

情報教育課題合格ログデータによる受講生の類型化

中村 永友¹ 松井 佑介² 石川 千温³ 渡辺 慎哉⁴ 小池 英勝⁵

要 旨

札幌学院大学の情報教育は、コンピュータ・リテラシ、ワードプロセッサ、表計算ソフトウェアのスキルを身につけるための数多くの課題を課しており、課題の合格は単位取得に直結している。このスタイルの授業を始めて以来問題となっているのは、課題の進捗状況を能動的に把握しない学生が途中で脱落したり、出席良好にもかかわらず単位未修得となる学生の存在である。時系列的な課題合格情報を用いて、受講生の類型化とそれを基にした教育指導のための基礎的な資料を作成するための基礎研究を行った。

キーワード：関数データ解析法、関数多次元尺度法、関数 k -means 法、コンピュータ・リテラシ教育

1 はじめに

札幌学院大学におけるコンピュータ・リテラシ教育は1996年より実施しており、1年生の1年間で電子メール、ワードプロセッサ、表計算、プレゼンテーションの各ソフトウェアの習得を目的としたカリキュラムを実施している(中村 他, 2015; 小池 他, 2010)。これまで、前期にワードプロセッサの Microsoft Office Word を、後期に表計算ソフトウェアの Microsoft Office Excel を中心に行い、これらのソフトウェアのスキルをつけさせるために、類似の内容を含む多くの演習課題を課している(前期に17, 後期に16課題)。膨大な提出課題を公平に採点をするために、Word と Excel の課題に対して、学生自らが課題の可否を確認するための課題採点ツール(以下、ツール)を2000年以來開発し、2005年度から本格的に運用している(石川 他, 2008; Koike *et al.*, 2007)。

本報告の目的は、課題の合格数に関する情報から、学生の課題提出の挙動を通して学習進捗状況を分類(類型化)し、単位の取得状況(可否情報)とつきあわせることでどのような情報を得ることができるのかを

探ることである。具体的には、授業が始まってからの課題の累積合格数を表す階段状のデータを作成し、ある種の平滑化手法で滑らかな曲線として近似したもの(関数データ)をまず作成する。それらに関数データ解析法を通して類型化して、各グループの特性を明らかにしていく。さらに、このようにして得られるある種のビッグデータを分析することで、学生一人ひとりの指導の処方箋を作るための資料としたい。典型的なビッグデータは、個人の商品購入履歴等が膨大に蓄積したものが挙げられ、本分析対象データは同類のものと考えられる。

本報告で用いる関数データ解析法(Functional Data Analysis: FDA)は、「関数」をデータとみなす一連の分析手法である。FDA は1990年代に Ramsay らによって提唱され、これまでの成果をまとめた成書も出版されている(Ramsay & Silverman, 2005)。FDA の一連の手法は従来の多変量解析法を関数データに対応(適応)させて拡張したものが多く、これまでに関数回帰分析、関数主成分分析、関数クラスター分析法など、多くの有用な手法が開発されている。

以下、データの形式、分析の方法、類型化の結果とその特徴、学科との関係について述べる。

2 データの形式

分析対象となるデータは、2013年度後期の全18課題

¹ 札幌学院大学経済学部; nagatomo@sgu.ac.jp.

² 名古屋大学医学部; matsui@med.nagoya-u.ac.jp.

³ 札幌学院大学経営学部; chiharu@sgu.ac.jp.

⁴ 札幌学院大学経営学部; wattan@sgu.ac.jp.

⁵ 札幌学院大学社会情報学部; koike@sgu.ac.jp.

