Visualization of the Relationship Between Lectures and Laboratories Using SSNMF

Kyoka Yamamoto
Graduate School of Informatics
Kansai university
Takatsuki, Japan
k166981@kansai-u.ac.jp

Ryosuke Yamanishi
Faculty of Informatics
Kansai University
Takatsuki, Japan
ryama@kansai-u.ac.jp

Mitsunori Matsushita
Faculty of Informatics
Kansai University
Takatsuki, Japan
mat@res.kutc.kansai-u.ac.jp

***Abstract*—This study aims to visualize the relationship between lectures and fields of specialization (laboratories) so that students can choose lectures with a future direction. Since the university curriculum is highly flexible; students choose their own lectures. Taking into account their own objectives, students select the basic knowledge necessary for their purposes. However, it is difficult for students without sufficient knowledge to understand their relevance from the syllabus. The purpose of this study is to propose a method for estimating the relevance between lectures and laboratories in an undergraduate school as a help to provide students with an objective analysis of lectures, i.e., not only knowledge but also examples of its use. The proposed method applies a semi-supervised non-negative matrix factorization to identify common factors of knowledge in each combination of lecture and laboratory. It is suggested that the proposed method calculates reasonable results for the relationship between lectures and laboratories.**

***Index Terms*—Syllabus Analysis, Specialization Analysis, Knowledge structuring, Data mining, Semi-Supervised NMF**

I. Introduction

Since the university curriculum is highly flexible, students may choose their own lectures. Taking into account their own objectives to study or build their careers, students learn each basic knowledge necessary for their purposes in lectures. However, it is difficult for students who are novices in specialty without sufficient knowledge to understand the relevance between the expertise and lectures with the syllabus.

Rafael et al. provide a systematic overview of the current state of the art, showing the importance and growth of research using text mining in education(ETM) with the development of the online educational environment [1]. As one of the ETM, the study focused on college course design, primarily on estimating the relationships among lectures. Acosta et al. state that there are two types of content-based filtering and collaborative filtering when analyzing and recommending relationships among content, such as relationships between lectures [2]. Collaborative filtering is a method based on the idea that active users are more likely to prefer items to like-minded users. This method is implemented by calculating a similarity score between the active user and all other users. Content-based filtering assumes that users tend to be interested in items that are similar to those they have viewed and similar to those they have shown interest in in the past, and defines similarities between items. As a method to examine the similarity of items included in the latter content-based filtering,Kitto et al. propose a mapping between subjects while stating the importance of curriculum analysis [3]. They explore the utility of skill-based curriculum analysis and the usability of mapping between subjects' descriptions. A syllabus is often used as a representative description of lectures. Kawintiranon et al. considered that the syllabus was inadequate as an information source because it did not cover all lectures [4]. They accordingly have introduced the course materials used in the lectures into their analysis. It is reasonable to say that their work focused on the knowledge learned by taking lectures. In contrast, this paper begins a lecture analysis focusing on the viewpoint from the students' side. This research aims to provide students with an objective analysis of lectures: not only the knowledge but the usage examples of those. We, in this paper, calculate the relationships between lectures and laboratories as the reference for students with limited knowledge; it is expected that students would consider what kind of specialties are used in which laboratories. It should enhance students to consider their selection of lectures and their own careers in an exploratory manner.

In this paper, as the first step to the goal of this study, we try to visualize the relationships between lectures and fields of specialization (i.e., laboratories): we believe that students are able to choose lectures with their future vision. Research about lectures mainly focuses on the relationship between lectures themselves. Most research does not deal with how to utilize the knowledge and skills learned in lectures for specialized fields, i.e., research activities. A specialized field should consist of a combination of basic knowledge and skills learned in multiple lectures. The relationship between a specialized field and lectures should be equivalent to the relationship between artifacts and components. We consider that representing which combination of lectures constitutes a specialized field would lead to showing the relationship between those. We propose a method for estimating the relevance between lectures and laboratories in an undergraduate school. The proposed method applies semi-supervised non-negative matrix factorization to identify common knowledge factors for each combination of lecture and laboratory. And we study the availability of the proposed method in structuring knowledge.

