

# 授業への取り組みを可視化するためのインタラクティブマイニングとヒートマップ生成

土橋 喜<sup>1</sup>

**概要:** 学習管理システム(Moodle)にアップロードした教材を活用して授業を行い、教材のクリックストリームと小テストの結果を収集し、インタラクティブなマイニングにより授業への取り組み状況を可視化するためのヒートマップを生成する試みを行った。教材のクリックストリームと小テストの得点データから偏差を求めて、履修者を以下の4つのグループに区分する散布図とヒートマップを作成した。その結果、(1)クリックストリームが高く得点も高いグループ、(2)クリックストリームは低いが得点は高いグループ、(3)クリックストリームが低く得点も低いグループ、(4)クリックストリームは高いが得点が低いグループに分けることができる。平均点と偏差を用いて散布図の基礎になるデータ作成しているため、授業回数の増加に伴い多くの履修者はクリックストリームと小テスト得点によって4つのグループを移動するが、一部の履修者は一定のグループに留まる傾向も見られた。これらのグループのうちグループ3とグループ4の下位部分には、学習につまずいている履修者が含まれていると思われる。本稿では小テストを13回実施した結果を集計してヒートマップを作成し、学習につまずいている履修者を発見できる可能性があることを示した。またこれらの情報は履修者への指導に役立てることが可能である。

**キーワード:** ヒートマップ, 学習分析, データマイニング, クリックストリーム, 小テスト得点, 学習のつまずき, Moodle

## Interactive Mining and Heat Map Generation to Visualize Engagement in Classes

KONOMU DOBASHI<sup>†1</sup>

**Abstract:** In this study, course materials, mainly in the form of PDF files, were uploaded to the Moodle learning management system and face-to-face lessons were conducted. Student clickstreams of the course materials and the results of quizzes were collected. The quizzes, which used a 5-alternative format, were prepared from the course materials. Data from the course material clickstream and the results of the quizzes were analyzed, and these were conducted from the viewpoint of educational data mining and learning analytics. A scatter chart and heat map was created to classify students into the following four groups based on deviations from the average value of clickstream and quiz scores: (1) those with high clickstream and high quiz scores, (2) those with low clickstream but high quiz scores, (3) those with low clickstream and low quiz scores, and (4) those with high clickstream but low quiz scores. However, students with poor quiz scores, students who fail to read the course materials, and students who are unable to understand even after reading the course materials tend to stay in the same group, which is not the path of a triumphant learner. Especially in groups 3 and 4, the lower part of the group appears to include students who have learning difficulties. It is shown that an analysis of the results of 13 quizzes was collected to create a heat map and student clickstreams can lead to the identification of disengaged students and signal the need for corrective intervention.

**Keywords:** heat map, learning analytics, data mining, clickstream, quiz score, unsuccessful student, Moodle

### 1. はじめに

教育データマイニングは以前から行われているアクションリサーチにおけるデータ収集と関係が深い。教育におけるアクションリサーチは、現職教師が自己成長を目指して自分で行った授業を振り返るためにデータを集めて行う調査研究である[1]。そのために教師は授業を分析したり、教育効果を確認したりするために、自らデータ収集を計画して実施する。そして収集したデータの分析結果を観察し、その結果に基づいて自分の教え方の内省を行うのである。授業中の履修者の学習行動の分析は、出席状況や小テストなどさまざまな項目を考えることができるが、教師が教えることと学習者が学ぶことに関するものであればアクション

リサーチの対象になる。また教員の教え方や授業の進め方についての調査が基本であるから、アクションリサーチの結果は授業評価に強く関連すると考えられるため極めて重要である。また多くの大学などで行っている授業評価アンケートなども組織全体で行うアクションリサーチの一種と考えられ、学生の反応を調べるために有効な方法の一つと言える。

一方近年におけるコース管理システム(CMS)や学習管理システム(LMS)あるいはe-bookと呼ばれるシステムの普及が急速に進んでいる[2]。これらのシステムにはアンケート調査機能が備わっているため、教員は手軽に従来のアクションリサーチを行い、分析に必要なデータを即座に入手することが可能である[3]。さらにこのようなアンケート機能に加えて、利用者の様々なデータが記録されるようになっ

<sup>1</sup> 愛知大学現代中国学部(情報科目担当)  
Faculty of Modern Chinese Studies, Aichi University

ている。例えば、Moodle では教材にアクセスを開始した日時、授業でログインしたユーザ名、どのネットワークから接続したかという情報、閲覧した教材名、アクセスしたデバイスの IP アドレスなどがシステムに記録され、システム全体の管理者や各コースの担当教員が常時閲覧できるように蓄積されている[4]。

これらのデータは CMS/LMS に教材をアップロードしたり、出席を確認したり、フォーラムに投稿したりしてシステムを活用すれば、自動的に収集されるようになっているのが一般的である。またこれらのデータ収集は毎日 24 時間行われており、担当教員が事前に利用者に知らせない限り、一般の利用者はログデータが収集されていることを意識することはほとんどないのである。これは事前に学生に周知して行うアクションリサーチとは大きく異なる点であり、有効な活用方法の開発が期待できる。また CMS/LMS には小テスト機能が備わっているので、小テストを行えばそれぞれの問題に解答した日時や、正解と不正解などの情報のほかに、テスト結果についての様々な分析結果を得ることが可能である。

加えて対面型の多人数の授業においては、履修者の机の間を巡回して行う机間指導は重要である。机間巡回によって履修者たちの学習の状態を常に観察し、必要なときに適切に指導する必要がある。CSM/LMS を活用して授業を行う利点の一つには、対面型授業の机間巡回による観察では得られない詳細なログデータの収集と蓄積である。授業改善に必要な履修者についてのデータを収集する場合に、個人別のデータとクラス全体が把握できるデータおよび教材の利用状況のデータ収集が必須である。また各履修者の教材閲覧が授業の進度に合わせて適切に行われていれば、授業に真剣に取り組んでいると判断され、授業の進度に合わせた小テストや学期末テストの結果によって、教材の内容を理解しているかどうかの判断が可能になる。

