

# 非負値行列因子分解を用いたデジタル教科書におけるスクラッチ機能の閲覧パターン分析

毛利考佑<sup>†1</sup> 島田敬士<sup>†2</sup> 殷成久<sup>†3</sup> 魚崎典子<sup>†4</sup> 金子敬一<sup>†1</sup> 緒方広明<sup>†5</sup>

**概要**： 昨今、e-Learning システムやデジタル教科書などから収集することができるログを用いて多種多様な分析を行い、学習や教育を改善に取り組む「ラーニングアナリティクス」が世界中で注目されている。しかしながら、現状のデジタル教科書のシステムでは、学習者が教科書を開いただけなのか、それとも適切に閲覧しているのかを見分けることができなかつた。こうした問題を解決するために、アイトラッキング技術を利用することで問題を解決することができるが、新たにコスト面での問題を抱えることになる。そこで、本論では、アイトラッキング技術無しに学習者がデジタル教科書のどの個所を読んでいるかを収集することができるスクラッチ機能の開発を行った。さらに、本提案するスクラッチ機能で収集したログと従来のデジタル教科書で収集したログから、非負値行列因子分解 (Non-Negative Matrix Factorization) を用いて学生の閲覧パターンの比較を行った。その結果、教科書を単に開いただけの場合と適切に閲覧している場合の二つの閲覧パターンを発見し、それぞれのパターンが学生のレポートや演習課題などの学習成果に影響していることを明らかにした。

**キーワード**： デジタル教科書、非負値行列因子分解、データマイニング

## 1. はじめに

近年、多くの教育機関において、講義の出席やレポートの点数などを収集することができる e-Learning システムが運用され続けている。また、2020 年までに日本政府は、小中高等学校にデジタル教科書を導入する計画を進めている [1]。デジタル教科書は、e-Learning システムと違って、これまで収集することができなかつた、学習者の教材の閲覧履歴を収集することを可能にした。昨今、その収集した大量のログを多様な方法で分析・可視化することで学習や教育の方法の改善に取り組む「ラーニングアナリティクス」が世界中で注目されている [2]。

我が国においても、e-Learning システムとデジタル教科書から収集したログの分析に関する研究が盛んに行われている。その多くは現状のシステムで収集することができるログを利活用するために多様な分析・可視化の方法に基づいて得た結果から学習・教育改善を行っている [3], [4], [5], [6], [7]。しかしながら、現状のシステムでは、「学習者が教科書を適切に読んでいるか」のログを収集することができない。なぜなら、現状のデジタル教科書は、学習者がどの教材のどのページを読んでいるかのログの収集で留まっている。したがって、学習者が教科書を開いただけなのか、それとも適切に閲覧しているのかを見分けることが困難である。この問題を解決するための一つの方策として、アイトラッキング技術がある。しかしながら、アイトラ

ッカーをすべての学習者に提供するコスト面を考えた場合、大量の初期費用が掛かり、導入するには現実的に難しい問題がある。

この問題に取り組むために、本論は、アイトラッキング技術無しに学習者が単に教科書を開いただけなのか、それとも適切に閲覧しているのかを見分けることができる、デジタル教科書システムの設計を行った。さらに、本提案と従来のデジタル教科書で収集したログから、非負値行列因子分解を用いて学生の閲覧パターンの比較を行った。

## 2. Smart E-Textbook Application

学習者がデジタル教科書を単に開いているだけなのか、それとも適切に閲覧しているのかを見分けるために、本研究では、SEA (Smart E-textbook Application) と呼ばれる新たなデジタル教科書システムを開発した [8], [9]。図 1 は、SEA のインタフェースを示す。SEA は、ハイライトやメモなどの従来のデジタル教科書の機能を備えている。従来のデジタル教科書と違って、スクラッチ機能を実装した。

スクラッチ機能は、教員が教科書上で重要な場所に対して隠ぺいすることができ、隠ぺいした場所を学習者がクリック操作やマウスオーバー操作をすることで該当箇所の閲覧ログを収集することができる。これにより、学習者がデジタル教科書を単に開いているだけなのか、それとも適切に閲覧しているのかを見分けることが可能となる。表 1 は、本システムで収集することができるログの詳細を示す。ID は、ログの主キーを示す。Book\_ID は、デジタル教科書の ID を示す。Author\_ID は、Book\_ID に対応するデジタル教科書を閲覧した学習者の ID を示す。Page\_no は、教科書のページ番号を示す。Block\_no は、デジタル教科書内で隠ぺいした個所の番号を示す。その隠ぺいした個所に対する X

