

# MOOC プラットフォーム edX 上で獲得される 学習データの可視化基盤の構築

森村 吉貴<sup>\*1</sup>, 横山 慎<sup>\*1</sup>, Isanka Wijerathne<sup>\*1</sup>, 岡本 雅子<sup>\*1</sup>, 酒井 博之<sup>\*1</sup>  
<sup>\*1</sup> 京都大学

## Construction of Visualization Framework of Learning Data provided by edX MOOC Platform

Yoshitaka Morimura<sup>\*1</sup>, Makoto Yokoyama<sup>\*1</sup>, Isanka Wijerathne<sup>\*1</sup>,  
Masako Okamoto<sup>\*1</sup>, Hiroyuki Sakai<sup>\*1</sup>  
<sup>\*1</sup> Kyoto University

京都大学は 2013 年より MOOC の提供プラットフォームである edX に加盟し 2014 年から講義を提供している。edX のプラットフォームでは講義提供者は評価に利用する成績情報と利用しないログ情報が取得できるため、これらの情報を統合することで学習の分析に用いることができると考えられるがその方法は確立されていない。筆者らはそのような分析を行うための前準備として学習者毎の成績情報とログ情報の関係性を可視化するプログラムを Python により開発し、その機能を Jupyter notebook から対話的に呼び出す環境を構築することで、柔軟かつ再利用可能な形でデータを可視化する基盤を構築した。本稿ではその方法と、実際の可視化結果について報告する。

キーワード: ラーニングアナリティクス, 学習履歴分析, MOOC

### 1. はじめに

2012 年に Coursera や edX といった主要プロバイダの設立をきっかけとして、MOOC (Massive Open Online Courses : 大規模オープンオンライン講義) と呼ばれる新たな教育・学習形態が世界的な広がりを見せた。MOOC は大学レベルのオンライン講義をインターネットを通じて原則無償で提供するもので、一定の成績を収めた受講者には修了証が発行される場合もある。MOOC は登録すれば誰でも受講できるため、世界中から数千から数万名の受講者を集める講義も存在する。それ以降も、フランスやスペイン、中国など、地域 MOOC と呼ばれる英語圏以外の言語圏での MOOC プラットフォームが次々に開設されてきた。日本においても 2013 年 10 月に JMOOC (日本オープンオンライン教育推進協議会) が設立され、2014 年 4 月に運用

を開始している。

京都大学は MOOC の主要プロバイダの 1 つである edX に 2013 年に加盟し、2014 年 4 月に最初の講義「The Chemistry of Life」を開講した。それ以降、これまでに 15 講義を制作し edX を通じて世界に向け配信してきた。2022 年 6 月現在、全世界から延べ約 28 万名の学習者がこれらの講義を受講してきた。講義のリニューアルや複数講義を統合した例外を除き、現在 11 講義を毎年継続的に配信しており、「The Chemistry of Life」は 2022 年度で 9 度目の配信となる。

edX は、講義受講のために学習者がアクセスする LMS と呼ばれる環境と、講義コンテンツのオーサリング環境である CMS に加え、各講義の受講者属性やアクセスログを用いて教員・スタッフ向けに受講者分布や教材への取り組み状況などを可視化する Insights

と呼ばれるダッシュボード環境を提供している。

Insights で得られる情報から開講中の講義における学習者の属性や受講状況を把握したり、講義終了後に教員に対するフィードバックの目的で利用したりすることができる。しかし、Insights では、再開講した講義と過去の講義の受講情報を比較できなかつたり、動画ごとの視聴状況や週毎の教材へのアクセス状況といった全受講者を対象とするマクロレベルのデータは提供されているが、個々の受講者に対する時系列上の学習活動をミクロレベルで状況把握することができないといった課題がある。

## 2. MOOC の分析に関する周辺状況

多数の学習者の活動データを含む MOOC の実証的研究においては、ラーニング・アナリティクス (LA) と呼ばれる手法を利用した学習データの分析が国内外でおこなわれている。LA は、Society for Learning Analytics Research (SoLAR) により「学習とその環境を理解・最適化を目的とした学習者とその文脈についてのデータの測定、収集、分析、報告」と定義されており<sup>(1)</sup>、大量の学習ログデータを収集できる MOOC において有効な手法となることが期待されている。

藤本ら<sup>(2)</sup>は、LA を利用した MOOC 関連の国内外の実証研究をレビューの上、「MOOC 学習者の行動と終了率の研究」「MOOC の学習継続支援についての研究」「MOOC の受講者間の相互学習についての研究」に分類して整理しており、学習ログデータとの親和性が高い受講者の学習行動の把握や修了率の改善に関する研究の関心が高いことを指摘している。

