

LMS のログから見える学生の学習行動分析

Analysis of Students' Learning Behavior from LMS Logs

吉 岡 卓*

YOSHIOKA Suguru

Keywords : Cluster Analysis, Data Analytics, Evidence-Based Education, LMS

概要

In recent years, the importance of evidence-based education has increased, and efforts to use data analysis as well as teachers' experience and practical knowledge for education have been widely implemented. In this paper, I present a method to dynamically analyze students' learning progress by focusing on the log data on the LMS. By analyzing the learning process rather than the results of learning, we aim to predict dropouts in advance, change the flow of the class, and verify the content of homework and assignments in advance. By using this method, I found that changes in the number of accesses to the LMS affect the learning progress. In particular, I show that the use of a quadratic approximation curve for the analysis of the number of accesses enables us to analyze the learning process in detail. As a result, it was found that in distance education, there were more than a certain number of accesses during the course hours, not late at night, and that students who changed the number of accesses according to the contents of the course materials performed better. Finally, a comparison with clustering analysis shows the usefulness of the analysis method presented in this paper.

1 はじめに

近年、教育分野においてエビデンスに基づいた教育 (Evidence-Based Education) の重要性が増している。従来の EBE はメタ分析や教員の実践知に対してエビデンスを産出する方法が主であり、その産出は研究者が主体であった。そのため実践面での定量・定性的な議論についてはこれからの発展が広く期待されている [6, 9]。

一方、人工知能の急速な発展やビックデータの利活用と合わせて、データアナリティクスが幅広く行われている。データアナリティクスは統計的手法、機械学習、様々なアルゴリズムを用いる事で、一連のデータに含まれる特徴的なパターンや相関関係を見出すことを目的としている [1, 2]。特にビックデータにおいては、一見無意味なデータ群に対してチャンキングの手法を取ることで、データに構造的な意味づけを与え具体化や抽象化の処理が可能となる。このことは、データ群の中から有意となるデータセットを抜き出

* 都留文科大学 学校教育学科

すことにつながり、実用的な応用を可能とする。

そこで本稿では、教育ビッグデータを分析する事で介入に有用となるエビデンスの産出を目的とする。本稿で扱うデータは本学で使用している学習管理システム WebClass のログデータを対象とする。令和2年度は遠隔授業が実施され、各回の授業資料に対する学生のアクセス時間や回数など様々なデータが WebClass 上に保持されている。そこで、これらのデータを解析することで、遠隔授業下における学生の学習行動を分析し、学生および教員の行動モデルを提案する。

2 エビデンスに基づく教育とデータアナリティクス

20世紀後半よりエビデンスに基づく教育がアメリカやイギリスを中心として広がってきた [3, 4]。教員の経験や勘だけではなく、教育に対するより客観的な根拠や検証の必要性が求められてきている。国内においても、国立教育政策研究所を中心として教育学にエビデンスを求める動きが広まっており、様々なケーススタディや提言が示されている。

従来のエビデンス産出についてはメタ分析や比較実験が主であった。そのため、学習過程ではなく学習結果からの帰結のみに着目しているという批判や、少人数学級や学年の違いによる効果の立証が難しいという実践面での課題も数多くあげられている。日本国内では GIGA スクール構想や ICT の活用、学校教育の情報化推進により、各小学校や中学校に対して PC 機器やソフトウェアの導入が進んでいる。ICT 機器の導入による一番のメリットは学習活動履歴であるログが収集できる事であり、授業外での活動に対しても分析が可能となる点にある。また令和2年度はコロナ禍によって多くの大学や高校でも遠隔による授業形態を余儀なくされた。このような状況の中、特に大学においても学生の学習ログから最適な教育手段を見つけようという動きが広まってきている。学習ログというビッグデータを対象として分析を行う事で、個別最適化された学習指針の提示を目的としており、いくつかの事例報告がなされている [7, 8]。

一方、情報科学の分野ではビッグデータに対するデータアナリティクスが広く行われており、様々な分析手法が提案されている。特に近年では、上記のような学習記録に対して分析を行うラーニング・アナリティクス分野が発展してきている [10, 11]。多くの大学では、FDの一環として学生アンケート等による評価データを収集しているが、ここから得られるものは授業の雰囲気や学習結果に対するアウトプット、成績との関連性であり、実際に学生がどのように受講したのか、教材をどのように活用したのかといった学習過程に対する分析が難しい。そのため、学習ログからそれらの情報を解析する事でより自由度の高い教育へと結びつけることが期待できる。

