Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education.
Asia-Pacific Society for Computers in Education

Topic-Based Representation of Learning Activities for New Learning Pattern Analytics

Jinghao WANG[a*], Tsubasa MINEMATSU[a], Yuta TANIGUCHI[b], Fumiya OKUBO[a] & Atsushi SHIMADA [a*]
[a]*Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan*
[b]*Research Institute for Information Technology, Kyushu University, Japan*
*wangjiha029@gmail.com, atsushi@limu.ait.kyushu-u.ac.jp

Abstract: In recent years, several kinds of e-learning systems, such as e-book and Learning Management System (LMS) have been widely used in the field of education. When students access these systems, their activities on the systems will be continuously and automatically recorded and stored as learning logs. As the learning logs are stored in association with students and indicate students' learning activities, most studies have been "student-based" learning log analyses focused on students and each student's learning behavior. However, the "student-based" learning log analysis focuses on each student's learning behavior during the entire lesson (for example, studied well or didn't study enough) and cannot show what they learned. Therefore, if there is a need to investigate students' learning behavior regarding each topic of the lesson, such as which topic is learned well and which not in order to optimize the syllabus, we cannot conduct "student-based" learning log analysis directly. Instead of "student-based" learning log analyses, this study describes a method of "learning-topic-based" learning log analysis. We will show how to convert a learning log associated with students into a learning-topic-associated one and shape the logs into a two-dimensional matrix of learning topics and learning activities. Then we apply Non-negative Matrix Factorization (NMF) to the matrix in order to extract the learning patterns by activity. In addition, we make a three-dimensional matrix (tensor) of students, learning topics, and learning activities by subdividing the learning activities of each learning topic by students. We then apply Non-negative Tensor Factorization (NTF) to the tensor to extract detailed learning patterns. The methods proposed in this study will help teachers to have a comprehensively view of students' learning behaviors towards each learning topic easily even if the learning log is in a large-scale, so teachers can adjust syllabus according to the attracted learning behaviors, which is helpful to increase learning efficiency.

Keywords: Educational big data, e-book, learning analytics, Non-negative Matrix Factorization (NMF), Non-negative Tensor Factorization (NTF)

1. Introduction

In recent years, in the field of education, with the introduction and practical application of various kinds of e-learning systems such as e-books, "learning logs" that show what kinds of learning activities students completed are automatically recorded and have become easy to use. Against this background, by analyzing the learning logs accumulated and feeding back the analysis results to teachers and students, it is possible to promote improvements in lesson plans and learning methods, increase learning efficiency, and reduce the burden on teachers and students. In the conventional learning environment, grasping and improving learning conditions is done through limited data such as exam results and teacher's daily class observation. However, by using the abundant learning logs on e-learning systems, we can expect to improve teaching and learning methods scientifically. The fields of application of learning logs are wide-ranging, including recommendation of teaching contents, prediction of grades and early detection of students with poor grades.

The different e-learning systems include Learning Management Systems (LMSs), e-book systems and learning diary collection system. The learning activities of each student are recorded in the database of each system. By making a request to the databases, we can access the accumulated learning logs. As learning activities in each system are conducted by students, the learning logs acquired in each database are accumulated in a form linked to each student. For example, from an e-book system, we can obtain learning logs such as "student A read page 10 of e-book for 100 seconds." Due to this characteristic of learning logs storage, most studies have been based on students ("student-based") and focus on each student's learning activities.

As a result of the "student-based" learning log analysis of the existing studies, there have been studies examining the patterns of the learning conditions of students that revealed the relevance between grades and learning patterns[1]. However, the "student-based" learning log analysis focuses on the learning activities of each student in the entire lesson, through which it is not possible to identify what they have learned. For that reason, in case of educational support from teacher's perspective, such as when teachers want to investigate what kinds of lesson content are well learned or review the teaching plan, a "student-based" learning log analysis is not useful.

Therefore, in this study, we seek to analyzing learning logs in the "learning-topic-base" of learning activities conducted by "learning topics." Learning topics are keywords that appear in the lesson. For example, in the subject of "Digital Signal Processing," keywords like "Fourier Transform," or "Discrete Signal" can be chosen as learning topics. In existing studies, "learning topic-based" learning log analysis is not seen.