†1 東京農工大学  
Tokyo University of Agriculture and Technology

†2 九州大学  
Kyushu University

†3 神戸大学  
Kobe University

†4 大阪大学  
Osaka University

†5 京都大学  
Kyoto University

表 1. SEA に関するログの詳細

ID	Book_ID	Author_ID	Page_no	Block_no	X 軸	Y 軸	Create_time
1	Epub_1	Epub_1	1	1	24.2	35.3	2018/4/21 18:47:00
2	Epub_1	Epub_1	1	2	32.4	78.2	2018/4/21 18:47:02
3	Epub_1	Epub_1	2	1	54.2	13.2	2018/4/21 18:47:03
4	Epub_2	Epub_2	3	1	23.3	35.2	2018/4/21 18:47:05
5	Epub_2	Epub_2	5	2	23.3	35.4	2018/4/21 18:47:05

軸 Y 軸の値を示す. Create\_time は隠ぺいした個所をクリックした時間を示す.

今回は, 学習者の閲覧パターンを分析するため, ページ内のすべてのテキスト上に自動的に隠ぺいを行った. 図 1 の mask は, テキスト上に隠ぺいしたことを意味している. 本論では, この隠ぺいした個所の単位をブロックとする.



図 1 SEA のインターフェース

### 3. デジタル教科書のログに関する非負値行列因子分解

従来のデジタル教科書のログ分析の研究では, 収集することができるデジタル教科書のログから学習者の閲覧パターンを分析する研究がある. その従来研究では, 学習者の閲覧パターンを発見するために, 公式 1 に示す非負値行列因子分解 (Non-Negative Matrix Factorization) が用いられている [10].

$$V \approx WH \quad (1)$$

その従来研究で取り扱う, 学習者×ページ番号のマトリ

ックス  $V_{i,j}$  をベースとした, 「学習者×パターン」及び「ページ番号×パターン」の行列分解を行っている. さらに, 学習者が閲覧ページに対して適切に読んでいるかの判定として, それぞれの学生のページごとの閲覧時間の閾値  $th$  を以下の公式 2 で設定を行った.

$$V_{i,j} \begin{cases} 1 & (if\ t_{i,j} > th) \\ 0 & (otherwise) \end{cases} \quad (2)$$

本研究では, 従来研究では収集することができなかった「学習者がデジタル教科書を単に開いているだけなのか, それとも適切に閲覧しているのか」に関してスクラッチ機能を用いて収集を行った. その収集したログを扱って, 学習者×ページごとのブロック番号のマトリックス  $V_{i,j}$  をベースとした, 「学習者×パターン」及び「ページごとのブロック番号×パターン」の行列分解を行った. また, 公式 2 の閾値  $th$  に従って, ブロック間の閲覧時間の設定を行った.

## 4. 評価実験

### 4.1 被験者とデジタル教科書の詳細

今回, 東京農工大学の学部 1 年生の 36 名を対象としたプログラミング演習に関する講義で SEA を利用した. その講義内容は, C 言語のポインタと構造体である. 教員は, 講義前に講義で使用するデジタル教科書を予習することを被験者に要求した. 表 2 は, 使用したデジタル教科書の詳細を示しており, 主に 5 つの章で構成されている. 最初の章では, 前回の講義の内容の要約を記載している. 第 2 章では, 構造体の定義や変数 (メンバ) のやり方, 構造体を用いた単方向リストの作成方法が記載されている. 第 3 章では, 構造体と単方向リストの実際のプログラムの一例を記載している. 第 4 章では, 構造体に基づいて単方向リストのプログラムを作成する演習課題を与えている. 第 5 章では, 第 4 章で与えられた演習課題を解くためのヒントを

記載している。

表 2 教科書の詳細

ページ	詳細
番号 1~7	前回の講義の内容 (ポインターの宣言)
番号 8~19	構造体の定義や構造体内の変数. 構造体を用いた単方向リストの詳細
番号 20~30	構造体と単方向リストのプログラムの例
番号 31~36	演習課題
番号 37~50	演習課題のヒント

