ってきた PCA と SOM を用いた解析手法が他の科目においてもある有効であるという結果が得られた。また、時間に関するデータ欠損に対する補完については、平均代入法で得られた値を代入する方法が適切であるという結果が得られた。今後、SOM によって得られた結果に対し下位者の判定を定量的に行うことで、ドロップアウト兆候がみられる学生の特定ができると推察する。

参考文献

- [1] 藤巻星香, 小川賀代, “オンライン授業における学習履歴を用いた学習支援の検討”, 研究報告教育学習支援情報システム, 2021-CLE-33, 14, 1-5 (2021).

表 1 検討するデータ欠損の取り扱い手法

	着手までの時間	解答に要した時間
パターン①	平均代入法	平均代入法
パターン②	25 時間	平均代入法
パターン③	25 時間	25 時間
パターン④	48 時間	平均代入法
パターン⑤	48 時間	48 時間

表 2 SOM の解析結果

	分離できた回数(回)
パターン①	45
パターン②	32
パターン③	33
パターン④	30
パターン⑤	36

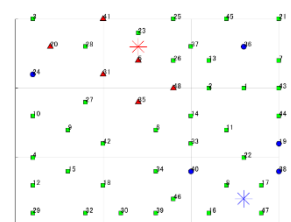


図 1 SOM の出力結果 (小テスト第 4～6 回)