本稿においては、教材閲覧のクリックストリームと小テスト結果から成績不振につながる異常値を検出し、学習につまずいている学生を発見する手法の提案を行うため、クリックストリームと小テスト結果のデータから偏差を求め、履修者の特徴から 4 つのグループに区分する散布図を作成し考察を行った。その結果、(1)クリックストリームが高く小テスト得点も高いグループ、(2)クリックストリームは低いが小テスト得点は高いグループ、(3)クリックストリームが低く小テスト得点も低いグループ、(4)クリックストリームは高いが小テスト得点が低いグループに分けることができる。これらのグループのうちグループ 3 と 4 の下位部分には、学習につまずいている履修者が含まれていると思われる、分析結果から容易に見出すことができる。本稿では小テストを 13 回実施した結果を分析し、学習のつまずきを発見できる可能性があることを示し、これらの分析結果は履修者への指導に役立てることが可能であることを述べる。

## 2. 関連研究

本稿における CMS/LMS に蓄積されたログデータの分析は、WWW におけるアクセスログの収集と分析に強く関係している[10]。現在では様々なシステムごとに様々なトランザクション形式で蓄積されるデータが増大しており、これらのデータに対してデータマイニングの技術を応用すると、それまでは気づくことがなかった新たな知見を発見する可能性を秘めている。

最近では CMS/LMS や電子ブックなどを活用してより教育効果を高める研究や、大量に蓄積される学習履歴の分析を支援するためのラーニングダッシュボードと呼ばれるシステム開発も盛んになっている[6]。ラーニングダッシュボードでは、教材の閲覧回数や閲覧時間あるいは小テストにおける結果を分析するなど、学習履歴から得られた学生の特徴を把握し、授業の運営や改善に役立てるために多くの研究が行われている[7]。さらに CMS/LMS などに蓄積した学習履歴データをどのように活用するかが重要な課題であると認識されるようになり、学習分析の観点からの研究も盛んになった[8]。また BookLooper は電子ブックを閲覧するためのシステムであるが、大学などではデジタル教材を閲覧するための教科書配信システムとしても使われている。このシステムは利用者が教材を開いた時刻や閉じた時刻の記録のほかに様々な画面操作を記録できる機能を備えており、学習分析に必要なデータを得ることができる。学生の学習時間やアクセス回数、マーカー数、アノテーション数などを集計する機能も実装されており、学生の学習履歴の詳細な分析を行うことが可能である[9]。

また Mazza らが開発した GISMO は Moodle のプラグインシステムとして公開されており、このシステムは講義型授業を対象に Moodle の教材や小テストへのアクセス履歴を時系列に分析することができる。表形式のヒートマップやヒストグラムを使い、クラス全体と個人別の分析結果を可視化する機能を備えている[10]。このように対面授業を対象にした CMS/LMS においては、教員と学生との相互作用に必要な支援機能として、どのようにして学生の学習状況のデータを取得し、いかにして効果的に学習状況を教員へフィードバックするかが要求される。

教育分野においても外れ値や異常値を検出する研究が行われており、Pytlarz らは潜在的に学習が危険な状態にある学生を早期に見出す研究を行っている。キャンパスネットワークのトラフィック量に基づいて、出席や授業への取りくみ、授業外での学習時間などから GPA を予測し、学生の授業への取り組みを予測する方法を提案した[11]。また Gitinabard らは学生の教材へのアクセスとフォーラムのログを活用してドロップアウトの予測を行った。これによって潜在的に学習にいきづまっている学習者を早期に特

定し、学習者への介入や指導に活用できることを示した [12].

さらに大学内に構築された学生管理や成績管理などのシステムと CNS/LMS のデータを統合して学生の学習支援を行うシステムの研究が行われている。Purdue 大学の Course Signals では、CNS/LMS の一つである Blackboard に蓄積された学生の成績や過去の学習履歴などから、成績不振者を見出すシステムを開発し、学生の学習行動を判定して色付きの信号で示し、メールを使ってリアルタイムに学生にフィードバックすることができる [13]。Krumm らは CNS/LMS に蓄積されたデータを活用して、学部生に学習アドバイスを早期に行うことを支援するシステムの開発と応用研究を行っている [14]。また Michigan 大学の E2Coach と呼ばれるシステムも、コースの成績データに基づいて学生にメッセージを送信するものである。これらのメッセージは学生に必要な行動を行わせるための動機付けを行うものであり、例えば試験の準備のために十分な時間を確保するように指示を出す [15]。

また Dawson は CNS/LMS で収集したデータを教員に通知し、オンラインで学生の取り組み状況を可視化できるようにし、いつでも教師が学生の学習指導などに活用できるようにした [16]。これらの研究開発のように、学習管理システムに蓄積された情報から、どのようなデータを取り出しどのようにして学生に公開するか、現在では極めて重要な研究課題になっている。

### 3. 履修者データ

#### 3.1 オンライン授業と異常値検出

授業における異常値検出の目的は、異常な行動を行う学習者をできる限り早期に検出し、授業から脱落することを防止することである。本稿では Moodle とオンライン教材を使ったブレンド型授業として、パソコン実習室で実施する一斉授業を対象にしている。具体的には授業中を例にすれば次のような状態で異常値が発生することを想定している。(1)教師が指示した教材を開かない、(2)教材を開いても読まないあるいは内容が理解できない。これらの状態が継続するとテストの得点が低くなりやすく、このような状態は学習につまずいている異常な行動とみなされ、異常値の発生につながりやすい。上記のうち(1)の状態については、Moodle に蓄積された教材閲覧履歴から検出が可能であるためクリックストリームを調べる。(2)については小テスト得点とクリックストリームを組み合わせで判断することで対応が可能である。また CNS/LMS に蓄積されるオンライン学習ログは観測変数であり、学習ログそのものをデータ分析のための特徴量として扱うことが可能である。

これまでに MOOCs のように多数の学習者の学習ログが蓄積されたとき、授業のつまずきの原因となる異常値や、重要な知識やデータを発見するためのデータマイニング手

法は数多く提案されている。代表的なデータマイニング手法には、相関ルール、決定木、サポートベクターマシン、ベイジアンネットワーク、クラスタリング、外れ値、異常値検出、情報論的アプローチ、テキストマイニングなどがある。ここでは教材の閲覧回数と小テスト得点を利用して、学習につまずいている学習者を、ヒートマップを作成して可視化する手法により見出すことを提案する。