There are two reasons why there has been almost no "learning-topic-based" learning log analysis. First, it is difficult to extract learning topics from teaching contents. Learning topics are not explicitly defined in the teaching contents, and no existing study has proposed a method for extracting learning topics. Second, it is difficult to acquire and organize the learning logs of each learning topic. As mentioned above, learning logs are stored in association with each student in the databases of e-learning systems. The subject of existing studies was how to convert or reshape the "student-based" learning logs into "learning topic-based" ones. We will propose a method of extracting learning topics and acquiring and merging "learning topic-based" learning logs in Section 3.

We apply Non-negative Matrix Factorization (NMF) and Non-negative Tensor Factorization (NTF) to analyze learning logs in this study. NMF and NTF have been proposed as effective analytical methods for extracting data characteristics from two-dimensional matrices and three-dimensional or higher-order tensors, respectively. The two-dimensional matrix of learning topics and learning activities can be formatted by reshaping learning logs that have been converted to "learning-topic-based" ones. Depending on the lessons, the number of types of learning topics or learning activities can reach the tens or even hundreds. Therefore, in this study, we try to reduce the ranks of the matrices of learning logs by applying NMF to a matrix to make it easier to extract learning patterns. Furthermore, by extending the above matrix method to a three-dimensional tensor of students, learning topics and learning activities and applying NTF to the tensor, we can extract and interpret learning patterns considering students, learning topics and learning activities at the same time.

This study established the following three points:

- We propose a method to acquire and reshape learning logs into "learning-topic-based."
- We propose a method to format a three-dimensional tensor of students, learning topics and learning activities.
- We apply NMF and NTF to the learning log matrices and tensors and extract and interpret learning patterns.

2. Related Studies

2.1 Studies of learning log analysis

Jo et al.[2] analyzed the length, frequency, and interval of log-in time to investigate how working learners develop time management strategies in online class. Based on the results of the analysis, the authors reported that learners with regular login intervals show higher learning performance. Gitinabard et al.[3] used students' access log to e-books and learning forum to predict dropout. Prediction results showed that learners at risk of potentially dropping out could be identified early. These studies are all learning log analysis in "student-base", and there are no attempts of "learning-topic-based" learning log analysis.

2.2 Studies applying NMF and NTF

Hasumoto et al.[4] extracted latent customer purchasing patterns by applying NMF to store purchasing data. Based on the extracted results, the authors clarified what kind of purchasing activities have a strong influence on "Customer Lifetime Value," which reflects future corporate profits, and proposed a new way to formulate corporate marketing strategies.

NTF, a multidimensional extension of NMF, analyses data that cannot be expressed in two-dimensional matrix. Kuwano et al.[5] reshaped search logs from bus route and timetable search site into a four-dimensional tensor of place of departure, destination, day of the week, and time of the day. By applying NTF to the tensor, the authors extracted patterns of bus use characteristics considering all the four factors above.

As described above, NMF and NTF have been applied in a wide range of fields for pattern extraction. However, few studies have attempted to apply them in the field of Learning Analytics.

3. Extraction of learning logs based on keywords

Since we choose NMF and NTF as the analysis methods of this study, we reshape learning logs into a two-dimensional matrix and a three-dimensional matrix (tensor).

Subsection 3.1 will give a brief introduction of e-learning systems from which the learning logs are obtained, and details of learning logs. As mentioned in Section 1, original learning logs are stored in association with students, so we need to convert them into learning topic-shaped ones. Subsection 3.2 will show how to extract learning topics, convert learning logs into "learning topic-based" ones and shape them into a two-dimensional matrix. Subsection 3.3 will show how to extend the matrix into a three-dimensional tensor.

3.1 Introduction of e-learning systems and learning logs

In this study, we will obtain learning logs from three e-learning systems: an e-book system, Learning article management system and Reflection management system.

3.1.1 e-book system

Students can view teaching contents everywhere using computer or smartphone via the e-book system. We collect learning logs from e-books as follows:

- The content name and page number that the student has read;
- the reading time of each page; and
- the history of learning support feature use. *

*As shown in Table 1, students can use following learning support features.

