Exploring the AI Topic Composition of K-12 Using NMF-based Topic Modeling

HoSung Woo\textsuperscript{a}, JaeHong Kim\textsuperscript{b}, JaMee Kim\textsuperscript{c}, WonGyu Lee\textsuperscript{d}

\textsuperscript{a}Department of Computer Science and Engineering, Graduate School, Korea University, Seoul 02841, South Korea
E-mail: hosung.woo@inc.korea.ac.kr

\textsuperscript{b}Major of Educational Information System, Graduate School of Education, Korea University, Seoul 02841, South Korea
E-mail: jaehong.kim@inc.korea.ac.kr

\textsuperscript{c}Major of Computer Science Education, Graduate School of Education, Korea University, Seoul 02841, South Korea
E-mail: celine@korea.ac.kr

\textsuperscript{d}Department of Computer Science and Engineering, College of Informatics, Korea University, Seoul 02841, South Korea
E-mail: lee@inc.korea.ac.kr

Abstract—Recently, artificial intelligence has become more prevalent due to the combination of more data, faster processing power, and more powerful algorithms. AI technology has been introduced into almost all industries and is also affecting the education sector. The objective of this study was to explore AI topics through an analysis of literature related to AI education for grades K-12 and provide implications for the composition of a system for AI education. For this purpose, 27 materials released at the 2018 and 2019 AI4K12 Symposiums were collected. Besides, artificial intelligence integration across subjects and artificial intelligence curriculum published by CBSE of India were collected for analysis. The frequency of words, word cloud, and topic modeling was performed for each collected document. According to the analysis, content on the necessary future direction for AI education and introductions to educational tools were extracted from the 2018 symposium, whereas the 2019 symposium contained more concrete discussions on how to conduct AI education in schools. Meanwhile, content involving the principles of integration for how to integrate AI with other subjects and AI-based teaching and learning methods were extracted from Artificial Intelligence Integration Across Subjects. Finally, Artificial Intelligence Curriculum covered the theories and principles of AI. This study has significance in that it analyzed how much discussion about AI education is being conducted in K-12 based on topic modelling and suggested future directions for AI education.

Keywords—K-12 AI curriculum; AI curriculum; topic model; topic analysis.

I. INTRODUCTION

Patterns obtained by processing vast quantities of data are used to derive new knowledge or present methodologies in various fields. According to Gartner’s 2019 CIO Survey, the number of firms that adopted AI in the workplace increased by 270% over the past four years [1]. This is because AI is used to efficiently solve numerous problems that were previously left unsolved [2]. Based on its potential to provide tremendous value across a variety of fields, the utility of AI is increasing as a national growth engine [3]-[5]. Education is one such field, in which discussions on how to reflect AI from the perspective of fostering talent have begun. Though AI was recognized as a field of higher education as it involves specialized knowledge, its scope has recently expanded to K-12.

The AI4K12 Initiative in the United States is providing an infrastructure for K-12 AI training by developing AI curriculum guidelines, building repositories for teaching, and learning materials for teachers, and creating collaborative communities [6]. China is making systematic efforts at the national level, including issuing AI textbooks, operating pilot schools, and cultivating teachers [7]. Through the Central Broadcasting Board (CBSE), India published the 9th Grade Artificial Intelligence Curriculum and AI Integration Across Subjects [8]. As basic knowledge for the digital society, Japan defined “mathematics, data science, and artificial intelligence” as standards for national literacy and established educational goals for these fields [9]. Though individual nations have proposed, including educational content for AI, they have not composed a concrete content system. In contrast, an educational system for curriculum guidelines, teaching and learning materials,
and the like has been created through various collaborative projects and is sometimes distributed to schools through textbooks. Unlike other content, though numerous countries agree on the necessity of AI education, there are few documents that systematically and concretely present what knowledge should be taught at what level in the curriculum. As education is based on curricula, the objectives, methods, and levels of AI education must be established to concretize the educational content [10].

This study aims to explore topics related to AI for K-12 and provide implications regarding the composition of a system for AI education. To achieve these objectives, this study extracted topics from AI-related documents for K-12 using NMF-based topic modeling. The topics extracted from topic modeling can help determine what topics should be taught in AI education, thus contributing to providing direction for topic composition.

II. MATERIALS AND METHOD

The following procedure was conducted to analyze AI-related topics and explore meanings through the relationships between terms.

![Research procedure](image)

**A. Data Collection**

This study collected documents provided at the national level or by credible institutions. First, data published at the AI4K12 symposium was collected [11],[12]. The 2019 symposium had a wider variety of topics than the 2018 symposium. The 2019 symposium involved discussions on the use of STEM education from the perspective of convergence. Also, it deals with the necessity of AI education as a field of CS, primary contents of preparing AI education, and curriculum. There were eight published materials in 2018 and 19 in 2019, totaling 27.