II. General Concept to Support Students for Selecting Lectures

University has several departments for fields of research, such as engineering, psychology, and computer science. Some universities have a more flexible department where students can have lectures on multiple research fields: the students do not have to fix their own specialty when they enter the university. During the study, the students try to find their own specialties and determine the supervisor for their graduate research (i.e., laboratory) through their studying in lectures. They have to know which lecture is related to which laboratory; however, it is so hard for students who are unfamiliar with each research field.

The problem of "which lectures are related to which laboratories" should be not a classifying problem but a representation of the relationships between lectures and laboratories. Students have to consider not one versus one relation but the multi versus multi when they take lectures and choose the laboratory. The candidates of the laboratory where they offer to join should differ based on the students' background, i.e., what kinds of lectures they have taken and what kinds of fields they have been interested in. The set of laboratories in the department might change because of the professors' transfer. So, the relationships between lectures and laboratories are not fixed but fluid. Accordingly, we propose a method capable of analyzing the multi versus multi relationships.

III. The Proposed Method

The basic idea of factorization is expressed as follows.

$$\text{set of observed variables} = \text{common factors} \times \text{unique factors}.$$

Features that are common to different observed variables are defined as common factors. Unique factors are defined as factors that can not be explained by common factors. The proposed method uses Semi-Supervised NMF (SSNMF) [5], which is an extension of Non-negative Matrix Factorization (NMF) [6]. Using SSNMF, the template pattern of the set of analysis results (e.g., frequency of sound sources and knowledge) can be given as a common factor for the set of observed variables, and each analysis result can acquire activated factors. We can express the activation of each observed variable for each analysis result based on the relative relationship between the objects included in the set of analysis targets. The task of this paper is to reveal the knowledge handled in lectures. By using the lectures as a template pattern, the proposed method figures out the laboratory that is relatively related to the lecture.

A. NMF

Factorization in NMF is defined by the equation(1).

$$y_{ij} \approx y'_{ij} = \sum_{k=1}^{K} h_{ik} u_{kj}, \tag{1}$$

where, $y \in \boldsymbol{Y}$,$h \in \boldsymbol{H}$, and $u \in \boldsymbol{U}$ each represent the observed variables matrix, the basis matrix, and each element of the activation tendency matrix, respectively. i and j denote the indices of the elements in the matrix. K denotes the number of bases; the dimensions of the underlying matrix $\boldsymbol{H}$. The given $\boldsymbol{Y}$ is approximated by $y'_{i,j}$:the product of $\boldsymbol{H}$ and $\boldsymbol{U}$. Observation matrix $\boldsymbol{Y}$ is decomposed into the basis matrix $\boldsymbol{H}$ and the activation tendency matrix $\boldsymbol{U}$. Here, y, h and u each represents the elements of $\boldsymbol{Y}$, $\boldsymbol{H}$, and $\boldsymbol{U}$, respectively. The i and j denote the indices of the elements in the matrix. The K denotes the base number.

Generally, NMF minimizes the errors between the matrix $\boldsymbol{Y}$ and $\boldsymbol{HU}$ obtained by NMF. In the NMF, the initial values of $\boldsymbol{H}$ and $\boldsymbol{U}$ are randomly assigned. The matrices $\boldsymbol{H}$ and $\boldsymbol{U}$ can be updated by minimizing the errors in $\boldsymbol{Y}'$, which is the product of the matrices $\boldsymbol{Y}$ and $\boldsymbol{HU}$, respectively. The matrix $\boldsymbol{H}$ is the set of feature patterns that can represent the observed variable matrix $\boldsymbol{Y}$ with a pre-specified number of bases. The matrix $\boldsymbol{U}$ represents the activity trend of the feature pattern $h \in \boldsymbol{H}$ in the observed variable matrix $\boldsymbol{Y}$. As the error is sufficiently minimized, it should be assumed that the set of feature patterns and their activation tendencies can be computationally obtained by multiplying the two matrices. There are several types of error functions used in minimizing the error of $\boldsymbol{Y}'$ and $\boldsymbol{HU}$. In this paper, we use the squared Euclid distance D_{Euclid} expressed in the following equation;

$$D_{Euclid}|(\boldsymbol{Y}, \boldsymbol{HU})| = ||\boldsymbol{Y} - \boldsymbol{HU}||^2. \tag{2}$$

B. SSNMF

In NMF, the initial values of the basis matrix and the activation tendency matrix are generated with random values. Both basis and activation tendency matrices are dynamically updated so that the product of the two matrices approximates the observed matrix. In SSNMF, the template pattern matrix is prepared in advance as a basis matrix. SSNMF updates only the activation tendency matrix by error minimization.