Table 1. *Learning support functions.*

Activity name	Description
Add marker	to highlight text
Add bookmark	to make a bookmark of contents on current page
Add memo	to take a memo of contents on current page
Getit	shows the student understood contents on current page
Notgetit	shows the student didn't understand contents on current page

3.1.2 Learning article management system

A learning article is an article that contains knowledge and learning methods generated by the student during the process of learning. Learning articles are written by learning topics and one article corresponds to one learning topic or multiple related learning topics. Students post the key points, knowledge, and learning methods used for understanding the learning topics in one Learning article using sentences and figures. The Learning management system is a system that allows sharing of Learning articles among students, and provides functions to post, view, and "like" articles. We collect learning logs from the Learning article management system as follows:

- The number of articles posted by each student;
- the number of characters and figures in each post;
- the number of views and likes of each post; and
- the articles posted (text data)

3.1.3 Reflection management system

A Reflection management system supports a learning diary that describes the student's reflections on what they learned in the lesson that day. A Reflection diary includes what the student understood and failed to understand, impressions of the lesson, and questions to the teacher, which are subjectively described from the student's point of view. We collect learning logs from the Reflection management system as follows:

- The number of articles posted by each student;
- the number of characters in each post;
- the articles posted (text data)

3.2 Converting "student-based" learning logs to "learning-topic-based" logs

As shown in Table 2, a learning log shows when, where, and by whom the learning activity was conducted. By combining learning logs obtained directly from e-learning systems, we can make a timeline of each student's learning activity and extract his learning behavior. However, with original learning logs obtained from e-learning systems, it is impossible to extract specific information about what kinds of learning topics the students learned. As a result, we need to reshape the original "student-based" learning logs into "learning-topic-based" ones.

Table 2. *An Example of learning logs obtained from the e-book.*

Student ID	Time	Learning content name	Page number	Activity
001	4:02:13 pm ,1 Feb 2021	2. Fourier Series	4	Add Marker
002	3:22:12 pm ,1 Feb 2021	2. Fourier Series	5	Add Memo
003	6:53:45 pm ,1 Feb 2021	2. Fourier Series	7	Notgetit

First, as learning topics are not defined in advance, it is necessary to define several learning topics. There are three steps to define learning topics.

1. *Extract nouns from the text of e-book using the morphological analysis tool "MeCab."*
2. *Remove alphanumeric characters and words that contain only one Japanese character (as one-character words usually don't make sense).*
3. *Remove words that have little relevance to course content.*

After extracting the learning topics, we will merge the learning logs of each learning topic. There are three steps to merge learning logs.

1. Sum the total amount of learning activities of each page of e-book. For example, as shown in Table 3, page 5 of the e-book contains learning activities conducted by student 001, 002, and 003. Then, the total amount of learning activities conducted on page5 will be shown in the 4th row of Table 3.
2. Make a text search as to whether each learning topic is included as a substring in the e-book text (check whether the learning topic appears as a substring on each page of the e-book) and create a list P for each learning topic of the pages where the learning topic appears. For example, assume the learning topic "discrete time signal" appeared on the 3rd and 5th pages of the e-book. Then P(discrete time signal)={3,5}.

3. Merge the learning logs of the pages included in list P of each learning topic.

Table 3. *Learning activities on page 5 of the e-book.*

Student id	Number of markers	Number of memos	...
001	14	5	...
002	3	2	...
003	8	4	...
Total	25	11	...

After these steps, we obtain a two-dimensional matrix of learning topics and each topic's learning activity as in Table 4.

Table 4. *Two-dimensional matrix of learning topics and learning activities.*

Learning Topic	(e-book) Number of marker	(e-book) Number of memo	...	(Learning article) Number of graphs
Periodic Signal	10	8	...	1
Aperiodic Signal	6	5	...	0
...	...	...	...	...

3.3 Tensor of students, learning topics and learning activities

The "learning-topic-based" learning log matrix introduced in Subsection 3.2 is a two-dimensional matrix of learning topic and learning activity. By applying NMF to this, it is possible to extract the learning behavior and grasp the learning condition of each learning topic. However, the two-dimensional matrix contains the learning activities of each learning topic conducted by all the students and the activities are not divided by student, so it is difficult to analyze it while considering students, learning topics, and learning activities at the same time, making it difficult to fully grasp the learning pattern with the two-dimensional matrix.

Therefore, in this study, we aim to grasp more detailed learning patterns by constructing a three-dimensional tensor of student, learning topic and learning activity and applying NTF to the tensor. Formatting into a three-dimensional tensor can be realized by dividing the learning activity of each learning topic for each student using the two-dimensional matrix in Subsection 3.2. For example, as shown in Figure 1, assume students A, B, and C have drawn 3, 5, and 2 markers for the learning topic "periodic signal," respectively. In the matrix, the learning topic is on the x-axis and the learning activity is on the y-axis. A three-dimensional tensor can be generated by dividing the 10 markers related to "periodic signals" for each student as 3 for A, 5 for B, and 2 for C and reflecting them in the z-axis.

