

学業成績データを用いた退学兆候者検出の検討 Detection of dropout candidates using the students' academic performance data

久保田 愛麗[†]
Airi Kubota

小川 賀代[†]
Kayo Ogawa

1. はじめに

大学や専門学校を含む高等教育機関への進学率は、平成 26 年度から増加傾向にあり、令和 5 年度には 86.9%と過去最高値を記録した[1]。なかでも大学進学率は 57.7%で、「将来の仕事に役立つ勉強を行いたい」「専門的な研究に取り組みたい」といった学問に前向きな動機が高校生の進学理由として上位を占めている[2]。しかしながら進学率の上昇と並行して、学生の退学率および休学率も増加傾向を示している。現在では全体の 2.17%に達しており、主な理由として、「学生生活不適応・修学意欲低下」が 16.5%、「学力不振」が 7.3%を占めている[3]。学問に前向きな動機で入学したにもかかわらず、学生生活の途中で勉強への意欲を失ったり、授業についていけなくなったりすることが主な要因と考えられ、対策が迫られている。そこで、教員が学生の状況を把握し、修学意欲が低下し退学に至る前に学生のケアやサポートをしていく必要がある。先行研究では、一つの授業における学習ログを自己組織化マップ (SOM : Self - Organizing Map、以下 SOM と表記) で可視化し、その結果から、当該授業でドロップアウト兆候のある学生を予見する方法を提案し、有効性を確認した[4]。しかし先行研究は、授業内のドロップアウト兆候者の検出に留まっている。そこで本研究では、一授業で有効であった解析手法を、複数授業の成績を含む半期ごとの成績データに適用して、退学兆候者の検出が可能かどうか検討を行う。

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、オープンソースリポジトリ UC Irvine[5]より“Predict Students' Dropout and Academic Success”を引用した。卒業に関する成績データセットであり、変数 37 個、学生 4424 名で構成され、欠損値は存在しない。本研究では、変数を成績に関する 6 個とし、これを表 1 に示す。この変数を前後期合わせて、計 12 変数で実施した。

表 1 成績に関する変数

修得単位数 credited	登録単位数 enrolled	評価単位数 evaluations
承認単位数 approved	成績 Grade	評価なし単位数 without evaluations

ラベルとして「到達度 : Dropout (退学)、Enrolled (留年)、Graduate (卒業)」を用いた。また、履修に関する規則や授業数などの条件を統一するため、志願形態を「第

[†] 日本女子大学大学院 理学研究科

Graduate School of Science, Japan Women's University

1 期一般選抜」、専攻を「看護学専攻」に限定した。看護学専攻の学生数は 415 名で内 Dropout 43 名、Enrolled 50 名、Graduate 322 名である。本研究では、履修者の中から退学兆候者検出するために、授業登録をしていない学生 13 名を除く 402 名を対象として解析を実施した。

解析手法には、先行研究同様 SOM を採用した。SOM は高次元データでも位相を保ったまま低次元空間に写像することができ、似た特徴を持つデータが近接して配置されるという特性があるため、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。本検討における SOM のシミュレーション諸元を表 2 に示す。初期値に乱数を用いているため影響を考慮して各 50 回試行することとした。今回は SOM の結果にラベル情報を付加し、前後期それぞれの Dropout と Graduate の分離の可否を目視で確認した。分離ができた場合の結果を図 1 に示す。赤い三角で示された Graduate と青い丸で示された Dropout がそれぞれ固まって分布していることがわかる。これにより SOM を用いた分離が可能であることが明らかになり、各到達度の学生の成績に傾向があることが確認された。

表 2 SOM シミュレーション諸元

学習回数	5000 回
学習率	0.1
マップ数	30×30 (看護)
データの分類	到達度をもとに分類 Dropout (●) Enrolled (■) Graduate (▲)

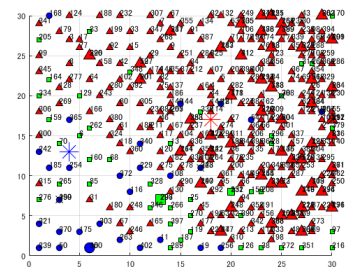


図 1 SOM 結果例

3. システムのパラメータ検討

看護学専攻における退学兆候者早期発見システムについて、半径、試行回数、判定条件の 3 つのパラメータを検討した。本システムは、SOM の結果にある Dropout 学生の重心値 (青の*) から半径 r 以内にプロットされた学生をテキストで出力し、 n 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退学兆候者として検出した。重心値は、Dropout 学生の前期 6 変数の各平均値を用いた。システム概念図を図 2 に示す。