Second, this study collected data from Artificial Intelligence Integration Across Subjects (AIIAS), a document published by CBSE in India for integrating AI into education. Rather than direct knowledge of AI, this is a guide for teachers in various subjects to use AI in the classroom to maximize learning effectiveness. Third, this study collected as in Table I from the Artificial Intelligence Curriculum (AIC) for 9th grade, published by CBSE in India.

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>CONTENTS BY DOCUMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category</strong></td>
<td><strong>Contents</strong></td>
</tr>
</tbody>
</table>
| **2018 Symposium** | • What is the "AI for K-12" initiative?  
• AI for K-12: Converging agendas  
• AI programming by children  
• Towards early AI for preK-2nd grade  
• AI and society  
• What to think about machines that think?  
• What can kid in grades 3-5 do?  
• - What can kid in grades 9-12 do? |
| **2019 Symposium** | • AI4K12 Initiative Update  
• From Computational Thinking to Computational Action  
• All means All  
• An Interdisciplinary Approach to Bring AI into Existing Curriculum  
• A High School Student's Perspective on Artificial Intelligence Education  
• Barriers to Convergence with AI in K-12  
• Empowering Educators to Teach Artificial Intelligence  
• INSPIRE CS-AI  
• - K-12 AI Outreach in Higher Education  
• STEM-based AI education in professional development for K-12 teachers  
• Building a Research Foundation for K-12 AI Education  
• Computing in English Schools: Lessons to Learn for AI education  
• Integrating Ethics into K12 AI Learning Experiences  
• Bringing AI to K-12 Education via Global STEM Classroom  
• CS in K-12 and supplement  
• What do you meme?  
• 30 Minutes to Introduce AI to Kids  
• Teaching AI in K-12 |
• Conversational Artificial Intelligence Development Tools for K-12 Education

**AIIAS**
- An Introduction to Artificial Intelligence
- How to Integrate AI in School Teaching– A Call to Teachers
- AI Integrated Activities – Exemplars

**AIC**
- Conceptual Framework
- The AI Curriculum
- Unit 1: Excite / Relate / Purpose / Possibilities / AI Ethics
- Unit 2: AI Project Cycle / Problem Scoping / Data Acquisition / Data Exploration / Modelling
- Unit 3: Neural Networks
- Co-Curricular Activities

**B. Data Preprocessing**

A preprocessing step for removing words unnecessary for semantic analysis is needed to select meaningful words in the data analysis. By using the collected unstructured data, preprocessing was conducted through the following procedure. First, 31 materials were converted to a corpus, after which uppercase letters were normalized to lowercase letters. Second, various forms of words derived from the same meaning (different tenses, singular vs. plural, etc.) were converted into one headword. Third, content that frequently appears in sentences but does not greatly contribute to analysis (numbers, special characters, pronouns, prepositions, names, and name of places) were removed. Based on the collected data, the number of words used in this study is shown in Table II.

<table>
<thead>
<tr>
<th>Category</th>
<th>Total word count</th>
<th>After preprocessing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total word count</td>
<td>Unique word count</td>
</tr>
<tr>
<td>2018 Symp.</td>
<td>4,788</td>
<td>4,287</td>
</tr>
<tr>
<td>2019 Symp.</td>
<td>20,014</td>
<td>18,134</td>
</tr>
<tr>
<td>AIIAS</td>
<td>37,145</td>
<td>35,803</td>
</tr>
<tr>
<td>AIC</td>
<td>21,118</td>
<td>20,416</td>
</tr>
</tbody>
</table>

**C. Analysis Method**

As methods for analyzing the AI-related documents, this study conducted topic analysis and frequency analysis, which are suitable for analyzing unstructured data. First, a word cloud applying the frequency of the words and a visualization package was used. Frequency analysis uses the total frequency of words that appear throughout the entire document. The most straightforward and most concise text mining method, this technique enables the examination of AI-related keywords. A word cloud is a visualization technique in which the sizes of the words differ with frequency. More frequent or essential words are expressed in larger text, whereas less frequent words are shown in smaller text. As such, a word cloud allows the viewer to quickly identify which words are emphasized in the document [13]. Second, Topic modeling is a text mining technique used to discover the hidden semantic structure of the text. It is useful for exploring topics or changing topic trends according to a time series for large amounts of unstructured data, such as social media and newspaper articles [14]-[16].