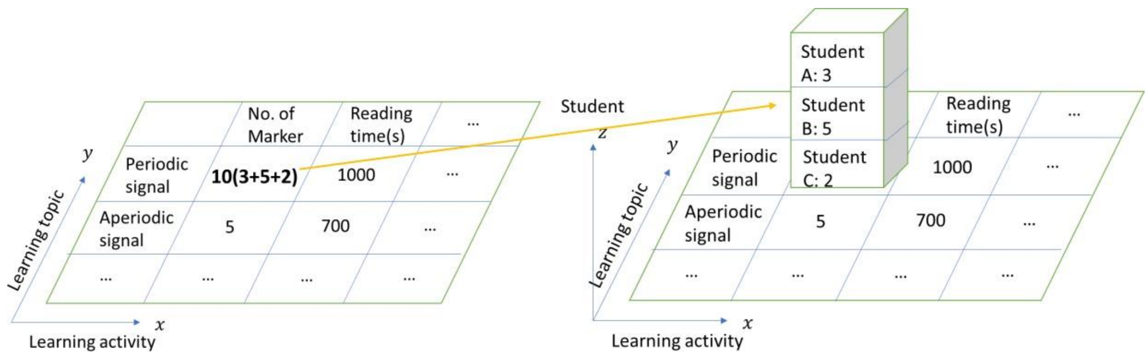

Figure 1. Formatting of the tensor.

4. Non-negative Matrix Factorization (NMF), Non-negative Tensor Factorization (NTF)

This Section outlines NMF and NTF, which are methods for analyzing the matrices and tensors of the learning logs created in Section 3. Subsection 4.1 introduces NMF, and Subsection 4.2 introduces NTF.

4.1 Overview of NMF

NMF[6] is one of the methods usually used for feature extraction from two-dimensional matrices. As mentioned in Section 2, NMF can be applied to any data that can be represented by a non-negative two-dimensional matrix and has been used in many studies in recent years. Since the learning log matrix created in Section 3 of this study is also a two-dimensional matrix composed of non-negative value data, NMF can be applied. NMF analysis yields several frequent patterns that represent the characteristics of the data. We will show the detailed results of the analysis applying NMF to the learning logs in Section 5.

A conceptual diagram of NMF is shown in Figure 2. The size of the original matrix X (the observation matrix, which means we will observe patterns from this matrix) is $m * n$. The result of applying NMF to X is to decompose it into the product of the matrix $T$of size $m * r$ and the matrix V of size $r * n$. Here, r is a parameter set in the initial stage of NMF called the "factor." By setting the factor to a value smaller than m and n, X can be decomposed into low-ranked matrices T and V, making it easier to observe patterns in the data. T and V represent the features of the vertical and horizontal axes of X, respectively, and are called feature matrices.

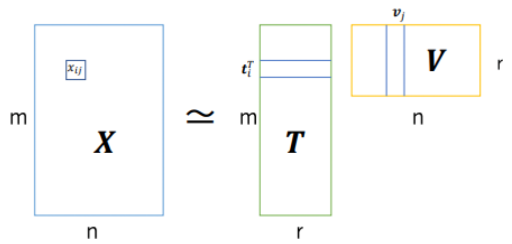

Figure 2. Conceptual diagram of NMF.

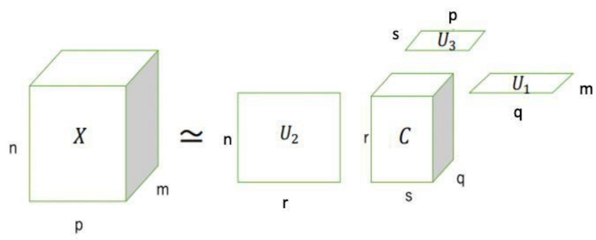

Figure 3. Conceptual diagram of NTF.

