

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

藤巻星香¹ 小川賀代¹

概要: コロナウイルス感染拡大の影響により、学習形態の在り方が改めて見直されている。オンラインでの教育は、自分のペースで学習を進めることができるという利点がある。一方、学習のモチベーションの維持が難しいといった欠点もある。また、コロナ禍においては、学生同士のコミュニケーションもオンラインのみに限定されることで、課題解決がしづらく、ドロップアウトしてしまう学生が発生してしまうことも課題である。ドロップアウトの兆候を把握できれば、その学生に対して支援が可能となる。本研究では主成分分析（以下 PCA）及び自己組織化マップ（以下 SOM）を用いて LMS に蓄積された 2020 年度前期の学習履歴データを分析し、着目すべきログについて検討を行う。更に、成績上位者と下位者のログを時系列に分析し、早期の支援につなげるために、ドロップアウトの兆候の検知可能な時期について検討を行う。

Examination of Learning Support Using Learning History on Online Classes

SEIKA FUJIMAKI^{†1} KAYO OGAWA^{†1}

Abstract. Due to the influence of the spread of coronavirus infection, the way of learning is being reviewed again. Online education has the advantage that you can study at your own pace. On the other hand, it also has the disadvantage that it is difficult to maintain motivation for learning. In addition, in the case of Corona, communication between students is limited to online only, which makes it difficult to solve problems and causes some students to drop out. If you can grasp the signs of dropout, you can support the student. In this study, we analyze the learning history data of the first half of 2020 accumulated in LMS using principal component analysis (hereinafter PCA) and self-organizing map (hereinafter SOM), and examine the logs that should be noted. Furthermore, we will analyze the logs of the top performers and the bottom performers in chronological order and examine the time when the signs of dropout can be detected in order to lead to early support.

1. はじめに

近年、インターネットの普及や情報技術の発展に伴い、オンラインでの教育が身近なものになってきている。特に 2020 年はコロナウイルス感染拡大の影響により、e ラーニングを活用した研修や遠隔授業など、オンライン教育の需要は急速に増加し、学習形態の在り方が改めて見直される年となった。全国の大学職員・教員 100 名を対象に実施されたアンケート調査によると、多くの大学では 4~5 月にオンライン授業の導入を行っており、6 月時点でその実施率は 97 % に上っている^[1]。

オンラインでの教育は、移動手段を問わないため、時間を効率的に使いながら学習でき、進捗を見ながら自分のペースで学習を進めることができるという利点がある。一方、自由が増えることで自己管理が必要となり、学習のモチベーションの維持が難しい、といった欠点がある。一律に情報や課題を発信できるものの、課題の量やレベルが適切ではないまま進んでしまうことや、授業のレベルに物足りない学生、またはついていけない学生などに対するフィードバックは行き届いていない。さらに、受講人数の増加に伴ってそのフィードバックは難しくなり、一方的な授業になりかねない。また、コロナ禍においては、学生同士のコ

ミュニケーションもオンラインのみに限定されることで、課題解決がしづらく、ドロップアウトしてしまう学生が発生してしまうことも課題である^[2]。教員側が学生の進捗や状況を把握し、授業のレベルに物足りない学生やドロップアウトしてしまう学生の兆候を検知することができれば、教員側も適切な授業や課題を提供できるようになる。このような教育の支援につなげてくためにはログデータの様々な観点からの分析が必要不可欠である。本研究では主成分分析（以下 PCA）及び自己組織化マップ（以下 SOM）を用いて LMS に蓄積された 2020 年度前期の学習履歴データを分析し、着目すべきログについて検討を行う。更に、成績上位者と下位者のログを時系列に分析し、早期の支援につなげるために、ドロップアウトの兆候の検知可能な時期について検討を行う。

2. 解析に使用するデータ概要

解析には筆者らが所属する学科 1 年生の必修科目である「情報科学概論」の LMS に蓄積されている学習履歴データを用いた。本講義は、全回オンデマンド形式（12 回）で、出席確認のための小テストが、正規の時間割の授業終了後 24 時間以内に受験することが課されている。受講者数は 137 名であり、使用学習履歴データは、小テストの点数、小テストに要した解答時間、小テストに着手するまでの時間、

¹ 日本女子大学
Japan Women's University

そして定期試験の点数とした。小テストは全 12 回行ったため、解析に用いたデータは全 37 変数である。小テスト未提出回の学生に関する欠損値データには、点数は 0 点とし、時間に関するデータに関しては平均代入法を用いてデータ整理を行った^[3]。解析に使用する学習履歴データの一部を表 1 に示す。表 1 のデータに対し、データの基準化を行い、解析に使用するデータとした。