すなわち学習結果だけでなく、学習ログという情報から学習過程に対するエビデンスを産出しその分析手法を示すことは、これからの教育学分野においても極めて高い実用性を有している。学習進度が分かる事によってドロップアウトする事を事前に予測したり、授業の流れを変えたり、宿題や課題の内容を事前に検証する事に活用できる。さらには GIGA スクール構想によって、児童生徒が手にする情報機器の活用法や教育の発展に強く寄与すると考える。本稿では、私が令和2年度に担当した教科のログを解析した結果を示す。まずは解析手法の提案を示し、その解析結果から得られた学生の学習行動および教員の行動モデルへの発展について示す。

3 学習ログに対する多角的な分析アプローチ

本稿で用いる学習ログは、私が令和2年度に受け持った科目のなかからコンピュータ基礎を対象とした。本授業は資料配布型の授業形態をとったため、動画配信によるオンライン型や双方向リアルタイムのオンデマンド型に比べてLMSへの依存が高く学習ログの分析としては有用であると考えた。LMSは本学で用意しているWebClassを用いて授業を行ったため、WebClassが保持するログデータを用いた。本授業の分析は、5月1日から7月31日までを対象とした。本授業は4月中から開講していたが、4月中は学生の受講状況などのアンケートや現状確認にあてたため、具体的な学習内容の取り組みとして上記期間14回分の講義を分析に用いた。

3.1 LMS上のデータ利活用と留意事項

WebClassはユーザ管理、教材の提示、掲示板の設置、アンケートやテスト項目の作成・分析などの機能を有している。特に、テストなどの採点結果だけでなく、学生が教材にアクセスした時間、回数、IPアドレスなどもログとして保持しており、任意の期間における集計が可能となっている。そこで本稿では、それらのログを用いて実験を行った。授業ごとのアクセス回数や経過時間を分析することで、学生の学習活動や進捗がどのように変移していったかを類推する。

ただし、WebClassの機能として一部未設定の部分があり、次の点に関する分析はできていない；(i) LMSの終了時間に関する未設定問題；(ii) Webブラウザ側の問題。(i)については、WebClassの任意の授業科目を閲覧し、その後他の科目に画面遷移した場合、最初に見ていた科目の終了時刻がnullとなっていた。つまり、科目ごとの経過時間を測定しているのではなく、WebClass全体の利用時間を測定したログとなっていた。そのため、1つの科目に要した時間が正確には解析できない問題がある（現在は管理者側で科目ごとの測定へと設定変更済み）。(ii)は、学生がWebClassにログインし、その後ログアウトを忘れてしまうと、次回以降正常に画面表示されない不具合が報告されている。ブラウザ側に前回のキャッシュが残ってしまいそれが影響していると考えられる。そのため、学生によってはログインしても教材が見えず、ブラウザのキャッシュをクリアする、ブラウザを変える、友人よりデータをもらう等をして対応したという報告がある。

上記の問題を考慮し、本稿では教材閲覧に要した経過時間を分析対象に含めなかった。もともと資料配布型の授業であった事もあり、資料をダウンロードしてから学習する事も想定されたため、分析に使用したのは教材閲覧回数やアクセス時間とした。

WebClassのログのように個々人の学習履歴やデータは個人情報にあたる¹。そのため、データの利活用に際しては特段の注意が必要であり学生からの同意を求める必要がある。本授業では受講学生数70名（1年生45名、2年生18名、3年生7名）の受講学生がいたが、同意を得られた人数は44名であった。また、聴講生が1名いたため、本稿では43名のデータを用いて分析した結果を示す。授業は令和2年度の前期に行い、同意を取ったのは令和2年度の後期終了後であったため、期間が開いてしまった点については検討材料といえ

¹ 大学ICT推進協議会 (AXIS) 「教育・学習データ利活用ポリシー」のひな形の作成について
<https://axies.jp/report/publications/formulation/>

