

# マウスカーソルの挙動データを用いた 学習者の特徴分析のための可視化手法の検討

池田 瑞穂<sup>1,a)</sup>

概要：欧米で Learning Analytics が提唱され 10 年近く経過し、e-learning における学習スタイルや利用方法、品質に関する研究が多く行なわれている。これまで、Web 教材コンテンツの品質管理のためコンテンツを構造化し学習ログとして、機器や環境に依存しないマウスカーソルの挙動情報を自動取得できるシステムを開発した。そして、学習者の特徴をローレンツ曲線、ならびにジニ係数などを用いてさまざまな観点から可視化し授業進行中においてリアルタイムに参照できるように機能強化した。このシステムを実授業へ適用し蓄積した学習ログ情報を用いて、リアルタイムに授業遂行において効果的に活用できる可視化方法について検討を行う。

## Learning Feature Analysis Using Mouse Cursor Tracking

MIZUHO IKEDA<sup>1,a)</sup>

### 1. はじめに

Learning Analytics(以下 LA と呼ぶ)に関するカンファレンスの開催が始まり 10 年近く経過した。EDM(Educational Data Mining) や LAK(Learning Analytics and Knowledge Conference) など国際カンファレンスが開催され、LA に関する研究が国内外の様々な専門分野において盛んに行われている。国内においては、e-JAPAN として教育の情報化の推進のためのアクションプランが策定されて以降、初等教育から高等教育の各機関において、教育の情報化が急速に推進され、LA に対する関心が高まっている。2017 年 LAK の報告書 [1] では、“数理モデルによる学習行動の解析と予測”、“省察”、“感情情報の利用”、“学習解析に関する倫理”、“組織としての対応”などの観点での研究報告があったことを伝えている。教育の情報化基盤の整備が推進されていく中で、学習に関して科学的根拠に基づく研究や実践がより重要視されていくと考えられる。

一方、2016 年 LAK の報告書 [2] は、国内の LA 研究の問題点として、分析対象が狭いこと、分析データの深さや

粒度について指摘をしている。分析対象がクラス単位や授業単位に限定されていたり、分析対象のデータが LMS 上の履歴データやアクセスログまでにとどまっているため、学習の本質的なメカニズムや有効な学習プロセスを導くための十分条件を明確にすることができない。統計データから何らかの一般法則を得ようとすれば膨大なデータの収集と処理、すなわち教育のビックデータの構築が不可欠であると指摘している。また、[3] では、匿名化した学習データの共有による LA の取組の活性化、[4] では、学習データ項目や形式、および、インターフェイスの標準化の必要性について指摘している。LA を進めていくには、膨大なデータを単純に構築、収集するだけでなく、分析や活用の目的に適したデータの構築が必要になる。そのためには、データの収集、分析、分析結果の評価、および、データベースへのフィードバックのサイクルを確立することが必要であり、このサイクルを継続的、日常的に行える環境の構築が重要である。

これまで、さまざまな学習ログを収集できるシステムを構築し、実授業に適用しながら、リアルタイムにシステムの改良を行ったりコンテンツの構成や内容をバージョンアップしてきた。新たに、より詳細な学習ログ収集のため、マ

<sup>1</sup> 関西学院大学理工学部  
School of Science and Technology, Kwansai Gakuin University  
<sup>a)</sup> gpikeda@kwansai.ac.jp

ウソカーソルの挙動情報を自動収集し、従来より詳細なレベルでの分析とその結果の可視化を行うことができる機能を強化した。この機能は、機器など環境に依存することなく、従来のシステムに追加可能であり、データの自動収集から分析、結果の可視化まで一貫して継続的に行うことができるシステムである。そして収集したデータをリアルタイムに分析し、可視化表示を行うことができる。授業進行中などリアルタイムに利用できるため、授業進行の助けとすることが可能となっている。そこで、マウスの挙動データを含めた学習ログを用いた可視化した手法と、分析結果である学習者の特徴分析結果について述べる。

