

ドロップアウト兆候者早期検出方法の汎用性の評価 Evaluation of the versatility of early detection method for dropout students

小糸 麻結[†]
Mayu Koito

小川 賀代[†]
Kayo Ogawa

1. はじめに

コロナ禍を経て、高等教育の授業形態は多様化している[1]。またデジタルを活用した教育高度化に向けての取り組みは全国の大学で普及している[2][3]。一方で授業についていけず、ドロップアウトの可能性のある学生が一定数いるという課題は依然として残っている。これまでも学習履歴データを用いた解析は多数行われているが[4]、解析手法や解析データの一般化には至っていない。

本研究は、一般的に入手可能な学習履歴データを比較的理解しやすいアルゴリズムで記述された解析手法で可視化し、ドロップアウトの可能性のある学生を早期発見・支援することを目的としている。著者らは、これまでに講義型授業で課される小テスト課題の学習履歴データを解析し、ドロップアウトの可能性のある学生の検出方法の提案およびその有効性を確認している[5]。

本論文では、小テスト課題の学習履歴データに加え、講義型授業で課されるレポート課題の学習履歴データの解析を行い、ドロップアウトの可能性のある学生が検出可能かどうかを検討する。また、小テスト課題の学習履歴データ解析結果とレポート課題の学習履歴データ解析結果を比較し、課題の種類が異なる授業において、ドロップアウト兆候者の検出方法に汎用性があるかについても評価を行う。

2. 解析データセットと解析手法

解析に使用したデータは、理学部数物情報科学科 1 年次向けの必修科目である「情報科学概論」の学習履歴データ 2021 年度・2022 年度分と同学科 3 年次向けの選択必修科目である「マルチメディアの基礎」の学習履歴データ 2022 年度・2023 年度分である。これらのデータは本学で使用している学習管理システム (LMS : Learning Management System) の一種である manaba から抽出した。表 1 に解析に使用したデータの受講者数を示す。

表 1 解析に使用したデータの受講者数

	2021 年度	2022 年度	2023 年度
情報科学概論	100 名	110 名	
マルチメディアの基礎		24 名	45 名

「情報科学概論」では計 13 回の小テスト課題を授業終了後 24 時間以内に提出することを課しており、第 14 回目の授業で期末試験を実施している。「情報科学概論」のデータ解析において変数に採用した学習履歴データは、小テスト課題各回の「点数」「取り掛かるまでの時間」「解くためにかかった時間」に加え、「期末試験の点数」である。

「マルチメディアの基礎」では、講義期間内に複数回、レポート課題を課している。2022 年度は計 8 回 (第 2 回、

第 4 回、第 5 回、第 8 回、第 9 回、第 10 回、第 13 回で実施、第 13 回では 2 つのレポート課題が課せられていた)、2023 年度は計 6 回 (第 13 回は未実施) 行われており、授業終了後 1 週間以内に提出するよう設定されている。こちらのデータも第 14 回目の授業で期末試験を実施している。「マルチメディアの基礎」のデータ解析において変数に採用した学習履歴データは、レポート課題各回の「点数」「受付開始時間と提出時間の差」に加え、「期末試験の点数」である。

いずれのデータセットも単位の異なるデータが含まれていることから、データはすべて基準化し、解析に用いた。

解析手法には、自己組織化マップ (SOM : Self-Organizing Map、以下 SOM と表記) を採用した。SOM は高次元のデータの位相を維持しつつ、2 次元上に表すことができるため、学生の学びの傾向を視覚的に捉える上で有効な手法である[6]。各検討における SOM のシミュレーション諸元を表 2 に示す。今回は SOM の結果図から、成績上位者の集団と成績下位者の集団がそれぞれ固まって分布しているかどうか、すなわち成績上位者の集団と成績下位者の集団が分離しているかどうか、目視で判別を行った。成績上位者の集団と成績下位者の集団が分離できた場合の SOM 結果図例を図 1 に示す。赤い三角で示されている成績上位者は上部に、青い丸で示されている成績下位者が下部に固まって分布していることが確認できる。また SOM は初期値に乱数を用いているため、その影響を考慮し、各検討でそれぞれ 50 回ずつ試行することとした。