NMF-based topic modeling was employed to explore the topics. NMF is an unsupervised approach for reducing the dimensions of matrices that do not contain negative values. As the semantic feature matrix and semantic variable matrix are composed of sparse matrices, it has characteristics that accurately represent partial information of the original data [17]-[20]. The NMF-based model has fewer parameters to select than the traditionally used LDA, making it easy to use. It is also fast due to the small number of calculations [20], [21]. The processing procedure for NMF-based topic modeling are select k topics to the cluster and generate matrix A using n corpora generated from n documents. m is the number of unique words among all words. Matrix A is shown in (1).

\[ A \in \mathbb{R}^{m \times n} \]  

Through NMF application, it is decomposed into non-negative semantic feature matrix (NSFM) W and H. The objective function (3) is learned to obtain the minimum value. r is generally set to a value smaller than m or n, and the size of W or H is set to less than that of A. The decomposed H row is interpreted as k topics for each term in the corpus vocabulary.

\[ A \simeq WH \]  

\[ \theta_j^t(W;H) = \| A - WH \|_F^2 = \sum_{i=1}^{m} \sum_{j=1}^{r} (X_{ij} - \sum_{r=1}^{t} W_{ir} H_{rj})^2 \]  

(3)

To update the element values of W and H, (4) and (5) are repeated until \( \frac{\|H_{t+1} - H_{t}\|_2}{\|H_{t}\|_2} \) becomes smaller than the convergence tolerance error or exceeds the specified number of repetitions.

\[ H_{t+1} = H_{t} - (H_{t} W_{t} W_{t}^T A_{t})_{t}^{-1} (W_{t}^T A_{t} - W_{t}^T H_{t} A_{t}) \]  

\[ W_{t+1} = W_{t} - (H_{t}^T H_{t})_{t}^{-1} (H_{t}^T A_{t} - W_{t} H_{t}^T A_{t}) \]  

(4)

The hyperparameters of the topic model were set as follows in this study. Frobenius norm was used for the objective function, and the number of repetitions to update the values of W and H was set to 100. To improve the quality of the clustered topics, initial values were generated using NNDSVD (Non-Negative Double Singular Value Decomposition) [22]. This is because the NMF algorithm is initialized with random factors, thus possibly leading to unstable results.

**III. RESULTS AND DISCUSSION**

This study extracted topics from AI-related documents for K-12 using NMF-based topic modeling, the analytical results
of which are as follows. As shown in Fig. 2, the visualization created by extracting the most frequently used words.

Among the top 20 keywords in the frequency analysis, the highest words are 'ai', 'student', 'data', 'learning', and 'machine'. This signifies that these words appear most frequently in the entire document. In the word cloud, 'ai', 'student', 'data', 'artificial intelligence' and 'machine' are shown with the largest font. Moreover, 'teacher', 'learning', 'help', and 'neural network' are shown at a confirmable level.

The emphasis on the words 'ai', 'student', 'data', 'learning', and 'machine' in the word frequency, and visualization results indicate that content on data and machine learning for educating students about AI frequently appears. The frequency and subject modelling results of the top 20 words per document are shown in Table III.

As shown in Table III, the most frequently used words at the 2018 symposium are 'ai', 'idea', and 'robot'. It is notable that in 2018, words related to education such as 'idea', 'group', 'working', 'session', and 'service' frequently appeared. According to the topic modelling results, Topic 1 consists of AI, Funding, Converging, Agency, Convergence, commercialization, and agenda. As this is the first symposium, these can be interpreted as topics related to the direction that AI education must take. Topic 2 includes the big idea for AI education, characteristics of AI, and social impact. Topic 3 and Topic 4 are introductions to tools for AI education. eCraft2Learn in Topic 3 is a platform for sharing projects in a cloud service environment to receive feedback from designers, engineers, and programmers. Topic 4 concerns popbot, an AI educational robot for children. Topic 5 involves sentencing software that issues judgements in place of a judge in court; as such, this topic is determined to be related to AI ethics. Thus, the 2018 symposium is shown to consist of discussions on the necessary direction of AI education, educational tools for AI education, and the impact of AI on society.

As shown in the Table IV, the most frequent words in the 2019 symposium were 'ai', 'student', and 'K-12' in order, with more words related to education compared to the 2018 symposium ('learning', 'education', 'teacher', 'curriculum', etc.).