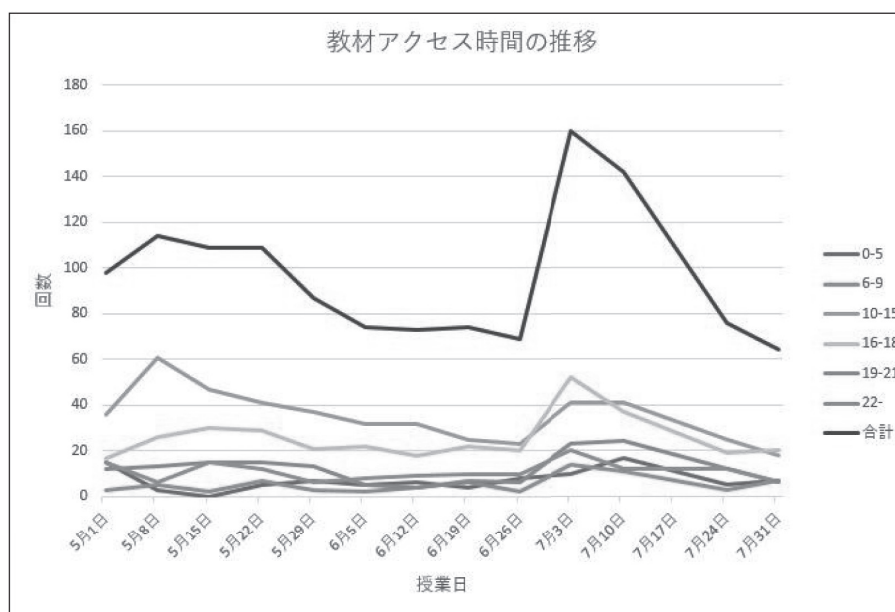


図1: 各授業資料へのアクセス時間と回数の推移 (N=43)

る。特に成績の良かった学生のほうが同意をしてくれやすいという側面があるため、できれば成績開示前に学習データの利用を告知するか、利用意図を正しく理解してもらう必要がある。データ利用の同意に関しては、Microsoft社のFormsを用い、学生には以下の点を告知したうえで、同意するか否かのアンケートを取った。

- ・個人が特定されることはない
- ・同意しても、いつでも取り下げられる
- ・希望があれば個々人のデータを開示する
- ・成績評価に関係しない

上記は公立大学法人都留文科大学「人を対象とする研究」倫理基準（平成27年12月9日基準第2号）第1条に沿うようにしており、またデータをどのように使用するかなどについても説明を行った。また、使用したデータは一定期間保持することとし、希望があればすぐに開示する事としている。

3.2 LMSのログを用いた分析

本節ではWebClassのログを多角的に分析した結果を示す。まず初めに、受講生がWebClassの各教材にアクセスした時間と回数を集計した。アクセス時間に関する変移を図1に示す。

図1では、学生のアクセス時間を6つの区分に分けており、0時～5時、6時～9時、10時～15時、16時～18時、19時～21時、22時～に分けた。学生によっては複数回教材を閲覧しており、43名のデータにも関わらず多い時で160回を超えるアクセス回数となっている。本授業は5時限目に開講した授業のため、16時～18時に閲覧するのが期待されていたが、教材は前日までにアップしており、学生は自由な時間に受講した事が分かる。一番多いのは10時～15時の時間帯であり、次に16時～18時となっている。合計を見てみる

と、授業が進むにつれてアクセス回数も減ってきており、学生が遠隔授業や受講方法に慣れてきた事が分かる。7月3日にアクセス回数が大幅に上がっているが、その日はプログラムに関する課題を出しており、各自の環境下において教材を見ながら受講する内容であったため、アクセス回数が増大したものと考えられる。

そこで図1の結果を踏まえ、以下の点についてさらに分析を進めた。

- ・ 受講日時の違いから類推する学習内容の理解度
- ・ 学生の順応過程と学習内容の理解度

上記に関する詳細な分析を次節にて示す。

3.2.1 受講日時の違いから類推する学習内容の理解度

学生は1回の教材を複数回閲覧している事が分かる。そこで、特に1回目閲覧したときのアクセス時間について分析を行った。大学での授業開講時間を考えると、10時～15時と16時～18時にアクセスするのが適当と考えられる。そこで、各学生について上記時間帯（以後、開講時間と表記）にアクセスした回数を集計した。その結果14回中、平均して7.4回は開講時間内に1回目のアクセスがあった事が分かった。すなわち、半分以上の授業回において開講時間外にアクセスしており、学生は必ずしも開講時間に縛られることなく自由な時間に受講したことになる。このことが学生にとって有意であったのかを確認するために、最終成績との相関を分析した。結果は次の通りである：