表 2 SOM シミュレーション諸元

学習回数	2000 回 (小テスト課題解析) 1000 回 (レポート課題解析)
学習率	0.1
マップ数	15×15
データの分類	期末試験の点数をもとに分類 成績上位者 (▲) 成績下位者 (●) それ以外 (■)

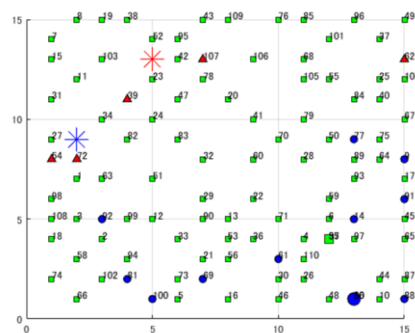


図 1 SOM 結果図例

[†] 日本女子大学大学院 理学研究科

Graduate School of Science, Japan Women's University

3. 欠損値補完方法の検討

課題未提出者はログデータが得られないため、欠損値になる。本論文では、2つの欠損値補完方法、パターンAとパターンBを比較しながら、最適な欠損値補完方法の検討を行った。パターンAでは、点数に0点、時間に関するログに平均値を代入し、補完した。パターンBでは、点数に0点、時間に関するログに課題受付期間の2倍の時間（小テスト課題の場合は48時間、レポート課題の場合は2週間）を代入し、補完した。パターンBでは、課題未提出者の特性をより強調するため、あえて課題受付期間を大幅に上回る時間を代入した。2021年度、2022年度小テスト課題ログデータの解析結果を表3、表4、2022年度、2023年度レポート課題ログデータの解析結果を表5、表6に示す。

表3 2021年度 小テスト課題ログデータ 解析結果

欠損値補完方法	分離できた割合 [%]	鮮明に分離できた割合 [%]
パターンA	86.00	10.00
パターンB	58.00	26.00

表4 2022年度 小テスト課題ログデータ 解析結果

欠損値補完方法	分離できた割合 [%]	鮮明に分離できた割合 [%]
パターンA	96.00	16.00
パターンB	72.00	48.00

表中の「分離できた割合」とは、成績下位者の集団の中に成績上位者が1~2名含まれることを許容したときの割合であり、「鮮明に分離できた割合」とは、成績下位者の集団の中に成績上位者が含まれることを許容しないときの割合である。

表3、表4より小テスト課題ログデータでは、パターンAの方が分離できた割合が高いことが確認できる。鮮明に分離できた割合が高い補完方法はパターンBであった。

表5 2022年度 レポート課題ログデータ 解析結果

欠損値補完方法	分離できた割合 [%]	鮮明に分離できた割合 [%]
パターンA	70.00	14.00
パターンB	68.00	28.00

表6 2023年度 レポート課題ログデータ 解析結果

欠損値補完方法	分離できた割合 [%]	鮮明に分離できた割合 [%]
パターンA	98.00	94.00
パターンB	98.00	98.00

表5より2022年度レポート課題ログデータでは、パターンAの方が分離できた割合が高いことが確認できる。表6より2023年度レポート課題ログデータでは、パターンAとパターンBは同じ割合で分離できたことが確認できる。一方、鮮明に分離できた割合が高かった補完方法は表5、表6ともにパターンBであった。また表5と表6を比較すると、2022年度レポート課題ログデータは分離できる割合が全体的に低くなっていることがわかる。2022年度レポート課題ログデータでは、成績上位者はSOM上で近い位置

で分布していたのに対し、成績下位者は離れた位置に分布していたケースが多く確認できた。普段のレポート課題では点数・時間ともに問題がなかった成績下位者とレポート課題の時から提出する時間が遅い傾向にあった成績下位者がいたため、成績下位者の中で課題の取り組み方が大きく異なっていたことが原因であると考えられる。

よって表3から表6の結果を総合すると、分離できた割合が高い補完方法はパターンA、鮮明に分離できた割合が高い補完方法はパターンBであることがわかった。またパターンAで成績上位者の集団に含まれることの多かった成績下位者とパターンBで成績上位者の集団に含まれることの多かった成績下位者は異なっていたため、それぞれの補完方法の特性を活かすことで、より多くのドロップアウト兆候者を検出できると考えた。よって今後の検討では、パターンAで補完したデータとパターンBで補完したデータの両方で検討を行うこととした。