#### 4.2 演習課題とレポートの評価基準

演習は、単方向リストの実装を C 言語で作成するプログラムを提出することを課題とした。図 2 は、課題で与えた単方向リストのデータ順序関係を示す。このリストを実装するために、ノードを先頭に追加、ノードを末尾に追加、ノードの削除及びプログラムを終了する関数の作成を行った。また、作成したリストを番号順、名前順にソートできる関数の作成を行った。

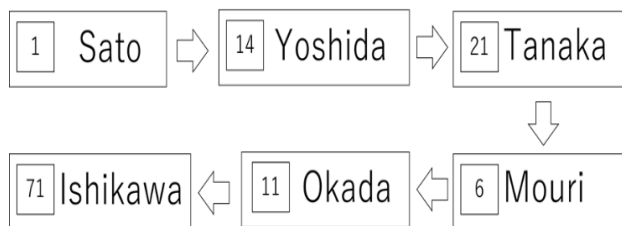


図 2 リストのデータ順序関係

レポートに関しては、演習課題の目的・実験方法・実行環境・実行結果・考察を提出することを求めた。表 3 と 4 は、本研究で行った演習課題及びレポートの評価基準を示す。

表 3 演習の評価基準

配点	評価の視点
0 点	課題の未提出
1~2 点	プログラムのコンパイルができない
3~4 点	コンパイルはできるが実行ができない
5~6 点	コンパイルができるが結果が間違っている
7~8 点	コンパイルができ、結果は合っているが、プログラムに不備がある(メモリー確保や解放ができていないなど)
9~10 点	コンパイル、実行、プログラムに不備がなく、結果が正しく出力される

表 4 レポートの評価基準

配点	評価の視点
0 点	レポートの未提出
5 点	目的が明確に記述されている
6~10 点	目的・実験方法が明確に記述されている
11~15 点	目的・実験方法・実行環境・結果が明確に記述されている
16~20 点	目的・実験方法・実行環境・結果・考察が明確に記述されている

#### 4.3 分析データ

講義で使用したデジタル教科書は、50 ページの 867 ブロックで構成されている。従って、本研究は、以下の 2 種類のマトリックスを作成した。

(1) 図 4 は、36 行×50 列のマトリックスを示す。縦軸がそれぞれの学生、横軸はページ番号を表す。学生がページを閲覧している場合は赤色、それ以外は青色とする。こちらのマトリックスは、従来のデジタル教科書で収集することが可能なログを表す。

(2) 図 5 は、36 行×867 列のマトリックスを示す。縦軸がそれぞれの学生、横軸はページ番号を表す。学生がページのブロックを閲覧している場合は赤色、それ以外は青色とする。こちらのマトリックスは、スクラッチ能で収集することが可能なログを表す。

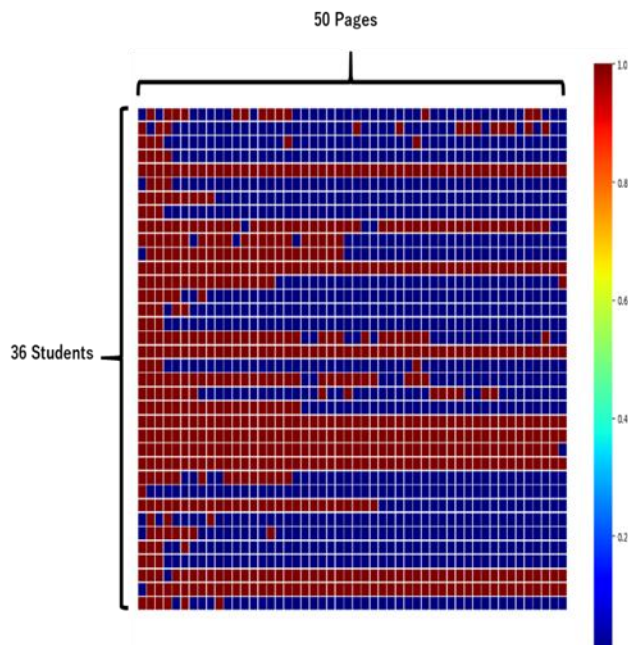


図 4 36×50 の行列 (縦軸：学生、横軸：ページ番号)