4.2 Overview of NTF

NTF[7] is an extension to higher dimensional matrices of NMF, which can only be applied to 2D matrices. A matrix of three or more dimensions is also called a tensor, and NTF is a method of extracting data characteristics from a tensor. NTF is formulated as $X \cong C \times U_1 \times U_2 \times U_3$. *A conceptual diagram of NTF is shown in Figure 3. If the size of the original tensor X (observation tensor) is* $m * n * p$, *we can set* q, r, s *smaller than* m, n, p *and decompose* X *into the feature matrices* U_1 $(m * q)$, U_2 $(n * r)$, U_3 $(p * s)$ *and a tensor* C *of size* $q * r * s$, *which is called a core tensor. The characteristics of the data can be grasped from the feature matrices and the core tensor. Specifically, data characteristics of three axes of* X *can be represented by* U_1, U_2 *and* U_3 *respectively. For example, in this study, X is a tensor of student, learning topic and learning activity, so* U_1, U_2 *and* U_3 *are feature matrices of students, learning topics and learning activities, respectively, obtained by NTF.*

The core tensor can be regarded as a compressed observation tensor, and the larger the value of an element of the core tensor, the more the combination of factors corresponding to that element appears. Specific examples will be described in detail in Section 5.

5. Experimental results of NMF and NTF

In this section, we applied NMF and NTF to the two-dimensional matrix formatted in Subsection 3.2 and the three-dimensional tensor formatted in Subsection 3.3, with consideration to the results of attempting to interpret learning patterns. Subsection 5.1 shows the results and considerations of *applying NMF to the matrix of learning topics and learning activities and Subsection 5.2 shows the results and considerations of applying NTF to the tensor of students, learning topics, and learning activities.*

In this study, we analyzed the learning logs acquired in the lesson "Digital Signal Processing" taken by 82 third-year students in the Department of Electrical and Information Engineering in the Faculty of Engineering of our university in the spring semester of 2020.

We extracted 47 types of learning topics and obtained learning logs of 12 kinds of learning activities consisting of the following items.

- e-book: number of markers, number of memos, number of bookmarks, number of Getits, number of Notgetits, reading time
- Reflection: number of posts, number of characters of each post
- Learning articles: number of posts, number of characters of each post, number of figures of each post, number of likes of each post

The data widths of different types of learning activities are not the same, as the number of markers ranges from several to several tens, while the reading time ranges from several tens to several thousand seconds. Therefore, we performed normalization of dividing each type of learning logs by the maximum value of that type so that the values for all kinds of learning logs fell within the range 0 - 1.

5.1 Experimental results of NMF

As a result of reshaping in Subsection 3.2, we obtained a matrix of learning topics (47 types) and learning activities (12 types) as Table 5.

Table 5. *Finally reshaped matrix of learning topics and learning activities.*

Learning topic	(e-book) Number of marker	(e-book) Number of memo	...	(Learning article) Number of likes
Periodic Signal	0.35396	0.33333	...	0.32955
Aperiodic Signal	0.05244	0.03333	...	0.14015
...	...	...	...	...
Window Function	0.05191	0.06667	...	0.12121

*We applied NMF with the number of factors set to 4, generating a feature matrix of learning activities (12*4) and a feature matrix of learning topics (47*4) as shown in Tables 6 and 7, respectively. We set the number of factors considering 2 points: (1) the error of NMF needs to be reduced. We conducted matrix multiplication to the two feature matrices to reproduce a matrix as the same size of that in Table 5 (47*12). Then we calculated the error between the two 47*12 matrices. We found that as the number of factors get larger, the error between two matrices reduced. (2) the ease of interpretation. Although setting the number of factors bigger helps to reduce error, learning behaviors can't be summarized efficiently and it's difficult to interpret these factors. So, we made a tradeoff decision between (1) and (2) in our experiments utilizing NMF and NTF.*

Table 6. *Feature Matrix of learning activity. B: e-book, R: Reflection, L: Learning article, F: Factor*

Learning Activity	F 1	F 2	F 3	F 4
(B) Marker	0.001	0.065	0.124	0.544
(B) Memo	0.026	0.000	0.000	0.051
(B) Bookmark	0.000	0.002	0.000	0.283
(B) Getit	0.010	0.003	0.417	0.000
(B) Notgetit	0.119	0.000	0.003	0.021
(B) Reading time	0.253	0.004	0.041	0.214

Learning Activity	F 1	F 2	F 3	F 4
(R) Number of articles	0.313	0.323	0.002	0.001
(R) Number of characters	0.144	0.417	0.000	0.000
(L) Number of articles	0.185	0.003	0.015	0.112
(L) Number of characters	0.227	0.000	0.001	0.127
(L) Number of graphs	0.040	0.000	0.018	0.005
(L) Number of likes	0.054	0.004	0.053	0.000

According to Table 6, Factor1 is a factor specific to the learning activities of Reflection and Learning articles. Factors 2,3,4 reflect the learning condition in e-book as Factor2 is specialized for Marker and Bookmark, Factor3 is specialized for Memo, and Factor4 is specialized for Getit. Based on these considerations, the following can be inferred from the feature matrix of learning topics (Table 7).

