

MOOCにおける大規模学習履歴データからの 受講者の学習様態獲得

Acquisition of learning style of MOOC learners from large-scale learning log-data

永田 裕太郎^{1*} 村上 正行² 森村 吉貴¹ 椋木 雅之¹ 美濃 導彦¹
NAGATA Yutaro¹ MURAKAMI Masayuki² MORIMURA Yoshitaka¹
MUKUNOKI Masayuki¹ MINOH Michihiko¹

¹ 京都大学 Kyoto University

² 京都外国語大学 Kyoto University of Foreign Studies

Abstract: In this paper, we consider a method for acquiring learning style of MOOC learners. Learning style is a pattern on how learners learn. It is useful for helping the learners in both not to dropout from the course and improving the learning materials.

When a learner learns on a MOOC system, the log-data are recorded on the system, such as the learner's operations and system's responses. We focus on transition events that occur when the learner moves from a page to another page. MOOC learning materials consist of two types of pages, that is video-pages and problem-pages. Each of transition events is described by "transition-feature" which consists of a 3-tuple; "quantity", "from page-type" and "to page-type".

We clustered the MOOC learners with transition-feature vector and compared the generated clusters based on the pass rate of the course. In the process we extracted the clusters of learners which a previous research had been suggested to be exist. The results of our analyses indicated: 1) there was a correlation between the pass rate and the number of transition from a video-page to the next video-page. 2) There also was a correlation between the pass rate and the number of transition from a video-page to the previous video-page, then we obtained the hypothesis that the learners who watch video-pages more were less likely to dropout.

1 はじめに

受講者の学習を支援するうえで、学習状況を把握し、それに応じて支援を行うことは、きわめて重要である。特に、eラーニングに代表される通信教育では、途中で学習をやめてしまう受講者(ドロップアウト)が多いことが問題となっている。通信教育の場合、対面授業と比較して受講者の学習状況を把握することが難しく、また状況に応じたフィードバックも困難であるため、そうしたことを考慮したうえでの学習支援が必要である。学習支援には大きく分けて、二つの方向性がある。一つは、ドロップアウトの兆候がみられる受講者に対して状況に応じた適切な介入を行い、ドロップアウトを防ぐという方向性である。もう一つは、受講者が実際にどのように学習したのかを分析し、分析結果を教材

コンテンツや、教育プロセスそのものの改善に生かすという方向性である。

そして、2010年代に入って急速に注目を集めているのがMOOC(Massive Open Online Course)である。MOOCの特筆すべき点は、従来のeラーニングに比べてきわめて大規模なことであり、コース受講者が数千人、数万人にのぼることも珍しくない。受講者が国籍の面でも、年齢層や学歴の面でも多様であることもまた、MOOCの特徴である。そのようなMOOCの有効性を高め、普及させていくためには、受講者の学習状況の把握、および、適切な学習支援が重要となる。

そこで本研究では、MOOCにおける学習支援を目的として、MOOC受講者の学習履歴データ(アクセスログデータ)から学習様態を獲得する手法について検討する。ここで学習様態とは、受講者が「学習にどのように取り組んでいるか」を類型化したものである。

本稿では、受講者の学習様態を分析するために、コンテンツ遷移に着目する。一般にMOOCの教材コン

*連絡先：京都大学 学術情報メディアセンター 美濃研究室
〒606-8501 京都市左京区吉田二本松町
E-mail: ynagata@mm.media.kyoto-u.ac.jp

コンテンツは講義動画とテストからなり、個々のコンテンツ間を受講者がどう遷移するかという情報は、学習様態獲得の手がかりとなる。コンテンツ遷移にドロップアウトの兆候がみられると判断できれば、システム上での介入を行うことができる。また、受講者間で多くみられるコンテンツ遷移を分析することで、よくみられるコンテンツ遷移に合わせてコンテンツを再配置したり、多くの受講者が再確認する(そこでつまづいている可能性が高い)コンテンツを改善したりすることができる。このようにコンテンツ遷移の分析は、介入・改善の双方に役立つ。