成績	開講時間内にアクセスした平均 (14回中)
秀	8.46回
優	7.39回
良	10.00回
可	null
不可	7.00回

本授業は共通科目であり、学習内容も基礎的な事項である事から学生の習熟度は高い。そのため、可評価や不可評価は絶対数が少ない。可評価の学生も一定数いたが、データ利用の同意に関して許諾を取らなかったため、ここではnullと表記されている。上記結果では開講時間内のアクセス回数が増えれば成績が良くなるわけではなく、不可となる学生のアクセス数が若干低くなっている。授業の到達目標に達するためには、授業回の半分以上は開講時間内に受講し生活リズムをつける事が重要であり、担当教員もしくはLMS側で測定する指針に使う事が期待できる。

次に、学生が1回目にアクセスした曜日について集計を取った。本授業は金曜日に開講していたが資料は授業前日に公開していたため、学生によっては規定の時間以前に閲覧する事も可能であった。そこで、規定時間の前日および当日（以後、開講日と表記）にアクセスした回数と成績との相関を次に示す：

成績	授業の開講日にアクセスした数 (14回中)
秀	9.08回
優	9.22回
良	9.60回
可	null
不可	9.00回

この結果においても、回数の差はゆるやかでありアクセス時間と同様、大きな違いが見られない。また、不可の学生も半数回以上の授業については当日中に閲覧していることから、開講日より開講時間を主軸にドロップアウトする学生を考慮するのが適切と考える。学生の生活サイクルと遠隔授業の関係を分析するのは極めて重要であり、学生がどのように遠隔授業に順応したかを次項にて示す。

3.2.2 学生の順応過程と学習内容の理解度

図1に示したように、各回の教材へのアクセス回数と学生が遠隔授業へ順応していった関係には相関がみられると考える。ただし、7月3日に出题した課題の影響が大きいため、6月26日までの9回の授業に焦点を当てて分析を行った。総計で考えれば回数は緩やかに減少しており遠隔授業への慣れに従って閲覧回数が減ったと考えられる。そこで、学生個人ごとのデータをさらに詳細に分析した。

分析には次の組 $T = \langle G, D, CD, CU, Diff \rangle$ を用いた。ここで、

G 学年

D 線形近似曲線の傾き

CD 4次近似曲線で上に凸となる頂点集合

CU 4次近似曲線で下に凸となる頂点集合

Diff $\max d - \min u$ ($d \in CD, u \in CU$)

である。

分析には数値データを用いたが、視覚的に表現した具体例を図2に示す。図2において、太い折れ線が学生の閲覧回数を示している。また、細い点線からなる直線はユークリッド距離で求めた線形近似曲線である。細い点線からなる曲線は4次式による近似曲線を表している。

例えば、図2左側の学生Aは授業が進むにつれて閲覧回数が減少している。それを示すためにDを用いる。すなわち、 $D > 0$ であれば上昇しており、 $D < 0$ であれば下降していると考えられる。なお、Dは小数点以下を切り下げて算出した。よって、線形近似式 $y = ax + b$ に対して、 $D = \lfloor a \rfloor$ となる。この時、例えば $0 < D < 1$ については $D = 0$ となるため、アクセス回数が一定の学生と区別がつかなくなる。そのため、 $0 < D < 1$ 、 $D = 0$ 、 $-1 < D < 0$ についてはそれぞれ $D = +0$ 、 $D = 0$ 、 $D = -0$ と表記する。

また図2中央において、学生Bのグラフを見てみると、折れ線は上下に移動している

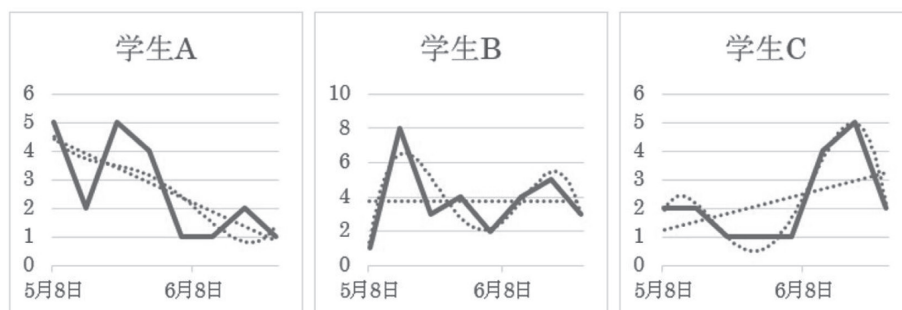


図2: 学生ごとの教材アクセス回数の推移例

にも関わらず $D = 0$ となる。これでは、毎回一定の閲覧回数を取った学生との差が出ない。そのため、4次式による近似曲線も計算した。授業回数が9回分なので、3次式では特徴を表せず4次式を用いている。x軸が9個なので、CDとCUの要素数は高々2個となる。