システム精度の目標として、到達度が Dropout である学生 30 名の半数以上が退学兆候者として検出できるようにす

ること（該当率 50%以上）、退学兆候者が全体の受講者数 402 名の半数程度を占めるようにすることの 2 つを設定した。

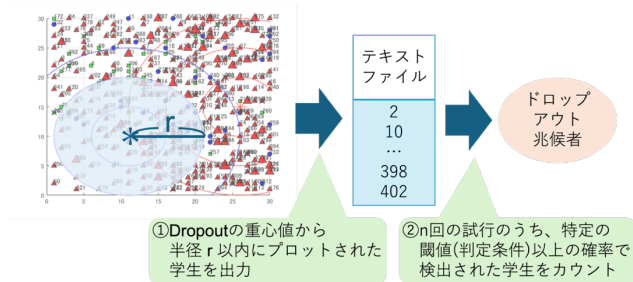


図2 システム概念図

3.1 半径の検討

半径を変化させることでシステムの精度にどのような影響を与えるか検討する。半径の検討において、試行回数は 50 回、判定条件は 3 割以上で固定した。

看護学専攻は、学生数が 402 名であることから、マップサイズを 30×30 に設定している。そこで、半径の候補として、 $r = 5, 10, 15$ の 3 つで検討した。結果を表 3 に示す。

表3 半径検討結果

	$r = 5$	$r = 10$	$r = 15$
円内の人数[人]	9	102	304
Dropout の人数[人]	5	18	27
該当率[%]	16.7	60.0	90.0

半径 5 では、試行回数 50 回の 3 割（15 回）以上検知された学生数が 9 名、内 Dropout が 5 名で、Dropout 学生全 30 名の 16.7%であることを示している。 $r = 10$ のときのみ、円内の人数（退学兆候者）が全体の受講者数の半数以内であり、かつ到達度が Dropout の学生の半分以上が検知できている。したがって、半径は $r = 10$ が最も適切であることが明らかになった。

3.2 試行回数の検討

前節で明らかになった適切な半径を用いて、試行回数を 1 回から 50 回まで増加させ、乱数の影響を受けなくなる回数を検討する。判定条件は 3 割以上で固定した。結果を表 4 に示す。

表4 試行回数検討結果

	10 回	20 回	30 回	40 回	50 回
円内の人数[人]	162	102	95	116	102
Dropout の人数[人]	20	18	17	18	18
該当率[%]	66.7	60.0	56.7	60.0	60.0

試行回数が 40 回を超えると、円内の人数が安定し、到達度 Dropout の学生数も安定した。表 4 では、10 回のときが最も多くの Dropout 学生を検知できているが試行回数が少ないため信頼性に欠ける。よって、最低でも 40 回は試行する必要がある、50 回行えばおおよそ安定した結果が得られることが確認できた。

3.3 判定条件の検討

これまでに明らかになった半径と試行回数を用いて、判定条件を 2 割から 4 割まで 0.5 ずつ増加させ、システムの精度を比較、検討する。結果を表 5 に示す。

表5 判定条件検討結果

	2 割	2.5 割	3 割	3.5 割	4 割
円内の人数[人]	191	131	102	57	45
Dropout の人数[人]	22	20	18	10	8
該当率[%]	73.3	66.7	60.0	33.3	26.7

表 5 より、看護学専攻においては判定条件が 2 割以上の場合、円内の人数が全体の受講者数の半分以上に抑えられ、到達度 Dropout の学生の 73%以上が検知できている。よって、2 割以上が適切であることが明らかになった。

以上より、パラメータの検討を実施した結果、最も精度の良いパラメータとなった組み合わせを表 6 に示す。これらのパラメータを用いて、73.3%の確率で到達度 Dropout の学生を検出することに成功した。

表6 看護学専攻 パラメータ組み合わせ

パラメータ	適切な値
半径	10
試行回数	50 回
判定条件	2 割以上

4. おわりに

本研究では、卒業に関する成績データを含むデータセットを用いて解析を行い、退学の可能性のある学生が検知可能かどうか検討した。その結果、学習履歴データとは形式や対象期間が異なるデータにおいても、SOMを用いた退学兆候者早期検出方法が有効であることが明らかとなった。

今後は、他専攻のデータセットにおいても有効であるかどうか汎用性の検討を行う。また、志願形態などの縛りをなくしたデータセットにおいても本手法を用いた検出が可能かどうか検討し、退学兆候者を早期に検出・支援するシステムの実用化を目指す。

参考文献

- [1] 文部科学省，“文部科学統計要覧(令和 6 年版)”，https://www.mext.go.jp/content/20240531-mxt_chousa01-000036236-22.pdf
- [2] 文部科学省，“進路を考える時の高校生の気持ち”，https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo10/shiryo/_ic_sFiles/afieldfile/2010/02/12/1289068_3.pdf
- [3] 文部科学省，“令和 5 年度学生の中途退学者・休学者数の調査結果について”，https://www.mext.go.jp/content/20240627-mxt_gakushi01-000013028_1.pdf
- [4] 小糸麻結,小川賀代,“ドロップアウト 兆候者早期検出方法の汎用性の評価”,情報科学技術フォーラム FIT2024,第3分冊,pp.89-pp.92(2024)
- [5] UC Irvine Machine Learning Repository,“Predict Students’ Dropout and Academic Success”,<https://archive.ics.uci.edu/dataset/697>