4. ドロップアウト兆候の検知可能時期の検討

ドロップアウト兆候者を最も検知できる授業回の検討を行った。2種のパターンで補完した、小テスト課題実施回ログデータとレポート課題実施回ログデータを時系列に沿って3授業回ごとに区切り、解析した。講義期間中のなるべく早い時期での支援を目的としているため、課題実施最終回を含むパターンについては解析を行わなかった。2021年度、2022年度小テスト課題ログデータの解析結果を表7、表8、2022年度、2023年度レポート課題ログデータの解析結果を表9、表10に示す。

表7 2021年度 小テスト課題ログデータ 解析結果

欠損値補完方法	データパターン	分離できた割合 [%]
パターンA	第1回~第3回	58.00
	第4回~第6回	4.00
	第7回~第9回	26.00
パターンB	第1回~第3回	82.00
	第4回~第6回	22.00
	第7回~第9回	10.00

表8 2022年度 小テスト課題ログデータ 解析結果

欠損値補完方法	データパターン	分離できた割合 [%]
パターンA	第1回~第3回	34.00
	第4回~第6回	8.00
	第7回~第9回	16.00
パターンB	第1回~第3回	52.00
	第4回~第6回	12.00
	第7回~第9回	8.00

表7、表8より小テスト課題ログデータでは、どちらの方法で補完した場合も第1回から第3回までのデータを用いたときに最も分離できたことが確認できる。授業回が進むにつれて専門性が高まっているにもかかわらず、授業序盤が最も成績上位者と成績下位者を分離できたことから、本授業において、課題に取り組む姿勢は大きな影響力を持っていることが考えられる。

表 9 2022 年度 レポート課題ログデータ 解析結果

欠損値 補充方法	データパターン	分離できた割合 [%]
パターン A	第2回・第4回・第5回	72.00
	第8回・第9回・第10回	96.00
パターン B	第2回・第4回・第5回	78.00
	第8回・第9回・第10回	76.00

表 10 2023 年度 レポート課題ログデータ 解析結果

欠損値 補充方法	データパターン	分離できた割合 [%]
パターン A	第2回・第4回・第5回	96.00
	第8回・第9回・第10回	96.00
パターン B	第2回・第4回・第5回	100.00
	第8回・第9回・第10回	100.00

表 9 より 2022 年度レポート課題ログデータでは、パターン A で補充した第 8 回・第 9 回・第 10 回のデータを用いたときに最も分離できた割合が高かったことが確認できる。また“パターン A で分離できた回数”と“パターン B で分離できた回数”を合算して考えたときに分離できた割合が高かったデータパターンは、第 8 回・第 9 回・第 10 回のデータであったが、第 2 回・第 4 回・第 5 回のデータを用いたときも 70%以上の割合で分離できたことがわかる。表 10 より 2023 年度レポート課題ログデータでは、パターン B で補充したデータはどちらのデータパターンも 100%の割合で分離に成功したことが確認できる。パターン A で補充した場合も非常に高い割合で分離できた。また、どちらのデータパターンを用いても分離できた割合は等しかった。よってレポート課題ログデータでは、どちらのデータパターンもドロップアウト兆候者を検知できる授業回として適しているといえる。

小テスト課題では、ドロップアウト兆候者を最も検知できる授業回は第 1 回から第 3 回であったが、レポート課題では、第 2 回・第 4 回・第 5 回と第 8 回・第 9 回・第 10 回のどちらもドロップアウト兆候者を検知できる授業回であったため、最もドロップアウト兆候者を検知できる授業回は特定できなかった。今回、小テスト課題では 3 変数で解析を行ったが、レポート課題では「解くためにかかった時間」を測ることができず、2 変数で解析を行った。小テスト課題と比較して解析に使用する変数が減少したことが原因であると考えられる。

5. ドロップアウト兆候者早期検出システムの提案

これまでの解析結果をもとに、ドロップアウト兆候者を早期検出するシステムを構築した。SOM の結果図にある成績下位者の重心値 (青い*) から半径 5 以内にプロットされた学生をテキストで出力し、100 回の試行のうち、特定の閾値以上の割合で検出された学生をドロップアウト兆候者として検出した。成績下位者の重心値は、小テスト課題やレポート課題から得られる、全てのログの各平均値をそれぞれ基準化し (さらに点数は逆数に変換)、足し合わせた値の上位者 10 名から作成した。提案システムのフローチャートを図 2 に示す。