ここで、図2の左にある学生Aと右にある学生Cの折れ線グラフを考えてみる。学生Aは試行錯誤を繰り返しながら減少傾向にあり、学生Cは局地的に上昇している事がわかる。この時、4次の近似曲線を見ると学生Aでは緩やかな曲線となっており、学生Cでは曲線の上下幅も大きくなっている。この特徴付けのためにDiffがある。元の折れ線グラフであれば、y軸の最大値と最小値の差は学生Aも学生Cも4となり9回の授業の中で4回のブレがあるという分析となるが、Diffを用いると学生Aの方が学生Cより小さくなる。すなわち、遠隔授業への適応度合いを測るためにDiffを用いる。

まず、 $\langle G, D \rangle \subset T$ による内訳は次の通りである。

G	D	人数
1	$\geq +0$	6
	$= 0$	1
	≤ -0	18
2	$\geq +0$	1
	$= 0$	2
	≤ -0	11
3	$\geq +0$	2
	$= 0$	0
	≤ -0	2

上記より、全ての学年において線形近似曲線は下降する割合が半数以上であり、授業回が進むにつれて遠隔授業に慣れていった事が分かる。ただし、1年生においては上昇する割合も高いため、大学の授業に慣れるために苦勞したことが分かる。また、アクセス回数が一定だったという学生は少ない結果が得られた。

そこで、次に成績との比較を行なった。成績での評価はGPA換算を基としたが、区分による差が大きすぎるため、3.2.1節で示した成績区分をより細かくした。具体的には、優、良、可に対してそれぞれ $+0$ と $-$ を評価を加えた3種類とし、GPAへの変換は ± 3 を通常のGPAに増減して換算した。結果は次の通りである。

D	成績 (平均)	G		
		1	2	3
$\geq +0$	2.6	3.09	1.5	1
$= 0$	3	2	3.67	null
≤ -0	3.02	3	2.92	3.4

この結果から分かることは、1年生にとっては傾きの上下による成績の差が少なく、むしろ均一のアクセスをした学生よりも試行錯誤して受講環境を模索した学生のほうが成績がよくなっている。一方、2年生以上の学生については、アクセス回数の減少している学生のほうが良い成績となっている。これは大学という環境における受講に慣れている側面が結果に反映されていると考えられる。全般的に言えば、前述のようにアクセス回数の減っている学生のほうが平均的に成績は良くなっているが学年による違いがかなり顕著にあらわれている事が分かる。

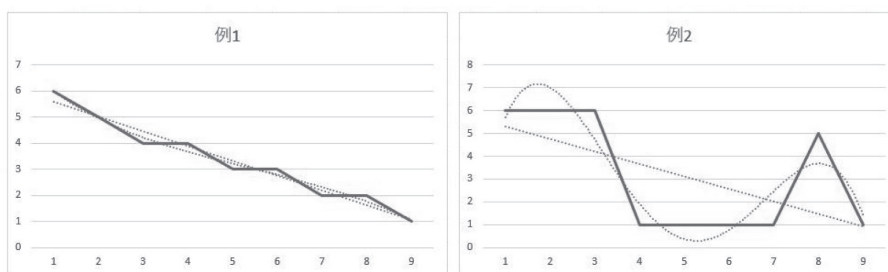


図3: 折れ線グラフと近似曲線の関係

特に、 $D \geq +0$ と $D \leq -0$ に関して詳細に調べた結果は以下の通りである。

D	成績 (平均)	G		
		1	2	3
$+0 \leq D \leq 1$	2.54	3.11	1.5	1
$1 < D$	3.00	3	null	null
$-1 \leq D \leq -0$	3.13	3.33	2.89	3.00
$D < -1$	2.93	2.80	3.00	3.67

上記表において、 $D = 1$ の場合、成績の平均は1.5で最低となり、 $D = -0$ の場合、3.36で最高となる。すなわち、線形近似曲線が上昇する学生においては、上昇率の高い学生のほうが成績が良く、一方下降する学生においては、下降率の少ない学生の方が成績が良いと言える。

