

# Learning Analytics における学習履歴の情報構造と粒度のあり方

中澤 真\*

後藤 正幸†

平澤 茂一†

## The Structure and Granularity of the Learning history in Learning Analytics

Makoto Nakazawa\*

Masayuki Goto†

Shigeichi Hirasawa†

**Abstract:** Recently, user can be performing almost all of the learning activities, the viewing of educational materials, examination, note-taking and communication, on the e-learning system. Newer standards on interoperability such as Tin Can been considered. Therefore, we can centrally collect and manage the learning history from a variety of devices and systems. However, it is not clear how to define the adequate granularity and structure of the learning history. In this study, we investigate the influence of the granularity on the results of learning analytics by means of demonstration experiments. In addition, I will report on the requirements for the structure of learning history data.

**Keywords:** Learning Analytics, Granularity, Student Model, Data Mining

### 1. はじめに

ICT を活用した授業において、学習者は教材の閲覧からテスト、ノートテイキング、コミュニケーションに至るまですべての学習活動を e ラーニングシステム上で行えるようになりつつある。特に高等教育においては、LMS (Learning Management System) や e ポートフォリオなどの導入により、学習者の学習活動履歴を詳細に記録できる環境が整ってきた。最近では Tin Can のような学習履歴データについての標準化も進められており、多様なデバイスやシステムに対して学習履歴を一元的に収集・管理が可能なシステムの構築も容易になり、また、MOOC (Massive Open Online Course) のような誰もが自由に受講できる開かれた講義も展開され、ビッグデータといえる大規模な学習履歴の蓄積も可能になった。

これらの学習活動の履歴や学習環境に関する多様なコンテキストは、学習者の置かれた状況を正確に把握して一人ひとりに適した学習支援や、インストラクショナルデザインに基づく教材や授業構成の見直しや学生指導に必要な情報を教員へとフィードバックすることを可能にする<sup>[1]</sup>。しかし、どのような学習履歴をどの程度の粒度で収集するか、また、その情報の構造をどのように定めるかべきかは明確になっていない。

そこで本研究では、学習履歴の粒度の細かさ、特に時間的な緻密さが Learning Analytics に与える影響について、早稲田大学の授業「Discussion Tutorial English」における個別学習と、会津大学短期大学部の対面型授業「情報ネットワーク」による実証実験で明らかにするとともに、学習履歴データの構造として求められる要件について報告する。

### 2. Learning Analytics と学習履歴

Chatti ら<sup>[2]</sup>は Learning Analytics (LA) を「学習やその環境を把握し最適化するために、学習者とそのコンテキストに関するデータの測定、収集、分析、報告をする取り組み」と定義している。LMS, MOOC, AIHS (Adaptive and Intelligent Hypermedia Systems), ITS (Intelligent Tutoring System) など学習や教育環境の ICT 化が進み<sup>[3]</sup>、学習者がこれらのシステム上で操作する教材の閲覧、ノートテイキング、コミュニケーション、小テストなどの幅広い学習活動のほとんどを学習履歴として蓄積できるようになりつつある。

Moodle などの LMS のログデータは学習履歴の典型的なものであり、クラスタリングのようなデータマイニング手法により分析されることが多い。この種のデータを分析処理するサイクルは図 1 のように示せるが<sup>[4]</sup>、学習者が学習活動を途中で中断・放棄するなど不完全なデータがあること、標準的な HTTP のみに基づくログの場合にはキャッシングやプロキシサーバの影響でノイズを含むデータとなっていること<sup>[5]</sup>、粒度の異なる多様なデータが存在することなどの理由から、前処理が分析結果に大きな影響を及ぼす。

この学習履歴の粒度について、古くは e-learning 環境モデル「RAPSDY」<sup>[6]</sup>において、IEEE LTSC が提案している LOM (Learning Object Metadata) を、それぞれの学習コンテンツに対し、コース、章 (ユニット)、ページ、の 3 段階で付与するモデルが提示されている。この他にも、学習者が Office やブラウザ、マルチメディア再生、コミュニケーションツールなどの操作した記録を、XML 形式で一元的に管理・記録できる CAM (Contextualized Attention Metadata) フ

\* 会津大学 短期大学部, University of Aizu

† 早稲田大学, Waseda University

フォーマットと呼ばれるフレームワークや<sup>[7]</sup>, 仮想化技術を用いて学習履歴を一括管理する方法<sup>[8]</sup>などが提案されている。

また, e-Learning コンテンツの標準規格のひとつ SCORM 2004 においては, 木構造に配置された学習の活動要素とトラッキングデータを紐付け, 学習者の学習進捗を管理する機能を定めている<sup>[9]</sup>。しかし, この場合の学習履歴は学習者単位あるいは活動単位がベースとなり, 図 2 に示すように粒度が十分細かく, また学習者の行動を正確に把握するのに十分なデータ量であるとは言いがたい。

最近では学習履歴収集の標準化の試みとして Tin Can API も注目されている<sup>[10]</sup>。Actor (主体), verb (動作), object (対象) に分けて, 誰が何をどうするのかを記述するフォーマットを定め, 様々なシステム上の学習履歴を一元的に収集・管理することを目的にしている。多様なデバイスやシステムに対応できるだけでなく, 従来よりも細かい学習履歴の記録と収集ができる特徴を持っている。しかし, 運用の柔軟性を考慮して, 粒度については利用者がそれぞれ定めることになっており, 適切な粒度の定義は明確になっていない。