Many studies have applied SSNMF to sound source separation [7], [8]. In these studies, activation tendencies of each sound source are obtained by providing the spectral structure of the sound source as a template pattern. Then, automatic music scoring from the sound source is realized by SSNMF. SSNMF is defined by the following equation.

$$\boldsymbol{Y} \approx \boldsymbol{HU} + \boldsymbol{FG}, \tag{3}$$

where, each $\boldsymbol{H}$ and $\boldsymbol{U}$ denotes the template pattern vector and its activation tendency matrix, respectively. And, $\boldsymbol{FG}$ denotes the noise term. Then, $\boldsymbol{Y}$, $\boldsymbol{H}$, $\boldsymbol{U}$, $\boldsymbol{F}$, and $\boldsymbol{G}$ represent matrices of the sizes defined by $l \times n$, $c \times n$, $l \times c$, $r \times l$, and $n \times r$, respectively.

C. Application of SSNMF to this research project

The application of SSNMF to the target problem in this paper is explained in this section. The target problem in this paper should be the manifestation of the knowledge dealt with in the lectures that are relatively relevant to the laboratory. By applying SSNMF and giving a set of lectures as a template

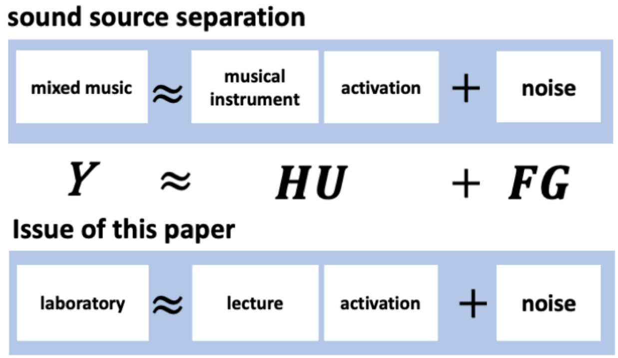

Fig. 1. Correspondence between sound source separation and the application of SSNMF in this issue.

pattern to a set of laboratories, we can obtain the output concerning which lecture is related to which laboratory as an activation matrix. The relationship between lectures and laboratories is shown as the activation matrix. The activation matrix is composed by considering both base and observed matrices. We believe this computation should be alike human thinking for decision-making in our life (that is, multi versus multi combinations) described in section II.

We detail the problem tackled in this paper by comparing it with other applications using SSNMF. Fig. 1 shows the correspondence between the acoustic source separation and the task in this paper. When SSNMF is used in acoustic source separation, it decomposes a matrix $\boldsymbol{Y}$ consisting of frequencies and times representing mixtures of multiple sound sources into a template pattern $\boldsymbol{H}$ that represents the frequency structure of the instruments to be extracted. As a result, a matrix $\boldsymbol{U}$ is extracted as the activation of each instrument at each time in polyphonic music. On the other hand, this paper targets which knowledge covered in the lectures is relevant to which laboratory where basic and the corresponding specialized knowledge are mixed. The problems are similar in structure of the problem, and it lets us apply SSNMF to analyze the relationships between lectures and laboratories. By decomposing the matrix $\boldsymbol{Y}$ representing laboratories by the matrix $\boldsymbol{H}$ concerning the lectures, we expect to obtain an activation matrix $\boldsymbol{U}$ showing the relationship between laboratories and knowledge handled in the lectures. As a resource for lectures and laboratories, we deal with matrix information that can be represented by non-negative values. This process will be detailed in section IV.

IV. Estimation of the Relationships between Lectures and Laboratories

We apply the proposed method to analyze the relationship between lectures and laboratories. This section describes the data and the parameters of the experiment.