ここで、グラフの特徴について考えてみる。簡単な具体例を図3に示す。図3において、2つの折れ線グラフが書かれているが、どちらも線形近似曲線の傾きは同程度と言える。しかし、例1では一定の割合でゆるやかに下降しており、例2では極端な増減となっている事が分かる。すなわち、線形近似曲線だけではグラフの特徴を正確に捉えたとは言いがたい。そこで、4次式における近似曲線を見てみると大きな違いの出ていることが分かる。そのため、 $\text{Diff} (\in T)$ を用いて分析した結果を以下に示す。

D	Diff	成績 (平均)
$1 < D$	≤ 3	2.5
	> 3	3
$+0 \leq D \leq 1$	≤ 3	null
	> 3	3
$-1 \leq D \leq -0$	≤ 3	3.16
	> 3	4
$D < -1$	≤ 3	2.93
	> 3	2.8

Diff については、 $1 < D$ かつ $\text{Diff} = 3$ の時が最小の2になる。そのため、 Diff については3を基準に分けている。成績の良かった分類としては、 $-1 \leq D \leq -0$ であることから、やはり線形近似曲線で緩やかに下降しており、特に4次の近似曲線において上下に揺れるようなアクセス回数が良いこととなる。これは、授業回が進むにつれて単調にアクセスが減るといった結果だけでなく、授業内容によって閲覧回数を自分なりに変えていった学

生の行動に結びついていると考える。また、線形近似曲線が上昇傾向にあっても、Diff > 3 の学生については成績も良いため、やはり授業内容と向きあう中で閲覧回数を変えた学生の習熟度上昇に関連していると考えられる。

すなわち、線形近似曲線によって、遠隔授業への慣れを推定することができ、また4次の近似曲線によって学生の試行錯誤の度合いを測ることができる。この時、試行錯誤の度合いのほうが習熟度により強い相関を持っていると考えられる。

4 クラスタリング手法による分析と比較

本章では、組 $T = \langle G, D, CD, CU, Diff \rangle$ に対してクラスタリングを行った結果を示し、3章で得られた結果との比較を示す。近年、ビックデータの活用に合わせてクラスタリングが広く行われている。クラスタリングとは、構造化されていないデータに対して類似度を計算し、内的結合と外的分離が達成される部分集合にデータを分割することである [12]。そのため、統計やデータマイニング、機械学習といった分野において広く用いられている手法であり、様々なアルゴリズムが存在する [1, 2, 5]。クラスタリングが目的とするのは、主観的な目的のために分類するのではなく、逆に分類された結果からどのような因子が重要になってくるのかを考える点にある。そこで、このクラスタリングの手法を用いて T を分類した。

まず、 T の要素はそれぞれの単位が異なるため基準化(標準化)を行った。すなわち全ての T の要素に対して、次の式から得られたものを T' とする。

$$\forall x, x'_i = \frac{x_i - \bar{x}}{\sigma}$$

ここで、

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

を表し、 \bar{x} は x の平均、 σ は標準偏差である。

基準化した $T' = \langle G', D', CD', CU', Diff' \rangle$ に対し、ユークリッド距離²を求め、階層クラスタリングと非階層クラスタリングを行った。分析には Vine Linux 6 CR 上の GNU R (Ver. 3.6.1) を使用した。

具体的な手順としては、WebClass より CSV 形式のログデータをダウンロードする。その後、同意を得た43名の学生についてデータ行列を得る。データ行列に対して基準化を行い、ユークリッド距離による距離行列を求めた後にコーフェン行列を求め樹形図を求める。樹形図の求め方や距離については様々な手法が開発されているが、本稿では次の手法を用いて分析を行った。

² 例えば、2次元平面の2点 (x_1, y_1) , (x_2, y_2) のユークリッド距離 d は $d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ で求める。