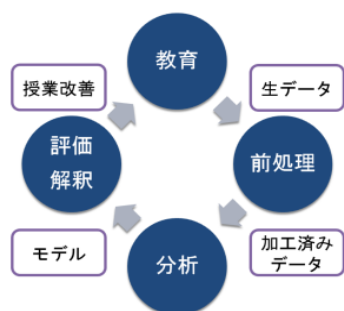


図 1 学習履歴の処理の流れ

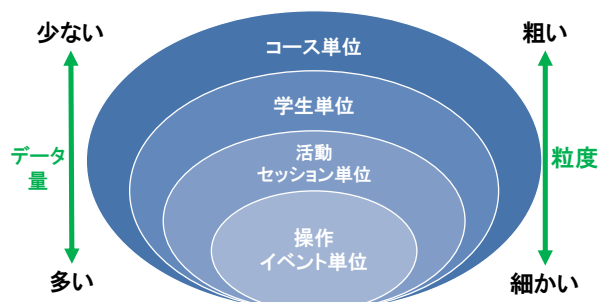


図 2 データ量と粒度の概念図<sup>[3]</sup>

粒度の細かい学習履歴の収集と分析のために, 筆者らはページ単位の閲覧行動やウィンドウのアクティブ・非アクティブを検出できる Web ベースの電子教材を用いてきた<sup>[11]</sup>。これは荒本らが開発した電子教材作成支援システムであり<sup>[12]</sup>, 学習者の閲覧行動について, 教材名, ページ番号, 閲覧開始日時, 終了日時, 閲覧秒数などの情報を Ajax により Web サ

ーバに自動送信されるようになっている。特に, 閲覧しているか否かの判断を JavaScript の focus を用いて実現しているため, 学習者が別のアプリケーションを操作したり, ブラウザのタブ切り替えで教材の閲覧を中断した場合でも, 学習者の行動を正確に把握することが可能であり, ブラウザのキャッシングの問題も解決することができる。本研究の実証実験では, PowerPoint で作成した資料をこのシステムによって Web 教材へと変換し, オープンソース LMS の Moodle 上に組み込んだ。この際, 教材組み込み部分のパラメータとしてユーザ ID を設定することで, Moodle のユーザ ID も学習履歴内に記録させ, 2 つのシステムにまたがる情報の紐付けをしている。

### 3. 学習履歴の粒度と前処理

本研究では二つの授業における学習履歴データを用いる。一つ目は早稲田大学の授業「Discussion Tutorial English」で実際に使用している英語教材を電子化したものを学生に事前学習させ, 学習時の操作履歴を記録したものである。教材内に選択問題や記述問題を組み込み, この解答時の行動も詳細に記録している。学習者は全 28 ページの教材で英語のリーディングをしつつ, この教材内に組み込んだ選択問題 6 問 (ページ 11~16) と記述問題 10 問 (ページ 17~20, 22~26, 28) に取り組むことになり, この正答率で学習効果を評価する。なお, 途中放棄したものや教材をほとんど閲覧していない学習者を除き, 35 名の学生を対象としたデータで分析する。

二つ目は会津大学短期大学部の専門科目「情報ネットワーク」において, 授業時間中に学習者が電子教材を閲覧した行動を記録したものである。教員は電子教材をプロジェクタに投影しながら授業を進め, 履修学生も演習室内の各自の PC 画面で自由に同じ教材を閲覧できるスタイルである。学習効果は授業時間内に取組ませた Moodle 上の練習問題の解答に要した時間および, 期末試験で対象回の授業の単元の得点により評価する。なお履修者数は 32 名である。

ここでは学習履歴の粒度について考える。先にも述べたように Moodle などの LMS は Web サーバベースのログを記録するが, その粒度はセッション単位や活動単位が基本であり, 教材の閲覧などもファイル単位でのアクセス履歴となる。このような粒度の学習履歴を用いた場合には, 学習効果との関係性を示すことが難しいことを著者らは既に示した<sup>[11]</sup>。このためページ単位での閲覧行動を記録し, ウィンドウのアクティブ・非アクティブを検出できる機構が必要となる。

しかし, このような細かい粒度のデータにはブラウザ操作の特性によるノイズが混入しやすい<sup>[5]</sup>。とくに, 本研究の電子教材はシーケンシャルに構成されているため, 閲覧行為ではなく単なるページ遷移

もログとして記録されてしまう可能性を考慮する必要がある。このため、データの前処理として慎重にクリーニングする必要がある。

そこで、二つの授業の学習履歴として記録されている一回の閲覧行動あたりの閲覧時間のばらつきについて調査した。これを度数分布としてまとめたものが図 3, 図 4 である。個別学習である「Discussion Tutorial English」、通常の対面型の授業である「情報ネットワーク」と学習スタイルはまったく異なる二つの授業であるが、どちらも 1 回の閲覧時間が 1 秒未満となるログが他の閲覧時間の頻度と比較して極端に多いことがわかる。このため、本研究では 1 秒未満の閲覧時間のログは解析対象から除外するクリーニングを前処理として行うことにした。なお、クリーニングの対象を 2 秒未満とすると、タイトルや指示内容のみが記載された短い閲覧時間で内容を把握できるページ (図 5 のページ 10, 21, 27) の履歴がほとんど除外されてしまうため適当でないと判断した。図 5 は 1 秒以上の閲覧時間のログのみを集計した結果であるが、ページ 17 以降の記述問題の平均閲覧時間が総じて長いことや、記述問題の中で最も平均点が高かったページ 18 の閲覧時間が短いことなど、電子教材の特性がよく現れていることが確認できる。