で、パワーポイントが2課題、エクセルが16課題である。元のデータの形式は、授業が始まった日を起点として、完成させた課題（正確には正規合格した課題）があるたびに1を加えていく。例えば、4日目に1課題、6日目に2課題、8日目に1課題を提出したとすると、

$$0, 0, 0, 1, 1, 3, 3, 4, 4, \dots$$

のように、 n 番目の数字が最初の日から n 日目までにいくつの課題を提出したかを表し、単調増加する数列が学生1人のデータである。また、休日などが間にある場合はそれを取り除くなどの処理をしている。

746人の学生が分析対象で、一人ひとりのデータは累積分布関数のような階段状をしている(図1上)。これをBスプラインによって平滑化し(図1下)、各個人の一連の累積データを関数データ化したものが、実際の分析対象となる。

3 クラスタ生成の方法

得られた個別の曲線を観測されたデータ、すなわち関数データとすることで、関数データ解析法の枠組みで種々の分析ができる。ここでは、関数多次元尺度法と関数クラスター分析法(関数 k -means 法)を用いて、一人ひとりの課題進捗状況を何種類かのパターンに類型化(分類)を行う。そして関数データ間の非類似度を計算し、それを利用して多次元尺度法により、2次元のユークリッド空間における布置を行う。これでクラスターの様子を観察できるようになる(図2)。

ここで、関数データ化する際に平滑化手法としてBスプライン法を用いたが、これは2回まで微分が可能なものとした。これによって、一人ひとりの課題作成の進捗速度(1回微分)と進捗加速度(2回微分)を得ることができ、関数データの特徴を測ることもできる。これらの詳細は松井他(2015)を参照。

4 結果

4.1 グループの特徴

図1下の一人ひとりの関数データ化した曲線を6つに類型化した結果が、図3の左側の「進捗曲線(平滑化データ)」の列である。同様に、中央の列は、進捗速度(1回微分)を、右列は進捗加速度(2回微分)を示す。

進捗曲線の類型化から、あえて名前をつけると

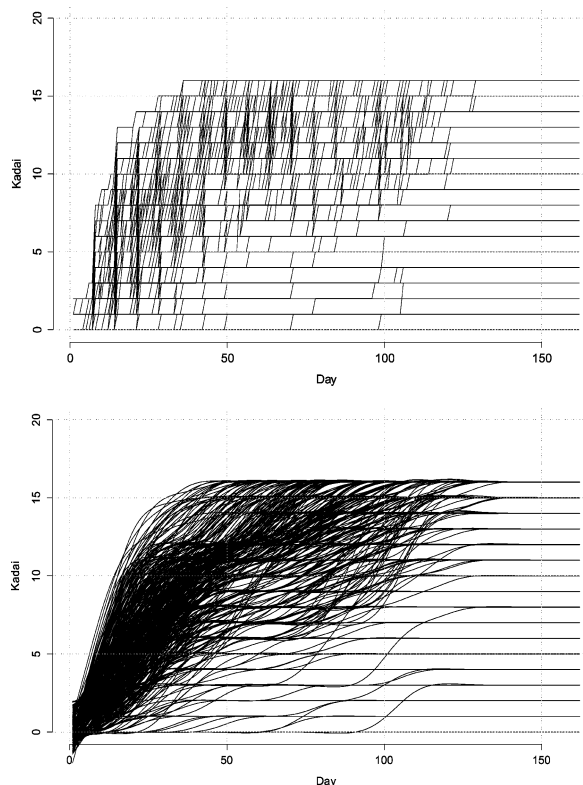


図1：分析対象のデータ(上)と平滑化した同データ(下)

類型Ⅰ 前半ダッシュ・合格型

類型Ⅱ 堅実ペース(スケジュール少し前倒し)・合格型

類型Ⅲ 堅実ペース(スケジュール通り)・合格型

類型Ⅳ 後半ダッシュ・ほぼ合格型

類型Ⅴ 中盤以降停滞・ほぼ不合格型

類型Ⅵ エンジンかからず・不合格型

となるであろう。これらの類型化は、関数 k -means 法を用いていることから、グループの間で重なり(オーバーラップ)を想定しないので、例えば類型Ⅰの進捗曲線の下方にある曲線と、類型Ⅱの上方の曲線は、ほぼ同じような性質を持つ関数データである。これはクラスター数を固定して、互いに素な分類を行うという手法の制約によるものである。図2の散布図は、グループ間で排反の様子を示すものである。