表 1 学習履歴データ一例

学生番号	小テスト第1回			...	小テスト第12回			定期試験点数
	点数	解答時間	着手までの時間		点数	解答時間	着手までの時間	
1	7	0:03:29	0:38:12	...	6	0:10:12	2:08:40	73.2
2	8	0:06:08	0:12:17	...	4	0:18:49	0:01:55	43.3
3	8	0:02:37	0:58:52	...	6	0:09:36	0:37:12	56.7
4	8	0:05:30	1:15:29	...	0	0:11:12	7:00:21	64.6
5	8	0:05:47	1:25:19	...	8	0:09:34	7:00:26	49.6

3. グループ分けの検討

3.1 PCA によるグループ分けの検討

前述の解析データを PCA にかけて、固有値 1 以上の成分の固有値・寄与率・累積寄与率を表 2 に示す。

表 2 固有値・寄与率・累積寄与率結果

	固有値	寄与率	累積寄与率
第 1 主成分	6.95	18.7%	18.7%
第 2 主成分	3.57	9.6%	28.2%
第 3 主成分	3.09	8.3%	36.5%
第 4 主成分	2.26	6.1%	42.6%
第 5 主成分	1.69	4.5%	47.1%
第 6 主成分	1.51	4.1%	51.1%
第 7 主成分	1.40	3.8%	54.9%
第 8 主成分	1.34	3.6%	58.5%
第 9 主成分	1.19	3.2%	61.7%
第 10 主成分	1.11	3.0%	64.7%

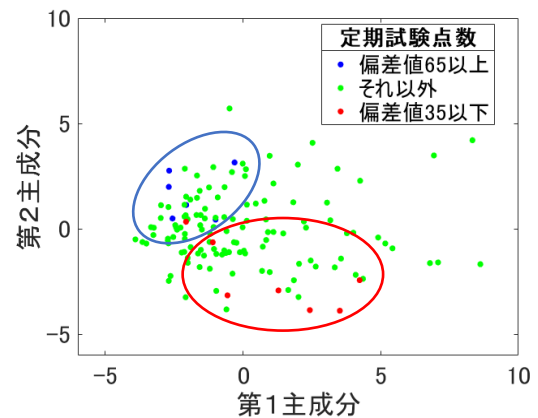
37 変数のデータが第 10 主成分までで主成分がまとまり、次元圧縮されたことが確認できる。さらに、PCA から得られた主成分得点や主成分の重みの大きさから主成分の解釈を行う。本研究では特に特徴の表れた第 3 主成分までを解釈した。

まず主成分得点の散布図を図 1 に示す。定期試験の結果を、偏差値 65 以上を青、35 以下を赤として図中に示し、成績上位者と下位者の分散値を表 3 に示した。図 1 (a) は横軸が第 1 主成分、縦軸が第 2 主成分を示している。この散布図から、成績上位者は、第 1 主成分の分散が小さく、第 2 主成分は正の値をとることがわかる。また、成績下位者は第 1 主成分の分散が大きく、第 2 主成分は負の値をとることがわかる。表 3 の分散値からも成績上位者より下位者のほうが、分散が大きい値を示している。また、散布図から、

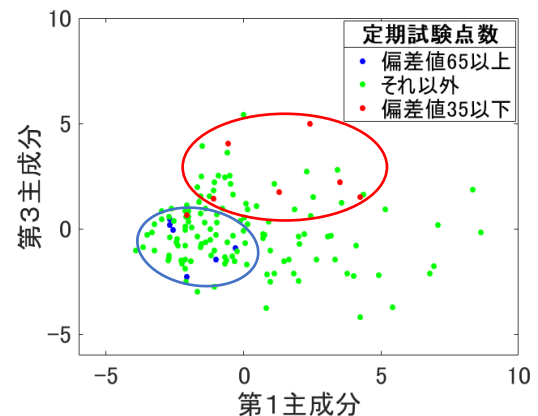
成績上位者と下位者は固まって分布していることが見てとれる。

図 1 (b) は横軸が第 1 主成分、縦軸が第 3 主成分を示している。この散布図により、成績上位者は、第 1 主成分の分散が小さく、第 3 主成分は負の値をとることがわかる。また、成績下位者は第 1 主成分の分散が大きく、第 3 主成分は正の値をとることがわかる。表 3 の分散値からも成績上位者より下位者のほうが、分散が大きい値を示している。分散に関しては図 1 (a) と同じ傾向だが、成績上位者と下位者の分布の正負は図 1 (a) と反対の傾向が見られた。