A. Data Preparation

This paper analyzes the Faculty of Informatics of Kansai University (the Faculty of SJ), which is the authors' affiliation. The faculty integrates humanities and sciences, where students learn various specialties in a cross-field manner. The Faculty of SJ offers courses in various specialized fields, such as basic theories of informatics: programming and algorithms (C course,) information processing in media and communication (M course,) and information processing in various fields including management, economics, psychology, and politics (S course.) The professors for each laboratory respectively offer lectures on information processing in the undergraduate department. The professor supervises students in their research for each corresponding laboratory.

1) Research fields: We used the graduation thesis outlines as the information resource for 43 laboratories, which were collected from the SJ undergraduate thesis outline collection in 2019. In this paper, a single laboratory is assumed as a specific field of research.

The entire text was extracted from the graduation thesis outlines for each laboratory. We assume the texts as the knowledge in the field of research. Here, the professor's name, student ID number/name, and references are excluded.

2) Lectures: The information source of lectures is the syllabus of the SJ faculty for the year 2020, collected from the university website. The number of lectures was 192 courses offered in FY2020, excluding foreign language and sports training lectures. The syllabus of lectures describes the title, style, professor's name, outline, plan, achievement objectives, methods, and grade evaluation. This paper focuses on the texts listed in the outline, plan, and achievement objectives as the content of the lectures.

3) Normalization: The text (the thesis outline and syllabus obtained in sections section IV-A1 and section IV-A2) was normalized for single-byte alphanumeric characters and symbols while line breaks and spaces were removed. The morphological analyzer mecab-python3 (ver. 1.0.1) [9] was used to segment the text into words, and only nouns were extracted. We used mecab-ipadic-NEologd [1] as a word segmentation dictionary. Then, the stop words [10] were excluded.

B. Numerical and Distributed Representation of Data

The data obtained in section IV-A were converted to numerical by using the bag of words (BoW) method. A numerical matrix combining lectures and laboratories was created, as shown in Fig. 2. Words that were less than one per a single document were removed for word weighting. And then, lectures and laboratories were numerically represented by the word frequency. From the resulting 235 (43 laboratories + 192 lectures) $\times$ 7560 (number of BoW words) matrix, a 500-dimensional variance representation was obtained by using NMF.

[1]https://github.com/neologd/mecab-ipadic-neologd

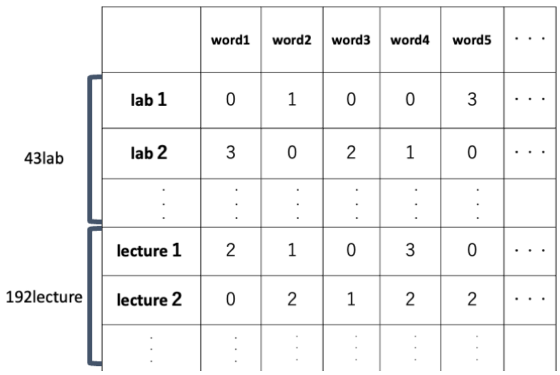

Fig. 2. Non-negative matrix showing lectures and laboratories obtained through the bag of words method.

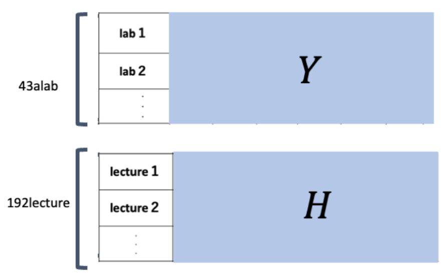

Fig. 3. The general idea of matrices $\boldsymbol{Y}$ and $\boldsymbol{H}$ of SSNMF in this paper.

C. Application of SSNMF

SSNMF was applied to the matrices obtained in section IV-B. As shown in Fig. 3, $\boldsymbol{Y}$ and $\boldsymbol{H}$ were obtained by splitting the matrix into two vector sets representing the matrix concerning lectures and laboratories. These two matrices satisfy two requirements: 1) a non-negative matrix and 2) representing the knowledge concerning the content. We use the matrix concerning lectures as the reference vector to factorize the matrix concerning laboratories in SSNMF. Here, the number of iterations for approximation was experimentally set to 100,000.

D. Evaluation Method

This paper quantitatively evaluates the effectiveness of the proposed method. We focus on manifesting the relationship between lectures and laboratories in the evaluation.