## 2. これまでの研究概要

### 2.1 Web 教材コンテンツの構造化

2004 年度より文科系学生に対する様々な情報科学教育の授業に携わり担当授業においては専用の Web 教材を構築し、e-learning システムを用いたブレンド型の授業を行っている。特に、Web 技術やコンピュータ言語である C 言語、Java に関する演習科目やデータ分析の授業において、Web 技術を用いた授業用テキストを作成し授業で利用してきた。そして、インストラクションデザインに基づいた Web 教材や授業の構成の改良を行ってきた [5]。例えば、操作性の改善としては、教材コンテンツの Web ページから新しいページを開く必要なくワンクリックでポップアップメニューを表示し、様々なデータの表示や授業進捗が入力可能なユーザインタフェースを実現した。また、授業中の学習者の状況や、学習者の理解度を教員のみならず学習者がリアルタイムに把握可能な機能を備えている。そして、学習者の授業に対する取り組みの改善を行ってきた。情報科学関連の教育は技術躍進や社会的概念のスピードが非常に速く変動するため、教員の技術力の向上だけでなく、Web 教材の見直しや改善を日々していく必要がある。このような状況のなかで、ブレンド型 e-learning システムとして高いパフォーマンスを得ることのできる Web 教材を実現するため、コンテンツ評価モデルが必要と考えられる。Web 教材の品質を 3つの階層（プレゼンテーション、ページ構成、ゴール構成）から構成されていると考え（図 1）、各層の構造化設計を行い、Web 教材を構築してきた。そして、学習ログとして、Web 教材システムのログイン、ログアウトの時間や各ページの閲覧開始時刻から終了時刻を自動で収集する機能を Web 教材システムに実装し、実際の授業に適用し、自動収集した学習ログを分析し、Web 教材の評価を行った [6]。

### 2.2 学習ログ収集の機能強化

Web 教材の品質は向上しているが、さらに学習者の学習理解度を上げ、より多くの知識を獲得してもらうため、より効果的な Web 教材を提供する必要がある。Web ペー

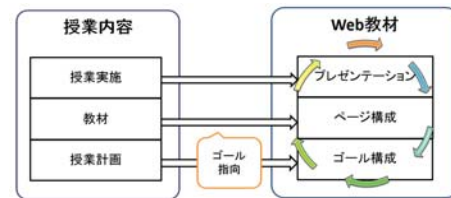


図 1 授業構成要素モデル

Fig. 1 Class component model.

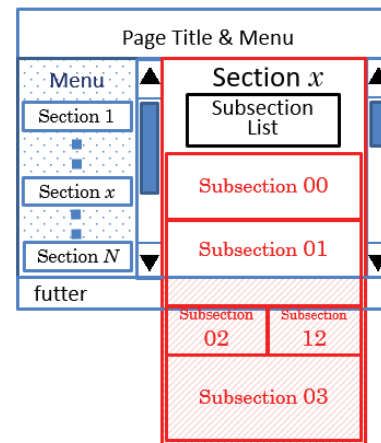


図 2 Web 教材の構成

Fig. 2 Composition of Web content used in actual classes.

ジ単位、つまりセクション単位での閲覧時間情報などの学習ログだけでなく、より詳細な学習者の挙動に関する情報を収集しようと試みた。

Web ページを閲覧するときの視線とマウスの動きの相関関係の研究 [8] において、Web ページを閲覧時に検索者は Web 画面上で興味のあるリンクを容易にクリックするためマウスの動きを目で追う傾向があると指摘している。これに基づき本研究では学習者のマウスの挙動を検出し、Web 教材のページ内の各サブセクションの挙動情報を収集できるように設計、実装し学習ログ収集の機能強化を図った。

機能強化のためのシステム要件を以下に整理する。

- 学習ログは、リアルタイムにサーバへ蓄積、参照できること
- 被験者は、授業に参加している学生全員とすること
- 学内共用システムのため他システムへの影響を及ぼさないようにすること
- 特別な機器、環境を要しないこと
- 教材コンテンツの修正、変更などに対して自動収集機能は影響を受けないこと