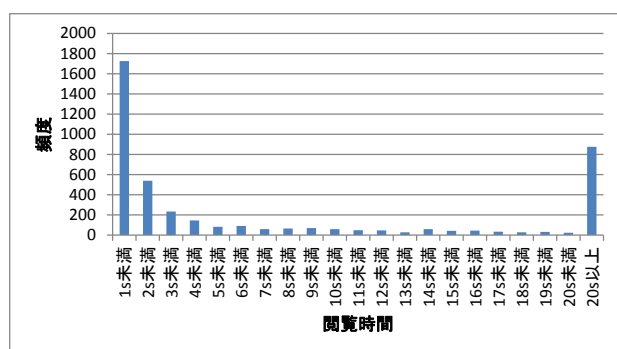


図 3 一回の閲覧行動あたりの閲覧時間の度数分布  
～Discussion Tutorial English

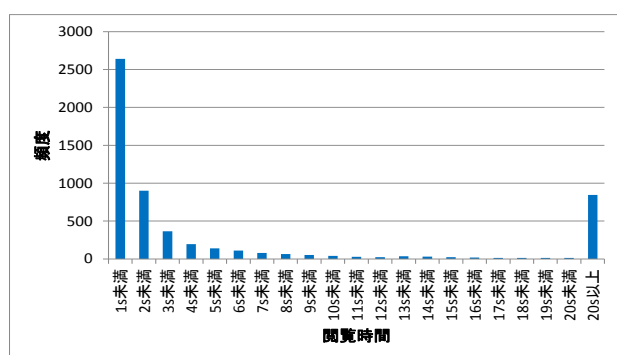


図 4 一回の閲覧行動あたりの閲覧時間の度数分布  
～情報ネットワーク

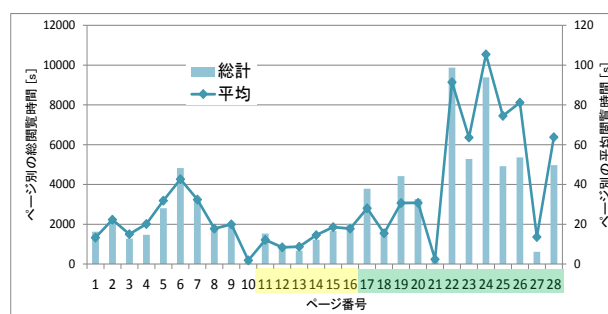


図 5 ページ別の総閲覧時間と 1 回あたりの平均閲覧時間  
～Discussion Tutorial English

### 3. 教材のページ遷移分析

先に示した図 5 のように、収集した学習履歴を単純集計するだけでも、教材の特性をある程度確認することができるが、履歴の時系列性に着目して分析するとさらに学習環境や学習者の行動特性などを明らかにしやすい。一般的な Web 履歴分析においては、簡単のため順序依存で時間独立な確率モデルを考える。つまり、1 次のマルコフ過程を仮定し、状態  $s_t$  の確率は状態  $s_{t-1}$  だけに依存し、それ以前の状態とは独立であるとする。このとき、全学習者の閲覧行動に対してページ  $i$  からページ  $j$  に遷移する確率は以下のように定義することができる。

$$P(s_t = j | s_{t-1} = i) = \frac{i \text{ から } j \text{ への遷移数}}{i \text{ からの全遷移数}}$$

この確率を遷移行列の  $i$  行  $j$  列の要素として示した結果が図 6 である。図中において、グラデーションが濃い部分ほど値が大きいことを表している。教材がシーケンシャルな構成のため、ページ  $i$  からページ  $i + 1$  に遷移する確率が特に大きくなっているが、前のページに戻る遷移や、同一のページでの自己遷移の確率が高いページの存在も確認できる。これは対面型授業である情報ネットワークの結果でも同様の傾向を示す。

$W$  ページから構成されている電子教材を用いる場合、 $W$  個の状態があるモデルとなるが、一般の Web サイトと異なり特定の授業ではアクセス数がかなり限定されるため、遷移行列において確率 0 となる要素が多数存在するスパースな行列になることが多い。受講者数が多い授業であったとしても、学習者個別の分析をする場合には、 $W \times W$  のすべての遷移が発生することは稀である。このため、各ページをそれぞれ状態に対応させてモデル化することは現実的ではない。これに対応するために、元々の  $W$  ページをグループ化し、このグループ単位でマルコフモデルの状態に割り当てると方法が考えられる。

本研究においては、先の遷移確率行列に基づき、現在閲覧中のページより前のページに戻った場合 (Back)、直後のページに遷移した場合 (Next)、2 ページ以上先のページに遷移した場合 (Forward)、他の作業をしてから同じページに自己遷移した場合の 4 つの状態で構成されたモデルを考える。まず、電子教材の各ページが 4 つの状態に遷移する確率をまとめたものを図 7 に示す。この中で特徴が明確に表れているものは記述問題のペ