提案手法では、アクセスした講義動画の割合、および、回答したテストの割合の2つの特徴を用いたフィルタリングを行ったうえで、受講者のコンテンツ遷移に着目してクラスタリングを行うことで学習様態を獲得する。さらに、修了率と学習様態の関係について分析を行い、特徴となる学習履歴について考察を行う。

2 関連研究

学習様態の一部については、以前から欧米では学習スタイルとして盛んに研究が行われてきた [1]。これらは教育・心理学・経営などの分野において、学びの個人差という観点から行われてきたものである。近年では大学教育のみならず、eラーニングへの適用・応用に関する研究も多く行われている [2]。大山ら [3] は、FELDERによってモデル化された学習スタイルをもとに、異なる傾向をもつ学習者を選定し、それぞれに教材構造の異なる2つのeラーニング教材に取り組みさせたうえでインタビューおよびログの分析・可視化を行った。その結果、活動的傾向のある学習者と内省的傾向のある学習者とは、異なる学習様態があらわれ、教材構造に対する嗜好にも違いがみられる、という可能性が示唆されたとしている。

上記のような学習理論に則ったアプローチでの研究とは別に、学習に関するデータから学習様態を発見・獲得しようというアプローチでの研究も増えてきている。孫ら [4] は、eラーニング受講者のコンテンツ遷移に注目し、ルールベースに照合することによって、受講者の学習様態を推定している。また、植野 [5] は、eラーニングにおける学習所要時間を用いて受講者の能力と課題の難易度を推定することで、飛ばし読み・他事をしている・行き詰まり状態などの異常学習プロセスを検知する手法を提案している。

荒ら [6][7] は、東京大学がCourseraへ提供した2コースから得られたデータを分析し、MOOC受講者の学習様態として、様子見的学習者(コース登録はしたものの、ほとんど何もしない受講者)・知識獲得型(講義動画はみるがテストには取り組まない受講者)・既有知識

確認型(講義動画をみずにテストにのみ取り組む受講者)・学習活動完遂型(講義動画・テストの双方に取り組む受講者)の4類型があることを示唆している。この4類型に基づいて考えれば、学習支援が必要なのは「学習活動完遂型」の、講義動画・テストの双方に取り組む受講者である。

3 学習履歴データに基づく受講者の分析手法

まず2次元の閲覧数特徴を用いて、受講者全体からなる集合をクラスタリングする(閲覧数によるフィルタリング)。アクセスした講義動画の割合、および、回答したテストの割合を用いてクラスタリングすることで、荒らの4類型に対応するクラスタを得る。

次に、得られたクラスタのうち「学習活動完遂型」に対応する受講者のクラスタから、コンテンツ遷移が多い受講者の群のみを取り出し、さらにクラスタリングを行う(遷移特徴ベクトルによるクラスタリング)。こうして得られたクラスタを解釈することにより、学習様態を獲得する。生成したクラスタの中心、すなわちクラスタごとのコンテンツ遷移の傾向と、最終的な修了率との相関を調べることで、学習様態についての考察を行う。

提案手法の処理の流れを、図1に示す。

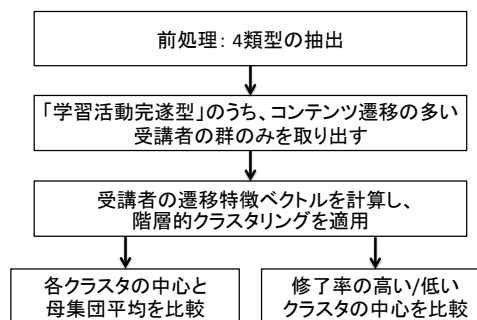


図1: 提案手法の処理の流れ

3.1 コンテンツ遷移に基づく学習様態の獲得

受講者がMOOCシステム上で学習を進めると、それに伴って生じたイベントと、それに対するシステムのレスポンスとが、アクセスログデータとして記録される。記録されている情報のうち、受講者があるページからあるページへ遷移したというイベント(コンテンツ遷移)に着目する。なぜならば、同じようなコンテンツ遷移をした受講者は、教材コンテンツに対して同じ