As shown in Table III, the most frequently used words at the 2018 symposium are 'ai', 'idea', and 'robot'. It is notable that in 2018, words related to education such as 'idea', 'group', 'working', 'session', and 'service' frequently appeared. According to the topic modelling results, Topic 1 consists of AI, Funding, Converging, Agency, Convergence, commercialization, and agenda. As this is the first symposium, these can be interpreted as topics related to the direction that AI education must take. Topic 2 includes the big idea for AI education, characteristics of AI, and social impact. Topic 3 and Topic 4 are introductions to tools for AI education. eCraft2Learn in Topic 3 is a platform for sharing projects in a cloud service environment to receive feedback from designers, engineers, and programmers. Topic 4 concerns popbot, an AI educational robot for children. Topic 5 involves sentencing software that issues judgements in place of a judge in court; as such, this topic is determined to be related to AI ethics. Thus, the 2018 symposium is shown to consist of discussions on the necessary direction of AI education, educational tools for AI education, and the impact of AI on society.

As shown in the Table IV, the most frequent words in the 2019 symposium were 'ai', 'student', and 'K-12' in order, with more words related to education compared to the 2018 symposium ('learning', 'education', 'teacher', 'curriculum', etc.).

Table of Contents

- Topic Modeling
- Word (Frequency)
- Topic (Frequency)
- Table IV 2019 AI4K12 Symposium Results
- As the numbers for eCraft2Learn, aiall, and k were removed during preprocessing, they are enclosed in parentheses.
sensors. Finally, Topic 5 comprised content on ReadyAI, an educational institution for AI.

This indicates that the 2019 symposium concerned more concrete discussions on how to conduct AI education in schools. AIIAS, which was published in India in 2019, is a guide for integrating AI with other subjects. As shown in the TABLE V, it includes words related to teaching and learning methods such as 'activity', 'story', 'problem', 'learning' and 'understanding'.

Topic 1 covered topics on symmetry in mathematics, with words such as symmetry, object, and line. A closer inspection of the curriculum revealed that it concerns an activity for predicting the finished image through an AI-based application that draws symmetric and asymmetric images. Topic 2 concerns principles or methods for teachers to integrate AI with topics across a variety of areas, and includes the words education, teacher, integrating, principal, etc. Topic 3 is an explanation of the AIIAS framework, including educational objectives, activities, thinking, and recommended content. Topic 1 is for the Mathematics subject, while Topic 4 concerns Science. More specifically, the analysis showed that Topic 4 contains content on the principles of sound in science. Topic 5 contains topics on teaching and learning methods such as story-, problem-, and question-based learning methods.

To summaries, AIIAS covers the principles of integrating AI with other subjects, explanations of AI-based teaching and learning methods, and examples of integration with other subjects. Unlike the other three documents, the frequency of 'data' in TABLE VI was the highest, followed by 'ai', 'problem', and 'student'. Words highly related to AI knowledge domains also appeared, such as 'layer', 'model', 'network', and 'tree'.

<table>
<thead>
<tr>
<th>Topic Modeling</th>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
<th>Topic 4</th>
<th>Topic 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>ai</td>
<td>ai</td>
<td>ai</td>
<td>ai</td>
<td>student</td>
<td></td>
</tr>
<tr>
<td>symmetry</td>
<td>education</td>
<td>data</td>
<td>sound</td>
<td>data</td>
<td></td>
</tr>
<tr>
<td>student</td>
<td>teacher</td>
<td>skill</td>
<td>machine</td>
<td>story</td>
<td></td>
</tr>
<tr>
<td>ask</td>
<td>subject</td>
<td>problem</td>
<td>student</td>
<td>ai</td>
<td></td>
</tr>
<tr>
<td>object</td>
<td>school</td>
<td>goal</td>
<td>class</td>
<td>problem</td>
<td></td>
</tr>
<tr>
<td>line</td>
<td>intelligence</td>
<td>activity</td>
<td>learning</td>
<td>ask</td>
<td></td>
</tr>
<tr>
<td>skill</td>
<td>teaching</td>
<td>thinking</td>
<td>teacher</td>
<td>activity</td>
<td></td>
</tr>
<tr>
<td>draw</td>
<td>artificial</td>
<td>recommended</td>
<td>education</td>
<td>statement</td>
<td></td>
</tr>
<tr>
<td>image</td>
<td>integrating</td>
<td>learner</td>
<td>intelligence</td>
<td>machine</td>
<td></td>
</tr>
<tr>
<td>learning</td>
<td>principal</td>
<td>session</td>
<td>data</td>
<td>understand</td>
<td></td>
</tr>
</tbody>
</table>