大量の学習ログデータから意味ある情報を生成するためには、表やグラフのような形で可視化することが重要となり、このための方法として、必要な情報にアクセスするためのダッシュボードの提供が有用である<sup>(3)</sup>。

MOOC の学習ログデータを対象としたダッシュボード環境の開発に関する研究がある。例えば、香港科技大学では、Coursera の講義の教員と連携し、講義ビデオへのアクセス状況や時間軸に沿った各教材へのアクセス状況などを提示する機能を備えた VisMOOC と

よばれるダッシュボードを開発している<sup>(4)</sup>。また、スペインのマドリッド自治大学 (UAM) では、MOOC のデータを利用し、SPOC (Small Private Online Courses) を対象とした Open edX 用の教員向けダッシュボードを開発している<sup>(5)</sup>。京都大学では、再開講した同名講義間の比較を容易にした制作スタッフ・教員向けのダッシュボードを開発しフィードバックや年次報告のために試行している<sup>(6)</sup>。学習ログデータの分析手法は未だ発展中の段階にあり、利用するデータの決定は探索的に行う必要があることから、このようなダッシュボードは利用目的に応じて提示する項目を柔軟にデザインする必要がある。そこで、そのような教員向けのダッシュボードのデザインに向けて、学習データを柔軟に可視化するための基盤を構築することを考える。このような基盤があれば、有用な項目や典型的なパターンを絞り込むことができ、ダッシュボードのデザインのみならず授業設計全般に有用であると言える。次章以降では、そのような基盤の構築を試みた上で、実際の学習データを適用した例を提示していく。

## 3. データ可視化プログラムの実装

筆者らは前章で述べたような可視化基盤の実現のために、学習者毎の成績情報とログ情報の関係性を可視化する Python プログラムを開発し、それを効果的に実行するための環境を構築した。

本プログラムでは、edX のシステムからコースの管理者が取得できるデータとして以下のものを利用した。

- 学習者毎の各課題の達成状況(0~1 に正規化されたもの)とそれから合成される評点(0~1 に正規化されたもの)
- 学習者毎の id とユーザー名
- 学習者毎の匿名化 id と id
- 時系列順の操作イベントと学習者名の関連
- 動画情報 (時間長や URL など)
- 時系列順の学習ログイベント (Web ページの移動や動画の再生操作、個別の問題の回答状況など) の詳細

これらは仕様として edX の公式サイト上に掲載されている<sup>(7)</sup>。

作成したプログラムは、これらの学習データを元に、個別の受講生に関するデータを抽出し、分析の点から必要と考えられる観点に再整理した上で、データベースの形にした上で出力する。作成されるデータベースは以下のようなものになる。

- 学習者毎の評点
- 学習者毎の課題点
- 学習者毎の学習イベントログ数
- 学習者毎の個別の問題の正答状況
- 学習者毎のタブ (Discussion や Wiki などの学習ツール) へのアクセス状況
- 学習者毎の動画の操作イベント(再生、一時的、停止等)

更に、そのデータベースを元に各学習者と課題の達成状況の関係性や、成績との学習イベントログ数の関係性を示し、コース管理者が全学生の受講結果を俯瞰的に把握するための可視化図を作成する。

- 学習者毎の評点と各課題点のヒートマップ
- 各学習者の成績と学習イベントログ数の対応の散布図

このようなデータベースと、そのデータベースを元にした可視化図を作成する Python プログラムについて次節で説明する。

### 3.1 データ分析プログラム

データ分析プログラムとして、Python を使ったスクリプトを作成している。利用パッケージとして、データ分析には NumPy, SciPy, pandas を利用した。また、可視化処理には matplotlib, seaborn, Plotly を利用した。

本プログラムでは基本的に処理したデータは pandas のデータフレームの形で格納されている。pandas のデータフレームはデータ分析の分野で普及率の高い枠組みであるため、これを利用することは後述の Jupyter notebook を利用した分析環境と相性が良く、処理結果や中間的に生成されたデータを柔軟に再利用することが容易であるという利点がある。また、他の環境で処理結果を利用するためにデータフレーム内容の csv 形式のファイルとして出力も行っている。

### 3.2 データ分析環境

教員やコースの管理者が単純に結果を確認したい場合と、分析結果を元にデータを柔軟に再処理して確認を行いたい場合の両方をスムーズに実施するため、Jupyter notebook を利用する。Jupyter notebook の利用により可視化に利用したデータフレームがそのまま notebook に残るため、そのデータフレームを再利用することで、対話的な分析が可能である。また、対話的な利用を促進するため、Jupyter notebook 用の ipywidgets(Jupyter Widgets)パッケージを導入し、分析に必要なデータを notebook 上で GUI で指定する機能を作成した。実際に作成した notebook の画面例を図 1 に示す。