ように取り組んだとみなせるからである。従って、コンテンツ遷移に着目することによって、MOOC受講者が教材コンテンツに対して「どのように取り組んでいるか」を類型化したもの、すなわち学習様態を獲得することができる。

受講者の個々のコンテンツ遷移にラベル付けを行う。このラベルは(遷移量, 遷移元ページ, 遷移先ページ)の3つ組によって表現されるものとする。 T 種類のラベルが受講者ごとにそれぞれ何回ずつ現れたかを用いて、受講者の遷移特徴ベクトルを定義する。遷移特徴ベクトルの集合をクラスタリングすることで、コンテンツ遷移の傾向が似たような受講者からなるクラスタが生成できる。つまり、クラスタリングによって、多くみられるコンテンツ遷移をしている受講者からなるクラスタや、あまりみられないコンテンツ遷移を同じようにしている受講者からなるクラスタが得られる。こうして得られたクラスタを解釈することにより、学習様態を獲得する。

3.2 閲覧数によるフィルタリング

アクセスした講義動画の割合、および、回答したテストの割合の2つの特徴を用いることで、4類型に対応するクラスタを得る。

ある単元 D において、受講者 i がアクセスした講義動画の個数を $v_{i,D}$ 、回答したテストの個数を $p_{i,D}$ とする。但し、同じ講義動画へ複数回アクセスした場合や、同じテストに複数回回答した場合には、重複しては数えず、1個として数える。単元 D に属する講義動画の個数を V_D 、テストの個数を P_D として、単元 D における受講者 i の閲覧数特徴ベクトル $r_{i,D} (\in \mathbb{R}^2)$ を次式で定義する。

$$r_{i,D} = \left(\frac{v_{i,D}}{V_D}, \frac{p_{i,D}}{P_D} \right)$$

単元 D における全受講者の閲覧数特徴ベクトル $r_{i,D}$ の集合 $R_D = \{r_{i,D} \mid 1 \leq i \leq N\}$ (N は受講者数) をクラスタリングすることによって、4類型に対応するクラスタが得られる。

3.3 遷移特徴ベクトルによるクラスタリング

受講者の個々のコンテンツ遷移にラベル付けを行い、受講者ごとにラベル L_1, L_2, \dots, L_T がそれぞれ何回ずつ現れたかを数えることによって、受講者 i の単元 D における遷移特徴ベクトル $s_{i,D}$ を定義する。

個々のコンテンツ遷移を、次の3つ組からなるラベルで表現する。

(遷移量, 遷移元ページ, 遷移先ページ)

本稿では、遷移量は {1つ次へ進む (ordered+1), 1つ前へ戻る (reverse-1), 2つ以上次へ進む (ordered+n), 2つ以上前へ戻る (reverse-n) } のいずれかであり、遷移元ページ・遷移先ページはそれぞれ { 動画ページ (video), テストページ (problem) } のいずれかであるとする。このとき、ラベルのとりうる値は $2 \times 2 \times 4 = 16$ 種類である。

受講者 i の単元 D におけるラベル $L_t (t = 1, 2, \dots, T)$ の出現頻度を $x_{i,D,t}$ とし、これを並べたものを、受講者 i の単元 D における遷移特徴ベクトル $s_{i,D} (\in \mathbb{Z}^T)$ とする。

$$s_{i,D} = (x_{i,D,1}, x_{i,D,2}, \dots, x_{i,D,T})$$

コンテンツ遷移の数が近い受講者 K 名の単元 D における遷移特徴ベクトル $s_{i,D}$ の集合 $S_D = \{s_{i,D} \mid 1 \leq i \leq K\}$ をクラスタリングすることによって、ラベルの数は似ているがラベルの出現傾向が異なるクラスタが得られる。

4 分析対象の概要

4.1 コンテンツ構成

以下では、京都大学がMOOCプラットフォームの一つであるedXに提供し、2014年4月から7月まで開講された“The Chemistry of Life”コースのアクセスログデータを分析する。