ージ群であり、他と比べて Back への遷移確率が高いものや自己遷移の確率が高い。英文のパラグラフが掲載されているページ群では、後半に Back への遷移確率がやや高くなっており、学習者が英語を最後まで読み進めた後に再度読み直しの操作をしていることが見て取れる。これは対面型授業の情報ネットワークでも同様の傾向が示され(図 8), 記述問題のページでは Back や自己遷移の確率が高く、説明が記されたページの終盤の Back への遷移確率が高い傾向が示されている。このようにいずれの学習スタイルにおいても、電子教材の各ページの特徴の差異を 4 つの状態の遷移確率で表現できていることから、この状態遷移モデルは適切であると考えられる。

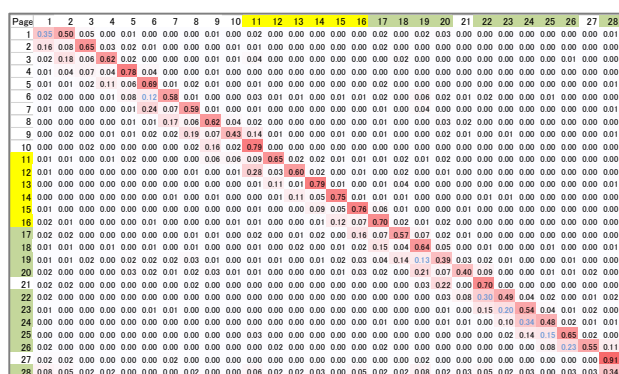


図 6 閲覧回数に基づくページ遷移確率  
～Discussion Tutorial English

| ページ | 種別    | Back | Self | Next | Forward |
|-----|-------|------|------|------|---------|
| 1   |       |      | 0.35 | 0.50 | 0.15    |
| 2   | パラグラフ | 0.16 | 0.08 | 0.65 | 0.10    |
| 3   | パラグラフ | 0.20 | 0.06 | 0.62 | 0.12    |
| 4   | パラグラフ | 0.12 | 0.04 | 0.78 | 0.05    |
| 5   | パラグラフ | 0.16 | 0.06 | 0.69 | 0.09    |
| 6   | パラグラフ | 0.11 | 0.12 | 0.58 | 0.20    |
| 7   | パラグラフ | 0.26 | 0.07 | 0.59 | 0.08    |
| 8   | パラグラフ | 0.19 | 0.06 | 0.62 | 0.12    |
| 9   | パラグラフ | 0.27 | 0.07 | 0.43 | 0.22    |
| 10  |       | 0.19 | 0.02 | 0.79 | 0.00    |
| 11  | 選択問題  | 0.16 | 0.09 | 0.65 | 0.10    |
| 12  | 選択問題  | 0.31 | 0.03 | 0.60 | 0.06    |
| 13  | 選択問題  | 0.12 | 0.01 | 0.79 | 0.08    |
| 14  | 選択問題  | 0.14 | 0.05 | 0.75 | 0.06    |
| 15  | 選択問題  | 0.11 | 0.05 | 0.76 | 0.08    |
| 16  | 選択問題  | 0.19 | 0.07 | 0.70 | 0.04    |
| 17  | 記述問題  | 0.27 | 0.07 | 0.57 | 0.09    |
| 18  | 記述問題  | 0.25 | 0.04 | 0.64 | 0.08    |
| 19  | 記述問題  | 0.42 | 0.13 | 0.39 | 0.06    |
| 20  | 記述問題  | 0.41 | 0.07 | 0.40 | 0.12    |
| 21  |       | 0.30 | 0.00 | 0.70 | 0.00    |
| 22  | 記述問題  | 0.13 | 0.30 | 0.49 | 0.08    |
| 23  | 記述問題  | 0.20 | 0.20 | 0.54 | 0.07    |
| 24  | 記述問題  | 0.13 | 0.34 | 0.48 | 0.04    |
| 25  | 記述問題  | 0.18 | 0.15 | 0.65 | 0.02    |
| 26  | 記述問題  | 0.12 | 0.23 | 0.55 | 0.11    |
| 27  |       | 0.09 | 0.00 | 0.91 |         |
| 28  | 記述問題  | 0.66 | 0.34 |      |         |

図 7 ページ別の 4 状態への遷移確率一覧  
～Discussion Tutorial English

| ページ | 種別   | Back | Self | Next | Forward |
|-----|------|------|------|------|---------|
| 29  |      | 0.29 | 0.07 | 0.49 | 0.15    |
| 30  | 記述問題 | 0.23 | 0.34 | 0.36 | 0.07    |
| 31  |      | 0.25 | 0.16 | 0.52 | 0.07    |
| 32  | 記述問題 | 0.33 | 0.42 | 0.22 | 0.03    |
| 33  |      | 0.28 | 0.15 | 0.47 | 0.09    |
| 34  |      | 0.36 | 0.05 | 0.56 | 0.04    |
| 35  |      | 0.22 | 0.09 | 0.56 | 0.13    |
| 36  |      | 0.30 | 0.06 | 0.56 | 0.08    |
| 37  |      | 0.27 | 0.04 | 0.54 | 0.14    |
| 38  |      | 0.44 | 0.06 | 0.46 | 0.05    |
| 39  |      | 0.42 | 0.04 | 0.54 | 0.00    |
| 40  | 記述問題 | 0.44 | 0.40 | 0.16 | 0.00    |