表1は類型化したグループごとの人数、単位取得者の人数(合格者数)、合格率、最終得点の平均(層別平均)、合格者数の累積数と比率、出席ポイントを示す。層別の平均の傾向と、さらに出席ポイントの大きさが合格率と符合していることがわかる。この出席ポイントとは、1回の授業で最初から最後までPCにサイン

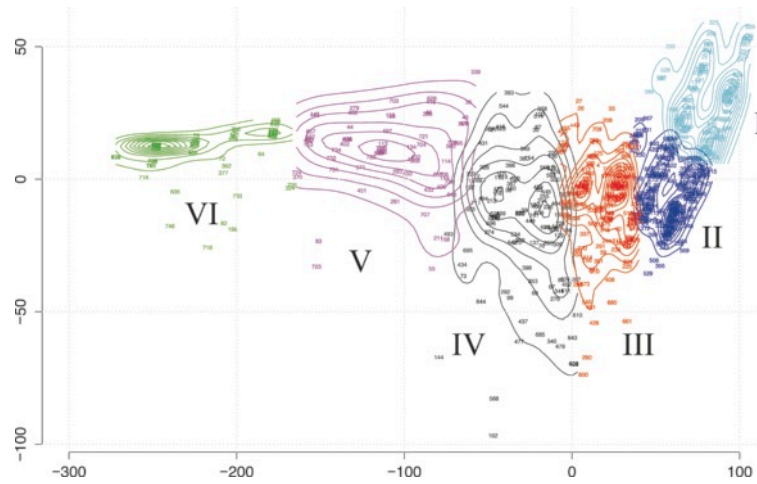


図 2：多次元尺度法による布置と類型化

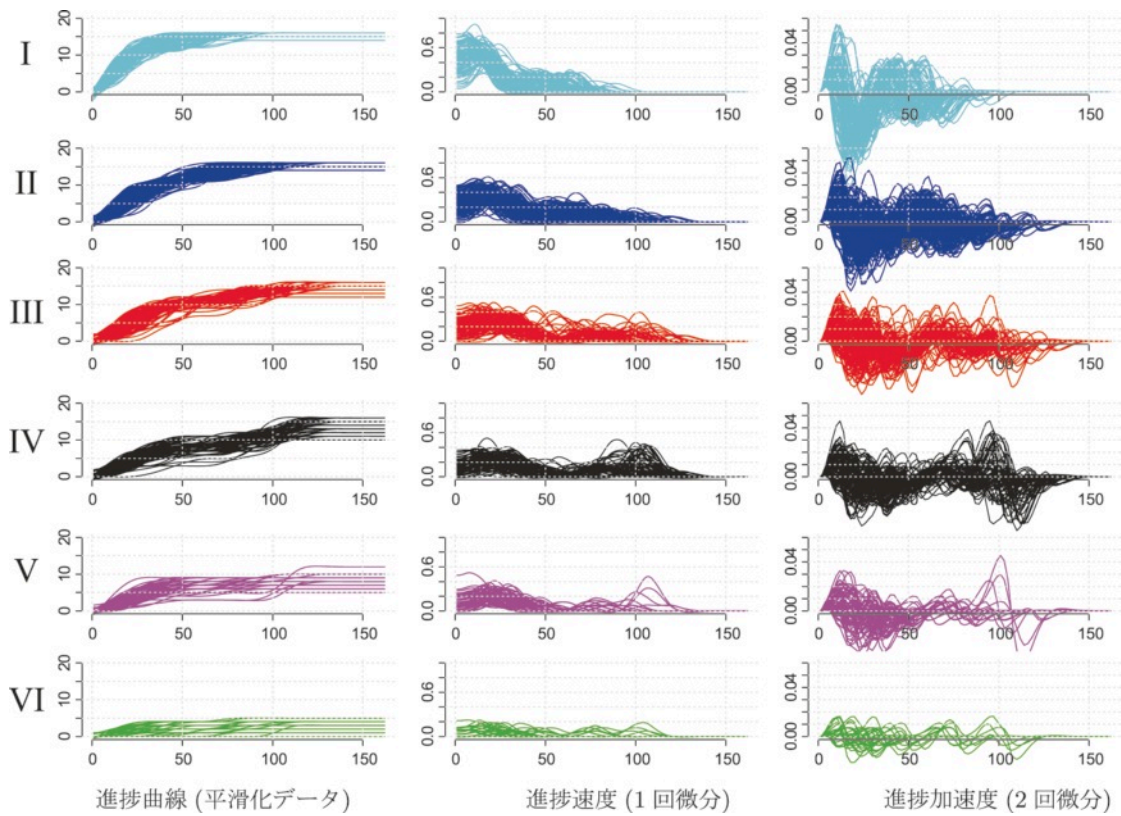


図 3：平滑化したデータの類型化

表 1：グループ別の統計量

グループ	層別平均	合格者数	人数	合格率	合格数 (累積)	累積比率	出席ポイント	グループの特徴
I	89.2	111	111	100.0%	111	14.9%	14.3	前半ダッシュ・合格型