“The Chemistry of Life”コースは、上杉志成教授(京都大学 物質-細胞統合システム拠点 / 京都大学 化学研究所 ケミカルバイオロジー)らによる“idea generation”と“integration of chemistry and biology”についてのコースである。教材コンテンツはすべて英語で提供されている。図2に示すように、講義動画の横には、教示者の発話内容に対応したテキストスクリプトが表示されており、動画再生時には対応部分が強調表示される。

本研究では、講義セクションのうち13セクションと、試験セクションのうち8セクション、あわせて21セクションを分析対象とする。この21セクションを、その内容と公開日に基づいて8セットに分割した。

各セクション、各ページには図2に示すようなナビゲーションがある。受講者は画面左端にある垂直方向のナビゲーションを利用し、学習したいセクションへ遷移する。セクション内では画面上部にある水平方向のナビゲーションを利用して、個々のページへアクセスすることができる。ページは講義動画コンテンツを中心とするページ(動画ページ)とテストコンテンツを中心とするページ(テストページ)のいずれかである。水平方向のナビゲーションには、動画やテストを示すアイコンが表示されており、受講者はいま自分がいるページがどのあたりなのかを把握できる。また、特定

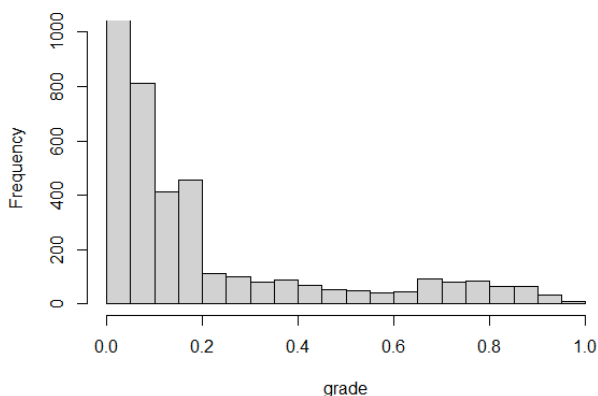
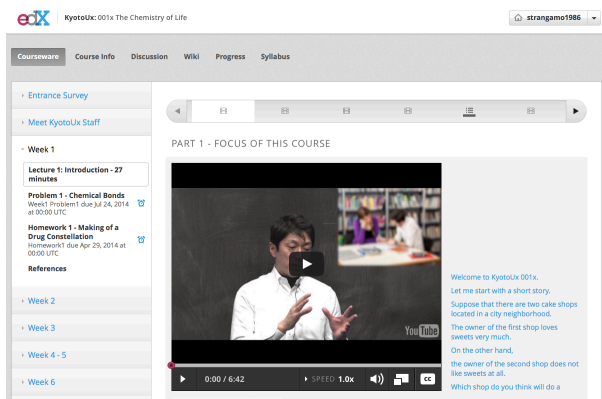


図 3: 最終成績の分布

の動画だけをみたり、テストだけをたどったりすることも可能である。

4.2 受講者数

”The Chemistry of Life” コースに登録し、最終成績が記録された受講者は 16,834 名であった。うち、最終成績が 20% 以上であった受講者は 1,113 名、修了認定の条件である 65% 以上に達した受講者は 447 名であった。最終成績の分布を図 3 に示す。横軸は最終成績 (5% 刻み)、縦軸は受講者数である。但し、最終成績が 5% 未満である受講者が 11,238 名と多かったため、その部分についてはグラフを打ち切っている。

5 分析

5.1 閲覧数によるフィルタリング

2次元の閲覧数特徴 (アクセスした講義動画の割合、および、答えたテストの割合) を用いて受講者全体をクラスタリングし、「講義動画はみるがテストには取り組まない受講者」「講義動画をみずにテストにのみ取り組