図 8 ページ別の 4 状態への遷移確率一覧  
～情報ネットワーク (抜粋)

次に、図 7 図 8 で求めた 4 状態への遷移確率のリストを各ページの特徴ベクトルと考え、これを k-means 法でクラスタリングすることにより、教材の遷移モデルを分類する。なお、k-means 法ではクラスタ数を 4 とし、ユークリッド距離を用いて最大 1000 回の繰り返し計算によりクラスタリングしている。ただし、1 ページ目と最終ページ、最終ページの前ページは、Back、Forward など遷移が起こりえないケースがあるため除外してある。また、Discussion Tutorial English のデータにおいては、タイトルや指示内容のみが記載された短い閲覧時間で内容を把握できるページ 10、21、27 も除外した。

このクラスタリングの結果を示したものが図 9 である。Back への状態遷移確率が最も高いクラス P4 は、平均得点が低かった記述問題のページと、リーディング対象の英文の最終ページで構成されており、難易度の高いページ群のモデルと考えられる。一方、Next へ遷移する確率が最も高いクラス P3 は、選択問題のページやタイトルページなどで構成されており、情報量が少ないあるいは難易度が低いページ群のモデルとなっている。また、クラス P2 の特徴は自己遷移確率が他よりも高い点であり、単語を調べるなど教材の閲覧以外の操作をしながら取り組む内容のページ群で構成されている。

これらのクラスに各ページを分類した結果はインストラクショナルデザインを見直すのに有用であり、教員が意図した授業設計や流れの通りに、電子教材の各ページが利用されているかどうかを確認することができる。例えば、前のページに戻って十分読み直しをしてほしいにもかかわらず Back への遷移確率が低かった場合には問題の作り方やヒントなどを見直すべきであり、逆にすんなり次ページへと読み進めると考えていたページで Back や Self の確率が高い場合には、説明の方法や難易度についての見直しを検討する必要があることと示唆してくれる情報となる。この際、電子教材のページごとにその種別に関する情報や、内容に関する関係性を表現するためにシーケンシングについての情報を学習履歴の構成要素として実装すると、より円滑で高度な解析へとつなげることができるであろう。

#### 4. 学習者のページ遷移分析

ここでは学習者の遷移モデルについて考える。4つの状態の遷移確率を成分として特徴ベクトルを考える点は先と同じであり、これを学習者ごとに選択問題の全ページでの平均遷移確率を用いたモデルと、記述問題の全ページでの平均遷移確率を用いた二つのモデルの二つをk-means法によるクラスタリングで求める。図10は授業「Discussion Tutorial English」の記述問題のモデルである。図中の各クラスタに示してある事前英語力は、学習前のアンケートからTOEIC、WeTECなどの点数を換算して授業担当者が算出した値をクラスタ平均したものである。クラスタの構成は図9の教材の遷移モデルと似た構成となっており、Backへの状態遷移確率が最も高いクラスタUE4は、英語力の低い学習者の遷移モデルとなっていることがわかる。一方、英語力の高い学習者のモデルはUE2であり、自己遷移確率が高い点が特徴である。中間的な英語力の学習者はクラスタUE1またはUE3に分類され、その違いは前ページの読み直しを少しするか、あっさり次ページへと進むかにあり、多少読み直しをしたほうが記述問題の得点が高くなる傾向が示されている。

この結果は学習プロセスとして過度の英文の読み返しをするよりも、英作文の推敲に力点を置き、検索サイトで英文フレーズの検索などをしたほうが適切であることを示している。ただし、この傾向は英語学習に限ったことではない。図11は授業「情報ネットワーク」における記述問題に対するクラスタリングの結果であるが、遷移確率に多少の差異はあるが、図10と類似のクラスタが構成されていることがわかる。特に、Backへの状態遷移確率が最も高いクラスタUEn4は、理解の低い学習者の遷移モデル、自己遷移確率が高いUEn2は理解の高い学習者のモデルとなっており、異なる分野であっても学習プロセスに共通の傾向が示されることが確認できた。

このようなモデル化は記述問題全体に対してではなく、ページ単位・問題単位での分析も可能である。例えば図12は「Discussion Tutorial English」の記述問題中、得点の分散が最も大きかったページ19に特化したモデルである。クラスタの傾向は記述問題全体のクラスタである図10と似ているが、クラスタUE<sub>(19)</sub>4はクラスタUE4と比べてBackへの遷移確率がかなり大きいことが示されており、英語力の低い学習者がかなり苦戦していることが明らかである。また、クラスタUE<sub>(19)</sub>3はBackやSelfへの遷移がほとんどなく、あまり努力せずに問題に取り組んだ学習者の集まりであり、記述問題全体に対するクラスタリングにはないタイプある。つまり、個別の問題に対する学習プロセスをこのモデルから明らかにし、学習者指導だけでなく問題の難易度調整や再設計にも用いることが可能である。また、このページ19の問題は教材のパラグラフページ5, 6を参照することを想定しているが、クラスタUE<sub>(19)</sub>4のBackへの高い遷移確率にもかかわらず、実際にページ5, 6へと遷移している確率は約0.3であり、学習者が適切な読み直し