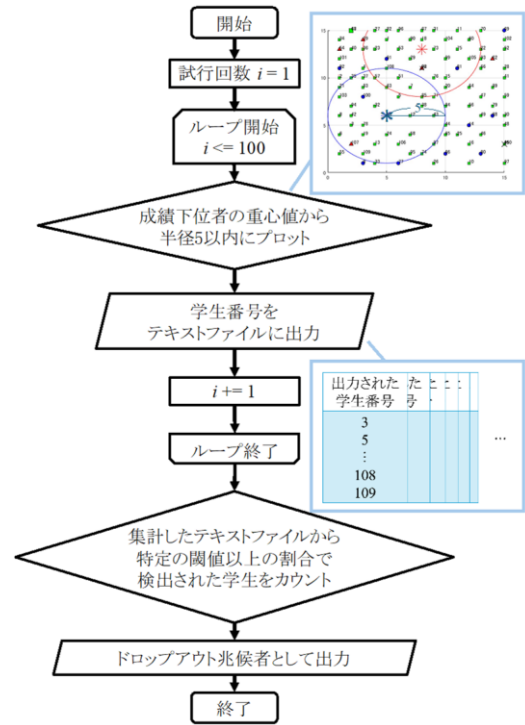


図 2 提案システムフローチャート

5.1 小テスト課題ログデータに提案システムを適用した結果

小テスト課題では、第 1 回から第 3 回までが最もドロップアウト兆候者を検知できる授業回であることがこれまでに明らかとなっている。そこで第 1 回から第 3 回までのログデータに提案システムを適用した。抽出条件は 20% から 40% まで 5% 間隔で調整した。2021 年度データの結果を表 11 に、2022 年度データの結果を表 12 に示す。

表 11 2021 年度 提案システム適用結果

抽出条件	20%	25%	30%	35%	40%
検出された人の割合[%]	52.00	47.00	42.00	34.00	26.00
下位者の割合[%]	77.78	77.78	55.56	55.56	44.44

表 12 2022 年度 提案システム適用結果

抽出条件	20%	25%	30%	35%	40%
検出された人の割合[%]	39.10	34.55	31.82	29.09	24.55
下位者の割合[%]	75.00	66.67	58.33	58.33	50.00

表中の「検出された人の割合」とは、ドロップアウト兆候者として検出された人数を全体の受講者数で割った値であり、「下位者の割合」とは、成績下位者のうちドロップアウト兆候者として検出された人数の割合である。「検出された人の割合」が低く、「下位者の割合」が高い場合が最も精度の良い結果である。

表 11 より 2021 年度データでは、抽出条件が 20%、25% の場合が最も成績下位者を検出することができ、かつ検出された人数も受講者数の半数程度であったことが確認できる。

表12より2022年度データでは、抽出条件が20%の場合が最も成績下位者を検出することができ、かつ検出された人数も受講者数の半数以下であったことが確認できる。

よって表11、表12の結果を総合すると、第1回から第3回の小テスト課題ログデータに抽出条件を20%に設定した提案システムを適用すると、75%以上の成績下位者を検出できることがわかった。

5.2 レポート課題ログデータに提案システムを適用した結果

レポート課題では、第2回・4回・5回のログデータと第8回・9回・10回のログデータのどちらのデータもドロップアウト兆候を検知できる授業回であることがこれまでに明らかとなっている。そこで第2回・4回・5回のログデータと第8回・9回・10回のログデータを用いて、それぞれに提案システムを適用した。抽出条件は20%から40%まで5%間隔で調整した。2022年度第2回・4回・5回データの結果を表13、2022年度第8回・9回・10回データの結果を表14、2023年度第2回・4回・5回データの結果を表15、2023年度第8回・9回・10回データの結果を表16に示す。

表13 2022年度 第2回・4回・5回データ
提案システム適用結果

抽出条件	20%	25%	30%	35%	40%
検出された人の割合[%]	50.00	45.83	37.50	29.17	20.83
下位者の割合[%]	50.00	50.00	50.00	50.00	50.00

表14 2022年度 第8回・9回・10回データ
提案システム適用結果

抽出条件	20%	25%	30%	35%	40%
検出された人の割合[%]	50.00	41.67	37.50	33.33	29.17
下位者の割合[%]	50.00	50.00	50.00	50.00	50.00