### 2.3 マウスの挙動データの収集の仕組み

マウスのトラックボールを用いて画面をスクロールする学習者が多いことから、マウスの位置は学習者が

閲覧している部分の周辺に存在すると考え、サブセクションの領域単位で収集することとした。そこで、大量のマウスカーソルの挙動を表す座標情報が存在するサブセクションの位置を検出することで、リアルタイムの分析とその可視化を可能とした。そして、継続的なデータの自動収集を実現できるように設計、実装を行い、学習者の特徴分析として、可視化を行い比較分析を行った。図 2 に Web 教材を分割した領域を示す。マウスカーソルがどの領域にあるかを収集し蓄積する。左のメニュー領域は目次であり、各コンテンツの Web ページ、すなわちセクションへのリンクがある。セクションは  $x$  方向、 $y$  方向それぞれ 10 個までのサブセクションで構成される。Web 教材は Subsection 00 から Subsection 30 までの 5 つのセクションで構成される例である。マウスカーソルの挙動データの収集はコンテンツ表示部分、すなわちセクション部分のみとしている。マウスカーソルが存在する Web ページ、すなわちセクションラベルと、そのマウスカーソルが入ったサブセクション  $x$  方向のラベル、 $y$  方向のラベルと滞留した時間 (秒) が記録される。

この機能強化したシステムを実授業に適用し学習ログの自動収集を行い、学習ログの粒度を細分化することによる学習者の特徴分析への一定の効果が得られた [7]。

### 3. 関連研究

LA は、大きく分類すると、(1) 学習データの収集、(2) 分析、(3) 分析結果の評価のプロセスから構成される。(1) の学習データの収集としては、LMS からの学習履歴データ、学習者の挙動を検出する観点からは生体情報 [9]、マウスカーソルの挙動 [10]、表情 [11] などのデジタルデータの検出、および、アンケート調査などの自己申告のデータを用いた研究が多く行われている。また、(2) の分析や (3) 分析結果と評価を行うために様々な可視化手法が提案されている。学習プロセスの可視化 [12] として、時間経過に従って、学習シナリオに沿った学習者の辿った経路を視覚化することにより、学習者に対して学習方法の見直しするための情報提供するシステムを考案し、教授者に対しては学習パターンの分析、および、コンテンツの見直しするための情報を提供している。プログラミング教育を対象とした編集履歴の可視化 [13] では、プログラミング作成過程を詳細に記録し編集履歴を可視化することにより典型的なつまづきパターンや思考過程の違いを解析している。また、概念マップを用いた研究 [14] では、Kit-Build Concept Map と重畳マップを用いて、学習者の理解度に関する研究を行っている。

LA 研究の多くは、被験者が限定されたり、特定の環境下での実験であり、実運用での適用や、継続的な評価を前提としている研究は少ない。授業遂行中にリアルタイムにデータ分析するには、システム全体のリソース環境設定の

変更による研究対象以外のシステムへの影響を考慮する必要がある。また、分析結果については、学習者個別の挙動に対する研究は多く行われているが、学習者全体あるいは、集団の挙動を可視化したものは少ない。分析結果として可視化されたものの比較、分析については、さらなる自動化を進めていく必要がある。

本研究では、継続的に Web 教材の品質評価を行うため、第 2.3 章で述べたマウスカーソルの挙動情報を取得する機能を用いた学習データの自動収集から分析までプロセスを実現するシステムを開発し、学習者の特徴分析を行った。このシステムは学習者個別の挙動、および、集団の挙動を可視化できるようにした。可視化されたものは、授業進行とともにモニタリングできるようにリアルタイムに処理をできるようにした。第 4 章では可視化の方針を明確にし、可視化システムの概要とその効果についての評価について述べる。第 5 章ではまとめと今後の課題について述べる。

## 4. 可視化手法による学習者の特徴分析

### 4.1 可視化の方針

可視化することによって、新たな発見を見出し、当初の目的が変遷していくものと考えられるが、可視化することによって、目的を明確にし、どのような条件下で分析を進めていくのか、多次元データをどのような手法でデータ整理したのかあるいは、次元縮約したのかを明確にする必要がある。本稿では、可視化したデータは、学習者、教授者、教授者の学習者サポートが利用することを目的とする。また、学習者個別の挙動、および、集団の挙動が把握できるように可視化を行う。