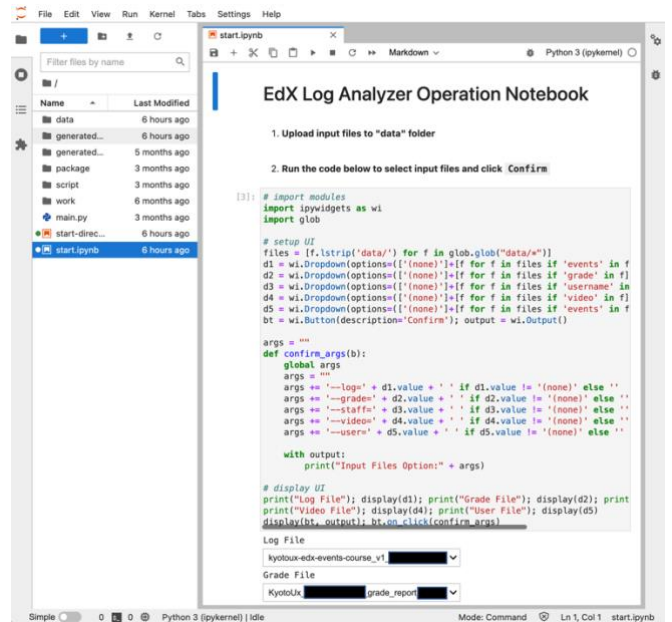


図 1 分析環境で利用する Jupyter notebook の画面例

このような notebook を作成することにより、コース管理者はプログラムの内容自体に触れずに notebook 上のコードを順番に実行するだけで分析結果と可視化図を得ることができる。一方で、このような環境ではコードやデータの透明性が確保されているため、コース分析者が自分自身で再分析を行いたい場合は notebook 上に存在するデータフレームに対してコードを追加したり、notebook が呼び出す python コードを修正することで柔軟に自分の望む結果を手に入れることができる。

また、教員やコースの管理者が必要となる Python のモジュールや Jupyter notebook の環境を 1 から構築することのないよう、コンテナ型の仮想環境である Docker イメージを作成し、手元の環境が Windows か Mac かの環境を問わず、必要な時にすぐ利用できるようにした。

ここで提案する可視化基盤の構成とデータ処理の流れについて、図 2 に示す。

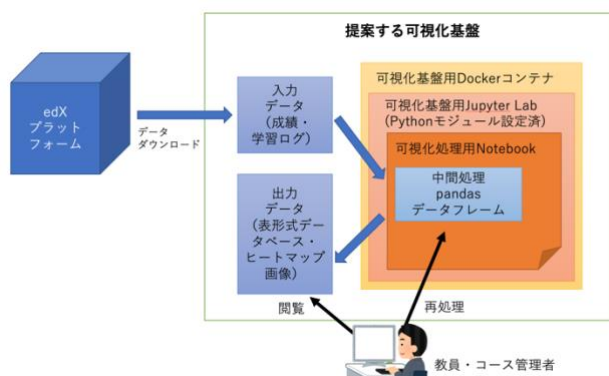


図 2 提案する可視化基盤の構成とデータ処理の流れ

#### 4. 可視化の実例と考察

2021 年度に実施されているコース 2 つについて、分析結果の実例を報告する。ここでは、2 つのコースのうち 1 つめのコースをコース A とし、2 つめのコースをコース B とする。

コース A の概要は以下のようになっている。

- 講義内容：化学系
- 登録者数：600 名程度
- 合格者数：25 名
- 講義の長さ：6 週

コース B の概要は以下のようになっている。

- 講義内容：化学系
- 登録者数：2500 名程度
- 合格者数：92 名
- 講義の長さ：12 週

それぞれのコースのデータは 2021 年 12 月 23 日時点で取得したものである。なお、両コースはデータ取得時点でも開講中であった。

#### 4.1 コース A の結果

コース A について、提案するデータ分析プログラムを適用した結果の、可視化図について以下に示す。

図 3 はコース A の学習者毎の評点と各課題点のヒートマップである。縦軸には学習者 ID であり、元の図では具体的な学習者名が入るが本稿では匿名化のためマスクしている。なお、分析結果においては、登録だけして具体的な行動を行っていない学習者 ID は省略されている。