#### 3.2 オンライン教材と Moodle コースログ

教育分野は広大なため対象となる学生や授業は多種多様である。本稿で分析のために学習ログを収集する対象は、筆者が大学で担当している「社会データ分析入門」という授業である。履修者は 4 年制大学でワードやエクセルなどの基礎を学び、パソコンの操作が自力でできるコンピュータリテラシーの能力を備えた学生である。1 年生から 4 年生まで受講可能になっており、履修者の男女の割合は男 62.7%、女 37.3%で、大部分の履修者の年齢範囲は 18 歳から 22 歳までである (2017 年秋)。授業の内容はエクセルを活用した統計学の入門である。実際の授業ではエクセルの操作方法の学習から始め、代表値、分散、標準偏差、シミュレーション、度数分布とピボットテーブル、属性相関、共分散、相関分析、回帰分析などを 15 週に分けて学習する。

教材は主に PDF ファイルで作成しており、12 章の構成、演習課題を含めて 112 節 (ファイル)、外部 URL は 10 あり、教材全体では B5 サイズの PDF ファイルで 154 ページである。Moodle には 112 のファイルに分割して、章と節に見出しを付け、Moodle のトピックモードでアップロードした。トピックモードでアップロードすると、入力したトピックの名称が Moodle のログに記録される。本稿においてはトピックの名称は、教材の章の名称を使っているので、学生の開いた教材名がログに表示される。

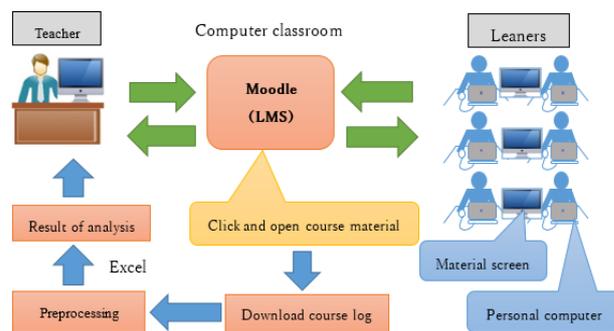


図 1 ログの収集と分析の流れ

Figure 1 Moodle log collection and analysis overview.

学生は授業中に教師の解説を聞きながら指示に従って教材を開き、エクセルによるデータ処理について学習する。授業の前半は Moodle 上の教材を閲覧しながら、同時にエクセルの画面を開き、履修者全員でパソコンの操作しながら

ら学習を行う。授業の後半は各自で演習課題を行い、履修者にはエクセルで行った演習課題の結果をまとめたファイルの提出を求めた。履修者は2人に1台ずつ教材を提示するモニターが用意されており、担当教員がモニターに表示する教材とパソコン操作のデモを見ることができる。また履修者は授業内外を問わずパソコンで Moodle の画面から教材を開き、各自で自由に閲覧したりダウンロードしたりすることも可能である。(図1)

### 3.3 クリックストリームの集計

ここでは教材の閲覧回数を集計するための前処理について述べる。なお本稿におけるクリックストリームは Moodle 上の教材を開いた回数のみであり、パソコンを操作するための全てのクリックストリームではない。また本稿において小テストの出題範囲に該当する教材閲覧履歴は、前回の小テスト終了時刻から、次回の小テスト終了時刻までの間に行われたすべての教材のクリックストリームである。

現在の Moodle に蓄積されているログデータは、時系列のテキストデータの一覧形式で提供されるだけであり、履修者の状態を把握するには十分ではないことが指摘され、有効に活用するには追加的な工夫が必要になっている。これまでも Moodle のコースログをエクセル形式や CSV 形式などでダウンロードして活用するエクセルマクロが開発されている。Dierenfeld らはエクセルの関数やピボットテーブル機能を使って、履修者ごとの教材の閲覧回数などについて、エクセルのピボットテーブルを活用した分析方法を提案している[17]。また Konstantinidis は Moodle のアクセスログの閲覧履歴を分析し、曜日ごとの時系列のアクセスグラフや履修者ごとのアクセスを一覧するグラフを作成するエクセルマクロを提案している[18]。

また一般的に CSM/LMS のログデータは、毎年蓄積されるためビッグデータ化する傾向がある。そのため役に立つ知見や傾向を発見するためのプロセスマイニング手法の研究が行われている。Moodle の閲覧履歴では、教師や履修者がいつどのコース教材を開いたかを記録しており、ログの項目とデータが一覧表の形式で時系列にまとめられており、ダウンロードして活用できる。筆者のクリックストリームの集計についての詳細は別稿にまとめたので、以下では概要を述べる[19]。

### 3.4 前処理とピボットテーブル

離散データを集計し、二次元の度数クロス表を作成するツールとしてエクセルに備わっているピボットテーブルを活用することができる。エクセルのピボットテーブルを使えば、複数の離散データに対して同時に度数分布を集計し、二次元のクロス表を生成することができる。またピボットテーブルを作成する機能は Python のライブラリにも備わっている。しかしここでエクセルのピボットテーブルを使

うのは、ピボットテーブル作成後もエクセルの関数やピボットテーブルのユーザインターフェースがデータ分析にそのまま使えるためであり、様々な処理を容易に追加できる[20]。

Moodle コースログのように集計できる項目が多数あるときは、複数の項目を選択してクロス集計する場合に、ピボットテーブルを有効に活用できる。ピボットテーブルには生成後のテーブルを操作するために、特定のデータの絞り込みを行うフィルタなど様々な機能が備わっている。これらの機能はテーブル生成後の分析に活用することができるため、授業の場合はクラスや教材の全体的な視点や個別の視点からの分析が容易に行える。

また Moodle のコースログに記録された分単位や秒単位の時刻データを使ってそのまま集計しようとすると、巨大なクロス表が生成されることが多く、パソコンの画面では結果を観察するときに困難が伴う。そのため上述したように、筆者はピボットテーブルで分析するために、Moodle のコースログに記載された時刻のデータに対して前処理を行う場合に、月、日、時、分などのように時間を表すカテゴリを使い、必要に応じて一定の時間間隔で時刻データを離散化し、複数の時系列データを生成している。これにより時系列分析を行う場合に、1分毎、1時間ごと、日ごと、週ごと、半年ごと、1年ごとの分析などのように柔軟に選択し、必要に応じて処理データの絞り込みを行うことができる。また多面的な分析を行うために、時間カテゴリを複数用意することにより、ピボットテーブル上に複数のタイムラインを表示させ、フィルタ機能として使うことが可能になる。例えば15分と1分のタイムランを組み合わせると、一部の15分間だけを1分毎に表示するような使い方が可能になる。さらにピボットテーブルのフィルタを使い、複数の時間カテゴリを選択すれば、様々な時間帯における集計と分析が可能になり、より詳細で多面的な分析を行う可能性が高まる。