### 4.2 可視化システム

自動収集された Web 教材のページ内の各サブセクションの挙動情報を用いて学習者の特徴分析としての可視化システムを構築した。可視化システムの要件は以下のとおりである。

#### (1) 可視化の目的

教授者、教授者の学習者サポートが利用するものとし、学習者対象のものは今回対象としない。

#### (2) 可視化の対象

学習者の個別の挙動、集団の挙動が把握できること

#### (3) 開発環境

データ処理、データ分析、可視化が可能な言語でシステム開発ができ、教授者や学習者サポート者も Web ページを通じて利用できる環境であること

#### (4) 処理の自動化

データの自動収集から可視化までのプロセスは自動化し、継続的に運用できること

#### (5) リアルタイム性

可視化された結果は、リアルタイムに Web ページに表示

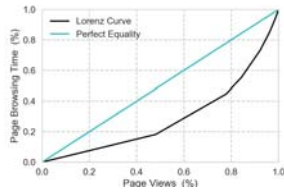


図 3 ローレンツ曲線 (1)Level2  
Fig. 3 Lorenz Curve (1)Level2.

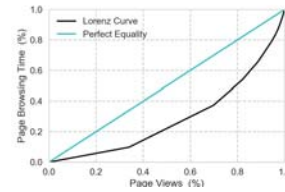


図 4 ローレンツ曲線 (2)Level5  
Fig. 4 Lorenz Curve (2)Level5.

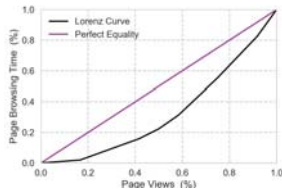


図 5 ローレンツ曲線 (1) 授業 1 回目  
Fig. 5 Lorenz Curve (1)Session 1.

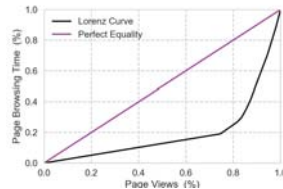


図 6 ローレンツ曲線 (2) 授業 4 回目  
Fig. 6 Lorenz Curve (2)Session 4.

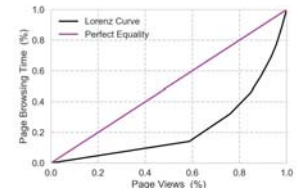


図 7 ローレンツ曲線 (3) 授業 8 回目  
Fig. 7 Lorenz Curve (3)Session 8.

できること

(1)の可視化の目的については、教師、および、学習者をサポートする視点で教材コンテンツの閲覧状況を把握することによって学習者の挙動の特徴を明確にする。(2)の可視化の対象として、集団の挙動が把握できるようにするため、閲覧分布の不均等性の分析に用いた統計量であるジニ係数を用いて時間遷移を可視化する。また、時間遷移において、利用する時間間隔を指定可能であり、時間幅による閲覧分布などの学習者の状況等の比較が行えるようにした。

(3)(4)(5)については、データ処理や分析だけでなくプログラミング言語としてアプリケーション作成も容易なオープンソースのPython3 (pandas, matplotlib ライブラリ等)、一部 R 言語を用いリアルタイム表示を実現した。

## 5. 学習者の特徴の可視化例と評価

### 5.1 授業へのシステム適用

文科系学部生 30 名が履修する共通科目である Web コンピューティングなどに学習ログ収集機能の強化を行ったシステムを適用した。この授業は WAMP 環境 (Windows, Apache, MySQL, PHP) にて PHP プログラミングと MySQL データベースシステムとの連携を学ぶものである。授業の最初に Web 教材の内容を教授した後、各学習者が演習を行うブレンド型授業である。この科目は、授業内容を飛ばして演習を行うことが難しい授業である。また、1 つの演習課題が比較的大きく、複数回に分けて説明する内容が多い。演習時は、教員と学生補佐が学習者を個別サポートする授業運営を行っている。