表15 2023年度 第2回・4回・5回データ
提案システム適用結果

抽出条件	20%	25%	30%	35%	40%
検出された人の割合[%]	37.78	35.56	28.89	28.89	24.44
下位者の割合[%]	100.00	100.00	100.00	100.00	100.00

表16 2023年度 第8回・9回・10回データ
提案システム適用結果

抽出条件	20%	25%	30%	35%	40%
検出された人の割合[%]	37.78	33.33	26.67	24.44	20.00
下位者の割合[%]	100.00	100.00	100.00	100.00	50.00

表13、表14より2022年度データでは、どちらの授業回データを用いても成績下位者を検出することができる割合は同じであった。今回提案するシステムは、講義期間中のなるべく早い時期に導入した方がドロップアウト兆候者をより早く効果的に支援できることから、提案システムに適用させるデータとしてより適しているのは、第2回・4回・5回データであるといえる。また抽出条件を変化させても成績下位者を検出できる割合は変わらなかった。

表15、表16より2023年度データでは、どちらの授業回データを用いても全ての成績下位者を検出することができた。したがって2022年度データと同様の理由から、提案システムに適用させるデータとしてより適しているのは、第2回・4回・5回データであることがわかった。また40%以外の抽出条件では成績下位者を検出できる割合は同じであった。

よって表13から表16の結果を総合すると、第2回・4回・5回のレポート課題ログデータに抽出条件を20%から35%に設定した提案システムを適用すると、50%以上の成績下位者を検出できることがわかった。

本論文では、4つの授業ログデータに提案システムを適用させたが、すべてのログデータで50%以上の成績下位者を検出することに成功し、今回提案したシステムに一定の汎用性があることを確認した。一方で、小テスト課題と比較するとレポート課題は、検出できる成績下位者の割合が年度によってばらつく傾向にあった。このような結果になった理由として、検出できた下位者の割合が低かった2022年度データに“成績下位者の中で課題の取り組み方が大きく異なっていた”という特徴があったため、成績下位者を安定して検出することが難しかったことが考えられる。

6. おわりに

本論文では、小テスト課題の学習履歴データとレポート課題の学習履歴データの解析を行い、ドロップアウトの可能性のある学生が検出可能かどうか検討した。その結果、課題の種類が異なる授業においても、著者らが提案したドロップアウト兆候者早期検出方法は一定の汎用性があることを確認した。一方で、課題の取り組み方が大きく異なる成績下位者がいた場合、多くの成績下位者を安定して検出することは難しい傾向にあることも明らかとなった。

今後は、講義型授業以外の授業形態で実施されている授業においても、今回提案した方法でドロップアウト兆候者を検出できるかどうか、汎用性の評価をさらに進めたいと考えている。また、ドロップアウト兆候者への通知方法や支援方法の検討を行い、ドロップアウト兆候者を早期に検出・支援できるシステムの実用化を目指す。

参考文献

- [1] 文部科学省, “令和4年度「全国学生調査（第3回試行実施）」の結果について（報道発表）”, https://www.mext.go.jp/content/20230712-koutou02-000001987_1.pdf (2024/06/14 閲覧)
- [2] 文部科学省, “「デジタルを活用した大学・高専教育高度化プラン」実施機関の取組概要(1)”, https://www.mext.go.jp/content/20210630-mxt_senmon01-000016115_1.pdf.pdf (2024/06/14 閲覧)
- [3] 文部科学省, “「デジタルを活用した大学・高専教育高度化プラン」実施機関の取組概要(2)”, https://www.mext.go.jp/content/20210630-mxt_senmon01-000016115_2.pdf.pdf (2024/06/14 閲覧)
- [4] Nabila Sghir, Amina Adadi and Mohammed Lahmer, “Recent advances in Predictive Learning Analytics: A decade systematic review (2012–2022)”, *Education and Information Technologies* (2022).
- [5] Mayu Koito, Kayo Ogawa, “Investigation of early detection methods for dropout students by using LMS log data”, *Proc. of CSEE&T2023*, pp.187-188 (2023).
- [6] Pitoyo Hartono, Kayo Ogawa, “Intuitive Analysis by Visualizing Context Relevant E-learning Data”, *IPJSJ Transactions on Computers and Education*, vol.3, pp.20-27 (2017).