をしていないということも抽出することもできる。

次に選択問題に対する学習者の遷移モデルをクラスタリングにより求めた結果を図13に示す。今回の教材ではクリックにより解答の正誤が瞬時に確認できることもあり、学習者は読み直しや他のツールを用いての調査などをほとんどしていないため、クラスタごとの差は記述問題と比べると曖昧である。ただし、クラスタUM3のモデルのように、英語力の低い学習者が読み直しも調べることがまったくせずに解答し、結果として正答率も低くなっている問題行動などを確認することは可能である。

最後に遷移確率を遷移回数ではなく、遷移先の閲覧時間に基づいて定義した場合のモデルについて検討する。この場合の遷移確率は次式のように定義できる。

$$P_T(s_t = j | s_{t-1} = i) = \frac{i \text{ から } j \text{ への遷移先の総閲覧時間}}{i \text{ からの全遷移先の総閲覧時間}}$$

図14はこの閲覧時間ベースの遷移確率に基づき、学習者の遷移モデルをクラスタリングにより求めた結果である。クラスタの構成は図10と似ているが、クラスタ間の遷移確率の差異はやや小さくなっており、閲覧時間よりも閲覧回数を基本としてモデル化は考えるべきであると結論付けた。

#### 4. 考察

今回の分析結果では、ページ単位での閲覧行動の遷移を秒単位の粒度で学習履歴とすることにより、教材の閲覧遷移モデルや学習者の閲覧行動のモデルを構築できることを示した。特に、図9から図14のクラスタリングの結果では、教材の内容や授業のスタイルあるいは分野によらずに、ほぼ同じ傾向の4つ遷移モデルとなることが明らかになった。これらのクラスタは教材のページ単位での特性を抽出したり、学習者の理解度を反映した学習プロセスを可視化するのにも有用である。このことは、学習履歴を構成する情報として、学習者を特定するIDと教材のページ単位のID、そして閲覧の時系列的順序を特定できる3種類さえあれば、かなり精緻な分析が可能であることを意味する。ただし、LOMのようなメタデータをページ単位で付与し、さらに分析の自動化のためにページ間の関係性を示す情報も追加しなければ、結果の解釈を利用者に委ねることになり適切であるとは言い難い。ページ遷移に着目したモデル化においては、ページ単位でのシーケンスを授業設計者があらかじめ用意し、学習者の行動がこれとどの程度ずれているのかを測定できる枠組みを考えていく必要があるだろう。

#### 5. おわりに

本研究で示した教材のページ遷移モデルを用いることで、ページ単位で教材の改善や再設計を容易に行うことが可能になる。また、学習者の遷移モデルは、学習プロセスに問題のある者を抽出し、学習指導の支援機能としての可能性を示している。例えば、

対面型授業のモデルを使えば、練習問題に取り組んでいる時間中に、教材の読み直しを頻繁に行っている学習者をリアルタイムに抽出し、教員へとフィードバックすることも可能である。

今回のモデルでは4状態の遷移モデルとしたが、学習プロセスをより詳細に把握するためには、コンテンツのページ特性を考慮した状態としたモデルも検討する必要があるだろう。また、一時マルコフという仮定のため、学習アクティビティ単位での特徴が失われているという問題もある。今後二次マルコフ以上のモデル化についても適切なサンプル数が確保できるかを考慮しつつ取り組まなければならない。

### 謝辞

本研究の実証実験授業を担当いただいた早稲田大学の中野美智子先生と関係者の皆様に厚く御礼申し上げます。

なお、本研究の一部は、独立行政法人日本学術振興会学術研究助成基金助成金(基盤研究(C)、課題番号:26350299)の助成による。

### 参考文献

- [1] J.A. Larusson, B. White (Eds.), Learning Analytics: From Research to Practice, Springer (2014).
- [2] Chatti, M.A., et al., "A reference model for learning analytics", Int. J. Technol. Enhanced Learn. 4(5-6), pp.318-331 (2012).
- [3] Romero, C., et al., "A Survey on Pre-Processing Educational Data", Educ. Data Mining Applications and Trends, pp.20-64 (2013).
- [4] Romero, C., Ventura, S., "Data mining in education," Data Mining Knowledge Discovery 3(1), pp.12-27 (2013).
- [5] Pierre B., et al., 確率モデルによるWebデータ解析法, 森北出版 (2007).
- [6] 岡本敏雄, 香山瑞恵, 人工知能と教育工学, オーム社 (2008).
- [7] Wolpers M., The attention metadata approach," Journal of Technology Education & Society, 10 (3), pp.106-121 (2007).
- [8] Romero Z., et al., "Monitoring student progress using virtual appliances: A case study," Computers & Education, 58(4), 1058-1067 (2012).
- [9] 青木久美子, eラーニングの理論と実践, 放送大学教育振興会 (2012).
- [10] Á del Blanco, et al., "E-Learning Standards and Learning Analytics," Global Engineering Educ. Conference (EDUCON), pp.1255-1261 (2013).
- [11] 中澤真, 小泉大城, 後藤正幸, 平澤茂一, "詳細な学習履歴を活用した学習者行動の分析," 情報処理学会第76回全国大会 講演論文集, pp.4-357-4-358 (2014).
- [12] 荒本道隆, 小泉大城, 須子統太, 平澤茂一, "PDFファイルをベースとした電子教材作成支援システム," 情報処理学会 第76回全国大会 講演論文集, pp.4-359-4-360 (2014).