図 3 では下に行くほど良い評点を持つよう学習者がソートされており、上に行くほど、つまり評点下がるほど右側が暗くなっており、基本的に全体の評点の良し悪しは、コースを通じてどの課題まで提出することができたかに大きく左右されることが示唆される。最後の方に良い課題点を取ったが前半の課題点は悪いといった学習者は稀であり、コースを通じた学習は積み上げ式となっており、途中で理解できなくなるとその後の理解も難しくなることが推測される。

次に、図 4 はコース A の各学習者の評点と学習イベントログ数の対応の散布図である。図から明らかであるように、評点と学習イベントログ数は一定の相関を持ち、MOOC を全く操作せずにより評点を得ている学生というのは稀であり、MOOC による学習以前に知識を持っていて課題だけを行い良い点を得るといった行動パターンはほぼないことが示唆される。

なお、この散布図は特定の 1 点、すなわち特定の学習者を表すデータにマウスオーバーすると、その学習者 ID と縦軸・横軸の具体的な数値が表示される機能を持っている。

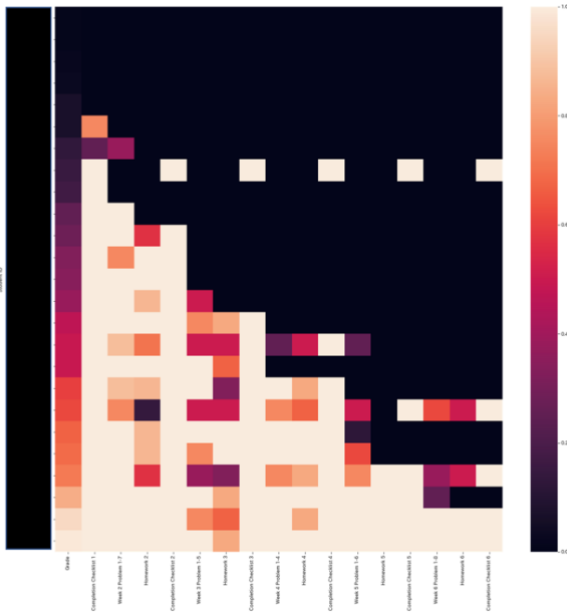


図 3 コース A のヒートマップ  
(縦軸：学習者 ID, 横軸：課題名,  
カラーチャート：課題点)

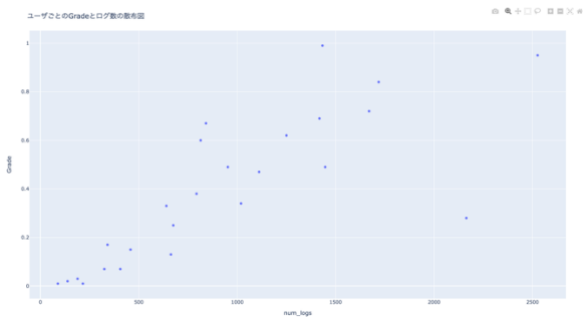


図 4 コース A の散布図  
(縦軸：評点, 横軸：学習イベントログ数)

#### 4.2 コース B の結果

コース B について、提案するデータ分析プログラムを適用した結果の、可視化図について以下に示す。

図 5 はコース B の学習者毎の評点と各課題点のヒートマップである。コース A と比べ学習者数が大きいいため縦方向の目盛りが圧縮されているが、全体の傾向としてはコース A と同様であることが見て取れる。

また、図 6 は各学習者の評点と学習イベントログ数の対応の散布図であるが、これも点の数は多いものの全体的な傾向はコース A と同様であることが見て取れる。

#### 4.3 考察

コース A の図 3 とコース図 5 のヒートマップについて、特に図 5 において「最初の課題だけ解いて、その後すぐに脱落するグループ」と、「ある程度課題を解いているが、ある課題（難易度が高いと思われる）の後に脱落しているグループ」が見られ、コースを通じた受講者の離脱状況の全体的な傾向が可視化により把握できるようになっている。

これらのヒートマップを見ると、特定の課題でヒートマップが暗くなっている列があり、そのような課題については、事後的に課題の難易度や内容について改善の余地がないかを検討するポイントになると考えられる。

また、散布図について、コース B の図 6 は全体的な傾向はコース A の図 4 と同様であるものの、右側に全体の比例関係から少し外れた点の一群があり、MOOC の操作に時間を費やしながらも相対的によい評点が得られなかった学習者群が存在することが示唆されている。

これらのように可視化により把握できる情報を通じて、学習者の行動や成果をより詳細に分析することで、MOOC 教材の改善を行う上でのヒントが得られる可能性を提供できると考えられる。