II	81.3	224	225	99.6%	335	44.9%	13.6	堅実ペース (スケジュール少し前倒し)・合格型
III	68.5	156	162	96.3%	491	65.8%	12.6	堅実ペース (スケジュール通り)・合格型
IV	56.2	90	116	77.6%	581	77.9%	11.2	後半ダッシュ・ほぼ合格型
V	29.7	11	66	16.7%	592	79.4%	8.2	中盤以降停滞・ほぼ不合格型
VI	10.0	2	66	3.0%	594	79.6%	5.1	エンジンかからず・不合格型

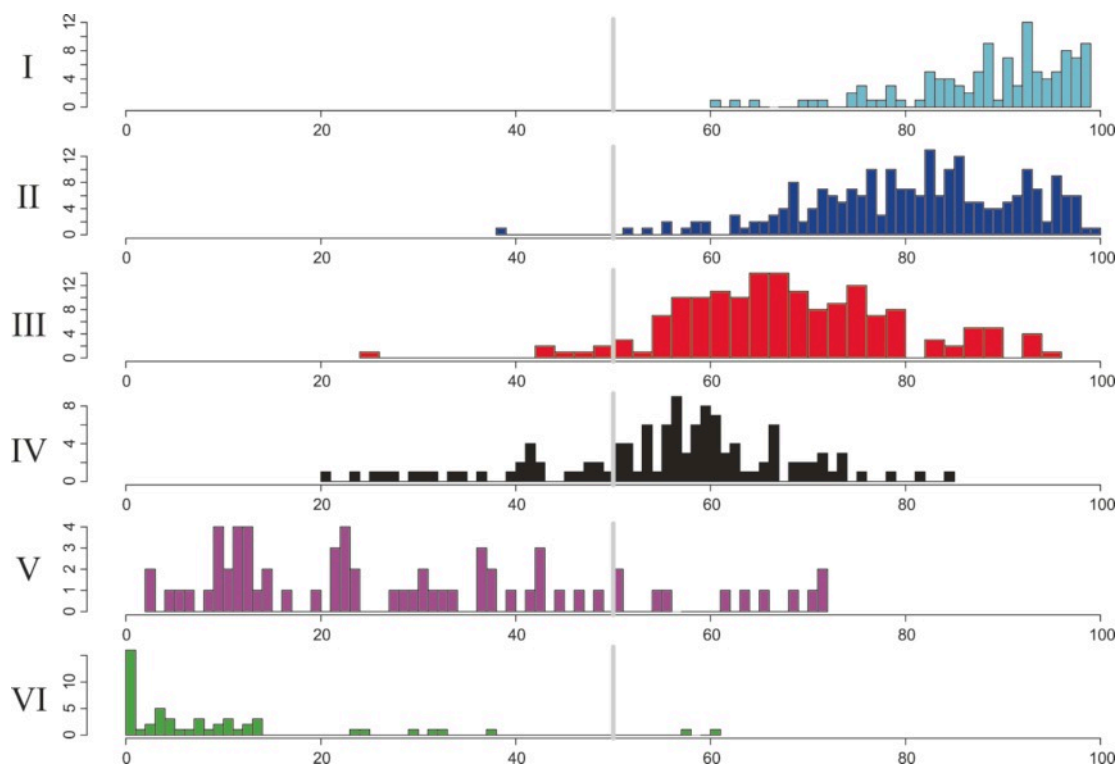


図4：類型化した各グループの得点分布

インしていれば1ポイント、遅刻や早退により短くなった時間に比例してポイントが下がる（小さくなる）。これを15回分の授業で累積したもので、満点が15ポイントとなる。