Table 7. *Feature Matrix of learning topic. F: Factor*

Learning topic	F 1	F 2	F 3	F 4
Periodic signal	0.246	0.207	0.214	0.374
Aperiodic signal	0.176	0.010	0.018	0.019
Fourier series	0.410	0.385	0.221	0.003
Fourier transform	1.309	0.294	0.318	0.610
Continuous time signal	0.167	0.583	0.121	0.212
Sampling	0.544	0.337	0.039	0.348
Discrete-time signal	0.260	0.722	0.229	0.387
Z-transform	0.791	0.122	0.091	0.410
Inverse Z-transform	0.254	0.013	0.156	0.067
Discrete time system	0.308	0.046	0.134	0.211
Linear time-invariant system	0.163	0.106	0.206	0.418
Difference equation	0.155	0.038	0.151	0.304
Frequency characteristic	0.249	0.069	0.018	0.151
Discrete Fourier transform	0.376	0.092	0.000	0.038
Fast Fourier transform	0.308	0.002	0.000	0.076
Digital filter	0.097	0.006	0.000	0.000
Impulse response	0.283	0.029	0.123	0.575
System function	0.083	0.032	0.000	0.062
Analog signal	0.143	0.663	0.189	0.080
Sampling cycle	0.008	0.170	0.015	0.286
Quantization	0.113	0.151	0.023	0.130
Image recognition	0.000	0.000	0.000	0.025
DC component	0.000	0.230	0.080	0.000
Euler’s formula	0.017	0.075	0.045	0.087
Learning topic	F 1	F 2	F 3	F 4
Fourier coefficient	0.000	0.365	0.162	0.166
Amplitude spectrum	0.000	0.000	0.063	0.143
Phase spectrum	0.000	0.000	0.063	0.143
Impulse function	0.084	0.011	0.141	0.127
Step function	0.000	0.009	0.043	0.046
Aliasing	0.081	0.017	0.000	0.013
Nyquist frequency	0.154	0.064	0.000	0.064
Laplace transform	0.157	0.061	0.000	0.013
Transfer function	0.306	0.082	0.054	0.279
Adder	0.049	0.042	0.003	0.069
Delayer	0.043	0.039	0.025	0.111
Phase response	0.090	0.009	0.000	0.091
Amplitude response	0.096	0.012	0.000	0.088
Discrete-time Fourier transform	0.249	0.133	0.002	0.043
Convolution	0.136	0.044	0.131	0.149
Conjugated	0.046	0.009	0.000	0.015
Conversion pair	0.006	0.146	0.179	0.234
Symmetry	0.006	0.036	0.544	0.000
Linearity	0.125	0.098	0.629	0.032
Complex	0.342	0.200	0.110	0.638
Filter	0.358	0.127	0.034	0.255
Linear phase	0.042	0.027	0.019	0.090
Window function	0.066	0.000	0.045	0.050

- Due to the high values of Factor 1 of “Fourier Transform” and “Z-Transform,” many Reflection and Learning articles on these two topics were posted.
- Since Factor 2 of “discrete time signal,” “analog signal” and “continuous time signal” were high, many markers and bookmarks were recorded on the pages of the e-book containing these learning topics.
- Since Factor 3 was high for “Symmetry” and “Linearity,” many memos were written on the pages of the e-book containing these learning topics.
- Since Factor 4 was high for “Complex,” “Fourier Transform,” and “Impulse Response,” many students pressed “Getit” button on these topics, which means these topics were understood well.