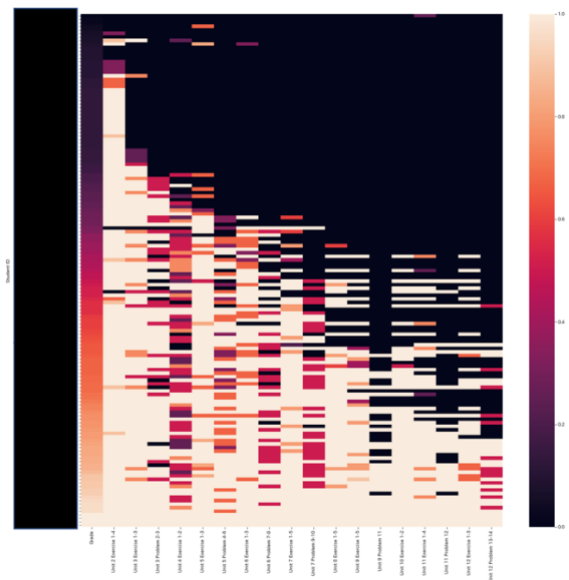


図 5 コース B のヒートマップ  
(縦軸：学習者 ID, 横軸：課題名,  
カラーチャート：課題点)

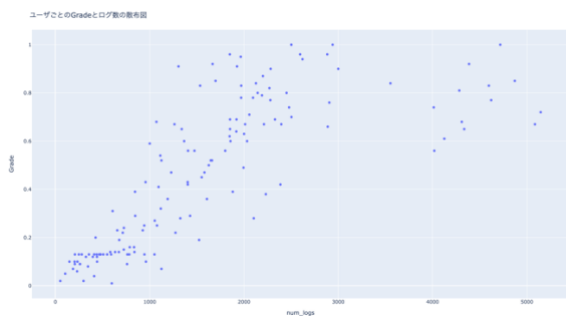


図 6 コース B の散布図  
(縦軸：評点，横軸：学習イベントログ数)

## 5. おわりに

MOOC の提供プラットフォームである edX において、講義提供者は評価に利用する成績情報と利用しないログ情報が取得できるため、これらの情報を統合することで学習の分析に用いることを目的として、学習者毎の成績情報とログ情報の関係性を可視化するプログラムを Python により開発した。また、その機能を Jupyter notebook から対話的に呼び出す環境を構築することで、柔軟かつ再利用可能な形でデータを可視化する基盤を構築し、実際に edX 上で開講された 2 つのコースに適用した出力結果について、可視化図としてヒートマップと散布図の実例を示し考察を行った。

今後の展開として、筆者らが同時に開発している Web システム型の MOOC データ管理ポータル<sup>(8)</sup>との統合や、可視化基盤を用いたより応用的なコース分析の提示が考えられる。

## 謝辞

本研究は科研費 (19K02972) の助成を受けたものである。

## 参考文献

- (1) Ferguson, R.: “Learning analytics: drivers, developments and challenges”, *International Journal of Technology Enhanced Learning*, 4(5/6) pp. 304–317 (2012)
- (2) 藤本徹・荒優・山内祐平: “大規模公開オンライン講座 (MOOC) におけるラーニング・アナリティクス研究の動向”, *日本教育工学会論文誌*, 41(3), 305-313 (2017)
- (3) E. Duval, E.: “Attention please! Learning analytics for visualization and recommendation”, in *Proceedings of LAK11*, pp.9-17 (2011)
- (4) Shi, C., Fu, S., Chen, Q. and Qu, Hu.: “VisMOOC: Visualizing Video Clickstream Data from Massive Open Online Courses”, *IEEE Pacific Visualization Symposium*, Hangzhou, China, 159-166 (2015)
- (5) Cobos, R., Gil, S., Lareo, A. and Vargas, F.A. *OpenDLAs: “An Open Dashboard for Learning Analytics”*, *L@S 2016: Proceedings of the Third ACM Conference on Learning@Scale*, pp.265-258 (2016)
- (6) Wijerathne, I. Okamoto, M. and Sakai, H.: “Centralized Lightweight MOOC Management System”, *Open edX Conference 2019*, San Diego (2019)
- (7) edX Inc.: “Using the edX Data Package - EdX Research Guide”, <https://edx.readthedocs.io/projects/devdata/en/latest/index.html> (2022 年 6 月 13 日確認)
- (8) Wijerathne, I. Okamoto, M., Morimura, M and Sakai, H.: “Development of MOOC Data Management Portal for Instructors and Production Team”, 第 21 回情報科学技術フォーラム (FIT2022) 講演論文集, *to appear* (2022)