しかし図 1 (a)と(b)のどちらの図も成績上位者と下位者の分布が固まっていることから、新たな合成変数によって、成績上位者と下位者の分布を分離できる可能性が得られた。



(a) 第 1 主成分と第 2 主成分の散布図



(b) 第 1 主成分と第 3 主成分の散布図

図 1 第 1~3 主成分の主成分得点散布図

表 3 第 1~3 主成分の成績上位者と下位者の分散値

	成績上位者	成績下位者
第 1 主成分	0.843	5.017
第 2 主成分	1.112	2.248
第 3 主成分	1.342	4.465

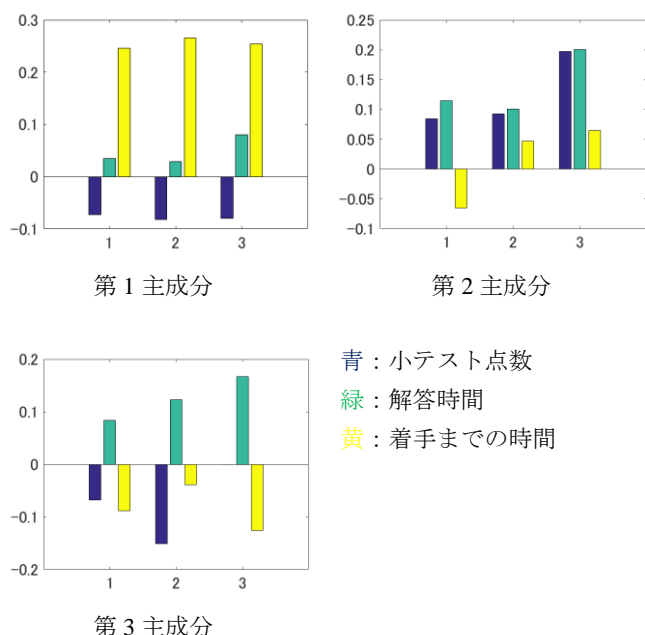


図 2 第 1～3 主成分の主成分負荷量

図 2 は第 1～3 主成分の主成分負荷量を示す。横軸は全 12 回行われたうちの小テストの第 1～3 回を抜粋して示している。第 1 主成分では小テスト着手時間の重みが 1 番大きい値を示している。第 2 主成分では小テストの点数や解答時間の重みが大きい値を示し、第 3 主成分では解答時間や着手までの時間の重みが大きい値を示している。第 1～3 主成分のどの主成分でも時間の重みが大きい値を示していることから、時間に関するログが、影響力の高い要素であることが分かる。

成績上位者の時間に関する学習ログを参照すると、小テストに要した解答時間や小テストに着手するまでの時間は短い傾向があり、下位者は反対の傾向が見られた。また、小テストの点数の平均点は、成績上位者のほうが高かった。

主成分得点と主成分負荷量の結果から、第 1 主成分を“解答時間の軸”、第 2 主成分を“小テスト点数の軸”、第 3 主成分を“解答・着手するまでの時間の軸”、と解釈した。また、第 1～3 主成分の重みの値の大きさや軸の解釈から、時間に関するログデータは成績上位者と下位者の取り組み方の違いに影響力の高いログであることがわかった。

3.2 SOM によるグループ分けの検討

PCA は次元圧縮され、作り出される主成分の図式化を可能とするが、2 または 3 変数までしか表現できない。一方、SOM は非線形な写像のため、すべての変数を加味し、データ間の距離を保ちつつ 2 次元マップ上への表現が可能である。3.1 節で使用した同じデータを、SOM により図式化した結果について述べる。SOM の設定や概要について表 4 に示し、SOM の出力結果を図 3 に示す。概ね、成績上位者

と下位者が分かれて分布していることが確認でき、成績上位者と下位者の学習の取り組み方には違いが見られることがわかった。しかしながら、成績上位者のグループの中に成績下位者である学生が 1 名いることがわかる。この学生について元の学習ログデータを参照すると、

- ・小テストの解答時間が平均的
- ・着手までの時間が短い
- ・小テストの点数の平均点が高い
- ・小テストの未提出回数が 0

という特徴があり、授業後に行われている小テストは毎回まじめに取り組み、高得点であったが、定期試験では低い点数を取った学生であった。この学生は、図 1 の主成分得点の散布図で成績上位者の丸の枠内に入っている成績下位者の学生でもある。上述の特徴は、成績上位者と同様の特徴であり、37 変数の内、36 変数が成績上位者と同様の傾向をとるため、分離がうまくいかなかったと考えられる。