### 5.2 学習者の特徴の可視化

閲覧時間分布の不均等性をみるため、縦軸がサブセクションの閲覧回数の累積相対度数、横軸はサブセクションの閲覧時間の累積相対度数によるローレンツ曲線を用いて可視化を行った。まず、学習者のレベルによる閲覧分布の比較を行った。本科目は知識蓄積型の内容であるため、最終の評価は課題到達度となる。その到達度から学習者のレベルを 5 段階評価し、グループ分けした閲覧時間分布として、レベル 2 を図 3、レベル 5 を図 4 に示す。ここで、レベル 5 は課題達成度が一番高いレベルを示す。ジニ係数はレベル 2 で 0.42、レベル 5 で 0.43 となりあまり差異は見られなかった。次に、履修科目である Web コンピューティングの 2017 年春学期 (全 14 回授業) において、授業 1 回目、4 回目、8 回目の全学習者の教材コンテンツの閲覧時間分布を図 5、図 6、図 7 に示す。ジニ係数は、授業 1 回目で 0.34、授業 4 回目で 0.60、授業 8 回目では 0.55 となった。授業 1 回目に比べて、授業が進むごとにジニ係数が増加傾向にある。これは、閲覧分布の不均等性が高くなっていることを示している。授業の後半に向けて、1 つのサブセクションを閲覧している可能性が高くなっていることを示している。

さらに、ジニ係数の時系列遷移を図 8 に示す。上述と同様に 2017 年春学期を対象に、横軸を授業回、縦軸をジニ係数でグラフ化した。ジニ係数が低くなると、閲覧分布の均等性が高くなり、すなわち、様々なサブセクションを一樣に閲覧していることを示している。ジニ係数が高くなると、閲覧分布の不均等性が高くなり、すなわち、1 つのサブセクションを集中的に閲覧していることを示している。レベル 5 の学習者は、他の学習者と比較して、ジニ係数が高く、あまり変動していない。特に、授業回の前半 (1 回から 6 回) 後半から最終回 (7 回から 14 回) は、ジニ係

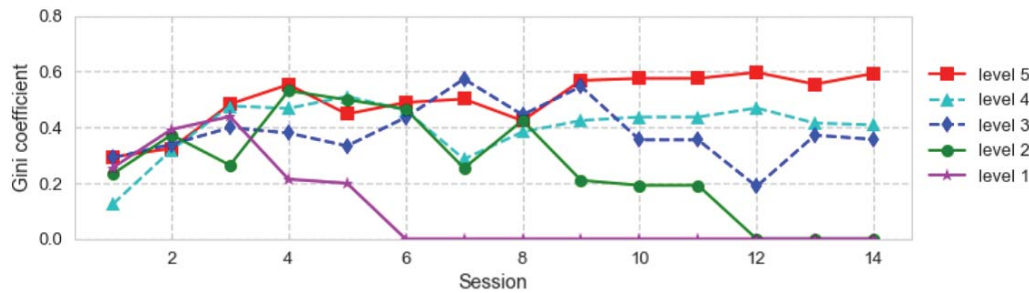


図 8 学生のレベル別各授業回のジニ係数の推移  
Fig. 8 Transition of the Gini coefficient of each session (each level of the students).

数は高い位置で定常状態を保っている。授業回の前半においては、教授者のサブセクションの閲覧遷移を追従しながら、他のサブセクションを閲覧することにより、閲覧分布は若干不均等性の変動が見られるが、授業回の後半になるに従って、1つのサブセクションに集中して閲覧していることがわかる。レベル2の学習者は、授業回の前半(1回から8回)までは、レベル5の学習者と同様に教授者のサブセクションの閲覧遷移に追従しているが、9回以降の授業回からは、均等性が非常に高くなり、教授者が閲覧しているサブセクション以外に、過去のサブセクションなど、様々なサブセクションを閲覧していることがわかる。

また、ジニ係数の時系列遷移図は、リアルタイムに表示することが可能なため、教授者は、授業を進めていくなかで、閲覧分布の不均等性を確認し、授業の進行のスピードを調整したり、場合によっては、過去サブセクションに戻って講義することが可能となる。さらに、教材コンテンツの内容精査、および、サブセクションの再配置を考えていく上で有用な情報になると考える。