## 4. クリックストリームと小テスト結果による散布図

授業のなかで学習につまずいている履修者を見出すことは、教師にとって極めて重要である。多人数の履修者を集めて行う授業では、個別に学習につまずいている履修者を見出すことは難しい。しかし教師としては授業の内容がうまく理解できないなどが明らかな履修者には早急に対応して指導する必要がある。そのため学習のつまずきによってドロップアウトする前に、潜在的につまずいている状態にある履修者を早期に見出す支援は極めて重要である。ここでは Moodle 上のクリックストリームと小テスト結果から、散布図とヒートマップを活用して授業につまずいている履修者を見出す支援について述べる。

#### 4.1 小テストの計画と実施

本稿の授業で実施した小テストは Moodle 上の教材を利用して各章ごとに作成した。主に文章の穴埋めを完成する問題に解答する形式にし、すべて 5 肢択一方式で出題した。作成した小テストの問題数は、各章ごとの平均が 12 問であり、全体で 146 問である。またこれらの問題は Moodle によって履修者ごとに選択肢の並べ替えが行われる。また学生が結果を確認してしばらくすると小テストは非公開となる。

本稿で対象とする授業は毎週一回実施しており、前週の授業内容の理解度を確認するため、各章の学習が終了したあと、翌週に小テストを実施した。授業開始時に Moodle の小テスト機能を使って、5 分間で 5 問に回答する小テストを、第 3 週から第 15 週まで 12 回行った。毎週の小テストでは各章ごとに 12 問から 5 問ランダムに出題した。

また学期末最後の授業では 30 問出題して学期末テストを実施した。授業内容全体の理解度を把握するため、学期末テストは小テスト 12 回分を連結して実施し、146 問から 30 問をランダムに出題した。そのため小テストで出題された問題が、履修者によっては学期末テストでも同じ状態で出題されることがある。

#### 4.2 学習者を 4 つのグループに区分

本稿では授業教材は担当教員が授業で解説し、履修者は Moodle から自由に教材を閲覧できるようになっており、授業期間中は 24 時間公開されている。ここではクリックストリームによって授業への取り組み具合を調べ、小テスト得点によって授業内容の理解度を把握することを試みている。また教材を閲覧して適切に学習した履修者のクリックストリームと、小テスト・学期末テスト得点の間には、相関関係が成り立つことが前提条件と考える。これらの間に正の相関が認められるとき、履修者は教材を閲覧し授業内容を適切に理解したと想定される。逆に負の相関があるときは、履修者の教材閲覧が不適切である場合や、授業内容の理解が不足している場合、あるいは教材の内容自体が不適切な場合などの原因が想定される。

ここではクリックストリームと小テストとの間に相関関係があることを前提として、クリックストリームと小テスト得点の偏差を求めて散布図を作成し、その結果から履修者を 4 つのグループに分け、潜在的に学習につまずいている履修者を見出す手法を提案する。そのため偏差を  $D_i$ 、観測値を  $x_i$ 、平均を  $m$  とし、次の式で偏差を計算する。

$$D_i = x_i - m$$

#### 4.3 散布図の作成

2017 年秋学期第 3 週に小テストを行った結果を例にして、教材のクリックストリームと小テスト得点から偏差を求め

て散布図を作成した例を示す (図 2)。クリックストリームと小テストのそれぞれの偏差から散布図を作成し、以下のように履修者を 4 つのグループに区分することができる。

- (1)グループ QI:教材クリック数が平均値より多く小テスト得点も高い。
- (2)グループ QII:教材クリック数は平均値より少ないが小テスト得点は高い。
- (3)グループ QIII:教材クリック数は平均値より少なく小テスト得点も低い。
- (4)グループ QIV:教材クリック数は平均値より多いが小テスト得点は低い。

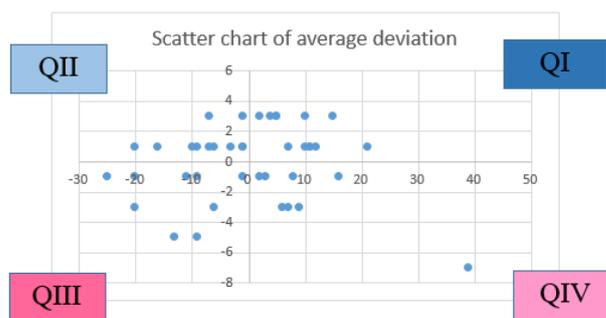


図 2 散布図の作成例 (第 3 週, 2017 年 10 月 4 日)

Figure 2 Example of creating a scatter chart for the third week (October 4, 2017).

また表 1 の作業用ファイルにおいて、A 列は匿名化した履修者名、B 列は小テスト範囲に該当する教材のクリックストリーム、C 列は小テスト得点、D 列はクリックストリームの偏差、E 列は小テスト得点の偏差、F 列は散布図上の該当するグループとセルの色をそれぞれ示しており、ヒートマップの色は手作業で設定した。

### 5. 結果と考察

#### 5.1 ヒートマップの作成

グループ QI は青色、グループ QII はライトブルー、グループ QIII は赤色、グループ QIV はピンク色で塗りつぶしている。散布図は D 列のクリックストリームの偏差、および E 列の小テスト得点の偏差を使って作成する。小テストの平均点と偏差を使って散布図を作成するため、常に 4 つのグループにデータが出限するが、グループ間に差が出る場合もある。ちなみに 2017 年度秋学期のグループ間の最大値はグループ QI において 0.439 であり、最小値はグループ QIV において 0.049 であった。また散布図のグループ QIV に外れ値が出現しているが、表 1 の Student20 であることが分かる。さらにグループ別に次のような特徴があると想定される。

グループ QI は、クリックストリームと小テスト得点が両方ともに平均値より高い傾向があり、教材を開いて読んでおり、教材閲覧が小テスト得点に影響を与えていると予想されるグループである。常にこのグループを維持している履修者は成績が良好で授業にしっかり取り組んでいると

表 1 偏差の計算結果とヒートマップの例 (第3週, 2017年10月4日, 小テスト受験者のみ記載)

Table 1 Calculation result of deviation and example of heat map (Week 3, October 4, 2017, Quiz candidates).