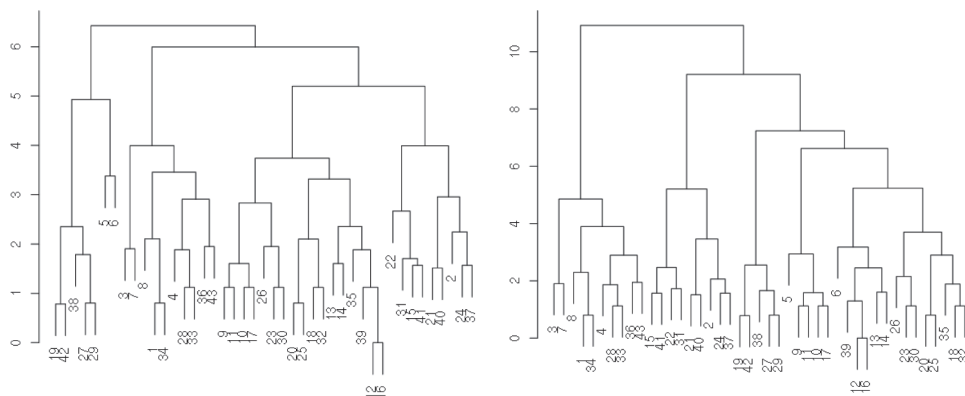


図4: 最長距離法 (左) とワード法 (右) によるクラスタリング

まず、最短距離法を用いてクラスタリングを行った。しかし本データに対して最短距離法を用いて分類するとチェイニング効果によって階段状の構造化が見られた。これは一般によく知られた現象であり、元データの質によるものではなく最短距離法自身の特性から生じる効果であるため、正しい分類とは言えないことが多い [2]。そこで、次に最長距離法 (完全連結法) とワード法 (最小分散法) を用いて分析した。最長距離法とワード法を用いた樹形図 (デンドログラム) を図4に示す。なお、学生の個人情報を伏せるために図4では単純な連番としている。一般的に、最長距離法は空間拡散の影響を受けて分割が多くなりがちであるが、本データにおいてはうまくクラスタリングが出来ていると考える。ワード法に比べてクラスタリングがうまくいっている最長距離法を基準として更に分析を行った。図4の左にある再長距離法では4つのクラスタに分かれると考え、クラスタの分析を行った。図5に示したのが、変数ごとにクラスターの平均を箱ひげ図で表示した結果である。

図5において、上下という要素があるが、WebClassよりダウンロードしてきたCSVファイルでは $D = +0$, $D = 0$, $D = -0$ の区別をつけるのが難しいため、線形近似曲線の傾きが上昇 (+) するか下降 (-) するかを上下という要素で示している。

この結果においても、クラスタとして特徴的になるのは4次式による近似曲線が主であり、3.2.2項で示したとおり線形近似曲線だけでなく4次式による近似曲線の有用度が示せたと考える。

さらに、同データに対して非階層クラスタリングを行った結果を図6に示す。階層クラスタリングは樹形図 (デンドログラム) からクラスタを類推する手法であり、始めから意図的にクラスタに分けることはしない。その一方、非階層クラスタリングの場合は先にクラスタの数を指定し、その上で分類する手法といえる。ここでは、K-means法を用いてクラスタに分けている。また、使用した計算アルゴリズムはHartigan-Wongである。

図6ではクラスタ4に関して分類精度が高いとは言えず、4次式の近似曲線以外の要素を取り入れることが必要と考えられる。ただし、K-means法は初期値によって結果が大きく変わるものであり、初期値の設定や他の手法を用いて分析していく必要がある。