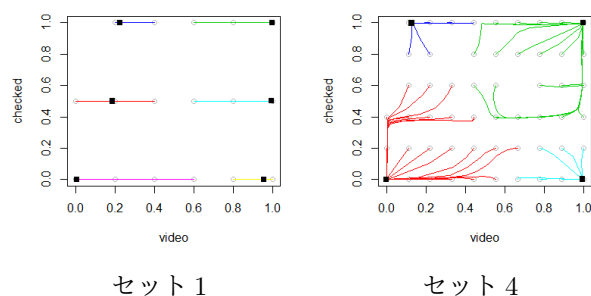


図 4: Mean Shift 法におけるクラスタリング結果

む受講者」および「講義動画・テストの双方に取り組む受講者」のクラスタを取り出すことを目指した。

全 16,834 名の受講者について、セット 1 からセット 8 までのそれぞれにおける閲覧数特徴ベクトルを計算し、それぞれのセットごとにクラスタリングを行った。クラスタリングには Mean Shift 法を用い、バンド幅パラメータは $h = 0.1$ とした。

結果、各セットは 4 個～8 個のクラスタに分かれた。クラスタリング結果のうち代表的なものとして、セット 1 とセット 4 の結果を図 4 に示す。図中の ■ がクラスタ中心を表す。図の横軸は各セットにおいてアクセスした講義動画の割合に対応しており、縦軸は各セットにおいて回答したテストの割合に対応している。他のセットにおいても下記の 4 点にはクラスタ中心があらわれ、このような受講者が多かったことがわかった。

- (0.0, 0.0) : 何もしていない [Neither 型]
- (1.0, 0.0) : 講義動画にはアクセスしているが、テストには一切回答していない [Video 型]
- (0.2, 1.0) : テストにのみ回答し、講義動画にはアクセスしていない [Problem 型]
- (1.0, 1.0) : 講義動画・テストの双方に取り組んでいる [Both 型]

これらは荒らの 4 類型に対応しているとみなすことができる。

5.2 遷移特徴ベクトルによるクラスタリング

セット 1 で Both 型クラスタに属した受講者のうち、コンテンツ遷移が多い受講者の群について、遷移特徴ベクトルを計算し、クラスタリングを行った。

セット 1 では Neither 型クラスタに 8,994 名、Video 型クラスタに 993 名、Problem 型クラスタに 239 名、Both 型クラスタに 4,414 名の受講者が分類された。このうち Both 型クラスタの 4,414 名のうち、コンテンツ遷移が多い群 (遷移が 15 回以上; 受講者 697 名, うち修

了者 102 名) について各受講者の遷移特徴ベクトルを計算し、階層的クラスタリングを行った。ここで、遷移特徴ベクトル同士の距離はユークリッド距離を用いて、集合間の距離はワード法を用いて定義した。階層的クラスタリングによって得られたデンドログラムを図 5 に示す。また、 $C_{1.1} \sim C_{1.5}$ の 5 クラスタに分けた場合の、それぞれのクラスタに属する受講者数 (s) と修了者数 (p)、および修了率 (p/s , 小数第 4 位以下を切り捨て) を表 1 に示す。

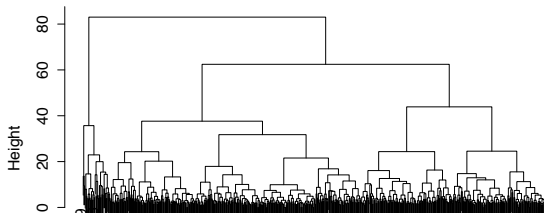


図 5: デンドログラム (セット 1 - 遷移数 15 以上の群)

表 1: 各クラスタにおける受講者の修了率 (セット 1 - 遷移数 15 以上の群)

	$C_{1.1}$	$C_{1.2}$	$C_{1.3}$	$C_{1.4}$	$C_{1.5}$	計
s	42	250	127	139	139	697
p	14	31	30	7	20	102
p/s	0.333	0.124	0.236	0.050	0.143	0.146

5.3 考察

受講者 697 名の群全体と比べて、修了率に有意な差がみられたクラスタ $C_{1.1}$ 、 $C_{1.3}$ 、 $C_{1.4}$ に関して考察を行う (t 検定, $p < 0.05$)。各クラスタの中心、および、受講者 697 名の群全体での平均値を表 2 に示す (小数第 2 位以下を切り捨て)。表の各列は、それぞれのクラスタに属する受講者が平均して何回の遷移をしたかを表す。