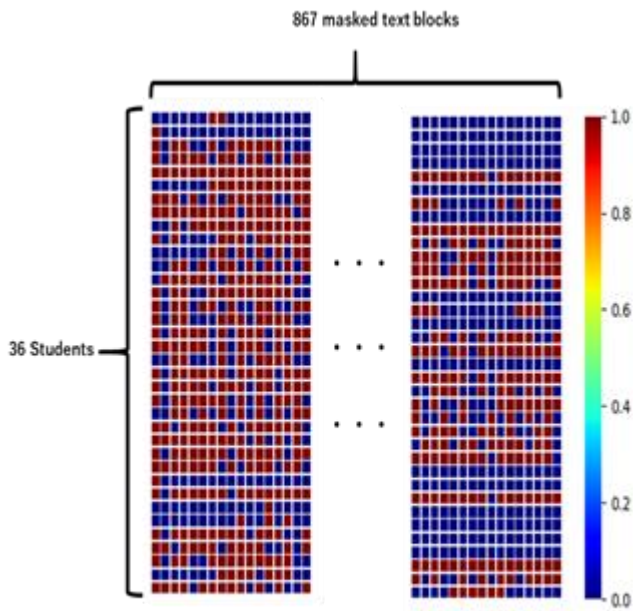


図 5 36×867 の行列 (縦軸：学生，横軸：ブロック番号)

#### 4.4 分析結果

##### 4.4.1 従来のデジタル教科書のログの NMF の結果

図 6 は，従来のデジタル教科書の閲覧ログに基づいて 36 人の学生×5 パターンに分解したマトリックスを示す。図 7 は，50 ページ×5 パターンに分解したマトリックスを示す。これらのマトリックス内の数値をヒートマップ化し，それぞれのパターンの分類を行った。表 5 は，その分類した結果を示す。表 6 は，その行列分解後のそれぞれのパ

ーンに基づいた学生の成績を示す。その平均値と標準偏差は，それぞれのパターンにおける上位 10 名の学生を抜粋し，計算を行った。

表 5 従来のデジタル教科書のログから発見した閲覧パターン

	詳細
Pattern 1	すべてのページを閲覧
Pattern 2	1~36 ページまでを閲覧
Pattern 3	1~30 ページまでを閲覧
Pattern 4	1~19 ページまでを閲覧
Pattern 5	1~7 ページまでを閲覧

表 6 それぞれのパターンに基づいた学生の成績の詳細

	演習		レポート	
	平均	標準偏差	平均	標準偏差
Pattern 1	9.6	0.8	19.5	0.67
Pattern 2	8.6	2.01	18.6	2.91
Pattern 3	8.4	2.49	17.5	3.07
Pattern 4	8.1	3.28	17.7	3.92
Pattern 5	8	3.09	17.6	3.37