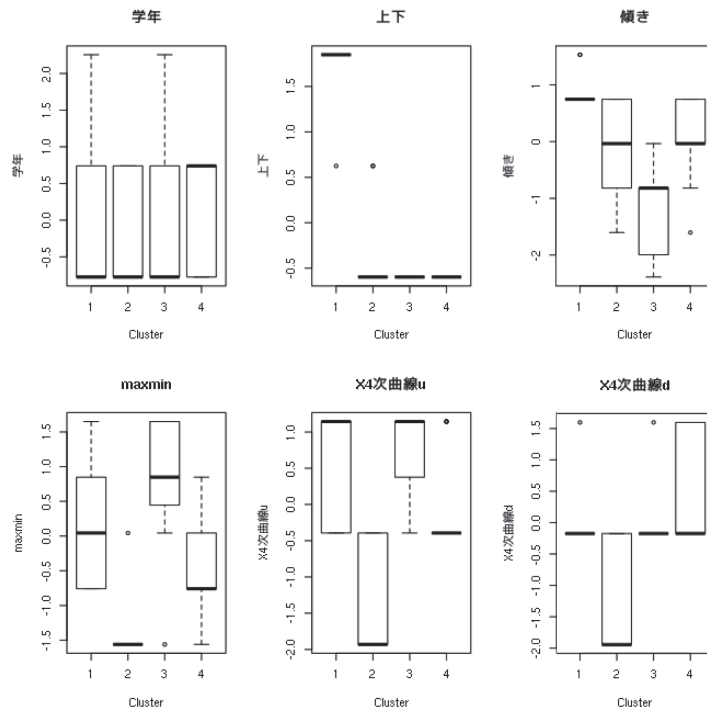


図 5: 変数ごとの箱ひげ図

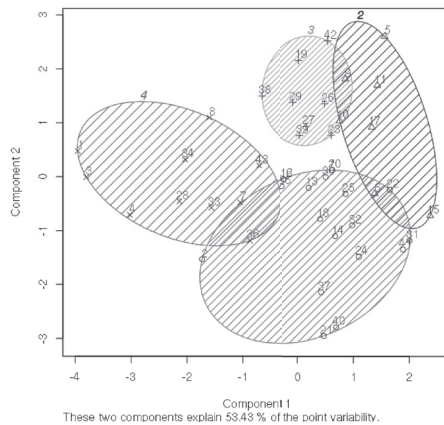


図 6: K-means 法 (クラス数 4) によるクラスタリング

5 おわりに

本稿では、LMS 上のログデータに着目し、学生の学習進捗状況を動的に分析する方法を示した。すなわち、従来型の授業結果からの類推ではなく、授業期間の途中においても分析が可能となる事を示した。特にアクセス回数と受講時間に着目し、非常に基本的なデータでありながら、その傾向によって学習進捗に影響がある事が分かった。進捗状況については 4 次式による近似曲線の有用性を示し、教材閲覧回数が単調に減る学生よりも、

授業内容において思考錯誤しながらアクセス回数を変えるほうが理解度も増えることが分かった。結果として、学生が遠隔教育にのぞむには深夜などの受講ではなく、開講時間中に一定数以上のアクセスを行うことが良いことが分かった。さらに、クラスタリング手法との比較を行うことで、4次曲線の有用性が示せた。検討課題としては、LMS上のデータを用いるための同意のとり方については全学的に検討できれば良いと考える。本稿のような分析によって学生の学習サポートにつながる事が考えられることから、データ利活用に関するポリシーを学生に周知していく必要がある。また、最後に示した非階層クラスタリングの結果では、さらなる分析要素の検討が必要であると考えられる。

参考文献

- [1] Cutting, D. R., Karger, D. R., Pedersen, J. O and Tukey, J. W.: Scatter/Gather: A Clusterbased Approach to Browsing Large Document Collections Proc. 15th Annual ACM SIGIR Conf. on Research and Development in INformatino Retiecal, pp.318-329, 1992.
- [2] Everitt, B. S.: Cluster Analysis, Edward Arnold, third edition, 1993.
- [3] Hagreaves, D. H. : The production, mediation and use of professional knowledge, Knowledge management in the learning society, OECD/CERI ed., pp.219-238, 2000.
- [4] Hargreaves, D.: Teaching as a research-based profession: possibilities and prospects, Educational research and evidence-based practice, Hammersley, M. ed., SAGE Publications Ltd., pp 3 -17, 2007.
- [5] Jain, A. K. and Dubes, R. C.: Algorithms for Clustering Data, Prentice Hall, 1988.
- [6] 惣脇宏：英国におけるエビデンスに基づく教育政策の展開，国立教育政策研究所紀要，第139号，2010.
- [7] 竹下浩，岡田行弘／同期型 e-learning における学習者特性とインタラクションの分析，日本教育工学会論文誌，Vol.32, No.2, pp.149-156, 2008.
- [8] 山本洋雄：e-learning での学習成績・学習時間・投資対効果などの効果測定，教育システム学会誌，Vol.19, No.1, pp.46-53, 2002.
- [9] 岩崎久美子：教育研究エビデンスの課題－知識社会における産出・普及・活用－，国立教育政策研究所紀要，第140号，2011.
- [10] 中山実，山本洋雄，Rowena, S.: 学習者特性がブレンド学習の行動に及ぼす影響 (e-Learning における学習評価／一般)，電子情報通信学会技術研究報告，ET，教育学，Vol.106, No.364, pp.49-54, 2006.
- [11] 植野真臣：eラーニングにおけるデータマイニング，日本教育工学会論文誌，Vol. 31, No.3, pp.271-283, 2007.
- [12] 神瀛敏弘：データマイニング分野のクラスタリング手法（1）：クラスタリングを使ってみよう！，人工知能学会誌，18巻，1号，pp. 59-65, 2003.
- [13] 今井康雄：教育にとってエビデンスとは何か：エビデンス批判をこえて，教育学研究，vol.82, no. 2, pp. 188-201, 2015.

Received : April, 26, 2021

Accepted : June, 9, 2021