A	B	C	D	E	F
	log duration	quiz day			
	9/20/17				
	10/4/17	10/4/17			
	clickstream	quiz score	devi.click	devi.score	quadrant
Student01	51	4	-6.2	-2.95	QIII
Student02	65	6	7.8	-0.95	QIV
Student03	60	6	2.8	-0.95	QIV
Student04	47	8	-10.2	1.05	QII
Student05	61	10	3.8	3.05	QI
Student06	32	6	-25.2	-0.95	QIII
Student07	62	10	4.8	3.05	QI
Student08	37	6	-20.2	-0.95	QIII
Student09	56	8	-1.2	1.05	QII
Student10	37	8	-20.2	1.05	QII
Student11	56	6	-1.2	-0.95	QIII
Student12	66	4	8.8	-2.95	QIV
Student13	48	8	-9.2	1.05	QII
Student14	50	10	-7.2	3.05	QII
Student15	63	4	5.8	-2.95	QIV
Student16	46	6	-11.2	-0.95	QIII
Student17	64	4	6.8	-2.95	QIV
Student18	59	10	1.8	3.05	QI
Student19	54	8	-3.2	1.05	QII
Student20	96	0	38.8	-6.95	QIV
Student21	44	2	-13.2	-4.95	QIII
Student22	56	10	-1.2	3.05	QII
Student23	72	10	14.8	3.05	QI
Student24	37	4	-20.2	-2.95	QIII
Student25	41	8	-16.2	1.05	QII
Student26	67	10	9.8	3.05	QI
Student27	48	6	-9.2	-0.95	QIII
Student28	48	2	-9.2	-4.95	QIII
Student29	73	6	15.8	-0.95	QIV
Student30	59	6	1.8	-0.95	QIV
Student31	78	8	20.8	1.05	QI
Student32	68	8	10.8	1.05	QI
Student33	69	8	11.8	1.05	QI
Student34	62	10	4.8	3.05	QI
Student35	56	8	-1.2	1.05	QII
Student36	67	8	9.8	1.05	QI
Student37	51	8	-6.2	1.05	QII
Student38	64	8	6.8	1.05	QI
Student39	68	8	10.8	1.05	QI
Student40	50	8	-7.2	1.05	QII
AVERAGE	57.200	6.950			
STDEV.S	12.696	2.480			
Data	40	40			

見なせる。

グループ QII は、クリックストリームは低めであるが、小テスト得点は平均値より高いグループである。このグループに属する履修者は、教材を読まなくても小テスト得点が高い傾向がある。そのため以前に教材内容と同じようなコンテンツを学習した経験があることが予想される。

グループ QIII は、クリックストリームが平均値より比較的少なく、小テスト得点も比較的低いグループである。そのため授業内容に関心が持てない傾向があると想定され、このグループの中に学習につまずいている履修者が含まれている可能性があると考えられる。

グループ QIV は、クリックストリームは平均より多いが、小テスト得点が平均より低いグループである。このグループには、教材を開いただけで読んでいない履修者や、あるいは読んでも理解できない履修者などが該当し、小テスト得点が低い傾向があり、グループ QIII と同じように授業に関心が少ない履修者が含まれていると予想される。

表 2 は履修者の授業への取り組みを可視化するため、2017年秋学期に行った授業について、第3週から第15週まで12回分の授業中のヒートマップを連結したものである。最右端は学期全体のクリックストリームと学期末テスト得点のヒートマップである。空白のセルは欠席を示している。授業に登録だけ行い、教材閲覧履歴がない履修者は削除した。

また履修者のクリックストリームと小テスト得点の平均値は毎回異なるので、すべての履修者は2つ以上のグループに出現しており、グループを移動していることが分かる。また表 2 では赤 (QIII) の出現回数を使い昇順にソートした結果を示しており、表 2 の上部には青色(QI)が多い履修者が存在し(Student44)、成績が優秀な履修者であることを示している。逆に表 2 の中ごろや下部には赤色(QIII)が多い履修者が混在していることが分かり(Student01 など)、これらの履修者は平均以下の成績の可能性が高い。

グループ QII の出現頻度で表 2 のヒートマップを昇順にソートすると、クリックストリームは低めであるが、小テスト得点は平均値より高いグループをヒートマップの下部に集めることができる。グループ QII の最多の該当回数は10回になっており、当該履修者の学期末テストの得点は22点と平均以上の成績であった。このような履修者はやはり以前に本稿の教材内容と同じような内容を学習した経験がありそうである。

## 5.2 学習につまずいている履修者

小テスト問題は五指択一式のため、どのグループも20%の確率で小テスト得点が上下することがある。そのため特にX軸の近辺に該当するデータは隣接グループに移動する可能性がある。またクリックストリームと小テスト得点の相関係数が高いほど、履修者全体が授業に集中していると考えられる。クリックストリームと小テスト得点の散布図を見れば、グループ QI により多くのデータが集まることが望まれ、QIII や QIV に該当するデータは少ないほうがよいと言える。クリックストリームだけでは授業内容を履修者が理解しているかどうかの判断は難しいが、本稿においても学期末テストと教材全体のクリックストリームとは相関関係(0.495,  $p < 0.05$ )があり、正の相関関係を示していた。第3週に小テスト第1回を行い、クリックストリームと小テスト得点からヒートマップを作成した。表 2 は QIII の出現回数で昇順にソート後のヒートマップであり、おおよそ次のことが分かる。

表 2 QIII の出現頻度で昇順にソートしたヒートマップ  
 (第1週と第2週は小テストなし, ブランクは欠席を示す。  
 2017年秋学期, 社会データ分析入門)

Table 2 Heat map, 2017 fall, Introduction to Social Data  
 Analysis. There is no quiz in the 1st and 2nd week. Blanks  
 indicate absence.