Table VI: AIIAS RESULTS

<table>
<thead>
<tr>
<th>Word (Frequency)</th>
<th>data (271)</th>
<th>student (267)</th>
<th>problem (155)</th>
<th>student (111)</th>
<th>learning (103)</th>
</tr>
</thead>
<tbody>
<tr>
<td>machine</td>
<td>activity (83)</td>
<td>project (73)</td>
<td>system (70)</td>
<td>layer (70)</td>
<td></td>
</tr>
<tr>
<td>skill</td>
<td>intelligence(65)</td>
<td>game (64)</td>
<td>artificial (60)</td>
<td>model (59)</td>
<td></td>
</tr>
<tr>
<td>image</td>
<td>story (54)</td>
<td>network (52)</td>
<td>tree (51)</td>
<td>get (50)</td>
<td></td>
</tr>
</tbody>
</table>

To summaries, AIIAS covers the principles of integrating AI with other subjects, explanations of AI-based teaching and learning methods, and examples of integration with other subjects. Unlike the other three documents, the frequency of 'data' in TABLE VI was the highest, followed by 'ai', 'problem', and 'student'. Words highly related to AI knowledge domains also appeared, such as 'layer', 'model', 'network', and 'tree'.

Topic 1 in AIC includes words related to the curriculum framework for teaching AI, such as teaching time, skill, project, and unit. Topic 2 concerns neural networks, with words such as layer, network, node, chit, data, machine, hidden, and input. A closer inspection of the curriculum revealed that "chit" relates to unplugged activities for understanding the principles of neural networks. That is, students can indirectly experience how neural networks work bypassing the words they wrote in a chit to the nodes of the input layer, hidden layer, and output layer. AIC also includes topics on the capabilities and skills that students must develop through AI education, future jobs, and topics on designing tree models based on given data. Topic 2, Topic 4, and Topic 5 are related to learning topics regarding the knowledge that learners should learn. A closer inspection of AIC showed that project-based methods were used as teaching and learning methods for topics such as decision trees and neural networks based on AI case studies. The curriculum was designed to attract the students' interest through gamification and focus on learning for topics difficult to cover for school-aged students. AIC was observed to cover basic theories and principles related to AI (neural, network, hidden, layer, etc.).

IV. CONCLUSIONS

As technology changes the paradigm of education, AI has emerged as a new methodology to solve a variety of problems in life. 2018 has seen proposals for AI as content in K-12 education, as its use in a variety of fields increases. Rather than to cultivate specialized labor, AI education in K-12 is considered socialization education for future generations who must live in the AI era. Accordingly, this study performed topic modelling to provide implications for composing a system for AI education. The analytical results of the four AI education-related documents are as follows.
First, educational objectives must be systematised. AI is rapidly changing compared to other academic fields. Everyone, therefore, recognises the necessity of AI education for K-12, though the systematic nature or hierarchy of the topics remain uncertain. The knowledge system constituting the curriculum will vary with how educational goals are set, and education based on this can help cultivate learners' capabilities. Educational goals should thus be set considering educational philosophy, social change, learning experiences, and what students should be capable of doing. The second implication concerns the composition of a curriculum that reflects academic sequence and continuity. Based on the structure of knowledge suggested by Bruner, a curriculum should establish sequence within K-12 education in addition to the level and scope of knowledge in relation to higher education. That is, AI education requires a curriculum considering sequence and continuity based on the knowledge covered in higher education. An AI curriculum published at the national level includes India's 9th grade AI curriculum. Other than this, there are very few to no standard curricula. An example of the curriculum composition direction is shown in Table VII.

**TABLE VII**

**EXAMPLE OF K-12 AI CURRICULUM COMPOSITION DIRECTION**

<table>
<thead>
<tr>
<th>Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-4</td>
<td>Education focused on play/activities without revealing AI knowledge domains</td>
</tr>
<tr>
<td>5-6</td>
<td>Education utilising programming tools to understand the basic concepts and operation of AI</td>
</tr>
<tr>
<td>7-9</td>
<td>Education to understand the principles of AI with conceptual diagrams (pictures) and basic algorithms</td>
</tr>
<tr>
<td>10-12</td>
<td>Education for using data from various fields in basic AI algorithms</td>
</tr>
</tbody>
</table>

Rather than designing high-level AI algorithms, the objectives of K-12 education for AI should be to understand the concepts and functional roles of AI and develop problem-solving skills. As discussed in this study, however, clear directions are needed for what students should be expected to learn how to do through K-12 AI education. This is because the content and methods of education will vary with the goals. The topic modelling results in this study can only intuitively identify topics and are limited in terms of deeper quantitative or semantic analysis. Furthermore, future studies must be conducted with more extensive data. This study is significant in that it analysed the extent of discussions on AI education in K-12 based on topic modelling and proposed future directions for AI education.

**ACKNOWLEDGEMENT**

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2019R1H1A2101061).

**REFERENCES**