## 6. まとめと今後の課題

マウスカーソルの挙動を自動収集するシステムを設計、開発し、実授業に適用し学習ログの収集し、蓄積した。さらに学習ログを用いて分析を行い、授業の進行に合わせてリアルタイムに可視化できるシステムを構築した。この学習ログを用いて、分析、および、結果を可視化することにより、Web ページ単位であるセクション単位のデータ分析と比較して、細分化されたサブセクション単位での学習ログのほうが学習者の挙動を明確にとらえることが可能であることがわかった。サブセクション単位の学習ログは、2015年秋学期から学期ごとにデータを蓄積している。さらに学習者の特徴分析を行い、学習プロセス、および、学習メカニズムへの分析を勧めていくため、この学習ログを用いて、数理モデルへの適用、および、統計量を用いた検定を行い、教材コンテンツの品質評価モデルの確立を行っていく予定である。

謝辞 本研究は JSPS 科研費 JP26330412 の助成を受け

たものである。

## 参考文献

- [1] 古川雅子, 中村泰之, 山川修他: ラーニング・アナリティクスの動向-第7回 Learning Analytics & Knowledge Conference(LAK'17), 情報教育シンポジウム, pp.31-36, 2017/08.
- [2] 安武公一, 中村泰之, 山地一禎, 古川雅子, 梶田将司, 山川修, 多川孝央:最先端の Learning Analytics 研究を目指して, 教育システム情報学会, 第41回全国大会, B5-1.pp.345-346, 2016.
- [3] 山川修:組織を超えた Learning Analytics の可能性:コンピュータ&エデュケーション, Vol38, pp.55-61, 2015.
- [4] 平田謙次: Learning Analytics による学習品質 - e-learning での教育・学習活動データの連携と技術標準化-, コンピュータ&エデュケーション, Vol38, pp.43-48, 2015.
- [5] 池田瑞穂:インストラクショナルデザインに基づく学習活動ログによる Web 教材コンテンツの利用形態分析と実装, 情報科学研究 (24), pp.17-24, 2010.
- [6] 池田瑞穂:Web 教材システムに基づいた教材コンテンツ評価モデルの作成と実装, 日本教育工学会研究報告集, 2012(4), pp.49-54, 2012.
- [7] 池田瑞穂:Web 教材の品質管理のための学習ログ収集の機能強化と学習者の特徴分析, 信学技報, 2018.
- [8] Paul B.Kantor, Endre Boros, Benjamin Melamed, Vladimir Menkov, Bracha Shapira, David J. Neu: Capturing HUMAN INTELLIGENCE Intelligence in the Net, ACM 08/2000, Vol.43, No.8, pp.112-115, 2000.
- [9] 矢島邦昭, 小川信之, 高附祐輔, 野村収作, 福村好美: e-learning 受講中の生体情報を用いたコンテンツの質の評価に関する実験, 情報処理学会研究報告, Vol.2012-CLE-7, No.6, 2012.
- [10] 杉田賢治, 福原知宏, 増田英孝, 山田剛一: Web ブラウザのスクロール操作に基づくユーザ注目箇所推定に関する分析, 人工知能学会, 26, pp.1-4, 2012.
- [11] 中村和晃, 角所考, 美濃導彦: e-learning 環境における学習者の観測に基づく主観的難易度の推定, 画像の認識・理解シンポジウム (MIRU2007), 2007/07.
- [12] Fionan Peter Williams, Owen Conlan: Visualizing Narrative Structures and Learning Style Information in Personalized e-Learning Systems, ICALT 2007, 2007.
- [13] 中澤真, 荒木道隆, 後藤正幸, 平澤茂一: 編集履歴可視化システムを用いた Learning Analytics, 情報処理学会, 第78回全国大会, pp.531-532, 2016.
- [14] 鍵山 貴一郎, 林 雄介, 平嶋 宗, :学習者集団および個人についての理解状態とその遷移の可視化の試み, 人工知能学会, 1C5-OS-13b-3, pp.1-3, 2016.