heat map	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Week	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	14th	15th	
Quiz	Chap.1	Chap.2	Chap.3	Chap.4	Chap.5	Chap.6	Chap.7	Chap.8	Chap.9	Chap.10	Chap.11	Chap.12	Final	
Student47	QI	QI	QIV	QI	QI	QI	QIV	QI	QIV	QI	QI	QI	QI	QI
Student14	QIV	QI	QI	QI	QI	QI	QI	QIV	QIV	QI	QI	QI	QIV	QI
Student39	QI	QI	QI	QI	QI									
Student44	QI	QI	QI	QII	QI	QI	QI	QI	QI	QIV	QI	QI	QI	QI
Student27	QII	QI	QI	QIV	QI	QI	QI	QI	QIV	QIV	QI	QI	QI	QI
Student48	QII	QII	QI	QIV	QI	QI	QI	QI	QI	QIV	QI	QI	QI	QI
Student25	QIV	QII	QII	QII	QIV	QI	QI	QI	QI	QI	QI	QI	QI	QI
Student20	QIV	QIII	QI	QI	QI	QI	QIV	QI	QI	QI	QI	QI	QI	QI
Student43	QII	QI	QIII	QI	QI	QIV	QI							
Student28			QIII	QIV	QI	QII	QI	QI	QI	QI	QI	QI	QIV	QI
Student41	QI	QI	QI	QIII	QII	QII	QI	QII	QI	QII	QII	QI	QII	QI
Student22	QII	QI	QI	QIII	QI	QI	QI	QII	QII	QII	QII	QI	QI	QI
Student32		QIV	QI	QI	QII	QII	QI	QIV	QII	QIII	QII	QII	QII	QII
Student16	QII	QIV	QII	QII	QII	QIII	QII	QII	QII	QII	QIV	QI	QI	QI
Student30	QIV	QII	QII	QI	QIV	QIII	QI	QI	QI	QI	QI	QI	QII	QI
Student21	QI	QI	QIV	QII	QI	QIII	QI	QI	QIII	QII	QI	QI	QIV	QI
Student46	QI	QIII	QIV	QII	QI	QI	QIV	QII	QI	QI	QIII	QII	QII	QI
Student02	QIV	QI	QII	QI	QI	QI	QII	QI	QI	QIV	QIII	QII	QII	QI
Student09		QIV	QIV	QIII						QII	QIII	QII	QI	QII
Student19	QIII	QII	QI	QI	QI	QII	QIII	QI	QI	QI	QIV	QIV	QIV	QI
Student06	QI	QIII	QIII	QII	QI	QII	QI	QI	QI	QII	QII	QI	QI	QI
Student13	QIII	QII	QI	QII	QI	QI	QI	QI	QI	QIII	QII	QII	QII	QII
Student45	QI	QI	QIV	QI	QI	QI	QI	QI	QI	QIII	QIII	QII	QII	QI
Student12		QI	QIV	QII		QII	QIII	QI	QI	QII	QI	QII	QIII	QI
Student29	QI	QIII	QI	QI	QII	QII	QI	QI	QI	QIII	QI	QII	QII	QI
Student10	QII		QII	QIII	QII	QII	QII	QII	QII	QIII	QII	QII	QII	QII
Student36	QIII	QII	QI	QIV	QIII	QII				QIV	QIV	QI	QI	QI
Student42	QI	QI	QIV	QIII	QII	QI	QI	QIV	QIII	QIII	QIV	QI	QI	QI
Student40	QI	QI	QII		QII	QIV		QII	QIII	QII	QIII	QIII	QII	QI
Student15	QII	QI	QIV	QI	QIV	QII	QII	QIII	QIII	QII	QII	QII	QII	QIII
Student26	QIII	QII	QIV	QIII	QIV	QIV	QI	QIV	QIV	QIII	QIV	QIV	QIV	QI
Student38	QIV	QI	QIV	QIV	QIII	QII	QII		QIII	QIV	QIII	QIV	QIII	QI
Student03	QIV	QI		QIV	QII	QIII	QI	QI	QI	QI	QI	QIII	QIII	QI
Student05	QII	QIII	QII		QIV	QII	QII	QI	QII	QIII	QIII	QII	QII	QII
Student37	QIV	QI	QIV	QI	QII	QIII	QIII	QII	QII	QII	QIII	QII	QII	QIII
Student18	QIV	QII	QI	QII	QIV	QIV	QII	QIII	QIII	QII	QIII	QII	QII	QI
Student17		QI	QIV	QIII	QIII	QIV			QIII	QIII	QIV	QI	QII	QII
Student31	QIII				QIII	QII	QIV		QIII	QIII	QI			QIII
Student34	QI	QIII	QIII	QII			QI	QIV	QIV	QIII	QI	QII	QIII	QII
Student23		QIII		QII	QIII	QIV	QIV	QIII		QI	QIII	QII	QII	QIII
Student11	QII	QI	QI	QIII	QII	QII	QI	QI	QI	QIII	QII	QIII	QII	QIII
Student04	QI			QIII	QIII	QII	QII	QII	QI	QII	QIII	QII	QII	QIII
Student07		QIV	QIII				QIII	QIII			QIII	QIII	QIII	QIII
Student33	QII		QIII	QIV	QIII	QIII	QIII	QII	QIII	QI	QI	QI	QIV	QI
Student24			QIII	QIV	QIII	QIII	QI	QI	QIII	QIII	QIII	QII	QIII	QII
Student08	QIII	QII	QIV	QIII	QIII	QIII	QIII	QIII	QIV	QIII	QIII	QII	QIII	QII
Student35	QIII	QIII	QII	QIV	QIII	QIII	QIII	QII	QII	QII	QIV	QIII	QIII	QII
Student01	QIII	QII	QIII	QIII	QIII		QIII	QIII	QIII	QIII	QIII	QIII	QIII	QIII

赤色(QIII)とピンク色(QIV)が継続して出現すると、平均点以下の状態が続くことになり、学習につまずいている傾向が高まる(Student01, 07, 08, 35 など)。

該当するグループに青色(QI)が一つもない場合も同様に学習につまずいている兆候であると判断される。例えば Student01 は第4週に水色(QII)に該当したが、第5週以降は赤色(QIII)が継続しており、学習につまずいているものと思われる。青色(QI)から他のグループに移動しても、復帰すれば良い学習状態が継続できる可能性がある(Student47)。しかし青色(QI)に復帰しないときは、良い学習状態とは言えないので観察や指導が必要になることがある(Student24, 26, 38, 40)。

また表2においてグループ QIV の出現頻度で昇順にソートすると、教材クリック数は平均値より多いが小テスト得点は低い履修者を、表2の下部に集めることができる。

最も多くこのグループに該当した履修者は7回であり、当該履修者の学期末テストの得点も低く 10 点であった(Student26)。

表 3 ヒートマップの該当グループと異常値

Table 3 Corresponding heat map group and outliers.