1) Comparative method: We used the word2vec approach as a comparative method. The word2vec approach is no longer a general method to obtain a distributed representation of words by machine learning [11]. The approach can represent the semantics of words as vector representations. By calculating the cosine similarity among the vectors, it is possible to know the semantic similarities among words. In this paper, we used word2vec with Facebook's trained FastText model [12] to obtain vectors of words. The averaged vector of the words composing a sentence was assumed as vector of the sentence.

TABLE I
MRR

data	word2vec	ssnmf
all	0.2584	0.2758
corseC	0.2624	0.2927
corseM	0.1470	0.3480
corseS	0.3659	0.1866

TABLE II
PRECISION RATIO

word2vec	ssnmf
0.34013	0.3509

Each sentence was represented as a 300-dimension vector. We studied the cosine similarity among sentences concerning lectures and laboratories as one of the ways to represent the relationships between those.

2) Evaluation Criteria: The lectures are labeled C, M, and S in the SJ Faculty to identify the lectures' specialties. Of the 192 lectures, 31 are in the C series, 25 are in the M series, 28 are in the S series, while 108 cross-disciplinary or unlabeled lectures exist. This paper labeled professors as well as the lectures considering the professors' teaching: e.g., if a professor teaches a lecture labeled C, the professor was also labeled C. These labels were used as the evaluation criteria for the estimation. We assumed that the relationship between a lecture and a laboratory was correctly estimated if the same label was used for both of them.

The Mean Reciprocal Rank (MRR) and the precisions were used for the evaluation in this paper. For the comparative method, the lectures were sorted by the cosine similarities between word2vec-based vectors of lectures and laboratories. Meanwhile, the proposed method showed the relative lectures for each laboratories as the activation matrix. We calculated the MRR and precision for each result considering the correct label. The higher the value of reciprocal rank is, the better the method estimates the lectures related to the laboratory. Precision refers to the percentage of correctly retrieved sentences among the sentences to be retrieved. Here, the percentage of correctly rated lectures out of the top 50 most relevant lectures by the proposed and comparative methods are shown for each laboratory, and the average value is obtained.

V. Results and Discussion

A. Discussion based on MRRs and Precisions

TABLE I shows the results of the MRRs. In addition to the average for all lectures, the average calculated for each course (i.e., C, M, and S) is also shown. Overall, the propose method showed a little higher effectiveness, but it was not significant. The problem should be only about 30 each applied to the course of the 192 lectures; 15% chance-level estimation. Accordingly, it is reasonable to say that both methods showed relatively effective results.

Let us study the results for each course. The proposed method showed higher effectiveness than the comparative method for courses C and M, while its effectiveness for course S was extremely low. Since the proposed method calculated the relevance by taking into account the surrounding information, we considered words that also appeared in other lectures, and laboratories were not considered as features. For example, “Diet” and “public administration” appeared in many lectures across the disciplines related to politics. Moreover, those words also appeared in the laboratories, such as for the budgetary accounting system, political systems, and rational behavior. Word2vec did not focus on the surroundings, and it might cause better results in such a case.

TABLE II shows the results of the precision. Both the proposed and comparative methods showed effectiveness. These results indicated that both methods were able to represent the relationship between lectures and laboratories reasonably.

B. *Multi versus multi relationships.*

This section discusses whether the proposed method was able to analyze the relationship between lectures and laboratories, which has a fluid relationship as described in section II. TABLE III summarizes lectures with higher estimation accuracy in a laboratory specializing in human-media communication design. TABLE IV shows the result with all lectures as the input, while TABLE V shows the result with only course S as the input.

In TABLE V, “environmental economics” was estimated as the most related lecture. However, in TABLE IV, different kinds of lectures were estimated as related lectures. This result indicated that environmental “economics” appeared to be more relevant with only course S, while other lectures were more relevant with courses S and M. In fact, this laboratory is for a human-media communication design, which is closely related to both courses C and M; it is reasonable that the lectures for course M should be more relevant to this laboratory. TABLE VI summarizes relevant lectures for a laboratory specializing in feature visualization. TABLE VII shows the result with all lectures as the input. On the other hand, TABLE VIII shows the result with only course M as the input. In this case, the theory of “information behavior,” which was fourth in TABLE VIII, came in second place in TABLE VII. Meanwhile, “media representation theory,” which was ranked first in TABLE VIII, was not ranked in the top 10 in TABLE VII; it was ranked 23rd. This might be caused by the fact that the features of the laboratory were similar to lectures of other courses. If the words appear in many lectures, such words could not be characteristic for a specific lecture. When the set of lectures would change according to the course, the characteristics of the lecture would change as well. This might be the reason why the results changed with reference to the input.