##### 4.4.2 SEA のログの NMF の結果

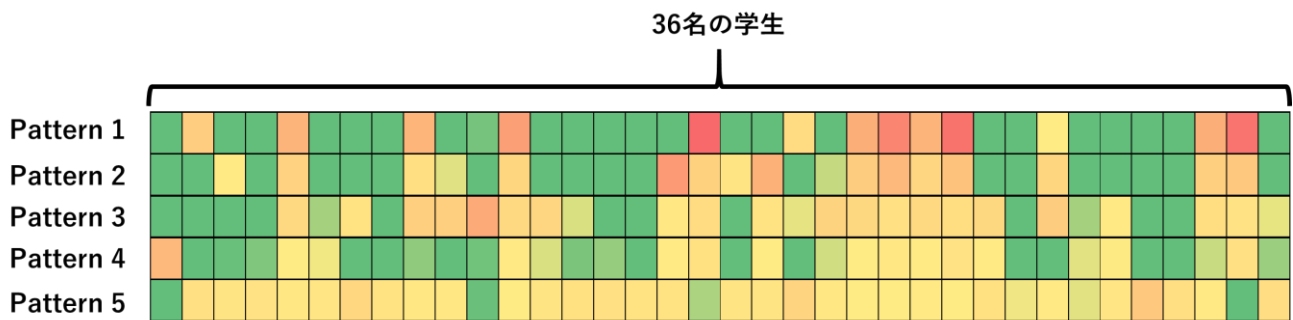


図 6 36 名の学生×5 パターンの行列

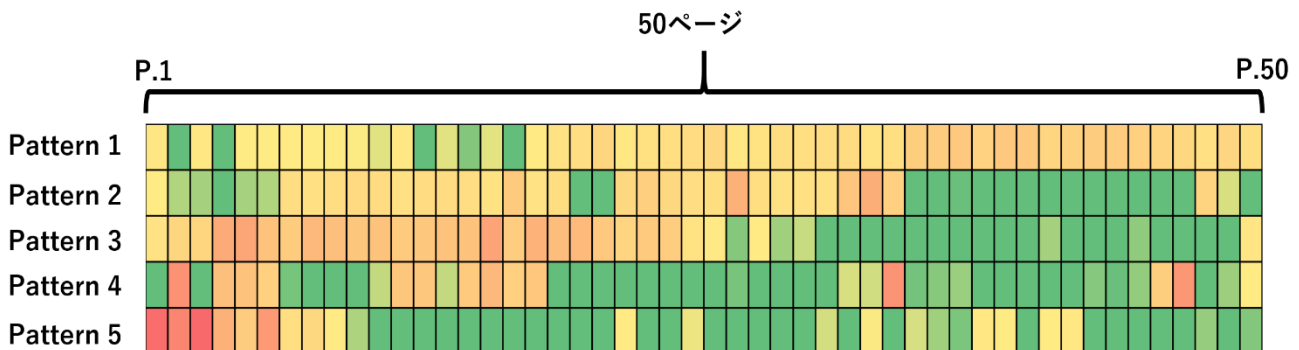


図 7 50 ページ×5 パターンの行列 (縦軸：パターン，横軸：学生)

図 8 は、SEA におけるデジタル教科書の閲覧ログに基づいて 36 人の学生×5 パターンに分解したマトリックスを示す。図 9 は、867 ブロック×5 パターンに分解したマトリックスを示す。同様に、これらのマトリックス内の数値をヒートマップ化し、それぞれのパターンの分類を行った。表 7 は、その分類した結果を示す。

表 8 は、その行列分解後のそれぞれのパターンに基づいた学生の成績を示す。その平均値と標準偏差は、それぞれのパターンにおける上位 10 名の学生を抜粋し、計算を行った。

表 7 SEA のデジタル教科書のログから発見した閲覧パターン

	詳細
Pattern 1	すべてのブロックを閲覧
Pattern 2	全てのページ内の最初から中間あたりまでを閲覧
Pattern 3	1~17 ページまでの全てのブロックを閲覧
Pattern 4	1~19 ページまでの全てのブロックを閲覧
Pattern 5	1~7 ページまでの全てのブロックを閲覧

表 8 それぞれのパターンに基づいた学生の成績の詳細

	演習		レポート	
	平均	標準偏差	平均	標準偏差
Pattern 1	9.2	1.83	18.2	2.89
Pattern 2	7.4	2.01	17.4	3.9
Pattern 3	8	2.49	17.5	4.84
Pattern 4	8	3.28	18.1	1.77
Pattern 5	7.8	3.09	17.6	2.08

## 5. 考察

本論は、従来のデジタル教科書で収集できる閲覧ログと SEA で収集できる閲覧ログのパターンの比較を行った。

表 5 と 6 の結果から、すべてのページを閲覧するパターン 1 が、一部のページを閲覧する他のパターンよりも学生の演習及びレポートの点数が高いことが分かる。しかし、これらの結果だけでは、学習者が単に教科書を開いただけ