クラスタ $C_{1.1}$ は、5 クラスタ中でも修了率が飛び抜けて良いクラスタである。クラスタ中心の値を群全体での平均と比べると、ほぼすべての遷移が多く、動画ページから動画ページへの遷移数が特に多い。一方で、動画ページからテストページへの遷移や、テストページから動画ページへの遷移は、それほど多くはない。このクラスタに属する受講者は、教材コンテンツに熱心に取り組み、特に動画ページを集中して閲覧する姿勢が、最終成績の良さにつながったと解釈できる。

クラスタ $C_{1.4}$ は、5 クラスタ中で修了率が最も低く、遷移数が多い割には伸び悩んでいるクラスタである。クラスタ中心の値を群全体での平均と比べると、動画ページから動画ページへの遷移が全体的に少ない。動画ページからひとつ前の動画ページへ戻る遷移

表 2: 各クラスタにおけるそれぞれの遷移回数の平均値 (セット 1 - 遷移数 15 以上の群)

クラスタ番号	$C_{1.1}$	$C_{1.3}$	$C_{1.4}$	群全体
ordered+1_vv	13.7	8.8	3.6	6.3
ordered+1_vp	5.9	2.9	4.2	3.4
ordered+1_pv	3.6	1.8	2.1	2.1
ordered+1_pp	0	0	0	0
reverse-1_vv	3.5	1.3	0.3	1.4
reverse-1_vp	1.5	0.5	1.2	0.9
reverse-1_pv	2.2	0.6	2.6	1.4
reverse-1_pp	0	0	0	0
ordered+n_vv	1.8	1.0	0.5	0.6
ordered+n_vp	1.9	1.2	1.0	1.3
ordered+n_pv	0	0	0	0
ordered+n_pp	0.5	0.2	0.4	0.3
reverse-n_vv	3.9	2.0	0.8	1.3
reverse-n_vp	0	0	0	0
reverse-n_pv	1.6	1.0	0.8	1.1
reverse-n_pp	0.1	0.1	0.4	0.2
計	40.8	21.8	18.5	20.8

(reverse-1_vv) や、テストページからいくつか前の動画ページへ戻る遷移 (reverse-n_pv) などは、5 クラスタ中で最も少ない。このクラスタに属する受講者は、理解が不十分なまま学習を進めてしまい、結果として最終成績が落ち込んだ可能性がある。

クラスタ $C_{1.3}$ は、修了率がやや高いクラスタである。クラスタ中心の値を群全体での平均と比べると、動画ページからテストページへの遷移や、テストページから動画ページへの遷移は群全体での平均より少ないものの、一方で、動画ページからひとつ次の動画ページへ進む遷移の数 (ordered+1_vv) や、動画ページからいくつか次の動画ページへ進む遷移の数 (ordered+n_vv)、動画ページからいくつか前の動画ページへ戻る遷移の数 (reverse-n_vv) は群全体での平均よりも多い。

これらの結果から、

- (ordered+1_vv) が多いほど、最終的な修了率が高い
- (reverse-1_vv) が多いほど、最終的な修了率が高い

という 2 つの傾向がみられた。 $C_{1.1} \sim C_{1.5}$ の 5 クラスタにおける相関を調べたところ、ordered+1_vv と修了率とが $r = 0.956$ 、reverse-1_vv と修了率とが $r = 0.875$ と強い正の相関があった。また、 $C_{1.1}$ と $C_{1.4}$ では、ordered+1_vv、reverse-1_vv ともに有意な差があった (t 検定, $p < 0.01$)。

他のセットにおいても同様に、Both型クラスタに属した受講者のうちコンテンツ遷移が多い群について各受講者の遷移特徴ベクトルを計算し、階層的クラスタリングを行った。それぞれ5クラスタに分け、同じくordered+1_vvと修了率との相関、reverse-1_vvと修了率との相関を調べた。得られた相関係数の値を表3に示す。ordered+1_vvについては8セットすべてで、reverse-1_vvについては8セット中7セットで正の相関があった。第7セットではクラスタリングの結果、受講者1人のクラスタと受講者2人のクラスタが生じともにreverse-1_vvが小さかったために、強い負の相関となった。