図4は類型化した各グループの最終的な得点(100点満点)のヒストグラムである。表1とあわせてみると、第I～IV群がほぼ合格で、第Vと第VIはほぼ不合格のグループであることがわかる。この事実から、第I～III群の合格者数は受講者全体の2/3を占め、このグループの学生はほとんど問題のない学生たちである。第IV群は合格率が78%で、この学生たちも課題を仕上げるのにそれほど問題は無い。第V群の学生は合格率が17%と非常に低く、特徴として「中盤以降停滞」していることから、授業中の指導にあたり、ここに分類される学生を早く見つけ出し、早めの手当をすることが必要である。第VI群は、よく調べると欠席も多く指導自体が困難な要因を含んでいる。

注意事項として、単位を取得するためには、ここでの分析対象となる関数データ化した進捗曲線以外の要素がある。それは最終的な合否(単位の取得の可否)の対象となる得点は、(1)各課題の完成時の得点の違い、(2)出席ポイント(授業時間に対する総ログイン時間の比率による得点)が7割以上ないと正規の課題得点が

得られないこと、等の要因である。

4.2 グループに対する学科の特徴

表2は類型化したグループ(I～VI)と学科のクロス表から、学科内での各類型に対する比率を表したものである。学科内でどのようなパターンの学生がどの程度いるのかその様子が見えてくる。数字の羅列を見てもその特徴を把握しがたいので、このデータに対して類型(I～VI)を変数、学科を個体として主成分分析を行い、その因子負荷量と主成分得点を散布図にしたのが図5である。横軸が第1主成分、縦軸が第2主成分で、第2主成分までで約76%の情報がある。この図が示すところは、原点から見てそのタイプの傾向が強い学科がある方向にあるということである。つまり、原点から類型Iの方向にG学科があり、同様に類型IとIIの方向にA、E、H学科が、類型IIとIIIの方向にC、F学科がある。さらに、類型IV、V、VIの方向にB、D学科がある。

5 考察

課題提出進捗データを関数データとして扱い、6つの類型化をして特徴を調べた結果、第V群に対して何らかの介入をする必要があることが示唆された。

表 2：各学科のグループ比率

学科	I	II	III	IV	V	VI	合計
A	22%	29%	24%	9%	4%	11%	100%
B	6%	19%	24%	21%	16%	13%	100%
C	9%	36%	27%	13%	9%	6%	100%
D	10%	32%	15%	22%	12%	10%	100%
E	19%	38%	21%	10%	4%	8%	100%
F	12%	35%	30%	16%	2%	5%	100%
G	33%	20%	14%	17%	10%	7%	100%
H	20%	40%	20%	10%	5%	5%	100%

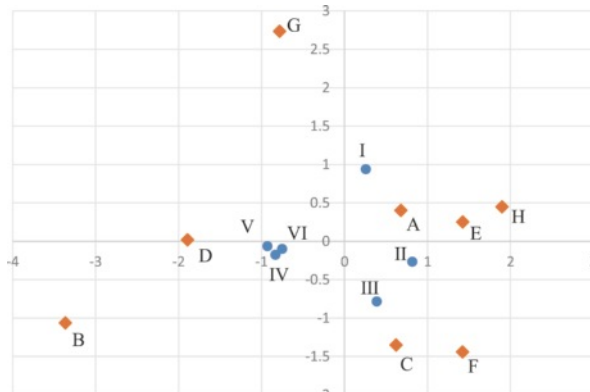


図 5：類型化（第 I～VI グループ）と学科（A～H）の傾向

松井他（2015）によれば、50日前後までに本授業の単位の可否が定まるという結果も出ている。図 3 の第 V グループの進捗曲線が50日くらいまでに平坦になり、また進捗速度も50日程度にかけて速度が落ちていることを見れば、この結果もうなずける部分もある。