It can be assumed that the learning topics such as “Fourier Transform,” for which the values of all four Factors were at a high level, were well understood as a result of the students studying them well. On the other hand, learning topics such as “Fourier Series,” of which the values of other factors were high but only Factor 4, specialized in “Getit,” was low, had many learning activities but students’ level of understanding was low. This suggests that the teacher's explanations of these learning topics were imperfect. However, learning topics with low values of all four Factors, such as “Aliasing,” might have been considered unimportant by teacher originally, so students hardly learned about it. Even though the value of Factor 4 was low, it is not necessary to enhance the explanation of these learning topics.

Table 8. *Difference of Factor 1-4.*

	Factor 1	Factor 2	Factor 3	Factor 4
Topics of the first half	0.198	0.179	0.137	0.172
Topics of the second half	0.177	0.047	0.043	0.154

In addition, there are different tendencies in the learning condition of the learning topics that appeared in the first half and the second half of the lesson. Specifically, as shown in Table 8, we plotted the average values of Factors 1 to 4 of learning topics of the first half and the second half of the lesson. From Table 8 it can be observed that the learning activities of the learning topics that appeared in the first half of the lesson were more active than in the second half. As an interpretation of these learning patterns, in the first half of the lesson, students were encouraged by teacher to actively post and draw markers, etc., but in the second half, students became bored or accustomed to the lesson, and their learning activities decreased due to the loss of motivation.

Feeding back the above results to the teacher may facilitate teachers' adjusting and improving the lesson content and lesson plan.

5.2 Experimental results of NTF

Although we interpreted several learning patterns using the two-dimensional matrix of learning topics and learning activities, we could not consider students, learning topics, and learning activities at the same time. Therefore, in this section, we applied NTF to the three-dimensional tensor in Subsection 3.3 and interpreted the learning conditions of the three axes of the tensor.

As a result of formatting, tensor X of students (82 people), learning topics (47 types), and learning activities (12 types) was created. By taking the strategy of balancing the error reduction and ease of interpretation as mentioned in Subsection 5.1, the size of the core tensor was set as 4 factors for learning activities, 4 factors for students and 3 factors for learning topics. NTF yielded feature matrices of learning activities, students, and learning topics of sizes 12*4 (Table 9), 82*4 and 47*4, respectively.

In the feature matrix of learning activities (Table 9),

Table 9. *Feature Matrix of learning activities. B: e-book, R: Reflection, L: Learning article, F: Factor*

Learning Activity	F 1	F 2	F 3	F 4
(B) Marker	0.001	0.065	0.124	0.544
(B) Memo	0.026	0.000	0.000	0.051
(B) Bookmark	0.000	0.002	0.000	0.283
(B) Getit	0.010	0.003	0.417	0.000
(B) Notgetit	0.119	0.000	0.003	0.021
(B) Reading time	0.253	0.004	0.041	0.214

Learning Activity	F 1	F 2	F 3	F 4
(R) Number of articles	0.313	0.323	0.002	0.001
(R) Number of characters	0.144	0.417	0.000	0.000
(L) Number of articles	0.185	0.003	0.015	0.112
(L) Number of characters	0.227	0.000	0.001	0.127
(L) Number of graphs	0.040	0.000	0.018	0.005
(L) Number of likes	0.054	0.004	0.053	0.000

- Factor 1 specializes in learning activities of Reflection management system and Learning article management system and reading time of e-book.
- Factor 2 specializes in leaning activities of Reflection.
- Factor 3 specializes in times "Getit" was pressed in the e-book.
- Factor 4 specializes in Markers and Bookmarks added in the e-book.

In the feature matrix of students, (The top 5 students were selected for each factor.)

- Factor 1 specializes in learning activities conducted by students with IDs 060, 038, 067, 057, 055.
- Factor 2 specializes in learning activities conducted by students with IDs 051, 041, 050, 049, 079.
- Factor 3 specializes in learning activities conducted by students with IDs 020, 008, 075, 040, 063.
- Factor 4 specializes in learning activities conducted by students with IDs 015, 063, 011, 031, 062.

In the feature matrix of learning topics, (The top 2 learning topics were selected for each factor.)

- Factor 1 specializes in learning activities conducted for "Fourier Transform" and "Z-transform."
- Factor 2 specializes in learning activities conducted to "Analog Signal" and "Sampling."
- Factor 3 specializes in learning activities conducted to "Discrete Time Signal" and "Complex."

Elements of the core tensor show the weights of the combinations of the factors of the feature matrices along the three axes of the original tensor. After extracting the meaning of the factors of each feature matrix as above, we examined the values of every element of the core tensor and revealed what kinds of combinations of learning activities, students, and learning topics appeared most.