36名の学生

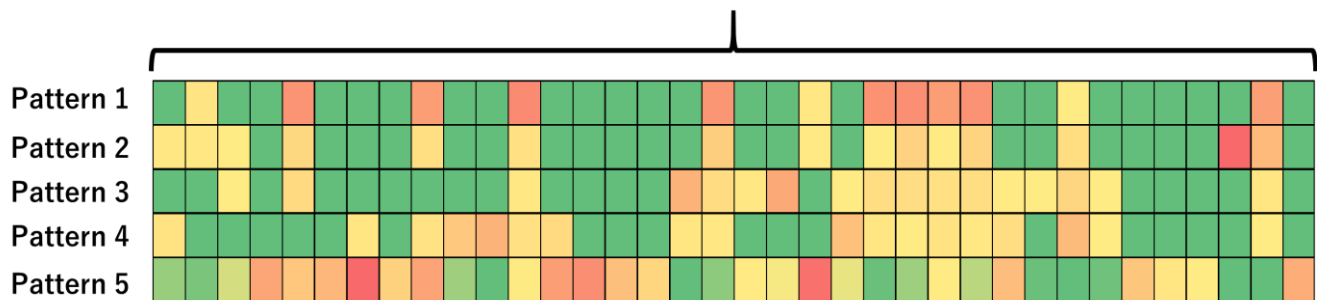


図 8 36名の学生×5パターンの行列

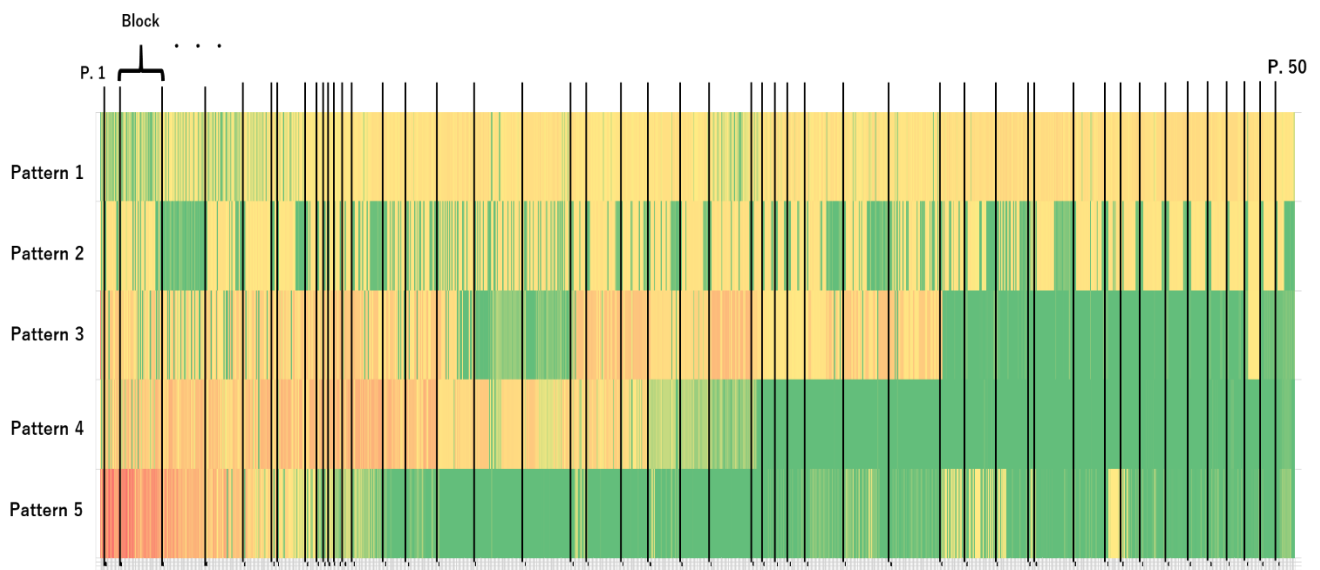


図 9 867 ブロック×5パターンの行列



なのか、それとも適切に閲覧しているのかを見分けることができない。そこで、表7と表8のスクラッチ機能のログの分析と比較を行った。

表7に関して、パターン1は、すべてのブロックを閲覧したことを示す。パターン2は、すべてのページ内のブロックを一部閲覧したことを示す。これらの二つのパターンは、学生がそのページを単に開いただけなのか、それとも適切に閲覧しているかを見分けることができる。この結果に基づいて、表8の学生の演習及びレポートの点数に影響しているかを見ると、パターン1は、演習とレポートの平均点が他のパターンよりも高いことが分かる。一方、パターン2は、すべてのページを閲覧しているが、適切に閲覧していないため、学生の成績がパターン1より低いことが分かる。

これらの結果から、適切に閲覧していない箇所が学生の課題とレポートに影響していることを見つけた。その適切に閲覧していない箇所に関して、手作業でページごとに確認を行った。その結果、適切に閲覧していない箇所に構造体の動的メモリーを確保・開放する内容が含まれていた。その内容を閲覧しなかったため、パターン2の学生らは、演習課題やレポートの内容で、メモリー確保・開放が記述されておらず、パターン1の学生らよりも点数が低くなったことが推察される。