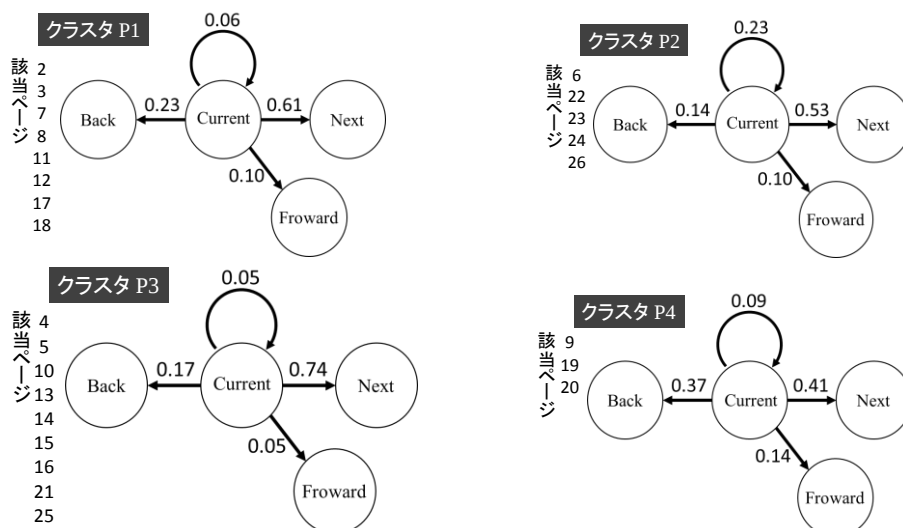


図 9 教材のページ遷移モデル ～Discussion Tutorial English

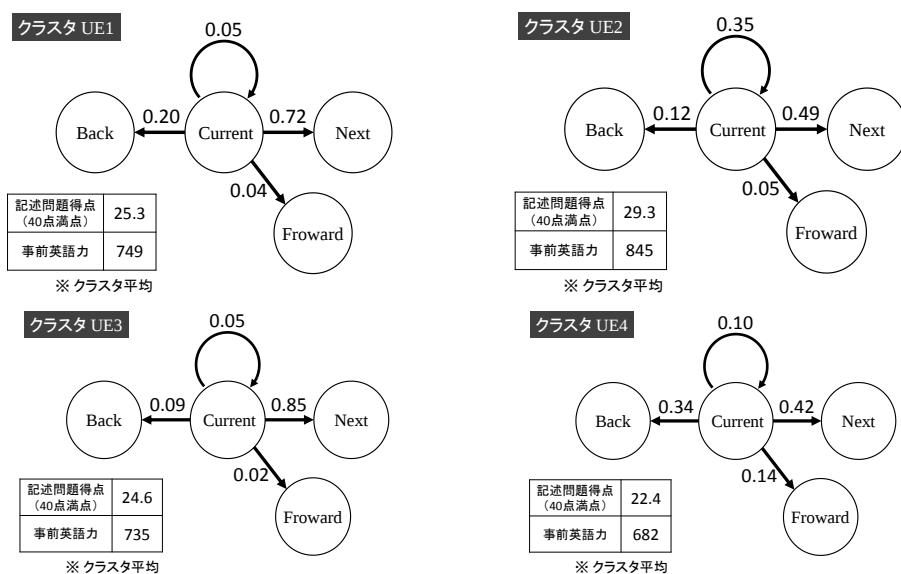


図 10 学習者の記述問題のページ遷移モデル ～Discussion Tutorial English

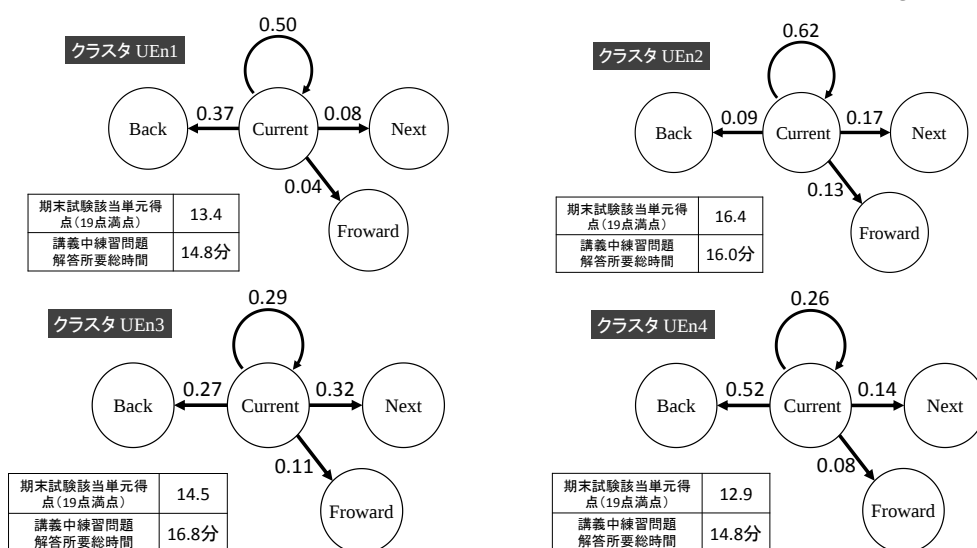


図 11 学習者の記述問題のページ遷移モデル ～情報ネットワーク