A	B	C	D	E	F	G	H	I	J	K	L
	pageview 9/20/2017	final test	Heat map Groups					Hotelling's T2 theory			
	2018/1/17	2018/1/17	QI	QII	QIII	QIV	abse	Click	Quiz	Click	Quiz
Student01	364	18	0	1	11	0	1	1.597	0.408	0.117	0.685
Student02	535	22	4	3	2	2	2	0.039	0.056	0.969	0.955
Student03	541	20	2	3	4	2	2	0.025	0.040	0.980	0.968
Student04	435	19	2	4	5	0	2	0.674	0.176	0.504	0.861
Student05	301	21	1	4	4	0	4	2.744	0.000	0.009	1.000
Student06	658	24	7	4	2	0	0	0.325	0.456	0.747	0.650
Student07	189	17	0	0	6	1	6	5.546	0.736	0.000	0.465
Student08	467	8	0	1	8	2	2	0.386	8.004	0.701	0.000
Student09	427	22	0	3	2	2	6	0.758	0.056	0.452	0.955
Student10	408	22	0	10	2	0	1	0.978	0.056	0.333	0.955
Student11	438	16	3	3	5	0	2	0.643	1.160	0.523	0.252
Student12	459	20	3	4	2	1	3	0.450	0.040	0.654	0.968
Student13	546	22	4	4	2	0	3	0.017	0.056	0.987	0.955
Student14	839	24	7	0	0	4	2	2.886	0.456	0.006	0.650
Student15	478	10	2	6	3	2	0	0.305	5.717	0.761	0.000
Student16	575	23	1	9	1	2	0	0.003	0.208	0.998	0.836
Student17	543	17	2	0	5	3	3	0.022	0.736	0.983	0.465
Student18	518	25	1	5	4	3	0	0.092	0.800	0.927	0.428
Student19	640	18	3	3	2	3	2	0.210	0.408	0.835	0.685
Student20	832	28	10	0	1	2	0	2.740	2.407	0.009	0.020
Student21	768	25	7	2	2	2	0	1.578	0.800	0.121	0.428
Student22	631	23	6	5	1	0	1	0.161	0.208	0.873	0.836
Student23	474	14	1	2	5	2	3	0.334	2.295	0.740	0.026
Student24	346	18	1	1	8	1	2	1.893	0.408	0.064	0.685
Student25	701	27	6	4	0	2	1	0.703	1.775	0.486	0.082
Student26	608	10	1	0	4	7	1	0.067	5.717	0.947	0.000
Student27	693	26	8	1	0	3	1	0.621	1.240	0.537	0.221
Student28	582	23	4	1	1	2	5	0.009	0.208	0.993	0.836
Student29	496	25	6	5	2	0	0	0.194	0.800	0.847	0.428
Student30	408	17	2	1	2	2	6	0.978	0.736	0.333	0.465
Student31	496	19	1	1	5	1	5	0.194	0.176	0.847	0.861
Student32	508	25	2	7	1	2	1	0.134	0.800	0.894	0.428
Student33	589	18	2	1	6	2	2	0.020	0.408	0.985	0.685
Student34	491	17	3	1	5	2	2	0.222	0.736	0.825	0.465
Student35	472	19	0	2	9	2	0	0.348	0.176	0.729	0.861
Student36	582	23	2	1	3	4	3	0.009	0.208	0.993	0.836
Student37	506	16	2	5	4	2	0	0.143	1.160	0.887	0.252
Student38	555	19	1	2	4	5	1	0.005	0.176	0.996	0.861
Student39	775	26	10	1	0	0	2	1.689	1.240	0.098	0.221
Student40	499	22	2	5	3	1	2	0.178	0.056	0.860	0.955
Student41	594	26	7	5	1	0	0	0.029	1.240	0.977	0.221
Student42	569	26	4	3	3	3	0	0.000	1.240	1.000	0.221
Student43	737	23	6	1	1	1	4	1.129	0.208	0.264	0.836
Student44	1034	26	11	1	0	1	0	0.497	1.240	0.000	0.221
Student45	538	24	5	4	2	1	1	0.032	0.456	0.975	0.650
Student46	639	25	6	3	2	2	0	0.204	0.800	0.839	0.428
Student47	948	22	10	0	0	3	0	0.658	0.056	0.000	0.955
Student48	765	24	8	2	0	3	0	1.531	0.456	0.132	0.650
Data(N)	48	48	48	48	48	48	48				
AVERAGE	566.6	20.9	4	3	3	1.8	1.75				
STDEV.S	160.3	4.6	3	2	2.5	1.4	1.76				
MAX	1034	28	11	10	11	7	6				
MIN	189	8	0	0	0	0	0				

### 5.3 異常値との比較

表3では履修者別に該当するグループ QI から QIV および欠席データを示し、さらにクリックストリームと小テストから異常値の検出を行った結果を示した。異常値の検出はホテリング理論に基づいており、異常値を a, 観測値を x<sub>j</sub>, 平均を m, 標準偏差を s として以下の式を用いた。

$$a(x_j) = \frac{(x_j - m)^2}{s^2}$$

表3からは、クリックストリームが少なく小テスト得点も

低いため、2人がホテリング理論でも異常値になっていることが判明した (Student07, 11)。また Student01 はクリックストリームが少ないため、Student08 は小テスト得点が低いため、Student09 と 10 はクリックストリームが少ないため、それぞれホテリング理論の異常値に該当した。