また、第 VI 群は比較的初期の段階であきらめがちな学生が多いことがわかる。それは、例えば図 3 の進捗速度や進捗加速度から、その変化の程度が他のグループに比べて小さいことからそのことが観察される。そもそも出席ポイントも低いことから、まずは出席を促すところから始まるのだろう。

6 今後の課題

今回分析した同じデータに対して、中村他（2015）ではロジスティック回帰モデルによってある時点における課題の合格数から、合否確率を算出した。この研

究の問題点は、データ全体の傾向からロジスティック回帰曲線を求めていることから、平均的傾向を求めたものであった。しかし、今回の分析では、個々の学習状況を表す曲線を基礎として類型化しているところが異なる点である。

また、今回の分析は課題の提出過程に注目して学生の類型化を行ったが、第 V、第 VI 群は出欠に関する情報をあわせて分析をした方が、個別の指導をする際の細かい情報が判別するのではないかと考えられる。これは今後の研究課題としたい。

謝辞 本学の情報教育の授業は多くの教育指導員（TA）によって支えられている。彼らに深く謝意を表す。

参考文献

- [1] 石川千温・皆川雅章・渡邊慎哉・中村永友・小池英勝・梅田 充 (2008). 自動採点システムを用いたコンピュータリテラシー教育の質保証, 教育システム情報学会第33回全国大会予稿集, 熊本大学.
- [2] Koike, H., Akama, K., Ishikawa, C., Nakamura, N., and Watanabe, S. (2007). Toward a Software Development Model for Automatic Marking Software, Proc. of ACM Special Interest Group on University and College Computing Services (SIGUCCS) Fall 2007 Conference, 190-193, Oct. 7-10, 2007, Orland, Florida, America.
- [3] 小池英勝・中村永友・石川千温 (2010). レポート・プレゼンが楽々！ パソコン活用術, ムイスリ出版, 東京.
- [4] 中村永友 (2009). 多次元データ解析法, 共立出版, 東京.
- [5] 中村永友・石川千温・渡邊慎哉・小池英勝 (2015). ロジスティック回帰による課題提出ログデータの解析, 札幌学院大学総合研究所紀要 (情報科学), Vol.2, 1-6, 2015.3.
- [6] Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis*, 2nd ed., Springer, New York.
- [7] 松井佑介・島村徹平・水田正弘・中村永友 (2015). 関数データクラスタリングにおける部分区間の同定法と学習ログデータへの適用, 日本分類学会第33回大会予稿集, 2015.3, 2-3.

Classification of Students by a Functional Data Clustering Based on the Submission of Learning Tasks into a Log System

Nagatomo NAKAMURA¹

Yusuke MATSUI²

Chiharu ISHIKAWA³

Shin-ya WATANABE⁴

and

Hidekatsu KOIKE⁵

Abstract

Education for usage of word processor and spreadsheet software are carried out at Sapporo Gakuin University. In order to improve the students' skills of these software, we impose a number of exercises. Passing this exercises is directly connected to the acquisition of credits. The objective of this report is to classify the students' advancement in the submission of tasks according to the number of passing grades. It is possible to carry out various analysis according to the functional data analysis based on the number of passing grades of individual exercises. We try to prepare the materials for the guidance of each student by investigating the characteristics of each group.

Keywords: Functional Data Analysis, Functional Multidimensional Scaling, Functional k -means, Computer Literacy Education.

¹Department of Economics, Sapporo Gakuin University; nagatomo@sgu.ac.jp.

²Department of Medicine, Nagoya University; ymatsui@med.nagoya-u.ac.jp.

³Department of Bussines Administration, Sapporo Gakuin University; chiharu@sgu.ac.jp.

⁴Department of Bussines Administration, Sapporo Gakuin University; wattan@sgu.ac.jp.

⁵Department of Social Infomation, Sapporo Gakuin University; koike@sgu.ac.jp.