In this study, the core tensor is a 4*4*3 three-dimensional tensor. So, there are 48 elements in total. The three-dimensional coordinates represent the factors of learning activity, student, and learning topic, in that order. For example, (1, 2, 3) is a combination of Factor 1 for learning activity, Factor 2 for student and Factor 3 for learning topic. Each element has a weight, and the larger the weight, the more often the combination appeared. We pick the top 2 of the 48 elements and explain their meaning.

- Element (1, 1, 1) Weight=10.971
  This is the most frequently appearing combination, where students with IDs 060, 038, 067, 057, 055 conducted learning activities including writing Reflection, Learning articles, and reading e-books on the learning topics of "Fourier Transform" and "Z-Transform."
- Element (2, 3, 1) Weight=6.477
  The second most frequent combination was students with IDs 020, 008, 075, 040, 063 writing Reflection articles about "Fourier Transform" and "Z-Transform."

In this way, it is found that NTF can realize learning pattern analysis that simultaneously considers the learning condition of three perspectives of students, learning topics and learning activities, As shown in Subsections 5.1 and 5.2, unlike "student-based" learning log analysis, the "learning-topic-based" learning log analysis shows how to investigate each learning topic's learning conditions and provides a new method of extracting learning patterns, which can help teachers to adjust and improve their teaching plans. Furthermore, by formatting the tensor of students, learning topics, and learning activities, it is easy for us to observe detailed learning behaviors.

6. Conclusion

In this paper, we proposed a method of converting original "student-based" learning logs into "learning-topic-based" logs, and then reshaped them into two-dimensional matrices of learning topics and learning activities and three-dimensional tensors of students, learning topics, and learning activities. Last, we applied NMF to the matrix and NTF to the tensor and extracted learning patterns. The results of analysis clarified that the characteristics of the huge amount of learning logs can be extracted and interpreted relatively easily using NMF and NTF.

These proposed methods will contribute to improving the learning efficiency of students and enhancing future lesson plans for teachers.

Future tasks include improving the extraction method of learning topics and the way to decide the number of factors before applying NMF and NTF. In this study, the number of nouns extracted from the "Digital Signal Processing" e-book was about 20000, the number of words remaining after duplicate nouns, alphanumeric characters, and single-character words were automatically removed was about 700. However, we removed those that were not related to the lesson content manually to obtain a final 47 learning topics. Since the learning topic selection criteria include subjective standards, the setting of quantitative criteria should be improved in future studies.

In addition, not only in this study but also in all studies applying NMF and NTF, it is necessary to determine the number of factors considering the ease of interpretation of the characteristics of the data while suppressing the error. However, quantitative criteria for determining the number of factors have not been formulated, which remains a task for future studies.

Acknowledgements

This work was supported by JST AIP Grant Number JPMJCR19U1, JSPS KAKENHI Grand Number JP18H04125 and JP22H00551, Japan.

References

Hiroaki Ogata, Chengjiu Yin, Kousuke Mouri, Misato Oi, Atsushi Shimada, Fumiya Okubo, Masanori Yamada, and Kentaro Kojima. Accumulation and analysis of learning logs for utilization of educational big data. *Journal of Japanese Society for Information and Systems in Education*, 33(2):58–66, 2016.

Il-Hyun Jo, Dongho Kim, and Meehyun Yoon. Analyzing the log patterns of adult learners in lms using learning analytics. *LAK '14: Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, pages 183–187,

Niki Gitinabard, et al. Your actions or your associates? predicting certification and dropout in moocs with behavioral and social features. *EDM2018*, 2018.

Kyosuke Hasumoto, Gendo Kumoi, et al. Predicting customer lifetime value in platform businesses using non-negative matrix factorization. *Journal of Information Processing*, 60(7):1283–1293, 2019.

Masashi Kuwano, Reo Kinoshita, Kei Fukuyama, Keishi Tanimoto, Kazunori Sugahara, and Takao Kawamura. Analysis of bus usage characteristics from search log data using non-negative tensor factorization. *Journal of Japan Society of Civil Engineers D3*, 73(5): I_569–I_578, 2017.

Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

Amnon Shashua et al. non-negative tensor factorization with applications to statistics and computer vision. In *Proceedings of the 22nd international conference on Machine learning*, pages 792–799, 2005.