表 4 SOM の概要

学習回数	3000 回
学習率	0.1
マップ数	15×15
データの分類	偏差値 65 以上(青) 偏差値 35 以下(赤) それ以外(緑) 重心値(*)

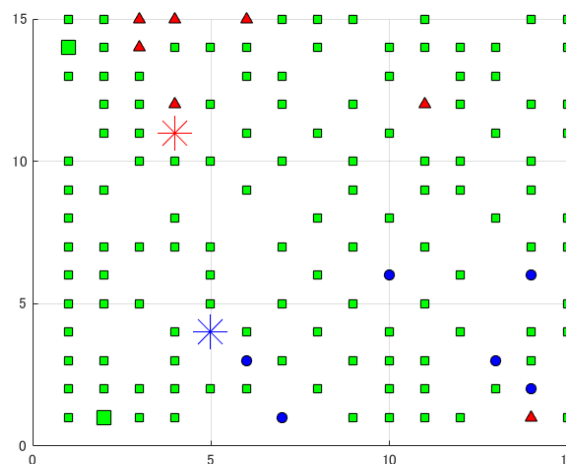


図 3 SOM の出力結果

4. SOMによるドロップアウトの兆候の検知可能時期の検討

ドロップアウトの兆候を検知するための時期を検討するために全授業 12 回を時系列で細分化し、SOM で解析を行った。その分割したパターンについて表 5 に示す。検討 1 では小テストの回数を 3 分割し、第 1~4 回、第 5~8 回、第 9~12 回をそれぞれ解析した。検討 2 では検討 1 と分割数は同じであるが、最初の 3 分の 2 のデータである小テスト第 1~8 回を解析に使用した。第 1~8 回としたのは、ドロップアウトの学生を授業終了前に検知し、支援することを目的としているため、授業期間の 3 分の 2 までの期間を対象とした。SOM の実行回数については、学習回数における入力ベクトルと参照ベクトルの平均二乗誤差 (Mean Squared Error : MSE) を調べたところ、図 4 に示すように、実行回数が増えるにつれ、MSE は低くなり、1800 回程度で安定したため、SOM の実行回数は 2000 回で十分であると判断した。

検討 1 の出力結果を図 5、検討 2 の出力結果を図 6 に示す。* は成績上位者と下位者のグループそれぞれの重心を示している。図 5 (1) は、小テスト第 1~4 回で解析を行った結果であり、成績下位者の重心の周りに成績上位者が付近していることが見てとれる。よって初期の段階において、ドロップアウトする学生の検知は難しいといえる。

小テスト第 5~8 回の結果を図 5 (2) に示す。小テスト中盤の第 5~8 回では、重心を中心に成績上位者・下位者がそれぞれ固まり、グループの分離が確認できた。

図 6 は、小テスト第 1~8 回の最初の 3 分の 2 のデータを使用した解析結果を示す。成績上位者の中に成績下位者が混在しており、成績上位者と下位者は分離していないといえる。また、成績上位者と下位者の重心の距離の差を示している表 6 を見ると、小テスト第 1~8 回より第 5~8 回のほうが重心の距離の差が大きい値を示しており、より成績上位者と下位者が分離していることがわかる。

このような検討結果から、最初からのデータよりも途中を切り出したデータのほうが、差が開いていることがわかった。授業の中盤で、学習のログに違いが出てくることから、ドロップアウトの兆候を検知できる可能性が得られた。

表 5 解析に使用する小テストデータの分割パターン

	分割数	使用したデータ
検討 1	3 分割	第 1~4 回 第 5~8 回 第 9~12 回
検討 2	3 分割 (全データの 3 分の 2 を使用)	第 1~8 回

表 6 成績上位者と下位者の重心の距離の差

	使用したデータ	重心の距離の差
検討 1	第 1~4 回	4.12
	第 5~8 回	13.60
	第 9~12 回	7.21
検討 2	第 1~8 回	5.83