The above result showed that the proposed method considered not only a single lecture in the input but also other surrounding lectures in the estimation. In other words, it was suggested that the proposed method visualized the key factors considering the surrounding information. So multi-versus-multi estimation should be realized by the proposed method.

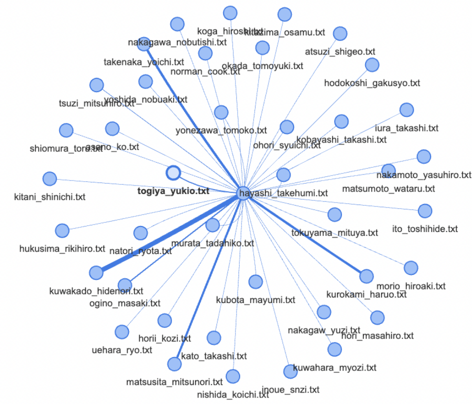

Fig. 4. An example of a network diagram visualizing the relationship between lectures and laboratories.

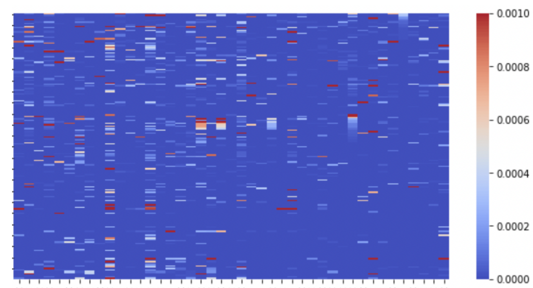

Fig. 5. An example of visualizing the relationship between lectures and laboratories in a heat-map diagram.

VI. Conclusion

In this paper, we proposed a method to estimate the relationships between lectures and laboratories. The proposed method applied semi-supervised non-negative matrix factorization to identify common knowledge factors for each combination of laboratory and lecture. The results of the quantitative evaluation suggested the effectiveness of the proposed method, especially for estimating multi versus multi relations. The results should be hints for students to find laboratories for their background, i.e., the history of lectures they have taken.

We will carry out the user tests after incorporating this relationship of the results, for instance, network diagrams as shown in Fig. 4 and heat maps as shown in Fig. 5. The future task will be investigating the effectiveness of the combination of the proposed method and visualization ways by studying the thoughts of students using the Thinking aloud protocol [13].

Acknowledgement

This research is, in part, supported by JSPS KAKENHI 22K00803. We show our best acknowledgments.

TABLE III

LECTURES RELATIVELY RELATED TO A LABORATORY SPECIALIZING IN HUMAN MEDIA COMMUNICATION ESTIMATED BY THE PROPOSED METHOD.

TABLE IV

TOP LECTURES WITH HIGH ACTIVATION TENDENCY VALUES WHEN ALL LECTURES ARE USED AS INPUT DATA.

Lectures	Value
Social Psychology	4.0E-323
Media Art	2.6E-322
Regional Media Theory	2.1E-322
Basic Mathematics	2.0E-323
Applied Mathematics	1.0E-323
Intellectual Property Law	1.0E-323
Cognitive Science	1.9E-322
Internet Journalism	1.0E-323
Intelligent Computing	1.0E-323
Psychology	1.0E-322
Environmental Economics	0.000136

TABLE V

TOP LECTURES WITH HIGH ACTIVATION TENDENCY VALUES WHEN S COURSE LECTURES ARE USED AS INPUT DATA.