要約すると、本研究は、表10に示すように、新たに2つの閲覧パターンを見つけた。その二つの閲覧パターンは、学生の成績に影響した。

表 10 閲覧パターンの要約

	従来の閲覧パターン	スクラッチ機能の閲覧パターン
すべてのページを閲覧	☑	☑
1~7ページを閲覧	☑	☑
1~19ページを閲覧	☑	☑
1~30ページを閲覧	☑	☑
1~36ページを閲覧	☑	☑
すべてのページ内のブロックを閲覧		☑
すべてのページ内の一部のブロックを閲覧		☑

## 6. おわりに

デジタル教科書の学習ログの分析に関する研究は、学習や指導のやり方を改善するために重要である。本稿では、従来のデジタル教科書では、収集することができなかった「教科書を単に開いているだけなのか」それとも「適切に

読んでいるのか」を見分けることができる新たなデータ収集の方法を提案した。その方法で収集したログから NMF を使って、学習者の閲覧パターンを発見した。教科書を適切に閲覧していない学生は、演習やレポートの点が低くなるのに対して、適切に閲覧している学生は、高くなる。予習の段階で双方の閲覧パターンの学生を見分けることができれば、学生ごとに閲覧していない重要な箇所を通知することで、学生の成績向上が期待される。また、今回は閲覧パターンを分析するために、全てにマスク処理を行ったが、重要なキーワードだけにマスク処理をすることで、閲覧しているかどうかを判断できる。今後の研究として、教員がキーワードを隠ぺいする方法と tf-idf 等を使ってシステムがキーワードを隠ぺいする方法では、学習成果にどのような影響があるかを評価する。

**謝辞** 本研究は JSPS 科研費の 17K12947, 16H03078, 16H06304, 18H04125, 18K0282 の助成を受けたものです。

## 参考文献

- 1) MEXT, Japanese Ministry of Education, Culture, Sports, Science and Technology. 2012. The Vision for ICT in Education, [http://www.mext.go.jp/b\\_menu/houdou/23/04/\\_icsFiles/afidfieldfile/2012/08/03/1305484\\_14\\_1.pdf](http://www.mext.go.jp/b_menu/houdou/23/04/_icsFiles/afidfieldfile/2012/08/03/1305484_14_1.pdf).
- 2) George, S. Learning analytics and educational data mining: towards communication and collaboration, *International Conference in Learning Analytics and Knowledge*, 2012, p.252-254.
- 3) Shimada, A., Mouri K. and Ogata, H. Real-time Learning analytics of e-Book Operation Logs for On-site Lecture Support, *International Conference on Advanced Learning Technology*, 2017, p.274-275.
- 4) Weyers, B., Nowke, C. Mouri, K., and Ogata, H. Web-based Interactive and Visual Data Analytics for Ubiquitous Learning Analytics, *Workshop on Learning analytics Across Physical and Digital Spaces*, 2016, p.65-69.
- 5) Ogata, H. and Mouri, K. Connecting dots for ubiquitous learning analytics, *International Conference on Hybrid Learning and Continuing Education*, 2015, p.46-56.
- 6) Mouri, K. and Ogata, H. Ubiquitous learning analytics in the real-world language learning, *Smart Learning Environment, Vol.2, No.1*, 2015, p.1-15.
- 7) Mouri, K., Okubo, F., Shimada, A. and Ogata, H. Bayesian network for predicting students' final grade using e-book logs in university education, *International Conference on Advanced Learning Technologies*, 2016, p.85-89.
- 8) Mouri, K., Shimada, A., Yin, C. and Kaneko, K. Discovering Hidden Browsing Patterns Using Non-Negative Matrix Factorization, *International Conference on Educational Data Mining*, 2018, in press.
- 9) Mouri, k., Uosaki, N., Shimada, A., Yin, C., Kaneko, K., Ogata, H. Redesign of a data collection in digital textbook systems, *International Conference on Learning Technologies and Learning Environments*, 2018, in press.
- 10) Desmarais, M. Conditions for effectively deriving a q matrix from data with non-negative matrix factorization, *International Conference on Educational Data Mining*, 2011, p.41- 50.