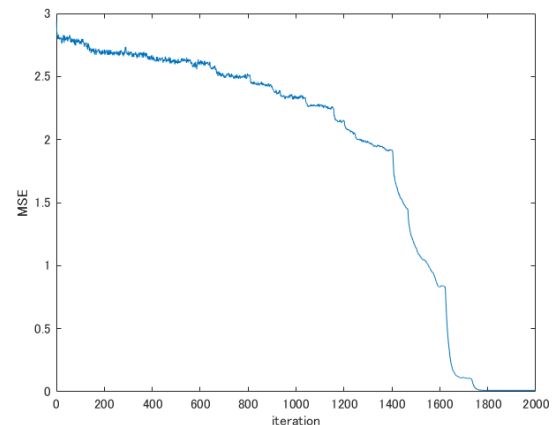
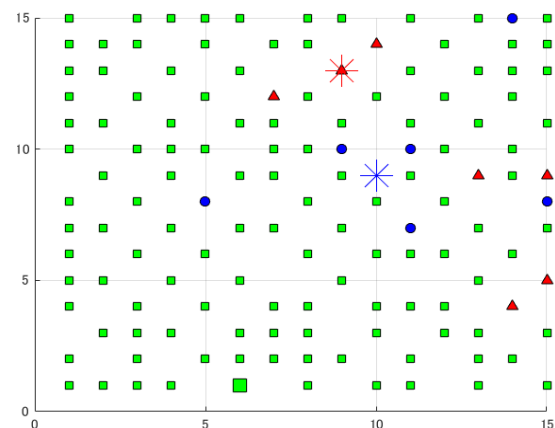
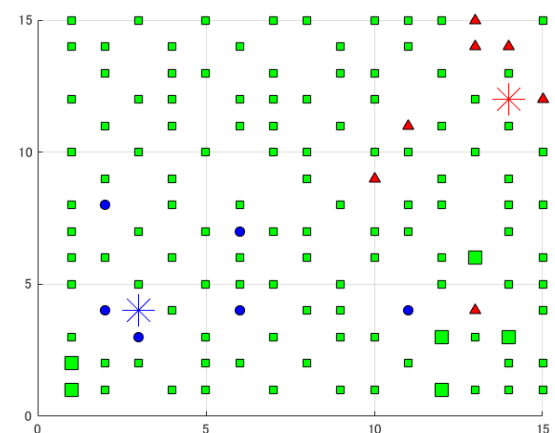


図 4 実行回数ごとの MSE

(入力ベクトルと参照ベクトルの差) の変化



(1) 小テスト第 1~4 回



(2) 小テスト第 5~8 回

図 5 検討 1 の SOM の出力結果

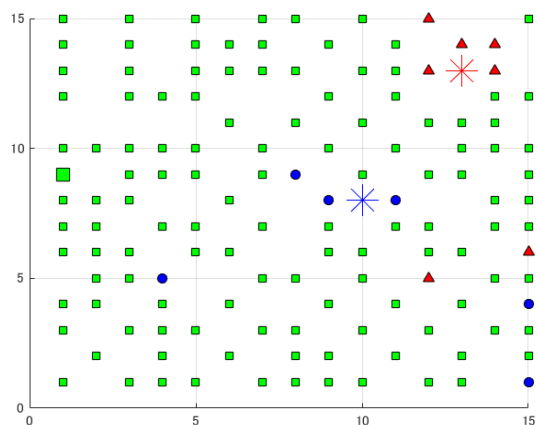


図 6 検討 2 の SOM の出力結果
 (小テスト第 1～8 回)

5. おわりに

本研究では、PCA を用いて時間に関するログが着目すべき項目であるという結果が得られた。また、PCA や SOM の結果から、成績上位者と下位者とは、授業の課題の取り組み方に違いがあることもわかった。さらに時系列で授業を 3 つに細分化し、SOM で解析を行ったところ、最初からのデータよりも途中を切り出したデータのほうが、成績上位者と下位者の間に差が開いていることがわかった。また、授業の途中回を切り出したデータからでも成績上位者と下位者の間に差が表れ、ドロップアウトしそうな兆候を検知することができた。このような解析結果は今後、ドロップアウトしそうな学生への支援システムにつなげていくことができると考えられる。

参考文献

- [1] “ReseED 教育業界ニュース, 大学のオンライン授業実施率は 97 %, 導入時期は 4-5 月が多数” .
<https://reseed.resemom.jp/article/2020/07/17/473.html>, (参照 2021-2-16).
- [2] 内閣府, “オンライン教育の重要性と課題” .
https://www5.cao.go.jp/keizai-shimon/kaigi/minutes/2020/0708/shiryo_01-3_2.pdf, (参照 2021-2-16).
- [3] 前波夏織, “欠損値を含む学習履歴データを用いた学習予測の検討”, 電子情報通信学会東京支部学生会 (2017)