Lectures	Value
Environmental Economics	8.4E-323
Management Strategy	5.0E-324
Nonprofit Organizations	5.0E-324
Public Administration	5.0E-324
Organizational Decision Making	1.0E-323
Microeconomic Modeling	1.0E-323
Political Institutions	1.1E-322
Economic Policy Simulation	4.39E-07
Risk Management	4.02E-07
Marketing Research	1.31E-07
Business Innovation	1.15E-08

TABLE VI

LECTURES RELATIVELY RELATED TO THE LABORATORY SPECIALIZING IN FEATURE VISUALIZATION ESTIMATED BY THE PROPOSED METHOD.

TABLE VII

TOP LECTURES WITH HIGH ACTIVATION TENDENCY VALUES WHEN ALL COURSE LECTURES ARE USED AS INPUT DATA.

Lectures	Value
Environmental Economics	7.0E-323
Information Behavior	7.1E-322
Mobile Computing Practice	6.0E-323
Mathematics Exercise (Analysis)	5.0E-323
Robot Brain Computing Practice	5.0E-323
Media Art	5.0E-322
Linguistics	5.1E-322
Business Database Practicum	5.0E-323
Information, Culture and Communication	4.0E-323
Systems Programming Practicum	4.5E-322
Fundamentals of Software Development	4.0E-323

TABLE VIII

TOP LECTURES WITH HIGH ACTIVATION TENDENCY VALUES WHEN M COURSE LECTURES ARE USED AS INPUT DATA.

Lectures	Value
Media Expression	4.0e-323
Information Media	1.0e-323
Design	1.7e-322
Information Behavior	4.83E-05
Information, Culture, and Communication	2.49E-05
Design Practice	1.23E-05
Design of Cognitive Artifacts	6.48E-06
Problem Setting and Assessment Methods in STEM	5.37E-07
Multimedia Education	1.80E-07
Ethics and Philosophy of Science	3.35E-08
Entertainment Theory	1.61E-08

REFERENCES

[1] R. Ferreira-Mello, M. André, A. Pinheiro, E. Costa, and C. Romero, "Text mining in education," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 9, no. 6, p. e1332, 2019.

[2] O. C. Acosta, P. A. Behar, and E. B. Reategui, "Content recommendation in an inquiry-based learning environment," in *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*. IEEE, 2014, pp. 1–6.

[3] K. Kitto, N. Sarathy, A. Gromov, M. Liu, K. Musial, and S. Buckingham Shum, "Towards skills-based curriculum analytics: Can we automate the recognition of prior learning?" in *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, ser. LAK '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 171–180. [Online]. Available: https://doi.org/10.1145/3375462.3375526

[4] K. Kawintiranon, P. Vateekul, A. Suchato, and P. Punyabukkana, "Understanding knowledge areas in curriculum through text mining from course materials," in *2016 IEEE international conference on teaching, assessment, and learning for engineering (TALE)*. IEEE, 2016, pp. 161–168.

[5] H. Lee, J. Yoo, and S. Choi, "Semi-supervised nonnegative matrix factorization," *IEEE Signal Processing Letters*, vol. 17, no. 1, pp. 4–7, 2010.

[6] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Advances In Neural Information Processing Systems*, 2001, pp. 556–562.

[7] H. Kameoka, M. Nakano, K. Ochiai, Y. Imoto, K. Kashino, and S. Sagayama, "Constrained and regularized variants of non-negative matrix factorization incorporating music-specific constraints," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 5365–5368.

[8] D. Kitamura, N. Ono, H. Saruwatari, Y. Takahashi, and K. Kondo, "Discriminative and reconstructive basis training for audio source separation with semi-supervised nonnegative matrix factorization," in *2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE, 2016, pp. 1–5.

[9] T. Kudo, K. Yamamoto, and Y. Matsumoto, "Applying conditional random fields to japanese morphological analysis," in *Proc. 2004 Conference on Empirical Methods in Natural Language Processing*, 2004, pp. 230–237.

[10] H. Ohshima, S. Nakamura, and K. Tanaka, "Slothlib: A programming library for research on web search," *The Database Society of Japan (DBSJ Letters)*, vol. 6, no. 1, pp. 113–116, 2007.

[11] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[12] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," *arXiv preprint arXiv:1607.01759*, 2016.

[13] C. Lewis, *Using the "thinking Aloud" Method in Cognitive Interface Design*, ser. Research report. IBM Thomas J. Watson Research Division, 1982.