以上の結果から、このコースでは「ひとつひとつ順番に進み」「きちんと戻って確認する」受講者は修了率が高い傾向がある、ということがわかった。

表 3: クラスタ中心の値と修了率の相関係数の値

セット番号	1	2	3	4
ordered+1_vv	0.956	0.625	0.948	0.465
reverse-1_vv	0.875	0.642	0.919	0.888
セット番号	5	6	7	8
ordered+1_vv	0.428	0.996	0.532	0.383
reverse-1_vv	0.456	0.693	-0.791	0.578

6 おわりに

本研究では、受講者のコンテンツ遷移に着目し、学習履歴データに類型としてあらわれる受講者の学習様態を獲得することを目指した。個々のコンテンツ遷移にラベル付けを行い、ラベルごとの生起頻度によって受講者の遷移特徴ベクトルを定義し、この遷移特徴ベクトルの集合をクラスタリングすることによって、学習様態を得る手法を示した。そして、修了率と学習様態の関係について分析を行い、特徴となる学習履歴について考察を行った。

分析ではまず、閲覧数によるフィルタリングを行った。この過程において、すべてのセットで共通して、ほとんどアクセスしていない受講者、講義動画へだけアクセスしている受講者、テストへだけアクセスしている受講者、講義動画・テストの双方にアクセスしている受講者、が多かったことをあらわすクラスタが得られた。この結果は、荒ら [6][7] による先行研究で示唆されていた4類型と対応するものである。

本稿ではさらに、より詳細な学習様態を獲得するため、それぞれのセットにおいて講義動画・テストの双方にアクセスしている受講者を対象とし、提案手法による分析を行った。その結果、このコースでは「ひとつ

ひとつ順番に進み」「きちんと戻って確認する」受講者は修了率が高い傾向がある、ということがわかった。

本稿ではコンテンツ遷移が多い受講者の群のみに絞って分析を行ったが、他の群については未検討である。また、他のMOOCコースではどのような学習様態が得られるのか、および、得られた学習様態を用いて実際にどのように受講者を支援するのかは、今後の課題である。

謝辞

本研究にはedXプラットフォームのログデータを利用しました。データの提供に御協力くださいました、京都大学の土佐尚子教授、飯吉透教授、上杉志成教授に謝意を表します。

参考文献

- [1] 青木久美子: 学習スタイルの概念と理論—欧米の研究から学ぶ, メディア教育研究, Vol. 2, No. 1, pp. 197–212 (2005).
- [2] 合田美子, 山田政寛: 海外のリメディアル教育におけるeラーニングの研究動向と適用・応用される学習理論, リメディアル教育研究, Vol. 7, No. 2, pp. 205–215 (2012).
- [3] 大山牧子, 村上正行, 田口真奈, 松下佳代: e-Learning 語学教材を用いた学習行為の分析: 学習スタイルに着目して, 日本教育工学会論文誌, Vol. 34, No. 2, pp. 105–114 (2010).
- [4] 孫勝国, 甘泉瑞応, Tongjun, H., AIGUO, H., 程子学: 学習者の学習順序や反応パターンに基づいた学習状態推論法を用いる Web-based 教育支援システム, 情報処理学会論文誌, Vol. 46, No. 2, pp. 327–336 (2005).
- [5] 植野真臣: eラーニングにおける所要時間データの異常値オンライン検出, 電子情報通信学会論文誌, Vol. J90-D, No. 1, pp. 40–51 (2007).
- [6] 荒優, 藤本徹, 一色裕里, 山内祐平: MOOC 実証実験の結果と分析: 東京大学の2013年の取り組みから, 東京大学大学院情報学環紀要, Vol. 86, pp. 83–100 (2014).
- [7] 荒優, 藤本徹, 一色裕里, 山内祐平: MOOC 受講者の多様性を考慮した教育効果分析観点の提案, 日本教育工学会第30回全国大会, pp. 747–748 (2014).