また一回だけ青(QI)に該当した履修者 8 人 (Student05, 16, 18, 23, 24, 26, 31, 38) について、ホテリング理論の異常値に該当するかどうかを調べたところ、Student16, 31, 38 の 3 人はホテリング理論の異常値に該当しなかったが、Student18 は学期末テストが高得点のため異常値になった。また Student05, 24 の 2 人はクリックストリームが少ないため異常値に該当することが明らかになり、Student23, 26 の 2 人は小テスト得点が低いため異常値になった。また学期末テストが 17 点以下の場合得点が低いためホテリング理論の異常値に該当したが、これらの履修者はグループ QI に 2~3 回該当していた (Student 11, 15, 17, 30, 34, 37)。

さらにホテリング理論では高得点の異常値も検出することから、本稿のヒートマップとの対応を調べたところ、クリックストリームと小テスト得点がともに高い値で異常値になった履修者 (Student20, 21, 25, 27, 39, 44) は 6 人であった。このほかに少なくともクリックストリームまたは小テスト得点のいずれか一方が異常値になる場合が 7 人であった (Student14, 29, 32, 41, 42, 43, 47, 48)。これらの異常値の傾向は青色(QI)の該当回数が多い履修者に見られた。

## 6. おわりに

本稿では、Moodle に蓄積した教材のクリックストリームと小テスト得点から、4 つのグループに分けて特徴を検討した。これらのうちグループ 3 と 4 の下位部分には、学習につまずいている履修者が含まれていると思われる、ソート後のヒートマップから容易に見出すことができる。特に赤色(QIII)とピンク色(QIV)が継続して出現する場合、担当教師は履修者に対して教材にしっかり目を通して授業に取り組むように指導する必要があると考える。またここで説明した分析の結果を学生に伝える方法や対応方法は、さらに注意を払う必要がある問題である。Moodle コースログから生成された散布図とヒートマップは、教師が教育へのアプローチを自ら見直して改善するために使用できる。

**謝辞** 本研究は JSPS 科研費 18K11588 の助成を受けたものです。また本稿の研究において討論とともに有益な助言いただいた愛知大学経済学部の杉浦裕晃教授に深く感謝いたします。

## 参考文献

[1] Taylor, M. Action research in workplace education. Quebec: Human Resources Development Canada. National Literacy Secretariat. 2002, 95p.  
 [2] Romero, C. and Ventura, S.. Data mining in education. Wiley

Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2013, vol. 3, no. 1, p. 12-27.  
 [3] Scheuren, F. June. What is a Survey? (2nd. ed.). Washington, American Statistical Association. 2004, 68p.  
 [4] “Moodle”. <https://moodle.org/>, (参照 2019-12-15).  
 [5] Sen, A., Dacin, P.A. and Pattichis, C.. Current trends in web data analysis. Commun. ACM. 2006, vol. 49, no. 11, p. 85-91.  
 [6] Verbert, K., Duval, E., Klerkx, J., Govaerts, S. and Santos, J.L.. Learning analytics dashboard applications. American Behavioral Scientist. 2013, vol. 57, no. 10, 1500-1509.  
 [7] Baker, R.S.J.D.. Data mining for education. International encyclopedia of education. 2010, vol. 7, no. 3, 112-118.  
 [8] Yassine, S., Kadry, S. and Sicilia, M.A.. A framework for learning analytics in moodle for assessing course outcomes. In 2016 IEEE Global Engineering Education Conference (EDUCON). IEEE. 2016, p. 261-266.  
 [9] “電子書籍配信サービス BookLooper”. <https://www.kccs.co.jp/ict/service/booklooper/>, (参照 2019-12-15).  
 [10] Mazza, R. and Milani, C.. Gismo: a graphical interactive student monitoring tool for course management systems. In International Conference on Technology Enhanced Learning. Milan, 2004, p. 1-8.  
 [11] Pytlarz, I., Pu, S., Patel, M. and Prabhu, R. What can we learn from college students’ network transactions? Constructing useful features for student success prediction. In Proceedings of the International Conference on Educational Data Mining (EDM) (11th, Raleigh, NC, Jul 16-20, 2018)  
 [12] Gitinabard, N., Khoshnevisan, F., Lynch, C.F. and Wang, E.Y.. Your Actions or Your Associates? Predicting Certification and Dropout in MOOCs with Behavioral and Social Features. In the 11th International Conference on Educational Data Mining (EDM 2018) (11th, Raleigh, NC, Jul 16-20, 2018).  
 [13] Arnold, K.E. and Pistilli, M.D.. April. Course signals at Purdue: Using learning analytics to increase student success. In Proceedings of the 2nd international conference on learning analytics and knowledge. 2012, p. 267-270.  
 [14] Krumm, A.E., Waddington, R.J., Teasley, S.D. and Lonn, S.. A learning management system-based early warning system for academic advising in undergraduate engineering. In Learning analytics, Springer, New York, NY.2014, p. 103-119.  
 [15] McKay, T., Miller, K. and Tritz, J.. What to do with actionable intelligence: E 2 Coach as an intervention engine. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. ACM. 2012, p. 88-91.  
 [16] Dawson, S.P., McWilliam, E. and Tan, J.P.L.. Teaching smarter: How mining ICT data can inform and improve learning and teaching practice. paper presented at The Ascilite 2008 (Melbourne, Australia, 30 November – 3 December, 2008).  
 [17] Dierenfeld, H. and Merceron, A.. Learning analytics with excel pivot tables. Proceedings of the 1st Moodle Research Conference (MRC2012). 2012, p. 115-121, Heraklion, Crete-Greece, Sept.2012.  
 [18] Konstantinidis, A. and Grafton, C.. Using Excel macros to analyse Moodle logs. Proceedings of the 2nd Moodle Research Conference (MRC2013). 2013, p. 33-39, Sousse, Tunisia, Oct.2013.  
 [19] Dobashi, K. Interactive Mining for Learning Analytics by Automated Generation of Pivot Table. In: Ahram T. (eds) Advances in Artificial Intelligence, Software and Systems Engineering. AHFE 2018. Advances in Intelligent Systems and Computing, Springer, Cham, vol. 787, 2018, p. 66-77.  
 [20] Jelen, B. and Alexander, M.. Excel 2013 pivot table data crunching. Pearson Education. 2013, p. 432.