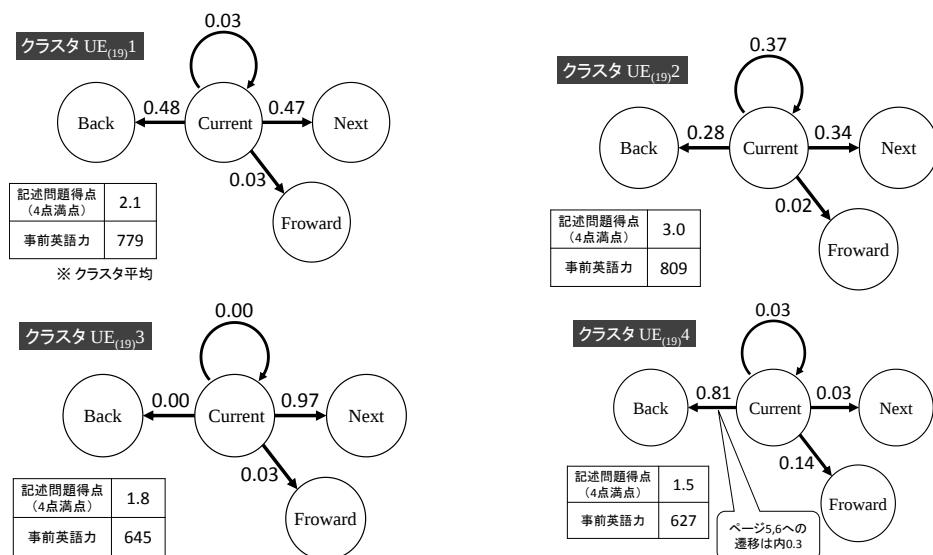


図 12 学習者の記述問題 19 のページ遷移モデル ～Discussion Tutorial English

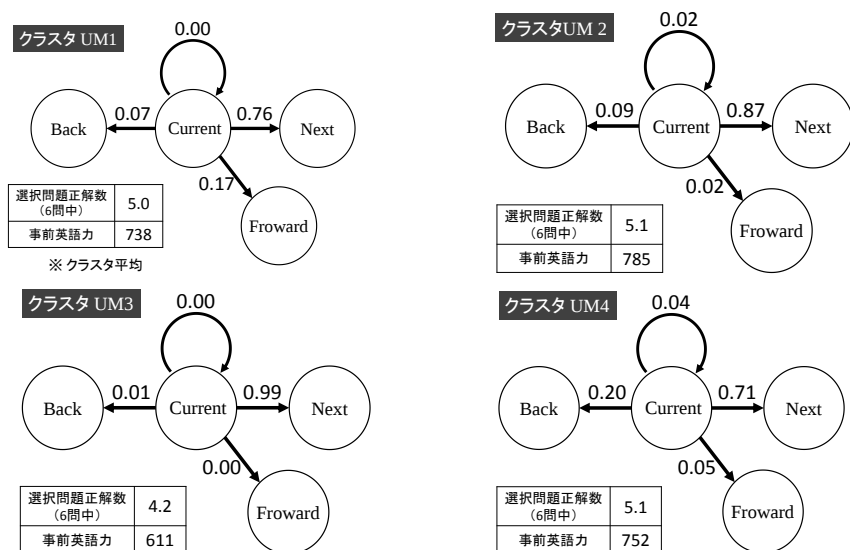


図 13 学習者の選択問題のページ遷移モデル ～Discussion Tutorial English

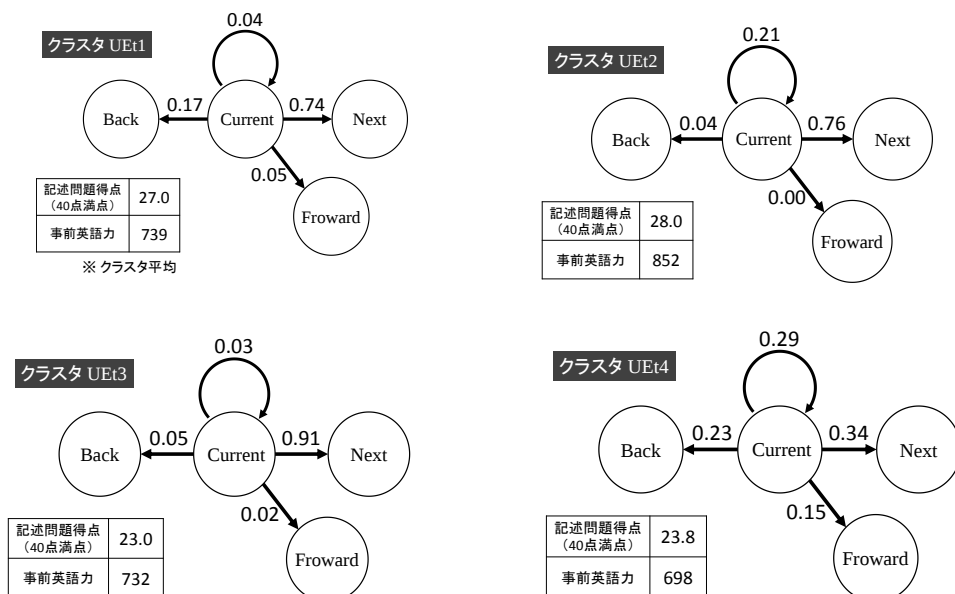


図 14 学習者の記述問題のページ遷移モデル ～Discussion Tutorial English (閲覧時間ベース)